

**Cristina Dotta Ortega**

**INFORMÁTICA DOCUMENTÁRIA:  
ESTADO DA ARTE**

Dissertação apresentada à Escola de Comunicações e Artes,  
da Universidade de São Paulo,  
como exigência parcial para obtenção do título de  
Mestre em Ciências da Comunicação,  
área de concentração Ciência da Informação e Documentação

Orientadora: Prof. Dra. Maria de Fátima G. M. Tálamo

**São Paulo  
2002**

## ERRATA

página	parágrafo	onde se lê	leia-se
4	3º	...dos referenciais teóricos da Ciência da <u>informação</u>	...dos referenciais teóricos da Ciência da <u>Informação</u>
13	6º	Mas parece ser a Ciência da <u>informação</u> que,...	Mas parece ser a Ciência da <u>Informação</u> que,...
16	4º	...em <u>1873</u> , Melvil Dewey publicou a primeira edição de sua Classificação Decimal, amplamente adotada	...em <u>1876</u> , Melvil Dewey publicou a primeira edição de sua Classificação Decimal, amplamente adotada
22	2º	...nas três áreas verifica-se <u>menos</u> divergência que arcabouço científico, o qual, no entanto, considera-se mais presente nos princípios modernos da Documentação que nas tradições bibliotecárias.	...nas três áreas verifica-se <u>mais</u> divergência que arcabouço científico, o qual, no entanto, considera-se mais presente nos princípios modernos da Documentação que nas tradições bibliotecárias.
26	4º	...leva, de fato, a um recrudescimento da passividade e da <u>obsolescência</u> nas bibliotecas.	...leva, de fato, a um recrudescimento da passividade e da <u>obsolescência</u> nas bibliotecas.
30	1º	...base BIDAC, cuja sigla indica as palavras Biblioteconomia, Informática Documentária, Documentação, Arquivo e Ciência da <u>informação</u>	...base BIDAC, cuja sigla indica as palavras Biblioteconomia, Informática Documentária, Documentação, Arquivo e Ciência da <u>Informação</u>
35	2ª linha	O quarto e atual estágio tem como objetivo realizar a <u>interoperacionalidade</u> entre sistemas de informações	O quarto e atual estágio tem como objetivo realizar a <u>interoperabilidade</u> entre sistemas de informações
36	2º	O termo “empréstimo entre bibliotecas” adotado inicialmente em inglês como <i>interlibrary loan</i> e <i>interlending</i> está sendo substituído por “ <u>entrega</u> ” de documentos” ( <i>document delivery</i> ).	O termo “empréstimo entre bibliotecas” adotado inicialmente em inglês como <i>interlibrary loan</i> e <i>interlending</i> está sendo substituído por <u>fornecimento</u> de documentos” ( <i>document delivery</i> ).
36, 37, 91, 180		... <u>interoperacionalidade</u> ...	... <u>interoperabilidade</u> ...

## ERRATA (cont.)

página	parágrafo	onde se lê	leia-se
62	1º	...ou programas de <u>entrega</u> de documentos ( <i>document delivery</i> ).	...ou programas de <u>fornecimento</u> de documentos ( <i>document delivery</i> ).
72		<u>entrega</u> de documentos ( <i>document delivery</i> )	<u>fornecimento</u> de documentos ( <i>document delivery</i> )
73, 76, 77, 79		... <u>entrega</u> ...	... <u>fornecimento</u> ...
181	5º	Gestão das entradas ( <u>catálogo</u> , análise, indexação);	gestão das entradas ( <u>catalogação</u> , análise, indexação);
184	10º	Todos <u>correm</u> em ambiente UNIX.	Todos <u>operam</u> em ambiente UNIX.
213	1º	...também foram os documentalistas europeus os pioneiros no reconhecimento do <u>documento como o suporte (impresso, sonoro, eletrônico etc.) que contém a informação.</u>	...também foram os documentalistas europeus os pioneiros no reconhecimento do <u>uso do termo “documento” como a base material da informação, independente do tipo de suporte (impresso, eletrônico) ou de informação (bibliográfica, cadastral, numérica, audiovisual).</u>

## ERRATA (cont.)

### Substituir *abstract* (p. iv) por:

ORTEGA, Cristina Dotta. “**Informática documentária**”: state of art. São Paulo, 2002. Masters in Communication – School of Communication and Arts, São Paulo University.

Due to the rapid technological development, of information recognition as object of study and work of Librarianship and Documentation Sciences and the process of turning these fields into Information Science, it becomes necessary to reorganize knowledge. Being that the case, this research has focused on the subject of *Informática Documentária* being a field that responds to the application of computerization and telecommunication in the various phases of production, processing, storing, retrieval and dissemination of the documentary information. In order to place the subject in context, it is shown a conceptual and historical chart of Librarianship, Documentation and of the Information Science. Following it, so that one can systematize theoretic and methodologic aspects of the *Informática Documentária*, there is a terminological and content research on the schools (production in the English, Spanish and French languages) and an analysis of the markets resulted by these schools. The adopted of research is the French model represented by the term *Informatique Documentaire*. The comparisons aim, mainly if the Brazilian reality is considered, to the distinct between automation of libraries services and *Informática Documentária*, being this one marked by the information accomplishment of the indexing process and the thesaurus management. Historical and conceptual aspects are presented in order to explain the limits between the French and North American models. It discusses the real meaning of the librarianship activity from its actual insertion on the activities related to the production of documents/ documentary informations. It concludes that the terminological diversity and reference models aim to the primacy of the North American model, which is evaluated as fragmented, needing to be revised so that the current challenges are appreciated. Finally it is emphasized that the appropriation of models in a partial way and out of context by cultures like the Brazilian one, results in the lack of built knowledge and explains to a certain level the low social recognition of the library professional, insatisfactory attention to the poor care for the demand and the difficulty of the market to filling positions.

Folha de aprovação

Banca Examinadora

---

---

---

São Paulo, de de 2002.

## Agradecimentos

À FAPESP pela bolsa de mestrado concedida.

À Prof. Dra. Maria de Fátima G. M. Tálamo, pela indicação de rumos,  
pelo rigor e pelo estímulo.

À Prof. Dra. Nair Yumiko Kobashi, pelas dicas constantes e pelo apoio.

À Virgínia, pela diagramação da pesquisa e pela amizade.

Ao Glauco, pela companhia, sempre.

## RESUMO

ORTEGA, Cristina Dotta. **Informática documentária**: estado da arte. São Paulo, 2002. Dissertação (Mestrado em Ciências da Comunicação) – Escola de Comunicações e Artes, Universidade de São Paulo.

Em função do acelerado desenvolvimento tecnológico, do reconhecimento da informação como objeto de trabalho e de estudo da Biblioteconomia e da Documentação e do processo de "cientifização" destas áreas pela Ciência da Informação, faz-se necessário reorganizar o conhecimento. Sendo assim, a pesquisa centrou-se no tema da Informática Documentária, enquanto área que responde pela aplicação da Informática e das telecomunicações às diversas fases de produção, tratamento, armazenamento, recuperação e disseminação das informações documentárias. Para contextualizar o tema, é apresentado um quadro conceitual e histórico da Biblioteconomia, da Documentação e da Ciência da Informação. Em seguida, com o fim de efetivar uma sistematização dos aspectos teóricos e metodológicos da Informática Documentária, realiza-se uma investigação terminológica e de conteúdo das vertentes encontradas (produção nos idiomas inglês, espanhol e francês) e uma análise dos mercados decorrentes destas vertentes. A linha adotada para a pesquisa é a francesa representada pelo termo *Informatique Documentaire*. As diversas comparações realizadas apontam, principalmente se considerada a realidade brasileira, para a distinção entre automação dos serviços de bibliotecas e Informática Documentária, sendo esta marcada pela concretização informatizada do processo de indexação e de gerenciamento de tesouros. São apresentados aspectos históricos e conceituais que procuram explicar a delimitação entre os modelos francês e estadunidense. Discorre-se sobre a re-significação da atividade bibliotecária a partir de sua atual inserção nas atividades de produção de documentos/ informações documentárias. Conclui que a diversidade terminológica e de modelos de referência aponta para a primazia do modelo estadunidense, o qual é avaliado como fragmentado e funcionalista, necessitando ser revisto para que os desafios atuais sejam contemplados. Finalmente, ressalta-se que a apropriação de modelos de forma parcial e descontextualizada por culturas como a brasileira, deflagra ausência de acúmulo de conhecimento e explica, em certa medida, o baixo reconhecimento social do profissional bibliotecário, o atendimento insatisfatório das demandas e a dificuldade do mercado no preenchimento do quadro de ofertas.

Palavras-chave: Informática Documentária; Automação de Bibliotecas; Recuperação da Informação; Biblioteconomia; Documentação; Ciência da Informação

## ABSTRACT

ORTEGA, Cristina Dotta. **Informática documentária**: estado da arte. [“Informática documentária”: state of the art]. São Paulo, 2002. Dissertação (Mestrado em Ciências da Comunicação) – Escola de Comunicações e Artes, Universidade de São Paulo.

Because of the quick technological progress, of the recognition of information as object of Librarianship and Documentation work and study, and the process of turning these fields into sciences by the Information Science, it becomes necessary to reorganize the knowledge. Being that the case, the research has focused on the subject of *Informática Documentária* while it is a field that responds for the application of computerization and telecommunication to the various phases of production, processing, storing, retrieval and dissemination of the documentary information. In order to contextualize the subject, it is shown a conceptual and historical chart of the Librarianship, of the Documentation and of the Information Science. Following it, so that one can systematize theoretic and methodologic aspects of the *Informática Documentária*, there is a terminological and content research on the schools found (production in the English, Spanish and French languages) and an analysis of the markets resulted by these schools. The adopted line for the research is the French one represented by the term *Informatique Documentaire*. The several comparisons aim, mainly if the Brazilian reality is considered, to the distinction between automation of the libraries services and *Informática Documentária*, being this one marked by the information accomplishment of the indexing process and the thesaurus management. The historical and conceptual aspects are shown in order to explain the limits between the French and North American models. It discourses about the real meaning of the library activity from its actual insertion on the activities related to the production of documents/ documentary informations. It concludes that the terminological diversity and reference models aim to the primacy of the North American model, which is evaluated as fragmented and functional, needing to be revised so that the current challenges are appreciated. Finally it is emphasized that the appropriation of models in a partial way and out of context by cultures like the Brazilian one, sparks the lack of built knowledge and explains in a certain level the low social recognition of the professional librarian, the poor care for the demand and the difficulties of the market in filling up posts.

Keywords: *Informática Documentária*; Library Automation; Information Retrieval; Librarianship; Documentation; Information Science

## RESUMÉ

ORTEGA, Cristina Dotta. **Informatique documentaire**: l'état de l'art. São Paulo, 2002. Dissertação (Mestrado em Ciências da Comunicação) – Escola de Comunicações e Artes, Universidade de São Paulo.

À cause de l'hâtif développement technologique, de la reconnaissance de l'information comme le sujet de travail et de l'étude de la Bibliothéconomie et de la Documentation, et du procès de la "scientifisation" de toutes ces branches par les Sciences de l'Information, il faudrait donc organiser toute la connaissance. Par conséquent, cette recherche-ci se plonge dans la branche de l'Informatique Documentaire parce qu'elle répond, par l'application de l'informatique et de la télécommunication, aux divers étapes de la production, du traitement, de l'stockage, de la récupération et de la dissémination des informations documentaires. Pour mettre en contexte ce sujet, on présente un cadre conceptuel et historique de la Bibliothéconomie, de la Documentation et des Sciences de l'Information. Ensuite dans le but d'effectuer une systematisation des aspects théoriques et méthodologiques de l'Informatique Documentaire, on réalise une recherche terminologique et des matières des branches trouvées (de la production en anglais, en espagnol et en français), et on fait aussi une analyse des marchés resultés des ces branches. Le principe adopté par la recherche suit la branche française reflétée par l'expression Informatique Documentaire. Les comparaisons réalisées signalent, surtout par rapport la réalité brésilienne, la distinction entre l'automatisation des services des bibliothèques et l'Informatique Documentaire, soyant celle-ci indiquée par la matérialisation informatisée du procès de l'indexation et de la gestion du thésaurus. On présente des aspects historiques et conceptuels qui chercent à expliquer la délimitation entre les modèles français et étatsuniens. On examine la re-signification de l'activité du bibliothécaire dès son actuel insertion dans les activités de production des documents/informations documentaires. On en conclue que la diversité terminologique et des modèles de référence signalent la primauté du modèle étatsunien, lequel est jugé comme fragmentaire et foctionnaliste, donc il faut qu'il soit réévalué pour satisfaire aux actuels défis. À la fin, on en détache que l'appropriation partiel des modèles et aussi hors du contexte par des cultures ainsi la brésilienne déclenche la manque de l'accroissement, et ceci explique de cette façon la petite reconnaissance social du profissionel de la bibliothèque, la réponse insatisfaisante des exigences et la difficulté de remplir le marché de travail.

Mots-clés: Infomatique Documentaire; Informatisation des Bibliothèques; Recupération de l'Information (Recherche Documentaire); Bibliothéconomie; Documentation; Sciences de l'Information.

# SUMÁRIO

Folha de aprovação.....	i
Agradecimentos.....	ii
Resumo.....	iii
Abstract.....	iv
Resumé.....	v

## INTRODUÇÃO

Objeto.....	1
Objetivos gerais.....	1
Objetivos específicos.....	1
Apresentação e justificativa.....	1
Hipóteses usuais.....	3
Hipótese alternativa.....	4
Delimitação do tema.....	6
Metodologia.....	7

## CAPÍTULO 1 - O contexto histórico da Biblioteconomia, da Documentação e da Ciência da Informação

1.1 A constituição do conhecimento e a Biblioteconomia.....	10
1.2 Breve história da Biblioteconomia, da Documentação e da Ciência da Informação.....	14
1.2.1 As primeiras bibliotecas (3000 a. C. ao século XVII).....	14
1.2.2 A Bibliografia e a Documentação (séculos XV ao XIX)...	16
1.2.3 Otlet e a Documentação (fins do século XIX e início do século XX).....	18

1.2.4 A Ciência da Informação (anos 50 em diante).....	20
1.3 Quadro atual da Biblioteconomia, da Documentação e da Ciência da Informação.....	24
<b>CAPÍTULO 2 - Informática Documentária: história, conceitos e mercados</b>	
2.1 Introdução à Informática Documentária.....	29
2.2 <i>Library Automation e Information Retrieval</i> .....	34
2.2.1 <i>Library Automation</i> .....	34
História da <i>Library Automation</i> .....	34
Sobre a história da <i>Library Automation</i> .....	37
O que é <i>Library Automation</i> : “Introduction to automation for librarians”, William Saffady, 1999.....	39
<i>Library Automation</i> , segundo William Saffady.....	53
Alguns currículos de ensino superior.....	55
“The electronic library”, Jennifer Rowley, 1998.....	57
Automação de Bibliotecas, Recuperação da Informação e bibliotecas eletrônicas, segundo Jennifer Rowley.....	77
2.2.2 <i>Information Retrieval</i> .....	80
O que é <i>Information Retrieval</i> .....	80
História da <i>Information Retrieval</i> : “The seven ages of Information Retrieval”, Michael Lesk, 1995.....	81
Estado da arte da pesquisa sobre <i>Information Retrieval</i> : a abordagem cognitiva.....	88
2.2.3 Breve comparação entre os termos <i>Library Automation e Information Retrieval</i> .....	90
2.2.4 Análise do mercado de sistemas de automação de bibliotecas e centros de documentação dos Estados Unidos..	90

2.3 <i>Informática Documental</i> .....	95
“Introducción a la Informática Documental: fundamentos teóricos, prácticos y jurídicos”, Carlos Costa Carballo, 1995.	95
Conteúdo complementar: organização curricular universitária.....	105
“Informática Jurídica Documental”, Mario Saquel, 2001.....	108
Comparação entre as abordagens para o termo <i>Informática Documental</i> .....	118
Outras referências sobre o termo <i>Informática Documental</i> ....	118
2.4 <i>Informatique Documentaire</i> .....	120
“ <i>Informatique Documentaire</i> ”, André Deweze, 1994.....	120
Comparação entre o termo <i>Informatique Documentaire</i> e os termos <i>Library Automation</i> e <i>Information Retrieval</i> .....	175
Outras referências sobre o termo <i>Informatique Documentaire</i> .....	179
2.4.1 Análise do mercado francês de sistemas de automação de bibliotecas e centros de documentação.....	182
Apontamentos sobre os mercados francês e estadunidense.	190
2.5 Tipologias de sistemas de automação para bibliotecas, centros de documentação e outros serviços de informação...	194
2.6 Comparação entre os termos <i>Informatique Documentaire</i> e <i>Informática Documental</i> .....	199
2.7 <i>Informática Documentária</i> .....	202

## CAPÍTULO 3 - Tentativas de preencher lacunas históricas

3.1 Biblioteconomia, Documentação e Ciência da Informação: abordagem especulativa.....	205
Sobre a distinção cultural estadunidense e francesa em <i>Informática Documentária</i> .....	210

3.2 Produção de documentos/ informações documentárias: apresentação/ representação do conteúdo e formas de acesso.....	212
CONSIDERAÇÕES FINAIS.....	217
Referências bibliográficas.....	231
Anexo 1 - Lista de siglas .....	I
Anexo 2 - Glossário.....	V
Anexo 3 - Bibliografia temática complementar.....	VIII

# INTRODUÇÃO

## OBJETO

O objeto da pesquisa é a Informática Documentária, área interdisciplinar responsável pela aplicação da Informática e das telecomunicações às diversas fases de produção, tratamento, armazenamento, recuperação e disseminação das informações documentárias.

## OBJETIVOS GERAIS

- Apresentar suporte para a análise do tema, a partir do desenvolvimento do quadro conceitual e histórico da Biblioteconomia, Documentação e Ciência da Informação.
- Levantar e analisar os conceitos, as metodologias e os usos que constituem a área interdisciplinar Informática Documentária.
- Apresentar sistematização conceitual e metodológica sobre Informática Documentária, com base no quadro histórico da Biblioteconomia, Documentação e Ciência da Informação e, em termos específicos, a partir dos conceitos e usos levantados sobre o termo Informática Documentária.

## OBJETIVOS ESPECÍFICOS

- Promover o discernimento entre automação dos serviços de bibliotecas e Informática Documentária.
- Analisar comparativamente as duas abordagens principais e distintas (norte-americana e francesa) sobre a área da Informática Documentária.
- Desenvolver historicamente o papel da produção de documentos e da representação do conteúdo de documentos pela atividade bibliotecária.

## APRESENTAÇÃO E JUSTIFICATIVA

A informação tomou um caráter primordial em praticamente todo o mundo, em especial nas últimas décadas, devido ao aumento da produção do conhecimento e a sua disseminação ostensiva em função da era da globalização. A Informática e as telecomunicações foram os facilitadores e propiciadores deste quadro.

Desde a implementação dos computadores de grande porte, mas principalmente com a introdução dos microcomputadores na década de 80, e da aplicação das telecomunicações, os serviços e produtos das bibliotecas e centros de documentação têm passado por significativas mudanças. Com isso, a atuação do bibliotecário apresenta novas exigências que implicam não apenas a implementação de novas ferramentas em sua rotina mas, uma nova concepção de planejamento e organização do trabalho que incorpore a noção de fluxo informacional. Acrescente-se ainda a vulgarização crescente

do acesso à informação e a ampliação e diversificação das atuações possíveis na área da informação, alterando radicalmente o universo que rodeia o bibliotecário e suas atividades e relações junto a outros profissionais e à comunidade em geral.

Questões relativas a transformações na prática do bibliotecário por conta das tecnologias eletrônicas são lugar comum na literatura e nas discussões sobre Biblioteconomia no Brasil, fornecendo ao tema um caráter de modismo e de pouca novidade. Haja vista que, já nas décadas de 70 e 80, produziu-se um número significativo de documentos sobre Informática aplicada a bibliotecas ou à Documentação, abordando estruturação e características de bases de dados bibliográficos, programas de automação de bibliotecas, lógicas de busca etc. Na década de 90, a literatura passou a enfatizar o acesso às informações via Internet, o hipertexto, a biblioteca virtual, publicações eletrônicas e formatos de registro bibliográfico com fins de intercâmbio de dados. Nos primeiros anos citados, foram abordadas, de forma embrionária, metodologias para a construção de bases de dados, mas o assunto deixou de ser desenvolvido na literatura sem que sua fundamentação teórica e metodológica tivesse se consolidado. Um dos fatores para isso está no alto índice de relatos de experiência de automação em detrimento da sistematização destas experiências, o que levaria, de fato, ao desenvolvimento de metodologias de construção de bases de dados. Passou-se dos catálogos manuais das bibliotecas, para a sua automação, e então, para a informação eletrônica, queimando-se etapas. No entanto, parece que este fato não tem ocasionado estranheza. Talvez porque, presenciamos no país a existência, não de automação de bases de dados, mas de automação de catálogos de bibliotecas, e não de informação eletrônica, mas de bibliotecas virtuais, entendidas como os catálogos das bibliotecas disponibilizados para acesso remoto. Não está em jogo a informação, seja eletrônica ou não, mas os documentos convencionais das bibliotecas, e o fato de que, a Informática disseminada e entendida no meio bibliotecário é aquela que organiza e trata o ciclo destes documentos em uma dada instituição.

Sendo assim, o impulso para esta pesquisa encontra-se na carência de uma elaboração consistente e amplamente divulgada sobre os processos do fluxo "informatizado" da informação no Brasil: é inexistente a sistematização sobre Informática Documentária (se é que este é o termo adequado para nomear esta área), que englobe o conhecimento e as atividades decorrentes dos processos automatizados de produção, processamento, recuperação e disseminação da informação. Um elemento que evidencia esta questão refere-se à superposição indiscriminada entre o significado da automação dos serviços/ atividades de bibliotecas (ou gestão automatizada de acervos) e da Informática e das telecomunicações aplicadas à efetivação do fluxo da informação, ou seja, a sua produção, tratamento, armazenamento, recuperação e disseminação. Pode-se dizer que, não está composto o conhecimento relativo aos novos conceitos e práticas advindos das tecnologias eletrônicas utilizadas nos diversos serviços de informação: o conhecimento está fragmentado e, práticas e conceitos superados convivem sem maiores questionamentos com outros ainda em elaboração.

Verifica-se, assim, uma ambigüidade: a mudança da sociedade industrial para aquela nomeada como sociedade da informação está presente na literatura e nas evidências diárias nas diversas áreas do trabalho e do lazer, em especial, nas grandes cidades. Mas não ocorreu, de forma ampla e satisfatória, a organização de conceitos e metodologias em Biblioteconomia e Documentação que contemplem tais exigências.

Constata-se aí uma lacuna entre as atividades de Informática e Biblioteconomia, onde profissionais atuam essencialmente na segunda e outros se empenham em

especial na primeira, carecendo do desenvolvimento de uma prática conjunta que aborde os conceitos de uma com os recursos da outra e vice-versa. O desenvolvimento tecnológico aplicado às atividades da Biblioteconomia vem sendo, historicamente, formulado e implementado externamente ao profissional desta área. Por outro lado, o crescente processo de “informatização” do bibliotecário desconsidera sua formação: envolve as linguagens da Informática mas não as “linguagens do conteúdo” da informação, ou seja, centra-se na ferramenta e não no seu objeto de trabalho. O desvirtuamento de seu objeto de trabalho, entre outros, leva a que, as inovações possíveis apresentadas pelos bibliotecários sejam emblemáticas da dificuldade destes com a desburocratização de suas atividades. Estes aspectos não contribuem para a concretização de atitudes e procedimentos para uma ação política e intelectualizada e, ao mesmo tempo, técnica e tecnológica.

## **HIPÓTESES USUAIS**

Acredita-se que, com o surgimento e desenvolvimento da Informática e das telecomunicações, o problema do tratamento e do acesso ao documento e à informação, estaria resolvido. Vários argumentos seguem nesta linha, quais sejam:

1 A informação contida no documento eletrônico tem sua recuperação automática, uma vez que todos os dados estão disponíveis integralmente para a pesquisa, prescindindo do tratamento intelectual efetivados pela catalogação e indexação dos documentos impressos. Em outros casos, algoritmos computacionais dados por lógicas matemáticas de busca respondem, de forma rápida e eficiente, a qualquer demanda informacional a partir de um conjunto de dados ou textos. De uma forma ou de outra, os recursos dos computadores são capazes de substituir automaticamente as lógicas de registro/ indexação e busca da informação desenvolvidas por bibliotecários e documentalistas há séculos.

2 Com o desenvolvimento e disponibilização de interfaces amigáveis de pesquisa ao usuário final, os profissionais intermediários no auxílio às pesquisas, como os bibliotecários, estão dispensados desta função. Sendo a Internet, uma mídia que unifica e dá rapidez e agilidade ao acesso a informações, futuramente, todos os tipos de informação estarão digitalizados. Desta forma, a biblioteca virtual, além de ser eficiente, rápida e democrática, substituirá a necessidade do espaço físico e real da biblioteca presencial, de difícil e cara administração e preservação.

3 Os bibliotecários, por sua formação de cunho operacional-técnico, não têm condições de executar, com competência, e portanto não lhes cabem, as atividades: de planejamento e gerenciamento de serviços de informação; de tratamento, busca, divulgação e promoção da assimilação do conhecimento em áreas especializadas ou com fins educacionais e culturais; e de desenvolvimento parcial ou integral de sistemas de informação em diversas linguagens computacionais e suportes. Por suas atividades de catalogação e gestão de acervos, os bibliotecários são considerados os responsáveis pela manutenção e preservação do patrimônio cultural da humanidade, o que lhes confere o título de agentes educacionais e culturais.

## HIPÓTESE ALTERNATIVA

A reformulação do significado e das atividades de tratamento, armazenamento, disseminação e promoção da assimilação da informação faz-se urgente, pois parece existir uma expectativa de que a Informática por si só, isto é, exclusivamente, possibilite o acesso a qualquer informação. No entanto, as tecnologias eletrônicas determinam novas formas individuais e sociais de pensar e agir, mas também são determinadas por estas formas de pensamento e ação.

A Biblioteconomia e a Documentação encontram-se inseridas em diversas linhas de pesquisa acadêmicas, muitas delas inscritas nos pressupostos teóricos da Comunicação. No entanto, esta inserção parece não ter se concretizado de fato. Além disso, como a Comunicação está intrinsecamente ligada ao desenvolvimento das tecnologias, a Biblioteconomia e a Documentação caminharão isoladas, inférteis e obsoletas, caso não se reformulem a partir dos aspectos comunicacionais.

Acredita-se que, com o aprimoramento do quadro conceitual destas áreas a partir dos referenciais teóricos da Ciência da informação, será possível, por meio das atividades de ensino, transferir o conhecimento que dê conta da atividade bibliotecária independente de uma ou outra inovação tecnológica em curso: a idéia essencial já estará traçada.

1 As lógicas biblioteconômicas tradicionais desenvolvidas para a organização de acervos foram funcionais para o que se propunham, em especial, considerando o contexto em que tiveram início: dadas as tecnologias calcadas no papel e na imprensa e a pequena produção bibliográfica, as soluções encontradas e disseminadas mundialmente apresentavam-se relativamente adequadas. Nas últimas décadas, o grande aumento da produção do conhecimento especializado e o decorrente desenvolvimento tecnológico exigiram, mais uma vez, novas e significativas mudanças das técnicas de organização da informação. Algumas destas técnicas já eram empregadas, de forma embrionária, há 5000 anos atrás, no período do surgimento da escrita propriamente dita. Neste sentido, o conhecido “ranço” bibliotecário pode ser parcialmente justificado pelo simples e óbvio motivo de que sua atividade data de milênios. Acrescente-se a isso o fato de que, a Biblioteconomia é intrinsecamente prática desde a sua origem, desenvolvendo-se pouco teoricamente e constituída, em grande parte, por profissionais em busca de garantia de sobrevivência e, ou da crença na proximidade com o universo cultural.

Apesar das inadequações e deficiências observadas na prática, as técnicas desenvolvidas até então para o tratamento e busca da informação pelos bibliotecários são complementares àquelas desenvolvidas pela Informática e respondem aos mesmos objetivos e anseios desta, no que se refere à garantia do fluxo informacional e do acesso à informação, seja para fins científicos, técnicos, educacionais ou culturais. De fato, estão nos conceitos e metodologias biblioteconômicas, algumas lógicas para o desenvolvimento das ferramentas informáticas e alternativas para os pontos nos quais a Informática ainda não conseguiu encontrar soluções.

2 Sabemos que os estoques de informação disponibilizados na Internet não possibilitam garantia de acesso, mesmo que devidamente organizados, pois a alfabetização tecnológica e o uso democrático de computadores interligados ainda são

distantes, em especial, nos países em desenvolvimento. Também o acesso rápido e eficiente não é evidente pois, para tal, deve fornecer informação pertinente e relevante, além de razoável em termos da relação custo-benefício do tempo para conexão, busca, seleção e aquisição.

Não se sustenta a hipótese de que a informação em papel e as bibliotecas presenciais estariam em extinção, pois a história tem mostrado que as tecnologias vêm surgindo e se desenvolvendo sem implicar, necessariamente, o desaparecimento das anteriores, mas em convivência que decorre no seu aprimoramento e complementação. Além disso, o papel e a imprensa representam um lugar altamente significativo no registro do pensamento linear, ao possibilitar o desenvolvimento do conhecimento e da própria noção de história que temos hoje, inexistente antes da escrita nas sociedades orais. Sendo os livros e as publicações impressas em geral, produtos seculares, ocupam lugares e valores incontestáveis em nossa sociedade.

O outro ponto, talvez o mais importante, refere-se à capacidade de compreensão e assimilação da informação pelo indivíduo, seja para fins utilitários, profissionais ou de pesquisa científica. Em tempos de banalização dos valores ligados à cultura e ao lazer, dados pela massificação dos meios de comunicação, afeitos a outros interesses que não o de comunicar informações relevantes à sociedade, o quadro apresenta-se perverso. Justamente nesta linha verifica-se que a oferta de informações de ordem utilitária e de ordem científica e tecnológica, disponível a partir de qualquer meio de acesso, é inversamente proporcional àquela demandada por cada uma delas.

3 Por outro lado, a referência ao bibliotecário como agente educacional e cultural encontra obstáculo na falta de determinação clara do objeto da Biblioteconomia. Definir este objeto supõe, por exemplo, identificar o bibliotecário como especialista em tratamento e recuperação de informação médica, mas não como especialista em Medicina. Ou ainda, implica a discriminação do bibliotecário como aquele que domina e promove o ensino-aprendizagem da busca e seleção de informações para a assimilação de conteúdos, para a produção de textos e outras formas de expressão e para a formação de leitores (distintamente do educador de Língua Portuguesa ou outra disciplina do currículo oficial de ensino). Da mesma forma, o bibliotecário recebe impropriamente o papel de guardião da memória cultural da humanidade pois, se ele “guarda” algo, trata-se somente da informação registrada, uma vez que, para manter a cultura deve-se, ao menos, transformá-la em fluxo informacional. Esta argumentação caminha no sentido da informação como objeto da Biblioteconomia.

Parte da nebulosidade do exposto acima pode ser explicado pela tentativa de humanização da Biblioteconomia ao diminuir sua importância técnica com vistas a dar conta das demandas educacionais e culturais da sociedade. No entanto, esta área apresenta uma base pragmática que a distingue de outras essencialmente teóricas, como as Ciências Sociais ou a Lingüística. Isto porque, para realmente atender a demandas educacionais e culturais de informação, há que se definir quais são os mecanismos que, por exemplo, em uma biblioteca pública, contemplem estas demandas. Alguns destes mecanismos exigem o domínio de habilidades gerenciais, técnicas e tecnológicas: se não houvesse estas exigências, um profissional de outra formação em Ciências Humanas estaria muito mais apto a esta atividade.

Sendo assim, a dificuldade dos bibliotecários em integrar-se no universo do fluxo “informatizado” da informação (ou de fazê-lo de forma desvirtuada), da resistência frente à conquista de um papel profissional intelectual e político significativo na instituição em que atua, além da baixa produção acadêmica, não vêm colaborando na reformulação desta profissão e área do conhecimento. Só com a constituição de uma linguagem própria, a área poderá contemplar necessidades e angústias de dois grupos da sociedade: bibliotecários e usuários da informação. Como consequência, profissionais e pesquisadores estarão preparados para responder às necessidades de informação da sociedade em ciência, tecnologia, educação e cultura, criando ou reformulando serviços de informação e elaborando e implantando estratégias e metas que se configurem em uma política nacional de informação.

## **DELIMITAÇÃO DO TEMA**

A sistematização do tema Informática Documentária aqui proposta está focada na sua interface com a Biblioteconomia e a Documentação e como disciplina constituinte da Ciência da Informação. Esta opção epistemológica está baseada na possibilidade de contribuição para a reformulação do escopo teórico e metodológico pertinente à formação inicial e continuada em Biblioteconomia e Documentação. Isto implica que não estão sendo considerados, por exemplo, a formação desejável a um profissional de Informática ou Ciência da Computação que dê conta das atividades de Informática Documentária, visto tratar-se de uma possibilidade. São contempladas essencialmente as características necessárias à formação e atuação de profissionais bibliotecários face às mudanças comunicacionais e tecnológicas e as conseqüentes demandas diferenciadas por informação, respaldadas pelos avanços em Ciência da Informação.

Como introdução ao tema propriamente dito, discorre-se brevemente sobre a história das áreas conhecidas por Biblioteconomia, Documentação e Ciência da Informação. Esta introdução é essencial para a contextualização da pesquisa pois foi encontrada na Documentação e na Ciência da Informação a origem e o desenvolvimento da área nomeada como Informática Documentária. No Brasil, a literatura e a atuação profissional centram-se na Biblioteconomia, apesar das diversas e crescentes demandas por serviços de informação pautados na aplicação das técnicas documentárias. Mesmo considerando relativa apropriação destas técnicas pela Biblioteconomia, além da tendência da aproximação destas áreas, constata-se um descompasso conceitual e prático, a partir do qual não é possível discorrer metodologicamente. Faz-se necessário e essencial, portanto, discernir sobre estes conceitos como fundamentação e ponto de partida.

Sendo assim, um dos aspectos investigados está na dicotomia evidenciada entre automação de serviços/ atividades de bibliotecas e a aplicação da Informática e das telecomunicações na produção, tratamento, armazenamento, recuperação e disseminação de informações. No primeiro caso, dá-se a gestão de acervos, e no segundo, a gestão informacional propriamente dita, as quais não são excludentes. Em termos específicos, a pesquisa concentra-se na questão de como se dá a distinção entre as atividades de estocagem e circulação de itens de acervo e os processos de organização e transferência de informações.

De fato, o tema apresenta-se difuso na literatura: a Informática Documentária contempla a Informática aplicada à Documentação e à Biblioteconomia? Se a Ciência da Informação proporciona a superação e a “cientificização” de ambas, então a Informática

Documentária é constituinte desta nova ciência? É relevante considerar que os conceitos para os termos Biblioteconomia e Documentação já não se referem na atualidade a sua definição original. No entanto, a prática da Biblioteconomia está arraigada em tradicionalismos inadequados para os tempos atuais e a Documentação apresenta apenas alguns aspectos que poderiam garantir essa modernização. Já a Ciência da Informação parece encontrar-se ainda em estágio embrionário para fornecer a fundamentação necessária às questões teóricas e metodológicas destas áreas.

Em suma, as três definições para os termos-chave da área indicam baixo comprometimento conceitual e ações pouco refletidas, deflagrando uma realidade caótica que merece atenção de profissionais e pesquisadores. É neste emaranhado conceitual e viciado por uma atuação muitas vezes engeguecida pela rotina, que se delineiam as noções sobre Informática Documentária.

## **METODOLOGIA**

Estudo teórico calcado inicialmente na contextualização do tema a partir do histórico da Biblioteconomia, da Documentação e da Ciência da Informação. Para tanto, apresenta breve relato histórico destas áreas e sua relação com o desenvolvimento e formalização do conhecimento humano no contexto dos movimentos socioculturais, localizando nestes movimentos cada um dos três conceitos como suporte referencial e teórico ao tema central da pesquisa.

Desta contextualização parte-se para a investigação do uso do termo Informática Documentária em outros países, uma vez que no Brasil o termo é incipiente e a sua utilização refere-se, salvo exceções, a questões genéricas sobre o bibliotecário frente às novas tecnologias do mercado, a relatos de experiência de automação de bibliotecas e, crescentemente, sobre as metodologias e sistemas informatizados envolvidos nestes processos de automação, assim como, questões sobre informação eletrônica e sua disponibilização. Realiza-se o levantamento dos conceitos e metodologias que envolvem o termo Informática Documentária: no idioma francês, onde tem sua origem; no idioma inglês por meio de termos equivalentes; e no idioma espanhol.

Sendo assim, foi verificado que, nos Estados Unidos e no Reino Unido, o tema é tratado pelos termos *Library Automation* ou *Information Retrieval*, abordando a aplicação de recursos de Informática à gestão de bibliotecas, no primeiro caso, e o estudo da recuperação de informação segundo ocorrência e frequência de dados, no segundo. Os termos na língua inglesa são analisados menos por compartilhamento conceitual que pela significativa influência que exercem em termos teóricos e práticos em todo o mundo.

No entanto, como não contemplam as expectativas mais abrangentes desta pesquisa, o referencial teórico segue o conceito embutido na expressão francesa *Informatique Documentaire*. Também é considerado o termo no idioma espanhol *Informática Documental*, cuja apropriação está baseada essencialmente na linha francesa.

Complementando a revisão teórica, são apresentados currículos de ensino superior e análise de mercados de *softwares*. Os currículos universitários de cursos sobre o tema em diversos países são utilizados como auxílio ao entendimento das formas de organização do conhecimento de uma dada cultura. As análises do mercado de sistemas de automação de bibliotecas e centros de documentação demonstram os modos de

concretização do conhecimento sobre organização, armazenamento e recuperação da informação.

As conclusões decorrem das análises comparativas entre as vertentes da Informática Documentária, que por sua vez, são consideradas com base no contexto histórico e conceitual apresentado sobre Biblioteconomia, Documentação e Ciência da Informação.

Foram utilizadas as seguintes fontes de pesquisa referentes ao período de 1960 a 2002:

Bases de dados:

- ISA (Information Science Abstracts);
- Infobila (UNAM - Universidad Nacional Autónoma de México);
- Biblioinfo e BIDAC (UDESC);
- DEDALUS (USP);
- Bases de dados sobre Ciência da Informação disponíveis no Dialog.

Bibliotecas:

- Biblioteca da Escola de Comunicações e Artes (ECA), da USP;
- Biblioteca da Universidade Federal do Rio de Janeiro (UFRJ);
- Biblioteca da Pontifícia Universidade Católica de Campinas (PUCAMP).

Revistas científicas:

- Ciência da Informação (Brasília);
- Transinformação (Campinas);
- Revista da Escola de Biblioteconomia da UFMG (Belo Horizonte);
- Revista de Biblioteconomia de Brasília;
- Revista Brasileira de Biblioteconomia e Documentação (São Paulo);
- Documentaliste : Sciences de l'Information;
- Documentación de las Ciencias de la Información (Madrid);
- Caderno Mais! – Folha de São Paulo.

Sites:

- Association des Professionnels de l'Information e de la Documentation (ADBS) (<http://www.adbs.fr>);
- *Sites* de universidades nacionais e internacionais com programação e bibliografia dos cursos sobre o tema da pesquisa;
- *Sites* de empresas estrangeiras e de especialistas na prestação de serviços de Informática Documentária e afins.

Outros:

- Anais e *papers* apresentados em eventos brasileiros de Biblioteconomia e Documentação;
- Literatura do idioma francês e espanhol sobre Informática Documentária;
- Literatura do idioma inglês sobre automação de bibliotecas e recuperação da informação;
- Teses e dissertações produzidas no Brasil sobre Informática e Biblioteconomia, Documentação e Ciência da informação;
- Bibliografia e material de aula das disciplinas da Graduação em Biblioteconomia e Documentação e do curso de Pós-Graduação;
- Livros e artigos localizados por meio de contatos com docentes e pesquisadores do país e do exterior.

Termos utilizados para pesquisa:

- *informática, informática documentária, automação de bibliotecas;*
- *informatique, informatique documentaire;*
- *library automation, computer for libraries, library technology, information retrieval, information science;*
- *informática, informática documental.*

# **CAPÍTULO 1 - CONTEXTO HISTÓRICO DA BIBLIOTECONOMIA, DA DOCUMENTAÇÃO E DA CIÊNCIA DA INFORMAÇÃO**

## **1.1 A CONSTITUIÇÃO DO CONHECIMENTO E A BIBLIOTECONOMIA**

Para traçar a história do registro do conhecimento e das idéias que o promoveram, necessário se faz discorrer sobre a constituição do próprio conhecimento no decorrer do tempo. O registro do conhecimento deu início à Biblioteconomia e as formas como se desenrolou este processo deflagram e são deflagradas pelos aspectos desta dimensão histórica.

A primeira segmentação ou especialização dos três domínios antes únicos, a arte, a ciência e a moral, e mais tarde, a lei e a política, marcaram a modernidade iniciada no século XVI. Esta divisão foi, de fato, a fragmentação da religião em domínios distintos e cada vez mais distantes dela, onde cada fragmento passou a ser especialidade/ propriedade de determinado grupo. Com isso, paulatinamente foi se traçando a separação entre os que contribuíam ativamente para o desenvolvimento de algum conhecimento ou forma de expressão, e o povo de uma forma geral, que ficou distante deste movimento (COELHO NETO, 1995).

Já neste momento, a atividade ou ação prática da organização de livros e outros documentos, também ficou marginalizada, distante do universo prestigiado da especialização/ constituição do conhecimento. Interessante notar, no entanto, que se esta atividade, depois denominada Biblioteconomia, sempre esteve próxima da produção decorrente desta especialização, ela própria não se tornou área de conhecimento, nem tirou frutos para si desta condição de proximidade.

O projeto da modernidade concretizou-se no século XIX, trazendo consigo algumas noções como: a dialética, entendida como algo que se define em oposição a outra coisa e não por comparação a si mesmo; o relativismo ditando que nada mais é absoluto; e o contexto como elemento chave caracterizador daquilo que está em questão. Em fins do mesmo século, surge o Positivismo que, apesar de influenciar o pensamento da modernidade, baseava-se em suas origens na explicação dos fatos pela observação empírica sem questionar as suas causas, não sendo consideradas a dialética, o relativismo e o contexto. Distante do projeto da modernidade, esta vertente positivista, marcada pelo funcionalismo, influenciou e marcou alguns princípios norteadores da Biblioteconomia.

Coelho Neto (1995, p. 18-19) fez equivalência do moderno, entendido como a consciência neurotizada da modernidade, com a novidade, a qual define como a representação neurótica do novo. A partir desta comparação cita que, uma época não se pensa tanto como modernidade quanto como moderna, e o que ela entende por moderno é mais a novidade do que o novo, pois o quadro da modernidade configura-se pela ação refletida.

Sendo assim, a não "cientifização" da prática biblioteconômica gerou: a permanência da tradição, menos por reflexão que por sua ausência, pois não se buscou nem a novidade nem o novo; ou a instauração da novidade por adoção impensada, seja como pretensão ao moderno (como mera negação do antigo), seja pelo ensejo de inclusão no modernismo, no sentido da lógica da fabricação, que se fundamenta na produção e no consumo massificados.

Se a atuação biblioteconômica fez por permanecer a tradição ou adotar a novidade, em ambos os casos, não tem refletido o pensamento da modernidade.

Apesar disto, a globalização e o desenvolvimento da tecnologia vêm alterando este anacronismo, ao dispor universalmente o projeto da modernidade. Torna-se então relevante contextualizar a Biblioteconomia nesta discussão, uma vez que a crítica à sua concepção e atuação tradicionais é tão recorrente e superficial quanto despojada de referenciais teóricos.

Nesse sentido, Rodrigues (1994) apresenta as categorias tradicional e moderno como representações de mundo (e não como tempos históricos). Considera que o moderno é sempre desejado em oposição à tradição pois esta não contribui para o domínio da linguagem, o mundo da cultura e a esfera da consciência individual. E ainda, que o moderno procura restabelecer a objetividade da experiência primeira, rompendo com uma herança que, embotada pelo hábito e pela rotina, perdeu o seu sentido originário. É sob esta abordagem que a Biblioteconomia pode ser classificada como tradicional, pois o hábito e a rotina prevalecem e não são capazes de indicar sua razão de ser.

Da mesma forma, a tecnicidade tradicional ou o saber prático tradicional é um saber de natureza mágica onde o homem não concebe a autonomia do mundo da linguagem em relação ao mundo das coisas naturais. Para a tradição, é a evocação e a invocação de um mesmo fundo coletivo herdado e aceite em função de sua autoridade que serve de fundamento legitimador dos saberes, comportamentos e discursos. Já no pensamento moderno, o indivíduo é essencial como elemento legitimador (RODRIGUES, 1994, p. 64). Na Biblioteconomia, a desconsideração da linguagem na organização do conhecimento produzido, se dá por conta do uso "mágico" dos instrumentos de trabalho, por mais que, em sua concepção, estes instrumentos contemplaram a natureza lingüística do conteúdo dos documentos. Parece que essa transposição ainda não foi satisfatoriamente percebida, pois verdades superiores previamente dadas são adotadas sem contestação ou questionamento, mesmo quando se trata de elementos com traços de modernidade, como a Informática e os instrumentos de classificação.

Esta noção de tradição em que os conhecimentos práticos são passados de geração a geração, num processo natural, fornece elementos para a compreensão da Biblioteconomia, no aspecto que se refere ao seu apego à atividade essencialmente operacional (registro e preservação), em detrimento do desenvolvimento intelectual e político.

Por sua vez, Rodrigues (1994, p. 57) conceitua o termo tradicionalismo como a modalidade patológica da tradição, que consiste na recusa de qualquer inovação e apego absoluto ao passado, predominante nos períodos em que valores da tradição estejam em crise. Entende-se que tradicionalismos ainda estejam presentes na Biblioteconomia, mas são aspectos em extinção.

Com os primeiros registros das narrativas míticas, estava iniciada a substituição do mito (essencialmente oral e circular) pela razão humana (dada pela linearidade possibilitada pela escrita). A partir daí, a organização e preservação dos escritos passou a ser realizada, acompanhada por formas de organização que foram sendo criadas e aprimoradas no decorrer do tempo. No entanto, é paradoxal constatar novamente a proximidade desta área com o desenvolvimento do pensamento e da cultura, sem envolvimento com os mesmos e sem que ela própria viesse a constituir pensamento e cultura próprios. Esta proximidade, de fato, foi tão evidente e ampla quanto meramente formal.

Nesse sentido, desprovida de maiores indícios da modernidade, a tradição biblioteconômica não reconhece as diferenças, vê o mundo de forma única, mítica e cíclica, sem oposição entre sujeito e objeto, ou sujeito e realidade. Ela transmite miticamente seu saber, sem distinguir as unidades com as quais trabalha. E não houve ainda objetividade suficiente (e eficiente) que promovesse o rompimento com essa tradição em termos significativos.

Ainda segundo Rodrigues (1994, p. 71), considerando que a história moderna tenha sido, de fato, a história dos fortes e dos poderosos em detrimento dos mais fracos, surgiu o movimento denominado pós-moderno, caracterizado pela ruptura perene frente à indecisão de tantos caminhos a tomar. A pós-modernidade pode ser definida pelo tradicional atualizado, ou ainda, pelo atual sem prescindir do tradicional.

A tradição, sustentada pelo senso comum, funciona pela indistinção entre sujeito e objeto. Na modernidade, dá-se o reconhecimento do objeto pelo sujeito, ou seja, o sujeito conhece e percebe a realidade. Já na pós-modernidade, sujeito e objeto (ou realidade) estão em constante interação e transformação mútua. Como já foi dito, a Biblioteconomia está significativamente calcada na tradição, apresenta alguns traços da modernidade (sistemas de classificação, estratégias de recuperação da informação), mas está distante da relação dialógica do projeto pós-moderno. Talvez a perpetuação da imagem que vincula bibliotecas e Idade Média tenha se originado na atribuição de guardiãs do patrimônio cultural da humanidade. Quem guarda não precisa inovar, apenas conhecer superficialmente a fôrma nova que deve guardar o conteúdo produzido por outrem. Também não precisa produzir visando a demandas externas: basta organizar e preservar. Ou seja, guardar e aguardar.

Por outro lado, as técnicas da Documentação implementaram alguns aspectos modernos, como o tratamento do conteúdo dos documentos e a mecanização da pesquisa documentária, mas ainda perduram e prevalecem princípios tradicionais da Biblioteconomia. Já a Ciência da Informação sofre para constituir-se ciência de fato, devido à fragmentação decorrente de sua própria interdisciplinaridade e, entre outros, por conta do anacronismo e do pragmatismo da Biblioteconomia.

Para o entendimento do processo de constituição de uma área, interessa reconhecer o conceito de epistemologia de uma ciência. Segundo Santos (1989, p. 20), epistemologia é a: “ciência que estuda os conhecimentos científicos, formulando as leis da produção e transformação dos conceitos de cada disciplina”.

As origens da epistemologia remontam à fragmentação da religião, quando o estudo sobre a natureza, as origens e as características da ciência era terreno exclusivo da filosofia. No entanto, acreditava-se que a ciência devia ser tratada sem especulações

filosóficas, ocasionando a distinção entre filosofia da ciência e epistemologia. A falta de univocidade continuou, contudo, quanto à questão da pureza científica (autonomia dos critérios epistemológicos) e as condições sociais de produção da ciência (determinantes em relação às condições teóricas). Esse último parecer encontra sustentação na constatação de que a epistemologia científica acompanha o desenvolvimento da ciência através de sua prática concreta, sensível à evolução da imagem de um mundo progressivamente conformado pela ciência e pela técnica (SANTOS, 1989, p. 26).

Assim, Santos apresenta três rupturas epistemológicas. A primeira é a ruptura da ciência com o senso comum. A segunda é a ruptura com a própria ruptura epistemológica, ou seja, a aproximação da ciência com o senso comum. Em seguida, a terceira ruptura procede a um trabalho de transformação tanto do senso comum como da ciência, criando uma forma de conhecimento que, sendo prática, não deixe de ser esclarecida e, sendo sábia, não deixe de estar democraticamente distribuída. A primeira ruptura é imprescindível para constituir a ciência, deixando o senso comum inalterado, a segunda transforma o senso comum com base na ciência, e a terceira decorrente dessa dupla ruptura, pretende um senso comum esclarecido e uma ciência prudente.

Com isso, os utilizadores desta ciência assim constituída passam a ser tanto os cidadãos quanto os cientistas e a reflexão epistemológica destina-se a aumentar a competência lingüística de ambos os grupos, assim como a comunicação entre si. Isso é possível hoje graças ao desenvolvimento tecnológico que a ciência moderna produziu, ou seja, a nova configuração do saber é a possibilidade de que o desenvolvimento tecnológico contribua para o aprofundamento da capacidade cognitiva e comunicativa (SANTOS, 1989, p. 42).

Pode-se fazer uma analogia entre as três rupturas e os conceitos de tradição, modernidade e pós-modernidade, já citados. Na primeira ruptura, o desenvolvimento da ciência pelo distanciamento com o senso comum, manteve este ainda mais ligado à tradição, como é peculiar a sua própria existência. A aproximação da ciência e do senso comum, evidenciada na segunda ruptura, permite que este observe aquela, incorrendo na distinção entre sujeito e objeto, dada na modernidade. E por fim, a terceira ruptura promove a transformação tanto da ciência quanto do senso comum, assim como, no projeto pós-moderno, sujeito e objeto interagem entre si numa relação dialógica.

Também é importante ressaltar que a segunda ruptura, verificada em especial, a partir do século XIX no pensamento ocidental, foi marcada pela tecnologia, a qual é conseqüência do desenvolvimento da ciência. A ligação remota entre tecnologia e ciência e a origem da primeira em virtude da evidência das condições de produção da segunda promovem o entendimento da dimensão histórica da tecnologia e do seu contexto de origem.

Já a terceira ruptura é o quadro da era da comunicação em que vivemos, marcada pela disseminação e pela primazia da tecnologia e da informação.

Assim, se a Biblioteconomia vinha atuando essencialmente na conservação dos documentos, os processos documentários reforçaram e implementaram a noção de exploração destes documentos conservados. Com isto, a Documentação permitiu introduzir a noção de difusão informacional à idéia já estabelecida de conservação de documentos. Foram as técnicas documentárias, calcadas na diversidade de documentos e no reconhecimento do conteúdo informacional, que possibilitaram a transferência da noção de acervo para a de acesso. Mas parece ser a Ciência da informação que, enquanto ciência pós-moderna, baseada no conceito de inter-relação entre sujeito e

objeto, acrescenta à área constituída pela Biblioteconomia e Documentação, a condição da interação entre o objeto informação e seus sujeitos produtores e consumidores. Só neste momento, estabelecida como área calcada na interação, pode ser concebida dentro dos princípios da Comunicação.

Acima, a Biblioteconomia foi caracterizada segundo seus traços tradicionais, com indícios da modernidade representados pelo advento da Documentação. Esta última que, ora faz parte da primeira, ora lhe faz oposição, colaborou para o início da distinção entre sujeito e objeto ao exigir de profissionais e usuários a interpretação das técnicas desenvolvidas a partir do reconhecimento do conteúdo intelectual dos documentos. Atualmente, Biblioteconomia e Documentação tendem a ser confundir e, mesmo que carecendo de projeto científico devido a uma precária passagem pela modernidade, buscam identidade e linguagem própria na pós-modernidade por meio da Ciência da Informação e a partir das tecnologias e da Comunicação.

## **1.2 BREVE HISTÓRIA DA BIBLIOTECONOMIA, DA DOCUMENTAÇÃO E DA CIÊNCIA DA INFORMAÇÃO**

### **1.2.1 AS PRIMEIRAS BIBLIOTECAS (3000 a. C. ao século XVII)**

A existência comprovada das primeiras coleções organizadas de documentos, ou aquilo que se poderia chamar de primeira biblioteca primitiva que se tem notícia, data de 5000 anos atrás. Trata-se da Biblioteca de Ebla, na Síria, descoberta em 1974, alterando a própria história conhecida sobre este país e o Oriente Médio no período. A coleção era composta de textos administrativos, literários e científicos, registrada em 15000 tábuas de argila, dispostas criteriosamente em estantes segundo o tema abordado. Também foram encontrados três grandes dicionários (considerados os mais antigos até então) e 15 tábuas pequenas de argila com resumos de conteúdo dos documentos. A escrita era a cuneiforme, porém não no seu idioma original (o sumério), mas numa língua desconhecida a qual se chamou eblaíta e os documentos encontrados apresentam as primeiras listas de vocabulário bilíngüe conhecidas pelo homem. Esta vem sendo considerada a origem da Biblioteconomia, antes contemplada às famosas bibliotecas gregas (SAGREDO, NUÑO, 1994).

Interessante notar que, segundo Kato (1987), data também do terceiro milênio a. C. (ou seja, há 5000 anos atrás), o início da escrita propriamente dita, como a escrita cuneiforme e os hieróglifos (respectivamente do idioma sumério e egípcio), em substituição aos sistemas pictográficos que ainda não eram representações da fala. A escrita demorou séculos para se constituir nestes sistemas de representação e seguir se aperfeiçoando até o que conhecemos hoje. Mesmo assim, pode-se dizer que a necessidade e a prática da organização de documentos nasce com a própria escrita.

Também na civilização mesopotâmica, no segundo milênio a. C., foi constatada a organização de documentos acompanhada de representações para fins de recuperação: tábuas de argila eram protegidas por espécies de envelopes nos quais estavam dispostas informações sob a forma de resumos (WITTY citado por KOBASHI, 1996).

Lemos (1998) cita outras grandes bibliotecas da Antigüidade que tiveram lugar entre os séculos VIII e VII a. C., como a de Assurbanipal, rei da Assíria. Os templos gregos também possuíam bibliotecas sendo as mais importantes aquelas referentes ao auge da cultura helênica, a partir do século IV a. C., como a que Aristóteles criou em sua escola de filosofia. Esta biblioteca teria sido o modelo para a fundação, no século III a. C., da famosa biblioteca de Alexandria, uma das maiores já conhecidas e que sobreviveu a muitos saques e desastres naturais, até ter seu fim definitivo. Manguel (1997) comenta que Alexandria, no Egito, era uma cidade excessivamente burocrática com um trânsito de documentos administrativos maior que sua capacidade para tal, o que demonstra disseminado domínio da escrita. No entanto, não se sabe se a biblioteca era reservada apenas aos eruditos ou a um público mais amplo. Quanto à constituição do acervo desta biblioteca, Ptolomeu III decretou que todos os navios que parassem em Alexandria tinham que entregar seus livros para serem copiados (após a cópia, às vezes devolviam-se os originais, às vezes as cópias).

Na Idade Média predominaram as bibliotecas ligadas a ordens religiosas tanto no Ocidente como no Oriente, as quais foram responsáveis pela preservação da antiga cultura greco-romana registrada. No entanto, na Europa, no século XIII, começaram a ser fundadas as bibliotecas das universidades, ao mesmo tempo em que surgiram os grandes colecionadores de livros entre a nobreza, cujas coleções viriam a formar o núcleo de algumas bibliotecas nacionais (LEMOS, 1998).

O período por volta do ano de 1440 é marcado pela invenção da imprensa por Gutenberg. Após vários aperfeiçoamentos e grandes somas de dinheiro, ele apresentou, na Feira de Frankfurt, uma Bíblia impressa, cuja tecnologia então desenvolvida foi basicamente a mesma até o início do século XX. Ao que consta, a tipografia foi, nos seus primeiros tempos, uma arte hermética de iniciados que prometiam segredo sob juramento por razões mais comerciais que intelectuais: era possível competir com o trabalho lento e caro dos escribas, pois os produtos impressos diferiam muito pouco dos manuscritos que tentavam imitar. Em 1462, com a tomada da Mogúncia, na Alemanha, os impressores foram dispersos pelos diversos países da Europa, permitindo a vulgarização da imprensa (MARTINS, 1996, p. 156). Esta revolução tecnológica acelerou a produção do livro, contribuindo para seu barateamento e sua distribuição, o que pode ser considerada a primeira explosão bibliográfica. Também propiciou o rompimento do monopólio que a Igreja exercia sobre a produção editorial. Apesar das bibliotecas serem muito anteriores aos livros e manuscritos, com a existência dos livros tipográficos passam a apresentar alguma diferença como elemento social, até então não verificada na Idade Antiga e Média quando se mantiveram praticamente inalteradas. Importante acrescentar que, até fins do século XVIII, museus, arquivos e bibliotecas eram a mesma entidade, pois organizavam e armazenavam todos os tipos de documentos.

A tecnologia da impressão, por mais remota, promove uma primeira modificação na atividade da organização e preservação de documentos, uma vez que, aos poucos, foi retirada da biblioteca a tarefa de reprodução de manuscritos realizada pelos copistas, que passou a ser feita em oficinas especializadas. Relevante se faz analisar os efeitos desta mudança que distanciou o bibliotecário da responsabilidade, cumplicidade e envolvimento com o conteúdo dos documentos.

Por outro lado, no século XVII, surge nos países mais adiantados da Europa e mais tarde nos Estados Unidos, o conceito de biblioteca pública moderna, constituída de acervos gerais de livros e aberta gratuitamente ao público em horários regulares. Eram bibliotecas financiadas por mecenas, cujo movimento continuou até praticamente o

século XX (LEMONS, 1998). Gabriel Naudé, em 1627, em sua obra “*Advis pour adresser une bibliothèque*”, escreveu os primeiros princípios da Biblioteconomia moderna. Já o termo “biblioteconomia” foi usado pela primeira vez em francês na obra de 1839 intitulada “*Bibliothéconomie: instructions sur l’arrangement, la conservation e l’administration des bibliothèques*” (LAHARY, 1997).

A biblioteca pública, espaço para acesso a acervo organizado, e não a Biblioteconomia, processo que possibilita aquela, é que fez parte do projeto da modernidade. De fato, em função do surgimento da biblioteca pública, geral e aberta e do crescimento dos periódicos e sua importância na divulgação científica, a Biblioteconomia trilhou novos caminhos, passando a dividir seus espaços com as atividades desenvolvidas pela Documentação.

## **1.2.2 A BIBLIOGRAFIA E A DOCUMENTAÇÃO (séculos XV ao XIX)**

Shera (1980), Shera e Egan (1961) e Bradford (1961) consideram como primórdios da Documentação, os trabalhos de construção de bibliografias universais realizados por Konrad Gesner, em fins do século XV, por Johann Tritheim, meio século depois. Após estas obras, seguiu-se uma proliferação de bibliografias especializadas e catálogos de bibliotecas particulares, a ponto de, em fins do século XVI, os estudiosos sentirem necessidade de sistematizarem tal volume de literatura registrada, demonstrando certa semelhança com os problemas atuais de organização bibliográfica. Surgiram então muitas bibliografias comerciais, precursoras da bibliografia nacional, mas pouco adequadas aos estudiosos.

A esta altura, era claro para os profissionais das bibliotecas, a necessidade de bibliografias para seu trabalho. Até 1840, nenhuma biblioteca tinha índice de assunto de seus acervos, até porque, nenhuma estava completamente catalogada: eram verdadeiros museus de livros. A partir desta data, catálogos de autor e assunto passaram a ser empregados e as bibliografias foram se aperfeiçoando. Alguns fatos explicam estes avanços: em 1841, no Reino Unido, foram criadas as “noventa e nove regras”, que estabeleceram as bases da catalogação durante várias gerações; em 1853, nos Estados Unidos, Jewett e contemporâneos propuseram um centro nacional de Bibliografia e Documentação; em 1873, Melvil Dewey publicou a primeira edição de sua Classificação Decimal, amplamente adotada; e, pouco tempo depois, a Library of Congress (dos Estados Unidos) tomou liderança na organização bibliográfica com seu sistema de distribuição de fichas catalográficas.

A crescente importância dos periódicos como veículo de publicação, que atingiu seu auge em 1850, e a necessidade do tratamento de suas unidades de informação para possibilitar sua recuperação, foram marcantes na história da Documentação, entre outros, porque as bibliotecas não deram conta de dar continuidade ao serviço de tratamento destas publicações. O catálogo e o esquema de classificação das bibliotecas foram baseados na monografia, o que as tornava inoperantes para trabalhar os periódicos: não havia preocupação com a diversidade intelectual do seu conteúdo, pois foram idealizadas para reunir em uma proximidade física os conteúdos semelhantes. Outro aspecto da cultura do trabalho em bibliotecas referia-se ao tratamento e busca de documentos apenas em seus espaços locais. Sendo as coleções de periódicos comuns às bibliotecas, seria conveniente criar índices coletivos dos artigos. Apesar deste projeto

ter sido iniciado, a organização entre várias bibliotecas independentes e distantes era complicada e a imaturidade profissional da organização dessas instituições àquela época impediu o desenvolvimento desse serviço.

A partir daí, esse segmento foi recolhido e desenvolvido pelo crescente grupo de documentalistas, deixando o bibliotecário de ter como atribuição uma parte relevante da literatura e a biblioteca reduzindo-se a sua antiga função de custódia de documentos. Acredita-se que este foi o início do enfraquecimento do prestígio da profissão do bibliotecário e da formação intelectual em Biblioteconomia.

Em fins do século XIX, a Biblioteconomia e a Documentação apresentavam um desenvolvimento em grande parte inseparável: surgiram em consequência das mesmas necessidades, empregavam processos básicos e instrumentos comuns (como as fichas de 7,5 por 12,5 cm e a Classificação Decimal de Dewey-CDD), tinham objetivos quase idênticos e em muitos casos deviam seu progresso aos mesmos homens. Havia, no entanto, uma tentativa dos documentalistas em evitar os instrumentos e até mesmos as nomenclaturas da Biblioteconomia, o que levou, muitas vezes, a que aqueles seguissem os caminhos já trilhados e até descartados por esta. A diferença da Documentação era que pretendia fazer uma análise de conteúdo mais profunda. Da mesma forma, os arquivos apresentavam problemas semelhantes de organização. A Biblioteconomia, a Documentação e a custódia dos arquivos, como já citado, eram a única e mesma coisa: muito cedo, no entanto, forças sutis começaram a dividir os que se interessavam por essas atividades em grupos separados, gerando uma atitude de intolerância entre eles.

Sendo assim, por mais de quatro séculos, a Biblioteconomia foi quase sinônimo de Bibliografia. Considerando a Bibliografia como o princípio da Documentação, pode-se dizer que esta esteve unida à Biblioteconomia desde o século XV até fins do século XIX, quando Otlet e La Fontaine sistematizaram e desenvolveram a Documentação, enquanto disciplina distinta da Biblioteconomia. Mais tarde, a Segunda Guerra Mundial acentuou o desenvolvimento da Documentação devido às necessidades específicas dos países envolvidos na recuperação de conteúdos a partir de tipos diversos de documentos, inclusive com tentativas rudimentares de recuperação mecânica da informação.

Ao que parece, uma circunstância que levou os bibliotecários a desviar suas energias dos problemas da análise e da organização do conhecimento registrado foi a fé na educação das massas que marcou a Idade do Iluminismo: com o crescimento do Mercantilismo e, mais tarde, da Revolução Industrial, percebeu-se a necessidade de um corpo de trabalhadores alfabetizados e treinados em tarefas manuais específicas. Tanto no Reino Unido como nos Estados Unidos, já na década de 1850 (portanto, no período do auge do crescimento dos periódicos), a biblioteca pública era considerada uma agência educacional das massas por meio da democratização da cultura. Com a absorção do bibliotecário tradicional pela função “educativa” ou o culto da educação universal e a negação do significado do serviço de informação, os documentalistas aproveitaram-se das técnicas da Biblioteconomia e as aperfeiçoaram e aprimoraram.

Em 1900, os problemas da organização bibliográfica se acentuaram novamente e os documentalistas não procuraram os bibliotecários, pois estes praticamente já não estavam preocupados em proporcionar o acesso aos assuntos dos documentos. Foi quando estas áreas passaram a se bifurcar seriamente.

Os bibliotecários continuaram a procurar uma utopia na qual até mesmo o homem mais pobre pudesse, através da biblioteca pública, tornar-se rico dos conhecimentos de todas as épocas, e os documentalistas passaram a trabalhar mais acentuadamente na

lacuna deixada pelos bibliotecários. Acredita-se que esta divergência teria encontrado soluções caso tivesse sido reconhecida a natureza diversa do acervo e do atendimento para ambos os casos: em uma biblioteca com fins educacionais são necessários títulos devidamente escolhidos para permitir uma orientação eficiente ao leitor. Esta característica evidencia-se, por exemplo, nas bibliotecas das universidades onde algumas obras são separadas da coleção para atender às necessidades curriculares dos programas de ensino. A necessidade de estender os privilégios da educação a todos, afastou as bibliotecas do rumo originalmente traçado pelos seus pioneiros, o que ocorreu sem que se tivesse formulado os mecanismos para garantir esse objetivo educacional e cultural, além de levar ao cancelamento dos trabalhos de bibliografia.

Assim, tão grande foi o cisma entre bibliotecários e documentalistas que as próprias associações o refletiram. Em 1908, um grupo de bibliotecários especializados nos Estados Unidos separou-se da American Library Association para formar sua própria associação. E assim como estes, vários casos se sucederam de dissidências de associações inicialmente de bibliotecários, que passaram a representar a Biblioteconomia especializada, a Documentação, a Microfilmagem, ou áreas temáticas como Biologia e Química. Nesta nova Biblioteconomia, estavam homens de ciência ou que simplesmente se preocupavam com a localização das publicações. Já os bibliotecários, que foram eruditos e bibliófilos desde a biblioteca de Alexandria, sofreram uma acentuada mudança.

A cisão entre Biblioteconomia e Documentação foi ficando cada vez mais profunda, sem levar necessariamente ao desenvolvimento e sedimentação de uma ou outra área. Este fenômeno tem particular interesse para a sociologia da profissionalização, que ocorre quando um grupo muda a terminologia de outro para parecer que se ocupa de uma nova disciplina, neste caso, com mais dignidade e caráter científico.

Shera e Egan acreditavam que registros do conhecimento erudito foram os primeiros e, por muito tempo, os únicos a serem efetivados em bibliotecas, levando a que muitos bibliotecários continuassem dominados pela forma de organização imposta pela tradição filosófica que marcou o mundo antigo e medieval. A sociedade moderna, porém, cresceu e transformou-se rapidamente, graças à tradição científica da pesquisa indutiva e empírica, que exigiu novas formas de organização do pensamento registrado. No âmbito das bibliotecas públicas, esta modernização foi evidente nos Estados Unidos, mas menos observável na Europa em função da erudição que ainda é presente em suas bibliotecas.

Uma frase aparentemente óbvia citada por Shera deixa claro, a partir da história da Biblioteconomia, o fato de os bibliotecários serem comumente taxados como os mais culpados neste conflito: unicamente tiveram mais tempo de sê-lo. Assim, o bibliotecário deve reaver a erudição que perdeu, porém uma nova forma de erudição. Talvez algo relacionado a uma intelectualidade particular, a partir de uma linguagem de especialidade própria para a área.

### **1.2.3 OTLET E A DOCUMENTAÇÃO (fins do século XIX e início do século XX)**

As atividades dos documentalistas foram se desenvolvendo simultaneamente ao surgimento das bibliotecas públicas. No entanto, foi Paul Otlet que, quase na virada do século, sistematizou a Documentação, cunhando este termo para significar, de forma mais ampla, aquilo que era denominado de Bibliografia, e dando um novo impulso a um

movimento que datava da época de Trithem e Gesner. Mais que isso, Otlet vem sendo considerado precursor e fundador da Documentação e da própria Ciência da Informação. Vários autores como Rayward (1997) e Sagredo (1994) desenvolvem o argumento da antecipação e previsão das tecnologias por Otlet, como os sistemas de hipertexto e hipermídia.

Segundo Rayward (1997) e Bradford (1961), o projeto teve início no encontro entre Paul Otlet e Henri La Fontaine, em 1892, e no reconhecimento de preocupações comuns quanto à organização bibliográfica da produção científica, para o que criaram o Escritório Internacional de Bibliografia, em Bruxelas, na casa de Otlet. Logo perceberam que, para organizar um índice universal, era necessária a cooperação internacional. Após longo percurso, fundaram o Instituto Internacional de Bibliografia (IIB), em 1895, apoiados pelo governo belga. Iniciativa pioneira dentre as associações internacionais no campo da informação, o IIB teve seu nome alterado para Instituto Internacional de Documentação (IID), em 1931, e para Federação Internacional de Documentação (FID) em 1938. Desde 1986, recebe a denominação de Federação Internacional de Informação e Documentação, com a mesma sigla, e com funcionamento na Royal Library, em Haia, Holanda.

O IIB tinha como pretensão desenvolver o repertório bibliográfico universal. A Classificação Decimal Universal (CDU) foi criada a partir da CDD, para contemplar as necessidades do conhecimento especializado. Devido às primeiras experiências de trabalho cooperativo, optou-se pela descentralização e pelo emprego de bibliógrafos especializados em ciências. Para tanto, foram estabelecidos centros de informação científica em um grande número de bibliotecas científicas ou em anexos destas. Em 1930, o repertório bibliográfico universal continha 16 milhões de entradas e, até o final da Segunda Guerra Mundial, doze países tinham Seções Nacionais do IIB, quais sejam, Alemanha, Bélgica, Dinamarca, Estados Unidos, França, Reino Unido, Holanda, Itália, Polônia, Rússia, Suíça e Tchecoslováquia, e mais seis estavam em formação.

No Brasil, no começo do século XX, a Livraria Civilização, em São Paulo, recebia assinaturas, encomendas de publicações e fichários padronizados. Em 1900, Oswaldo Cruz introduziu a CDU na biblioteca do instituto de pesquisas que fundou e hoje tem o seu nome. E em 1911, o diretor-geral da Biblioteca Nacional, no Rio de Janeiro, criou o Serviço de Bibliografia e Documentação, “em correspondência com o Instituto Internacional de Bibliografia de Bruxelas”, segundo o ato que o regulamentou. No entanto, sendo a CDU um sistema europeu, passou a ser rechaçado pela influência norte-americana que começou a ser sentida por volta de 1930 na Biblioteconomia brasileira.

O IIB prestava serviços de pesquisa a partir das fichas catalográficas, contemplando a estratégia e a contextualização da busca. Para a determinação dos assuntos trabalhava-se com mapas conceituais que mostravam as intrincadas relações dos conceitos dentro das várias áreas. No tratamento bibliográfico, utiliza-se o princípio monográfico, descrito como a divisão física do livro de acordo com a divisão intelectual de idéias, o que consistia na fragmentação do documento para chegar ao seu conteúdo. Supõe-se que estas duas idéias ensaiavam, respectivamente, as noções de tesouro e de tratamento documentário das partes dos documentos, o que provavelmente contribuiu para o refinamento dos processos documentários conhecidos atualmente.

Otlet fazia uma crítica às bibliotecas por conta das políticas de seleção (advindas do princípio jesuítico), da resistência às inovações técnicas e à prestação de serviços. Além disso, o projeto de Otlet era essencialmente político, característica pouco presente

na Biblioteconomia. Isto reforça o que já foi descrito sobre a relação entre documentalistas e bibliotecários.

Quanto ao Instituto, em 1924, foi reestruturado, mudou seu foco e a CDU tornou-se uma de suas únicas tarefas. Em 1995, no centésimo aniversário da agora denominada FID, foi firmada a Resolução de Tóquio por 35 organizações não governamentais em informação, comunicação e conhecimento como uma aliança estratégica para melhor servir à comunidade mundial. Também foi organizado o Fórum de Conhecimento da FID, a fim de reunir a experiência comum e o conhecimento acumulado em diferentes comunidades e torná-los disponíveis. A FID continua organizando eventos em nível mundial.

#### **1.2.4 A CIÊNCIA DA INFORMAÇÃO (anos 50 em diante)**

Shera e Egan (1961) citam que o funcionamento das bibliotecas em comunidades isoladas e sem cooperação entre elas deixou uma herança de conseqüências sociais, legais e psicológicas que tende a perpetuar esta forma “paroquial” de funcionamento. No entanto, mesmo o desenvolvimento da Documentação assistiu a uma crescente tendência de separação dos departamentos e serviços documentários da concepção de comunicação, pois atendiam apenas a seus centros. Ou seja, a organização bibliográfica não se sustentou por uma estrutura de comunicação bibliográfica. A Documentação e a Biblioteconomia evoluíram como artes práticas com o fim de atenderem necessidades imediatas. Tanto as bibliotecas quanto os centros de documentação trabalharam de forma isolada, decorrente de um período caracterizado por poucas bibliotecas, acervo reduzido e clientela homogênea e erudita que pouco se relacionava entre si.

Sendo assim, por mais que a Documentação possa aparentar um caráter mais moderno que a Biblioteconomia, no final da década de 60, já com a Ciência da Informação nos Estados Unidos, a Documentação chegou a ser considerada até mais antiquada que a Biblioteconomia. O mesmo não se deu na Europa continental em que o termo Documentação é amplamente utilizado e se confunde, muitas vezes, com a idéia expressada pela Ciência da Informação. Em especial, na França, a divisão profissional entre bibliotecários e documentalistas deu-se de forma marcante, o que pode ser observado pelos cursos de formação específicos e pela existência dos centros de documentação. Historicamente, os primeiros centros de documentação surgiram neste país entre as duas Grandes Guerras, mas multiplicaram-se realmente depois de 1945, ou seja, depois das bibliotecas universitárias e de leitura pública. O desenvolvimento dos centros de documentação ilustra o papel crescente que desempenha a informação na ajuda à pesquisa e à decisão dentro das organizações. Na área escolar especificamente, os centros de documentação e informação (CDI), existentes na maioria das escolas francesas do ensino médio, substituíram, a partir de 1974, os antigos serviços de documentação e informação (SDI), assim nomeados em 1966. Os CDIs englobam biblioteca, sala de leitura, audioteca e videoteca, entre outros, efetivando o tratamento documentário dos materiais, a pesquisa, a promoção de leitura e a ação cultural. Neles trabalham os professores-documentalistas.

Por outro lado, em especial, nos Estados Unidos, os CDIs são representados pelas midiatecas (*media centers*) ou bibliotecas escolares, assim como, os centros de documentação são geralmente denominados de bibliotecas especializadas ou centros de

informação (*information centers*). De fato, a linha existente entre centros de documentação e bibliotecas especializadas ou midiatecas não é evidente e tende a diminuir em função do tratamento de documentos eletrônicos e seu acesso (PALLIER, 1997) e (VERNOTTE, 1997). Já na América Latina, em virtude da reconhecida influência dos Estados Unidos, os modelos instaurados pelos centros de documentação foram pouco disseminados.

Assim, a tendência predominante contou com a Ciência da Informação sendo empregada para designar a Biblioteconomia do tipo não tradicional. Passou-se do problema da diferença entre Biblioteconomia e Documentação para a relação entre Ciência da Informação e Biblioteconomia. Afinal, se a Documentação surgiu de uma cisão da Biblioteconomia e impulsionou a formação da Ciência da Informação, seria coerente considerar as duas últimas como divergentes. No entanto, nas três áreas verifica-se menos divergência que arcaçou científico, o qual, no entanto, considera-se mais presente nos princípios modernos da Documentação que nas tradições bibliotecárias. Sendo assim, segundo Shera (1980), na Medicina ou no Direito, o estudante aprende os princípios gerais da área e quando se especializa sabe que isto faz parte de um conjunto mais amplo. A Ciência da Informação não apresenta um conjunto de noções gerais comuns que garanta a orientação a campos mais específicos de atividade e estudo. Isto demonstra falta de maturidade e carência de acumulação teórica da área.

Neste sentido, a história da Ciência da Informação apresenta menos dados factuais devido a sua existência recente, mas muito sobre sua controversa natureza conceitual e sedimentação teórica enquanto campo de conhecimento.

Para Pinheiro (1999, p. 156), a Ciência da Informação “parte do reconhecimento de sua interdisciplinaridade, de sua natureza social, forte e profundamente relacionada à tecnologia da informação e do novo papel da informação na sociedade e na cultura contemporâneas, características essenciais da área”. A Ciência da Informação tem suas raízes na bifurcação da Documentação/ Bibliografia e da Recuperação da Informação. É uma ciência social cujo objeto é a informação, tendo início no campo da informação científica e tecnológica, passando a atuar também com a informação para fins educacionais, sociais e culturais. Apresenta interfaces com a Biblioteconomia, Ciência da Computação, Ciência Cognitiva, Sociologia da Ciência e Comunicação, entre outras áreas.

Outras tentativas de constituição da Ciência da Informação incluem ainda áreas do conhecimento como a Administração que fornece formas racionais de organização do fluxo de informações e as Artes Gráficas, utilizadas na produção editorial. Também podem ser citadas a Lingüística, Psicologia, Estatística e Economia.

Há que se fazer distinções, no entanto, entre interdisciplinaridade e aplicações. Goffman citado por Pinheiro (1999, p. 164-165), aborda a questão da interdisciplinaridade da Ciência da Informação com a Biblioteconomia e a Ciência da Computação. Entende que as bibliotecas foram as primeiras instituições que operam com informação a utilizarem o processamento automático e que a Biblioteconomia necessita de uma ciência que forneça a respeitabilidade acadêmica que lhe falta. No caso da Ciência da Computação, são apontados também a definição de problemas e a busca de legitimidade acadêmica. No entanto, não pode a Ciência da Informação restringir-se nem a uma nem a outra e deve transcender os sistemas físicos das bibliotecas ou dos computadores, embora possa utilizar suas aplicações e se beneficiar tanto da Biblioteconomia quanto da Ciência da Computação. A ligação da Ciência da Informação com estas áreas é tão forte que, em seu início, na década de 60, eram confundidas com ela.

Outra nomenclatura para a Ciência da Informação é Informatologia, utilizada pela primeira vez em 1962 em uma universidade de Estocolmo, na Suécia, para evitar o termo Documentação, oferecendo durante vários anos um curso de cem horas com este nome (CHERNYI, GILYAREVSKII, MIKHAILOV, 1973). Também a Enciclopédia Mirador (1994) considera que a Biblioteconomia deu origem à Bibliografia, que fundamentou a Documentação, que por sua vez, forneceu insumos à constituição da Ciência da Informação, também nomeada Informatologia. A Ciência da Informação é entendida como a preocupação com a unidade fundamental do saber, através de estudos interdisciplinares e de métodos como o estrutural. Engloba o conjunto das disciplinas voltadas para a produção, comunicação e consumo da informação que, chamadas por isso de ciências da informação, passaram a ser consideradas como uma só ciência da informação.

A adoção do termo Ciência da Informação no idioma inglês, pode ser verificada já em 1958, com a criação, em Londres, do Institute of Information Scientists, e dez anos depois, com a mudança de denominação do North American Documentation Institute para American Society for Information Science, o mesmo ocorrendo com escolas de Biblioteconomia e publicações da área. Marco considerado importante nesta nova ciência é a Encyclopedia of Library and Information Science que teve publicados, até 1972, cinco volumes dos doze planejados, e em cujo prefácio considera esta ciência como integrada da Biblioteconomia e informação. A própria Enciclopédia Britânica publicou em 1968, longo verbete intitulado *Information processing*, considerando a necessidade de muitas disciplinas, tecnologias e atividades sobre o processamento da informação para sua organização, armazenamento, comunicação e uso.

Na antiga União Soviética, a construção deste conceito teve início alguns anos depois de fundamentada e desenvolvida pelos Estados Unidos. A equipe de Mikhailov, concebeu alguns enfoques diferentes e criou o termo *Informatika* para defini-lo. Para os russos, essa nova disciplina é responsável pelo estudo da estrutura e das peculiaridades da informação científica (cujo sentido possui abrangência maior e diferenciada daquele por nós adotado), assim como as leis que regem essa atividade, sua teoria, história, métodos e organização. O assunto fundamental da *Informatika* é a informação considerada no contexto comunicativo dentro do sistema social interessado pelas mudanças no estado dos conhecimentos: deve produzir alterações no pensamento das pessoas, o que significa que a informação é comunicação que se recebe e se assimila (CHERNYI, GILYAREVSKII, MIKHAILOV, 1973) e (MORENO GONZÁLEZ, 1995).

As áreas que compõem a *Informatika* foram definidas por: *Informatika* teórica (estuda os sistemas abstratos de informação); *Informatika* de gestão (relaciona-se com a gestão de sistemas de informação); *Informatika* científica (investiga os sistemas de informação automatizados); e *Informatika* bibliotecária (estuda os sistemas de informação bibliotecária). É realizada uma distinção entre o serviço de informação científica do trabalho ativo com o leitor, a propaganda com o livro e a orientação para a leitura, como uma crítica aos desenvolvimentos dos “bibliotecários burgueses”. Citam que estes, por pensarem as atividades da biblioteca como meramente técnicas, baseadas na aquisição de livros e sua organização para os leitores, acreditam que livros serão substituídos pelos microfilmes e as bibliotecas e os bibliotecários estariam em extinção. Na União Soviética, a biblioteca pública tinha um papel pedagógico e formativo, como meio para edificação da sociedade socialista (CHERNYI, GILYAREVSKII, MIKHAILOV, 1973) e (MORENO GONZÁLEZ, 1995).

No entanto, ocorreu um processo que levou à adoção ampla da concepção e do termo *Information Science*, elaborado nos Estados Unidos, em detrimento da *Informatika*, pensada pelos russos no mesmo período. O fenômeno é explicado pela ambigüidade do termo Informática no Ocidente e pelos avanços na constituição conceitual e nas aplicações da corrente americana, sustentada pelo predomínio teórico e comercial. Com o esgotamento do regime socialista, teve fim também uma das principais correntes teórico-práticas da Ciência da Informação (MORENO GONZÁLEZ, 1995).

Teórico de produção relevante no campo da Comunicação, Saracevic considera o objeto da Ciência da Informação como o comportamento, as propriedades e os efeitos da informação em todas as suas facetas, tanto quanto os vários processos da comunicação que afetam e são afetados pelo homem. Para ele, a Ciência da Informação estuda: a dinâmica e a estática do conhecimento, ou seja, suas fontes, organização, criação, dispersão, distribuição, utilização, expressão bibliográfica e obsolescência; os aspectos comunicacionais relacionados ao homem enquanto produtor e usuário de informação; os problemas da representação simbólica da informação como na classificação e indexação; e, por extensão, o funcionamento de sistemas de informação como as bibliotecas e os serviços de armazenagem, recuperação e processamento de dados. Enfatiza ainda que é importante distinguir a Ciência da Informação propriamente dita do treinamento na tecnologia eletrônica aplicada à mecanização dos serviços de bibliotecas; este comentário é essencial nesta pesquisa, uma vez ser recorrente esta visão (ENCICLOPÉDIA MIRADOR INTERNACIONAL, 1994, p. 6114-6115).

Percebe-se que a tendência é de cruzamento das áreas que trabalham com informação no decorrer do delineamento dos seus meios e fins. Com base em Saracevic, é desejável que o amadurecimento da Ciência da Informação traga consigo uma interdisciplinaridade eqüitativa e consciente, a ponto de contemplar adequadamente a conjuntura social. Em outras palavras, ressalta que, o ponto que mais importa é aquele que se refere às formas de prover melhores serviços de informação aos usuários.

Wersig (1997) argumenta a favor da natureza da Ciência da Informação como uma ciência pós-moderna. Isto implica um novo entendimento da noção de conhecimento, por meio, entre outros, do abandono da idéia de paradigmas como categorias únicas de compreensão científica. Diferente da ciência clássica que é baseada na pesquisa do completo entendimento de como as palavras trabalham, a ciência pós-moderna trata do desenvolvimento de estratégias para resolver em particular aqueles problemas causados pelas ciências clássicas e pelas tecnologias. Wersig define Ciência da Informação como o conjunto de modelos, desenvolvidos sob o ponto de vista do problema do uso do conhecimento nas condições pós-modernas de informatização.

O desenvolvimento do conhecimento científico deu origem à tecnologia, que por sua vez, transformou e transforma o conhecimento e as suas formas de disseminação. Possuindo como contexto a pós-modernidade, o período contemporâneo é marcado: pela informação (unidade do conhecimento ou o dado agregado de valor comunicacional), pela comunicação (que vulgariza o conhecimento por meio de sua disseminação massiva possibilitada pela tecnologia) e pela tecnologia (instrumento difusor e transformador do conteúdo). Para Mendonça (2000, p. 65), o campo da construção teórica da Ciência da Informação está situado entre o tecnológico e o humano, pois os avanços tecnológicos afetam o conceito e o uso da informação, que por sua vez influem na estruturação do conhecimento.

### 1.3 QUADRO ATUAL DA BIBLIOTECONOMIA, DA DOCUMENTAÇÃO E DA CIÊNCIA DA INFORMAÇÃO

A Biblioteconomia surgiu simultaneamente aos primeiros registros reconhecidos como formas de escrita não pictográfica e desenvolveu-se à medida que o próprio conhecimento humano foi se constituindo formalmente. Mas o decorrer histórico contribuiu para que a Biblioteconomia ficasse afastada do conhecimento que organizava e preservava, pois:

- A disseminação da imprensa, a partir do século XV, paulatinamente excluiu da prática de organização e preservação de documentos, a atividade de produção e reprodução, tornando o bibliotecário de então isento e sem responsabilidade de possíveis intervenções na sua forma ou conteúdo. A não cumplicidade foi impulsionada quando este profissional passou a receber o livro pronto e acabado, para ser devidamente alocado no acervo. Por outro lado, a biblioteca moderna surgida então, foi marcadamente diferente das anteriores, pois era aberta ao público e, mal ou bem, respirava os ares do Renascimento.

Apenas atualmente, a Editoração, as Artes Gráficas e a marcação de textos para recuperação da informação, incluídas como aplicações da Ciência da Informação e da prática do bibliotecário, podem estar reaproximando este profissional da produção do documento, em especial, quando a informação passa a surgir e permanecer no meio eletrônico. Outra evidência deste fenômeno está na nomenclatura utilizada para o suporte físico da informação com que o bibliotecário trabalha: depois de cinco séculos de primazia do livro, volta a necessidade do termo documento para indicar a diversidade de suportes como os bibliográficos, audiovisuais e eletrônicos.

- A crescente importância e a ampliação da publicação de periódicos, desde a metade do século XIX, decorrente do desenvolvimento científico, demandaram uma reformulação da atividade biblioteconômica. Neste mesmo período, estava sendo implementada a biblioteca pública no Reino Unido e nos Estados Unidos com base no ideal da educação das massas e face à industrialização que impeliu o processo de democratização pelo sufrágio universal masculino e pela disseminação da capacidade de consumo. Ocorreu uma mudança na atuação das bibliotecas, que se desviaram dos princípios propostos por pensadores da Biblioteconomia, mas não ficaram claras as estratégias que deveriam estar por trás desta nova atuação. A esta altura, a lógica do trabalho das bibliotecas já estava sedimentada no tratamento de monografias como um todo e na localização física (nas estantes) e local (somente naquela biblioteca) dos assuntos dos livros. Os documentalistas ocuparam a lacuna dos trabalhos de Bibliografia deixada pelos bibliotecários; mas não ocuparam o imaginário social ainda pertencente à Biblioteconomia, envolvida pela idéia mítica de profissão guardiã da cultura da humanidade e responsável pela sua difusão, o que lhe atribuiu, de início, um certo *glamour*, substituído com o tempo pela idéia de profissão meramente técnica, burocrática e de baixo reconhecimento social.

Mais recentemente, a tentativa de dar à Biblioteconomia um *status* humanístico e menos técnico, foi e ainda tem sido, ao menos no Brasil, uma bandeira dos que pregam em nome da biblioteca pública. Essa posição é consequência direta da história descrita e sofre dos mesmos males: ausência de definição das estratégias, processos e métodos de trabalho nas bibliotecas públicas. Para exemplificar, nos Estados Unidos, especialmente nos anos 70, empresas comerciais assumiram o serviço de indexação de artigos de

periódicos e sua disseminação, tornando definitivamente desnecessária esta atividade nas bibliotecas. No entanto, em países como o Brasil, a importação arbitrária e segmentada de modelos, formou um quadro em que bibliotecas organizam acervos e é quase inexistente a oferta de serviços de artigos indexados.

- A automação das funções biblioteconômicas e a transmissão acelerada e ágil dos dados pelas redes eletrônicas, vêm substituindo o papel da técnica baseada em suportes impressos pela tecnologia eletrônica. A atividade biblioteconômica caracteriza-se historicamente pela organização e controle dos documentos por meio do cumprimento de tarefas baseadas em procedimentos e instrumentos técnicos. Estes instrumentos técnicos, criados à medida que a Biblioteconomia se desenvolvia distanciaram-se, com o tempo, de seus princípios geradores, passando a adquirir valor de autoridade e considerados como um fim em si mesmos, como indica a noção de tradição, segundo a qual, o sujeito não reconhece o objeto. Essas tarefas, tradicionalmente exercidas pela Biblioteconomia e baseadas em técnicas burocráticas dadas como inquestionáveis, estão sendo substancialmente substituídas pelas atividades técnicas baseadas na Informática e nas telecomunicações. A transposição inadequada transforma a Biblioteconomia de profissão de rotinas burocráticas impressas em eletrônicas pois, estando arraigada no princípio do cumprimento de tarefas, a mudança de instrumento não alterou significativamente sua atuação. Considere-se, no entanto, que sendo a Informática mais flexível e rápida que todas as possibilidades que o papel pode oferecer, processos e produtos tendem a ficar mais requintados e a exigir, para tal, alguma reflexão e habilidades específicas, o que implica, no mais das vezes, algum salto qualitativo. Por outro lado, recursos avançados exigem mais entendimento e interação, para o que, a atitude tarefaira e o comportamento acrítico nada contribuem.

Freqüentemente, nem bibliotecários nem usuários, percebem o quanto e como os serviços oferecidos ainda são tucanos frente a tamanhas possibilidades de intervenção dadas pelo espectro informacional disponível e pelas tecnologias que lhes dão acesso e forma.

- A erudição dos antigos bibliotecários, caracterizada por sólida formação cultural e livresca, não foi substituída por algo que lhe faça jus. A dotação de parâmetros teóricos para a área está, muito provavelmente, na constituição da Ciência da Informação, enquanto garantia de uma linguagem própria para a área e a decorrente substituição da erudição dos antigos bibliotecários por profissionais intelectualizados, mas não sisudos, e versáteis e ágeis, mas não sem prumo.

Além disso, a tecnologia deflagrou a informação enquanto objeto de trabalho e estudo da Biblioteconomia, em oposição ao foco em seu suporte (o livro, para citar o mais tradicional). Isto implica o reconhecimento da informação como unidade socializada do conhecimento e, portanto, a necessidade do envolvimento e domínio da natureza da informação de uma dada área (o que é distinto do domínio de fontes de informação desta área). Esta abordagem inclui ainda a natureza e as características da informação institucionalizada, ou seja, a informação no seu contexto institucional.

Pode-se considerar que a Biblioteconomia apresenta uma longa história de atividades, desde o início da escrita até a época moderna, quando recebeu novo impulso com a invenção da imprensa e o Renascimento científico e cultural. Depois sofreu uma ruptura, na segunda metade do século XIX acentuada por desenvolvimentos na Segunda Guerra Mundial no século seguinte, que decorreu na área conhecida como

Documentação, cujas atividades e concepções contribuíram para o surgimento da Ciência da Informação. De todos os méritos da Biblioteconomia, o maior talvez tenha sido a preservação do conhecimento registrado acumulado pela humanidade e a possibilidade (não a efetividade) de sua recuperação e uso. E a maior ressalva está no seu grande empirismo e na sua dificuldade de constituição científica enquanto área do conhecimento, assim como, sua resistência em encontrar formas de atuação prática menos caseiras e mecânicas, em detrimento de uma postura profissional intelectual e política, que considere os modelos teóricos em contínua interação com a prática.

Paradoxalmente, por mais que seja contestável o uso e a aceitação do argumento do "bom senso" na prática biblioteconômica, é o que parece fazer falta, ao menos num primeiro momento, para a percepção e entendimento das atuais demandas da sociedade por informação. O bom senso é o meio do caminho entre o senso comum e a racionalidade científica. Neste sentido, o bom senso é mais desejável que o senso comum, ditado pela tradição e o hábito, mas ainda está distante da racionalidade e do rigor científico que cabe a uma área do conhecimento.

A Biblioteconomia busca *status* científico na Ciência da Informação. No entanto, a Ciência da Informação teve início, em parte, calcada nos princípios da Documentação. Por este motivo, entre outros, acredita-se que as bibliotecas devam pautar-se nas atividades e nos modelos da Documentação e no instrumental teórico proposto pela Ciência da Informação, constituindo-se em verdadeiros serviços de documentação e informação. Mesmo considerando os avanços neste sentido, em algumas bibliotecas universitárias brasileiras, por exemplo, observa-se o movimento oposto, por meio do cancelamento dos serviços de disseminação seletiva da informação (DSI) e do suporte às estratégias de busca junto aos usuários em suas pesquisas. A justificativa costuma ser a falta de tempo e a potencialidade das ferramentas informáticas e de telecomunicações como elementos que possibilitam a realização destas tarefas pelos próprios usuários. Acredita-se que esta questão mereça uma análise cuidadosa pois indica que o avanço tecnológico que evidencia os serviços de informação leva, de fato, a um recrudescimento da passividade e da obsolência nas bibliotecas.

A Lingüística Documentária e a Informática Documentária são disciplinas que representam uma forma especial e especificamente elaborada de Lingüística e de Informática aplicadas aos processos documentários. Estando em sua origem ligadas à Documentação e suas denominações assim o indicarem, expressam conteúdos essenciais para a composição da Ciência da Informação por considerarem, respectivamente, a natureza lógico-semântica do conteúdo dos documentos e os requisitos tecnológicos a serem utilizados no seu tratamento e disseminação.

Aqui cabe comentar a lacuna observada na produção sobre Ciência da Informação quanto às disciplinas documentárias já constituídas, como as acima citadas. Ou seja, se a Ciência da Informação surge, entre outros, a partir dos princípios da Documentação, por que estes princípios são tão pouco contemplados? Uma explicação possível está na constatação de que, tanto a Lingüística Documentária quanto a Informática Documentária, teriam sua produção fracamente disseminada. Isto, no entanto, não justifica que a construção teórica da Ciência da Informação não opere pela assimilação destas disciplinas interdisciplinares, de alguma forma já constituídas e aplicadas. Ou seja, por que são consideradas a Lingüística e a Informática (ou Ciência da Computação) como constituintes possíveis da Ciência da Informação, e não a Lingüística Documentária e a Informática Documentária? Ao que tudo indica, estas disciplinas

documentárias possuem baixo reconhecimento, sendo mais desenvolvidas em suas partes e não consideradas no todo.

A partir desta constatação, acredita-se que a história tenha sido pouco justa com a Documentação em favor da presença social marcante da Biblioteconomia. Isto porque, diferente desta, a Documentação desenvolveu métodos racionais e ferramentas atualizadas, mesmo que críticas possam ser feitas, por exemplo, à restrita perspectiva comunicacional efetivada pelos centros de documentação. No entanto, esta deficiência é característica de ambas, tendo a Biblioteconomia a única vantagem de contar com a intensa disseminação de seus espaços físicos, uma vez considerando-se a biblioteca como um equipamento cultural. Este fato não decorre, contudo, em maior capacidade comunicativa pois falta, em tantos espaços físicos, a implantação dos mecanismos que permitam a divulgação e a assimilação do material informacional que os compõem. Mesmo os trabalhos cooperativos das bibliotecas, por meio da constituição formal de redes, estabeleceram laços entre as bibliotecas mas não, necessariamente, destas com seu público usuário. Assim, mais potencialmente comunicáveis são os processos de organização que a Documentação propõe, centrados no conteúdo dos documentos, gerando centros de trabalho vinculados às instituições em que estão inseridos e, apontando para as inovações tecnológicas, a tal ponto que vieram a inspirar uma nova ciência.

No entanto, se mesmo na Ciência da Informação, os princípios da Documentação não têm a presença que merecem, pressupõe-se que: ou a história será mais uma vez injusta, provavelmente porque o apelo dos avanços tecnológicos obscureça a informação que veiculam, ou a Ciência da Informação esteja ainda tão crua e embrionária que não consiga apropriar-se, de fato, de áreas que lhe deram origem. Esta segunda suposição pode parecer contraditória, uma vez estar na Documentação parte da origem da Ciência da Informação. No entanto, como a Recuperação da Informação (advinda do termo inglês *Information Retrieval*) também participa desta nova ciência, supõe-se que esta contribuição esteja se dando em termos mais abrangentes e significativos. Reforçando esta visão, artigos da revista francesa *Documentaliste* (1993, n. 4-5, p. 248) que tratam da história da Informática Documentária na França e do desenvolvimento das linguagens documentárias em sua relação com os sistemas de pesquisa de informação são apresentados com a seguinte pergunta: pode-se estudar a história da Ciência da Informação sem abordar as técnicas documentárias? Reconhecendo a origem europeia da Documentação e a origem estadunidense da Recuperação da Informação, algumas análises são realizadas nos capítulos seguintes sobre o conceito da Informática Documentária com base na distinção entre estas duas culturas.

Por fim, conclui-se que a Biblioteconomia, a Documentação e a Ciência da Informação são áreas que se relacionam conceitual e historicamente. A Biblioteconomia tem origem efetiva na atividade de preservação e conservação do conhecimento registrado pela escrita, alterando-se com o tempo por meio da democratização do acesso à educação, mas sem constituir área cientificamente fundamentada no seu todo. A Documentação, uma dissidência da anterior mas também componente dela, é marcada pelo tratamento do conteúdo dos documentos, pela diversidade dos tipos de registro com que trabalha e pelas primeiras iniciativas de mecanização de seus processos, atingindo significativo grau de sistematização de suas atividades e princípios. Deu insumo à Ciência da Informação que, pautada na ciência pós-moderna, portanto interdisciplinar e sem vinculação a paradigma único, reflete a mudança instaurada no século XX pela comunicação, pela tecnologia eletrônica e pelos fluxos de informação.

Sendo a Biblioteconomia, a atividade mais antiga de registro e organização de informação, encontra na Ciência da Informação a possibilidade de construção de referenciais teóricos e de conquista de *status* científico, enquanto esta tem na Biblioteconomia a história e o empirismo daquilo que vem elaborando a partir de diversas disciplinas e aplicações. Já a Documentação, considerada em separado da Biblioteconomia em seus princípios e práticas, encontra-se pouco difundida, em especial, no Brasil. Aqui, o modelo da Biblioteconomia disseminado pelos Estados Unidos está sendo acrescido dos princípios da Ciência da Informação sem que esta tenha, de fato, reconhecido os desenvolvimentos documentários.

Atualmente, verifica-se a produção de tesouros, serviços de indexação e resumos e organização de informações na Internet por diversos profissionais que não os da área de Biblioteconomia e Documentação. Este fato reforça a ocupação histórica (desde metade do século XIX, pelo menos) por outros profissionais nas lacunas deixadas pela Biblioteconomia. Corporativismo profissional a parte pois se acredita na riqueza da diversidade, não faz sentido que os historicamente conhecidos profissionais da informação e do conhecimento não atuem ampla e significativamente nas atividades de pesquisa e de prestação de serviços desta área.

# CAPÍTULO 2 - INFORMÁTICA DOCUMENTÁRIA : HISTÓRIA, CONCEITOS E MERCADOS

## 2.1 INTRODUÇÃO À INFORMÁTICA DOCUMENTÁRIA

Para esta pesquisa, a definição mais satisfatória encontrada para Informática Documentária é: “o conjunto de aplicações da Informática à Documentação, técnica que se refere às intervenções da Informática nas diversas fases de produção e utilização de documentos: produção de textos, difusão pelo editor, gestão pela biblioteca, análise e indexação para a constituição de bases de dados bibliográficos e para a difusão seletiva, e *softwares* para a pesquisa nestas bases de dados” (DEWEZE, 1994, p. 1). Segundo o autor, estas aplicações especificamente documentárias permitem ao estudante de Ciência da Informação fazer a ponte entre o ensino da Documentação e as técnicas informáticas, além de contribuir para colocar e analisar corretamente os problemas de Documentação em vista de sua resolução informática.

Não é por acaso que a definição escolhida consta de um trabalho francês, uma vez ter ocorrido na França o surgimento do termo. A expressão *Informatique Documentaire* foi proposta por Chonez em seu Cabinet d'Études et Recherches Documentaires, fundado em maio de 1968, e empregado, também por ele, na ADBS (Association des Professionnels de l'Information et de la Documentation) no projeto de uma comissão técnica e no Ciclo de Informática Documentária de fevereiro de 1969. Além disso, o termo aparece regularmente na revista *Documentaliste* desde o primeiro número de 1969, mas já vinha sendo citado desde 1968 (CHONEZ, 1993).

Chonez nomeia como um equívoco a crença de que as primeiras experiências de automação documentária remontem aos fins da década de 60. Apresenta um histórico destas pesquisas citando como referência mais antiga na França o trabalho, com uso de cartões perfurados, do engenheiro químico Yoland Mayor, apresentado em 1937, no Congresso Mundial da Documentação Universal, em Paris, e intitulado “Les possibilités de mécanisation des recherches bibliographiques”. Em bibliotecas, foi também em meados nos anos 30, nos Estados Unidos, que técnicas de cartões perfurados foram utilizadas para registrar e agilizar o empréstimo de materiais bibliográficos. No entanto, pode-se indicar o período correspondente ao final dos anos 50 e início dos anos 60 como aquele em que ocorreu a disseminação da automação tanto na reprodução de fichas catalográficas como na aplicação da pesquisa documentária, pioneiramente nos Estados Unidos e na França.

Sendo assim, com origem na França, o termo *Informatique Documentaire* foi adotado em outros idiomas como o espanhol (*Informática Documental*) e o português (Informática Documentária). Já em inglês, não há uma expressão ou termo similar.

No caso da língua portuguesa, o emprego do termo no Brasil, refere-se de uma forma geral, a comentários sobre o bibliotecário frente às novas tecnologias do mercado, muitos relatos de experiência de automação de bibliotecas e de bibliotecas virtuais, cada vez mais, estudos de caso sobre características de *softwares* e recursos desejáveis dos mesmos, além da representação descritiva automatizada realizada pela adoção de formatos de registro bibliográfico com fins de intercâmbio de dados, seguindo os modelos

encontrados em língua inglesa. De fato, não é um termo exaustivamente utilizado no país.

Ao menos três bases de dados levantaram e trataram informações relativas à automação de bibliotecas brasileiras e informatização de processos documentários em geral, sendo nomeadas como bases de dados sobre Informática Documentária. Na base de dados Biblioinfo, da Universidade do Estado de Santa Catarina (UDESC), Ohira (1994) cita a definição para o termo Informática Documentária indicada no tesouro de Ciência da Informação do IBICT: “é usado para documentos que tratam de uma forma geral sobre as aplicações de informática às atividades de informação”. Mais tarde, a mesma Universidade elaborou a base BIDAC, cuja sigla indica as palavras Biblioteconomia, Informática Documentária, Documentação, Arquivo e Ciência da Informação, e inclui registros da produção intelectual dos profissionais nas áreas citadas em Santa Catarina de 1976 a 1996. Já na Base de Dados da Literatura Brasileira de Informática Documentária (BIBID) da Universidade Federal Fluminense (UFF), Souza (1994) define Informática Documentária como a: “área interdisciplinar da Ciência da Informação responsável pela aplicação da tecnologia dos computadores ao processamento eletrônico da informação documentária”. Estas definições expressam conceitos genéricos e menos precisos que aqueles propostos pelo francês André Deweze.

Quanto aos cursos de Biblioteconomia e Documentação do país, sabe-se que os conceitos e as atividades relacionados à Documentação (aspecto apontado como relevante neste trabalho) são pouco desenvolvidos. São dadas noções de Informática aplicadas a bibliotecas, faz-se conhecer alguns programas para automação de bibliotecas e, em menor escala, são fornecidos critérios de escolha, diagnóstico e planejamento para implantação destes sistemas. Poucas escolas apresentam os formatos de registro bibliográfico existentes e suas características e utilização, o que tem mudado radicalmente com a divulgação do Formato MARC. Em geral, o enfoque está nas ferramentas para a gestão de acervos, em detrimento da automação dos processos documentários, e falta um passo para introduzir noções de automação em todas as outras disciplinas. A introdução do método automático do tratamento da informação no ensino modifica, não o conteúdo, mas a forma de entendimento deste processo.

Sendo assim, pode-se inferir que não há sistematização relevante e ampla o suficiente para considerar uma consolidação destes conhecimentos no Brasil, podendo-se avaliá-los como incipientes e excessivamente preliminares. Já a literatura em espanhol e inglês apresenta massa crítica passível de análise.

Em espanhol, o uso da expressão *Informática Documental* segue a linha francesa. No entanto, as referências ao conhecimento acumulado sobre a área pelos Estados Unidos são bastante recorrentes. São analisados, o chileno Mario Saquel que trabalha com a Informática Jurídica Documentária e desenvolve o tema essencialmente a partir da confrontação de diversas definições da literatura, e o espanhol Carlos Costa Carballo, menos conceitual mas contribuindo por meio da indicação de um panorama sobre Informática Documentária.

Já a literatura em língua inglesa não possui termo que corresponda à Informática Documentária. Os termos encontrados que, somados, aproximam-se desta expressão são *Library Automation* (Automação de Bibliotecas) e *Information Retrieval* (Recuperação da Informação). No caso da Automação de Bibliotecas, os sistemas pesquisados não costumam apresentar tesouros ou outros vocabulários controlados; apenas mencionam a existência de um campo destinado aos descritores. Difere, portanto, substancialmente

dos sistemas apresentados nos idiomas francês e espanhol, nos quais, a implementação das ferramentas informáticas tem importância também nos processos documentários de análise do conteúdo e de gestão do vocabulário controlado (ou tesauro) que, portanto, seguem além da criação de campo para descritores.

Dentre os autores de produção em língua inglesa, William Saffady (1999) vem publicando o livro "Introduction to automation for librarians" desde 1980, abordando os fundamentos gerais e básicos de Informática, a automação dos serviços de uma biblioteca e os novos suportes e recursos tecnológicos. Não há menção ao tratamento temático da informação. Jennifer Rowley (1998) que também produz um livro sobre o tema desde 1980 abarca, além dos fundamentos gerais de Informática, a Recuperação da Informação (*Information Retrieval*) e os sistemas de gerenciamento de bibliotecas. Estes dois autores são analisados, sendo o primeiro com produção nos Estados Unidos e de interesse por centrar-se exclusivamente na Automação de Bibliotecas, enquanto a segunda, autora de produção inglesa, é apresentada por abranger tanto a Automação de Bibliotecas quanto a Recuperação da Informação.

Exclusivamente sobre Recuperação da Informação (*Information Retrieval*), foram apresentados textos com dados conceituais e históricos para que se possa compreender a área. Basicamente, pode-se encontrar aí o debate, desde sua origem nos anos 50, sobre a validade de pesquisas centradas na recuperação estatística dos dados e não na recuperação conceitual da informação, assim como na utilização de linguagem natural e não na linguagem controlada.

Voltando à obra de origem francesa publicada por Deweze e adotada como referência, pode-se indicar seu conteúdo pelos seguintes itens: fluxo documentário (ou cadeia documentária), funções e componentes do computador, estruturação dos dados, entrada de dados, produção dos documentos, gerenciamento eletrônico de documentos, informatização de bibliotecas, análise de conteúdo de documentos, concepção e implementação de tesouros, interrogação de bases de dados, guia de especificações de uma aplicação informática, custo de aplicações e escolha de soluções. Ou seja, aborda-se o fluxo informacional, assim como a tecnologia e a lógica que envolvem o conceito e o processo do fluxo, como o tratamento temático elaborado. Já a informatização de bibliotecas é apenas uma parte, muitas vezes necessária, do processo documentário.

Na obra de Deweze, o essencial da Informática Documentária não é a tecnologia a ser utilizada para a informação mas, ao contrário, a informação em primeiro plano (objeto) e os recursos tecnológicos que contemplem sua natureza e fluxo como pano de fundo (instrumento). Como este autor está pautado nos pressupostos da Documentação e da Ciência da Informação, parte do princípio do fluxo da informação e apresenta as soluções informáticas para este fim. Sendo assim, não fornece elementos para a gestão de bibliotecas, mas para a produção, tratamento e disseminação da informação e seus suportes, pela biblioteca ou centro de documentação. Ou seja, desenvolve o conceito e as técnicas necessárias para a gestão de informação pela biblioteca e não para a gestão das atividades da biblioteca.

Aqui cabe uma importante diferenciação entre os sistemas de recuperação de informação e os sistemas de gestão de bibliotecas. Por mais que esta distinção esteja se diluindo, cabe dizer que um bom sistema de gestão de bibliotecas deve desempenhar com eficiência e qualidade a recuperação da informação, mas um sistema de recuperação de informação não necessariamente tem por característica a gestão de uma biblioteca ou de um centro de documentação. No entanto, apenas atualmente os sistemas de gestão de bibliotecas vêm apresentando recursos mais aprimorados para

tratamento e recuperação da informação, enquanto, muitos dos assim chamados *softwares* documentários, têm historicamente acoplado módulos de gestão de biblioteca.

Quanto às ferramentas e recursos que executam total ou parcialmente a análise de conteúdo e a gestão do tesouro ou vocabulário controlado, são questões que aparecem sistematicamente na literatura francesa sobre automação, em oposição a maior parte da literatura inglesa. Nesta última, importam principalmente os dados ou os estoques a serem reunidos, selecionados, codificados, armazenados e disseminados. No entanto, as atividades essenciais de distribuição, disseminação e divulgação da informação ocorrem pela transferência da informação condicionada a elementos contextuais cognitivos, deflagrando a importância no tratamento dos temas/ categorias da informação.

Outro aspecto marcante nesta análise refere-se aos formatos. A literatura em francês cita o Formato MARC como uma especificidade do trabalho cooperativo das bibliotecas, sendo que, para as bases de dados documentários, o aproveitamento de registros é tarefa difícil. Não foram aqui investigadas, mais profundamente, as tendências das pesquisas realizadas em idioma francês sobre estas questões. Por outro lado, toda a literatura em inglês trata ampla e sistematicamente do Formato MARC como padrão de tratamento de dados para fins de compartilhamento entre as instituições. Neste sentido, os formatos de origem européia advindos dos Princípios UNISIST, como os formatos CEPAL, LILACS e CCF, foram criados com base em processos documentários e não de gestão de bibliotecas. Sendo assim, priorizam os itens que melhor atendem à catalogação analítica (tratamento de partes de documentos), incluem os documentos não convencionais (documentos audiovisuais, patentes, *folders*, *papers*, teses e outros) e oferecerem a possibilidade para a utilização de diversos tipos de descritores (primários e secundários, onomásticos, de região e cronológicos). Posteriormente, o Formato MARC também incluiu estas características, as quais foram pouco adotadas pelas bibliotecas.

Importante comentar que, nos três idiomas pesquisados (inglês, francês e espanhol), o termo mais amplamente utilizado é “formato de registro bibliográfico”, e não “formato de intercâmbio”, como adotado no Brasil. Acredita-se ter havido uma adaptação inadequada para o termo: considera-se que o fator relevante seja a metodologia de tratamento dos dados, legitimada e fundamentada, que constitui o formato de registro bibliográfico, sendo o intercâmbio de registros apenas uma das possibilidades conseqüentes da adoção de formatos comuns entre bases de dados. Vale investigar a origem do primeiro formato constituído (Formato MARC) que tinha inspiração na possibilidade do compartilhamento de registros, explicando talvez o enfoque na questão do intercâmbio no Brasil. Acrescente-se ainda que, com a informação eletrônica, outros formatos vêm sendo implementados como aqueles utilizados para textos, sons, imagens e multimídia.

Outro dado que reforça a dificuldade do idioma inglês na representação da idéia expressada pela Informática Documentária traçada pelos franceses aparece no levantamento de traduções de títulos de artigos franceses para o inglês em bases de dados acessadas pelo servidor Dialog. Foram encontradas as seguintes traduções para *Informatique Documentaire*:

- *documentary information science, computerization in information, computerization in information science* (CHONEZ, 1993);
- *computerized information, library automation* (KHOURI ST-PIERRE, 1989);

- *documentation and information technology, information and documentation technology, documentary information* (JAEGGER, 1971);
- *computers in documentation, information science, information processing, information retrieval* (DACHELET, 1990).

Do italiano *informatica documentaria*, traduziu-se para o termo em inglês *documentation*.

Entre os descritores dos registros destas bases, é necessária a composição de vários termos para dar conta de artigos que tenham como tema principal a *Informatique Documentaire*. Alguns dos termos utilizados são: *automation of library process, computerized information work, information retrieval, information science, library mechanization*.

Nos resumos, o termo original em francês não é traduzido, aparecendo entre aspas. Já o termo *documentaliste* tem sido traduzido por *documentary information scientist* ou até desconsiderado como termo relevante para tradução como no antigo nome da ADBS, Association Française des Documentalistes e des Bibliothécaires Spécialisées, traduzido em resumo por French Special Librarians Association.

Já a base de dados francesa Pascal apresenta os descritores das referências bibliográficas em inglês e francês sistematicamente. Assim, o descritor *Informatique Documentaire* é sempre traduzido pelo termo em inglês *documentation data processing*.

Ou seja, na literatura em língua inglesa, não foi localizada produção que discorresse sobre o tema com a mesma abrangência que a francesa, nem um termo equivalente a *Informatique Documentaire*. Como forma aproximada, foram encontrados os termos que designam características de *hardware* e *software* para automação das rotinas de biblioteca e os métodos desenvolvidos para recuperação da informação a partir das palavras de referências bibliográficas, resumos e textos integrais. A literatura inglesa aborda, quase sempre, os processos informatizados para a análise documentária em seus aspectos externos ou formais (representação descritiva) sendo bastante extensa, por exemplo, na questão relativa aos formatos de registro bibliográfico. Já os idiomas espanhol e francês parecem privilegiar o tratamento automatizado também nos processos referentes à análise do conteúdo (representação temática), abordando a gestão de vocabulários controlados, por exemplo.

Sendo a análise de conteúdo, a área que se tornou responsável pela evidência da informação enquanto linguagem, é essencial que o tratamento automático ou não da informação passe por esta constatação, para que o insumo maior da atividade documentária não passe despercebido: isto comprometeria seus resultados e sua própria razão de ser. Neste sentido, é bastante significativo que a análise de conteúdo esteja sendo considerada como elemento constituinte e essencial para a formulação da Informática Documentária. Pois um olhar possível para a constituição desta área é o que considera a Informática como instrumental para a Biblioteconomia e a Documentação e que transforma e amplia tanto a primeira quanto as segundas, fornecendo subsídios à constituição da Ciência da Informação.

Segue abaixo a análise das diversas tendências encontradas na literatura sobre Informática Documentária.

## **2.2 LIBRARY AUTOMATION E INFORMATION RETRIEVAL**

### **2.2.1 LIBRARY AUTOMATION**

#### **HISTÓRIA DA LIBRARY AUTOMATION**

A área reconhecida no idioma inglês, nos Estados Unidos e no Reino Unido, como *Library Automation* é também tratada pelos termos *computers in libraries*, *computers for libraries*, *automation for librarians*, *library technology* e, mais recentemente, *digital libraries* e *electronic libraries*.

Borgman (1997) apresenta a história dos trinta anos da automação de bibliotecas dos Estados Unidos e do Reino Unido. Acredita que, apesar de a história definitiva da automação de bibliotecas ainda estar sendo escrita, os eventos que a marcaram nestes países são os eventos da própria história da automação de bibliotecas. Os Estados Unidos e o Reino Unido compartilham um conjunto de padrões e práticas de operações de biblioteca, particularmente centrados na catalogação de dados, que serve como base para os sistemas de automação de bibliotecas, não só nestes países. Além disso, estão entre os mais avançados países em tecnologias da informação, com alta penetração de telecomunicações e tecnologia de computação em todos os setores econômicos e sociais. Assim, ao mesmo tempo em que as bibliotecas dos países do continente europeu têm muito em comum com as bibliotecas dos Estados Unidos e do Reino Unido, especialmente desde o advento dos programas cooperativos na União Européia, também há muitas diferenças entre elas: no continente europeu, muitos países empregam suas próprias regras de catalogação nacionais e as variações nas estruturas de registro bibliográfico complicam o intercâmbio de registros bibliográficos entre eles. Já os Estados Unidos e o Reino Unido são reconhecidos por fornecerem contribuições para outros países e para instituições internacionais.

As bibliotecas dos Estados Unidos e do Reino Unido iniciaram experiências com computadores nos anos 50, mas automatizaram de fato seus processos internos nos anos 60. Algumas das condições que facilitaram a automação de bibliotecas nestes países foram a tradição de compartilhamento e distribuição da catalogação e o acesso rápido a uma alta infra-estrutura de telecomunicações. Além dos modelos globais construídos e disseminados por estes países para a automação de bibliotecas, também foram pioneiros no estabelecimento de um mercado para sistemas de automação de bibliotecas, assim como, de associações de bibliotecas e instituições de serviços bibliográficos, além de fornecedores de produtos de automação e pesquisadores sobre automação de bibliotecas.

A questão “por que automatizar uma biblioteca?” provocou grandes debates nos Estados Unidos e no Reino Unido nos anos 60 e 70, no entanto, perguntas como “por que automatizar?” e “o que automatizar?” são difíceis de separar das questões “qual é a missão da biblioteca?” e “qual é a missão da organização em que está inserida a biblioteca?”.

Os objetivos da automação das bibliotecas dos Estados Unidos e do Reino Unido envolveram três fases: eficiência das operações internas, acesso aos recursos locais da

biblioteca e acesso aos recursos fora da biblioteca. O quarto e atual estágio tem como objetivo realizar a interoperacionalidade entre sistemas de informações necessários para construir uma infra-estrutura global de informação.

### 1 Eficiência das operações internas

A automação de bibliotecas tomou impulso primeiramente nos anos 60 com a expansão do ensino superior e o aumento das verbas para coleções de biblioteca. Com este crescimento, as bibliotecas perceberam que não poderiam adquirir e processar materiais pelos sistemas tradicionais e que automatizar poderia também ajudar a controlar os custos.

Assim, as bibliotecas ganharam eficiência por meio da automação em função do aperfeiçoamento do fluxo de trabalho dentro da biblioteca e do compartilhamento de dados entre bibliotecas. Bibliotecários e programadores aplicaram técnicas de análises de sistemas para identificar as operações que poderiam ser mecanizadas, para determinar relacionamentos entre as tarefas e para isolar os pontos nos quais a intervenção humana era necessária. Automação de bibliotecas, sistemas de controle de inventário, automação industrial e automação de escritórios tinham muito em comum nestes primeiros tempos.

Quanto ao compartilhamento dos dados, a idéia era que, sendo a catalogação uma das operações mais caras em bibliotecas e materiais comuns existiam em bibliotecas distintas, o registro de um material realizado por uma biblioteca poderia ser aproveitado por várias outras bibliotecas. Contudo, materiais únicos tais como aqueles de coleções especiais, arquivos e museus poderiam ser catalogados usando as mesmas regras. A mudança para o compartilhamento de registros na forma legível por computador é considerada como sendo, provavelmente, a transição mais significativa da preocupação local para a global na história da automação de bibliotecas. Atualmente, bibliotecas podem construir suas bases de dados locais ou transferir seus acervos para o meio automatizado, obtendo os registros dos sistemas de catalogação cooperativa existentes.

### 2 Acesso aos recursos locais da biblioteca

No começo da automação, os computadores eram visíveis aos usuários apenas na tarefa de circulação. Com a queda do preço dos computadores na década de 80, as bibliotecas passaram a operar todas as funções *online*, aperfeiçoadas pelos sistemas integrados. Este período foi marcado pela criação de um mercado de sistemas de automação de biblioteca, implementação de catálogos *online* e conversão retrospectiva de registros. A existência de grandes bases de dados bibliográficos *online*, de padrões de representação e de conteúdo de dados para intercâmbio de registros e de ligações entre estas bases de dados formam a fundação para a criação do acesso universal a registros de todos os materiais publicados.

### 3 Acesso aos recursos fora da biblioteca

Depois da expansão do ensino superior nos anos 60, as verbas para aquisição de materiais de acervo começaram a diminuir mas, ao mesmo tempo, a produção de recursos de informação em todos as mídias continua em expansão. Tendo como algumas das motivações, o fator econômico e a disponibilidade de tecnologia, as bibliotecas começaram a empregar tecnologias de comunicação para acessar coleções de outras bibliotecas.

Como resultado da automação de bibliotecas, três desenvolvimentos no acesso aos recursos externos às bibliotecas são apontados: novas formas de identificar, localizar e obter documentos; intercâmbio de dados *online*; e integração de coleções locais com outros tipos de recursos de informação.

Nas novas formas de identificar, localizar e obter documentos estão em questão os novos desenvolvimentos de empréstimos entre bibliotecas. O termo “empréstimo entre bibliotecas” adotado inicialmente em inglês como *interlibrary loan* e *interlending* está sendo substituído por “entrega de documentos” (*document delivery*). Este termo, mais geral, reflete o uso do *e-mail*, a transferência de arquivos por redes de computadores e o envio de documentos por aparelhos de fac-símile, mais que o simples pedido de empréstimo de materiais a outra biblioteca e o envio pelo correio.

Quanto ao intercâmbio de dados *online*, a gradual convergência de tecnologias de computação e de comunicações vem aperfeiçoando o acesso aos recursos de informação dentro e fora das bibliotecas. Agora, todos os maiores sistemas de automação de bibliotecas vendidos nos Estados Unidos possibilitam a transferência direta de registros bibliográficos entre os serviços bibliográficos e os sistemas integrados locais.

No aspecto da integração de coleções locais com outros tipos de recursos de informação, os catálogos *online*, os sistemas de acesso público a bases de dados e muitos outros recursos *online* para os quais as bibliotecas fornecem acesso podem ser considerados bibliotecas digitais. Embora o termo “bibliotecas digitais” seja datado do final dos anos 80, é especialmente problemático porque obscurece o complexo relacionamento entre coleções de informações eletrônicas e bibliotecas como instituições. As bibliotecas digitais são definidas como os conjuntos de recursos e capacidades técnicas associadas para criação, pesquisa e uso de informação, contendo informação coletada e organizada a favor de uma comunidade de usuários e incluindo funcionalidades que atendam as necessidades e usos de informação desta comunidade. Podem consistir de dados bibliográficos apenas ou da combinação destes dados com o conteúdo dos recursos de informação que descrevem ou outras formas de informação digital ou digitalizada. Cada vez mais, as bibliotecas fornecem acesso a outros recursos de informação que vêm sendo denominados de bibliotecas digitais como catálogos *online*, sistemas de acesso público *online*, bases de dados em CD-ROM e outras bases de dados.

#### 4 Interoperacionalidade de sistemas de informações necessários para construir uma infra-estrutura global de informação

Trata-se da evolução da atividade de intercâmbio de dados bibliográficos para a interoperacionalidade entre os diversos sistemas de informação (ou bibliotecas digitais), implicando em uma nova etapa para as bibliotecas na transição da preocupação local para a global no acesso à informação.

Embora seja possível trocar dados bibliográficos entre sistemas de computador, cada processo de intercâmbio é planejado com cuidado. Enquanto os formatos MARC determinam o mapeamento entre campos para troca de dados e são reconhecidos padrões técnicos, o conteúdo destes campos é determinado por regras de catalogação, que são antes diretrizes que padrões técnicos. Assim, as diferenças quanto à interpretação destas regras são minimizadas com o compartilhamento de registros, uma vez que as formas de interpretação tendem a se unificar.

No que se refere às interfaces de consulta, tradicionalmente os sistemas locais exigem determinado conhecimento para uso enquanto cada sistema externo fornece novas formas de consulta. Apenas recentemente tornou-se possível o relacionamento genérico entre sistemas de informação do tipo cliente-servidor. O mais avançado é o padrão Z39.50 que está sendo implementado para catálogos *online* e outros sistemas de recuperação bibliográfica. Este padrão habilita um sistema local a atuar como um cliente de outro sistema, de tal forma que o usuário local possa pesquisar o sistema remoto por meio da interface local. Estes padrões ainda não possibilitam a pesquisa em múltiplos sistemas ao mesmo tempo e a mescla de resultados em uma mesma apresentação dos dados.

Para a interoperacionalidade entre bibliotecas digitais, a quantidade e o tipo de informação que está sendo trocada é mais complexa. Além da troca de registros catalográficos altamente estruturados ou de registros bibliográficos contendo dados de resumo e de indexação, agora estão sendo trocados recursos de informação de texto integral com metadados associados, além de dados como características técnicas de computadores relativas aos parâmetros de troca, pesquisa, apresentação e autorização de usuários. A preocupação está mudando da troca de registros bibliográficos entre sistemas integrados locais e entre sistemas locais e serviços bibliográficos para a interoperacionalidade entre bibliotecas digitais.

A definição da interoperacionalidade de sistemas de computadores em rede é, no entanto, uma questão aberta. Embora bibliotecas sejam líderes em interoperacionalidade, tendo estabelecido padrões no final dos anos 60 para troca de catálogos e de outros registros bibliográficos na forma legível por máquina, ainda é o início do entendimento e da resolução do problema da interoperacionalidade *online* e global. Contudo, o movimento no sentido dos catálogos *online* para os sistemas de acesso público *online* e então para as bibliotecas digitais, é também o movimento do gerenciamento de dados bibliográficos para o gerenciamento do conteúdo dos recursos de informação na forma digital.

O desafio da interoperacionalidade reflete a necessidade de que as bibliotecas pensem globalmente, trocando dados com outras instituições como bibliotecas, serviços bibliográficos, vendedores de sistemas, editores, governos, outros fornecedores de informação e pesquisadores individuais. Esta tarefa requer sistemas que satisfaçam padrões internacionais e que sejam baseados em cooperação, troca e compartilhamento de recursos. Ao mesmo tempo, as bibliotecas necessitam desenhar sistemas que contemplem as necessidades de informação de suas comunidades locais e atendê-las de forma eficiente.

### **Sobre a história da *Library Automation***

Esta história da automação de bibliotecas fornece um quadro dos eventos que a constituíram e, ao mesmo tempo, indica os desafios atuais, realizando considerações que apontam para o cerne destes problemas.

Neste sentido, a autora considera que as bibliotecas são líderes em interoperacionalidade por terem estabelecido padrões no final dos anos 60 para troca de catálogos e outros registros bibliográficos na forma legível por máquina, mas ainda assim, há muitos problemas a resolver. Alguns destes problemas provavelmente têm

relação com a tendência de que o gerenciamento de dados bibliográficos esteja se transformando no gerenciamento do conteúdo dos recursos de informação na forma digital. Esta afirmação deixa dúvida quanto a que tipo de conteúdo se refere: o conteúdo intelectual ou formal do documento. Se for o primeiro caso, então parece estar havendo um reconhecimento da necessidade da representação de conteúdos para possibilitar acessos efetivos.

Mas, em termos gerais, o grande desafio é apontado como a dificuldade evidente da incompatibilidade entre as tendências globalizantes de organização da informação e o atendimento de necessidades locais por informação. O modelo de comunicação de massa que é reforçado pelo pensamento e ação globais da atualidade apenas reforça ou evidencia o pensamento e ação dos segmentos locais representados pelas comunidades.

Os sistemas cooperativos de catalogação entre bibliotecas dos Estados Unidos, além dos consórcios de bibliotecas para aquisição de materiais e outras formas de organização em grande escala, foram estabelecidos por meio de uso compartilhado e comum de padrões e sistemas, conformando-se segundo a teoria da comunicação de massa. Esta teoria foi estudada e disseminada pela Escola de Chicago (que será tratada no capítulo 3.1) e sua aplicação caracteriza a forma de funcionamento da sociedade estadunidense. É também um modelo em expansão para boa parte do mundo, influenciando nos serviços de informações dos diversos países.

Já a informação institucionalizada é baseada na prestação de serviços de informação dirigidos a segmentos de públicos determinados e a objetivos e necessidades institucionais, implicando em tratamento e disseminação documentária pautada nestas especificidades.

No âmbito dos Estados Unidos, as estratégias envolvidas na comunicação da informação documentária se apresentam nos repositórios nacionais de registros catalográficos legíveis por máquina e nas bases de dados oferecidas por instituições especializadas das diversas áreas do conhecimento. Os registros catalográficos são, essencialmente, a representação descritiva do conteúdo de livros e periódicos impressos tratados no todo, cada vez mais, incluindo os diversos tipos documentais em suas mais variadas formas, inclusive a eletrônica. O tratamento superficial do conteúdo tem sido predominante, apesar de o Formato MARC, que é predominante, contemplar maior complexidade neste aspecto. Estas informações, assim tratadas e disseminadas, contemplam a comunicação de massa pois se referem, quase sempre, a publicações tradicionais e de grande uso em todo o país, como os itens de bibliografias dos cursos universitários. Ao mesmo tempo, as bibliotecas que tratam seus acervos pela catalogação cooperativa dispõem a seus usuários acesso a muitas outras bases de dados com tratamento documentário mais elaborado, sendo que, muitos destes registros referem-se a materiais existentes nas próprias bibliotecas. Estas bases de dados contêm as assim chamadas informações institucionalizadas, no sentido de que focam as necessidades de determinada instituição ou grupo de usuários em uma área do conhecimento, como educação ou energia nuclear, ou em área de atuação específica como a de negócios.

Assim, no trabalho das bibliotecas, o tratamento deve ser massificado (ou generalizado) para suprir grandes volumes de informação e um conjunto amplo de usuários, justificando, com isso, investimentos e esforços cooperativos. Também são as bibliotecas que realizam a guarda e a preservação de acervos. Já as bases de dados *online* possibilitam o acesso a informações diversas (inclusive aquelas contidas nos documentos das bibliotecas). O catálogo da biblioteca permite o acesso a temas

genéricos ou a itens de acervo já conhecidos e as bases de dados especializadas possibilitam a pesquisa retrospectiva e específica sobre um tema. Este quadro denota uma fragmentação entre organização de acervo (comunicação de massa) e organização da informação (comunicação segmentada).

O questionamento fica por conta do nível possível de compatibilização entre estes distintos universos comunicacionais, mesmo com as facilidades que a evolução tecnológica vem fornecendo. Isto porque, a contextualização cognitiva do repertório de informações necessária para o acesso pertinente com esforço e tempo razoáveis, ainda não é realizada exclusivamente por máquinas.

Já em países como o Brasil, os catálogos de bibliotecas crescentemente automatizados são, muitas vezes, decorrências de esforços locais. As bibliotecas, em sua maioria, estão longe das características de um centro de documentação e inexistem serviços de indexação e resumos a não ser aqueles oferecidos no exterior ou decorrentes de sistemas internacionais de informação dos quais alguns centros no país fazem parte. Como consequência das atividades de tratamento de dados e de atendimento ao público realizadas de forma massificada, são poucos os serviços de disseminação seletiva da informação e demais serviços ou formas de organização que atendam a perfis de usuários. Além disso, muitas informações estão sendo disponibilizadas aleatoriamente na Internet. Considerando a contínua influência do modelo dos Estados Unidos e o quadro nacional acima, que serviços de informações estão sendo implementados no Brasil?

### **O QUE É *LYBRARY AUTOMATION*: “INTRODUCTION TO AUTOMATION FOR LIBRARIANS”, William Saffady, 1999**

A obra de Saffady, na quarta edição, de 1999, abarca de forma sistemática e abrangente os aspectos tradicionalmente envolvidos na automação de bibliotecas. Saffady, professor da Palmer School of Library and Information Science, da Long Island University, cita que a intenção é atender bibliotecários, especialistas de informação, estudantes de Biblioteconomia e outros que desejam um panorama destes aspectos da tecnologia da informação que são mais significativos para as operações em bibliotecas. Entre outros, o livro apresenta produtos, sistemas e serviços específicos como exemplos de abordagens particulares de automação. O autor enfatiza que o livro trata somente de tecnologia omitindo, intencionalmente, alguns aspectos que considera importantes na automação de bibliotecas como a relação entre automação e recursos humanos e as implicações da automação para o treinamento e desenvolvimento da equipe.

#### **Introdução**

O trabalho foi organizado em duas partes principais. A primeira parte trata dos fundamentos gerais sobre computação e tecnologias correlatas, englobando características de *hardware* e de *software* e conceitos sobre gerenciamento e comunicação de dados e sobre sistemas de automação de escritório. Discute aspectos fundamentais de computadores e tecnologias correlatas, enfatizando conceitos e termos usuais considerados essenciais para bibliotecários que devem se comunicar com pessoal

de processamento de dados, vendedores de equipamentos, desenvolvedores de *softwares*, especialistas de telecomunicações ou outras pessoas envolvidas com desenho e implementação de sistemas automatizados de informação. A segunda parte cobre os conceitos sobre automação de bibliotecas e discute os sistemas e serviços que automatizam as operações de biblioteca. Trata da catalogação descritiva, dos sistemas integrados de biblioteca (sistemas automatizados de controle de circulação, de aquisições e controle de publicações seriadas), do serviço de referência automatizado e das bibliotecas digitais.

## **Parte I - Aspectos fundamentais de computação e tecnologias correlatas**

### ***Hardware***

Um sistema de computador é composto de três grupos principais de componentes: *hardware* (equipamento), *software* (programas que fazem com que o equipamento execute operações específicas) e dados (informações que o *hardware* e o *software* processam, armazenam e recuperam, ou seja, que manipulam). O **hardware** é constituído de um processador central (*central processing unit*, ou CPU) e de dispositivos periféricos associados.

O **processador central** é composto internamente de uma seção de controle, uma seção aritmética/ lógica e uma seção de memória. A seção de controle dirige as outras duas seções, assim como, a relação do processador central com seus dispositivos periféricos associados. A seção aritmética/ lógica contém o circuito eletrônico especializado essencial para a computação, sendo que nas aplicações de armazenamento e recuperação da informação orientadas a bibliotecas, as operações lógicas são mais importantes que os cálculos aritméticos. Ou seja, as aplicações de bibliotecas não requerem muitos cálculos numéricos mas precisam efetuar repetidamente operações lógicas que supõem a verificação de determinadas condições ou a comparação de dados. A seção de memória fornece armazenamento para dados e programas no próprio processador central.

Historicamente, as tipologias de computadores segundo seus processadores incluem as categorias: computadores de grande porte, minicomputadores e microcomputadores, dependendo de tamanho físico, poder de processamento, aplicações e preço. Contudo, os avanços na tecnologia eletrônica foram tornando menos claras estas categorias.

Enquanto o processador central efetua as tarefas computacionais propriamente ditas, os dispositivos periféricos realizam tarefas específicas de entrada, saída e armazenamento auxiliar.

Os **periféricos de entrada** convertem a informação compreensível para seres humanos em informação que o processador central possa interpretar. Os periféricos mais utilizados para aplicações de bibliotecas e processamento de dados de negócios são o teclado e as tecnologias de *scanner*. Como alternativa aos procedimentos de digitação, ou seja, de entrada de dados caractere a caractere, as tecnologias de reconhecimento ótico escaneiam os documentos, convertendo seu conteúdo na forma legível por máquina automaticamente. O reconhecimento ótico de caracteres (*optical character recognition* ou OCR) usa a luz refletida para identificar os caracteres individuais contidos nos

documentos. Para tanto, os documentos devem estar digitalizados pelo processo de conversão de documentos textuais ou gráficos para imagens eletrônicas.

Os **periféricos de saída** convertem os dados legíveis por máquina em informações que os seres humanos possam processar. Em aplicações de biblioteca, os periféricos mais utilizados são aqueles em que a informação resultante pode ser impressa em papel, registrada em microfilme ou apresentada em tela. Em todos eles, as habilidades são personalizadas: alguns são capazes de apresentar caracteres acentuados, alfabetos não romanos e outros símbolos encontrados em dados bibliográficos. Há impressoras especialmente projetadas para a saída de gráficos, os *plotters*, adotados em aplicações científicas e de engenharia. O *computer output microfilm* (COM), ou saída de computador em microforma, é uma forma variante de tecnologia de saída sem impacto na qual a informação textual ou gráfica é registrada em microfilme. Os visualizadores de vídeos mostram os resultados do processamento computadorizado em uma tela. Muitos modelos incorporam um tubo catódico de televisão (*cathode ray tube*, ou CRT) como mecanismo de apresentação e, dentre os usos desta tecnologia, estão as telas de cristal líquido (*liquid crystal displays*, ou LCDs), grandemente adotadas nos computadores portáteis.

Os **periféricos de armazenamento auxiliar** são designados para reter informação na forma legível por máquina para processamento ou reprocessamento, funcionando como extensão para a memória limitada do processador central. Enquanto os primeiros sistemas de computador usavam papel como meio de armazenamento, fitas e discos magnéticos têm dominado o armazenamento auxiliar nas últimas décadas. Os discos magnéticos como os *hard disks* (discos duros) são os preferidos como meio de armazenamento, especialmente, quando dados devem ser continuamente disponíveis para rápido acesso, enquanto os *floppy disks* são discos magnéticos removíveis usados para simples instalações em computadores. *Drives* de fitas magnéticas eram utilizados antes dos *hard drives* como o principal dispositivo de armazenamento auxiliar dos primeiros computadores; as fitas magnéticas continuaram sendo utilizadas desde então para *backup* de segurança e arquivamento de dados, além da distribuição e intercâmbio de dados. Discos óticos, apenas para leitura ou para leitura e escrita, destacam-se por sua alta capacidade de armazenamento, como os CD-ROMs (Compact Disc-Read Only Memory) bastante utilizados para edição de bases de dados e outras aplicações relacionadas a bibliotecas.

### **Software**

Entende-se por **software** o programa ou seqüências pré-definidas de instruções que um computador executa para realizar operações de processamento da informação. Estas instruções são baseadas em **linguagens de programação** que são apresentadas à frente por Costa carballo (1995).

Vários **programas de microcomputador** são considerados de interesse para bibliotecas e aplicações de negócios:

- pacotes de programas de processamento de texto (ex.: Microsoft Word, WordPerfect da Corel, Word Pro da IBM);
- pacotes de programas pré-escritos para *desktop publishing* (ex.: PageMaker e Photoshop da Adobe Systems, Microsoft Publisher, além de outros editores para edição de páginas na World Wide Web);

- planilhas de cálculo (ex.: Excel da Microsoft e Lotus 1-2-3 da IBM);
- pacotes de gerenciamento de dados que permitem a criação e manipulação de registros legíveis por máquina armazenados em arquivos de dados. Entre os mais usados estão os programas Access e FoxPro da Microsoft, FileMaker Pro da FileMaker, Oracle da Oracle Corporation, Paradox da Corel e a linha de produtos dBase da Borland International. Enquanto **programas de gerenciamento de bases de dados convencionais** operam sobre registros estruturados que são organizados em campos, um grupo de pacotes de *software* de microcomputadores pode manipular segmentos não estruturados. Tais programas são conhecidos como **sistemas de armazenamento e recuperação de textos, sistemas de gerenciamento de dados baseados em textos, sistemas de informações textuais e sistemas de recuperação de texto integral (full-text)**. Os segmentos de texto podem consistir de documentos completos ou de pequenas representações de documentos como resumos ou anotações. Os índices permitem pesquisar os textos integrais pela identificação rápida de segmentos de textos que contém cadeias de caracteres especificados. Exemplo deste tipo de programa é o Inmagic DB/TextWorks da Inmagic Incorporated;
- programas de apresentação de gráficos;
- pacotes de *softwares* de comunicação que possibilitam a um computador pessoal funcionar como um terminal com o fim de se comunicar com um dispositivo remoto. Os mais amplamente utilizados são Microsoft Internet Explorer e Netscape Navigator;
- pacotes de *software* integrado combinam processadores de texto, planilhas de cálculo, gerenciamento de dados e capacidade de comunicação em um único produto, como o Microsoft Works (não confundir com os sistemas integrados de biblioteca).

Enquanto o *software* de sistema sempre foi adquirido junto com o *hardware*, o desenvolvimento do *software* de aplicação era considerado de responsabilidade do usuário. Em quase todos os casos, o desenvolvimento de um *software* começa com um estudo, por um analista de sistemas, das características e requisitos da aplicação. Normalmente daí resulta um relatório com recomendações para melhorar ou substituir o sistema estudado, incluindo especificações de *hardware* e *software* indicados. Estas especificações do *software* são desenhadas, implementadas e testadas por programadores. Com o uso do programa, falhas e erros aparecem, por não terem sido testados o suficiente, após o que, o programa é analisado e as instruções são corrigidas. Este processo, chamado manutenção, é facilitado se houver documentação adequada do programa, efetivada ao se escrever o mesmo. No entanto, como alternativa ao enorme custo econômico de desenvolvimento destas aplicações específicas, pacotes de *softwares* passaram a estar disponíveis no mercado, como processadores de textos, gerenciadores de dados e aplicações para apresentação gráfica, comunicação e outras tarefas específicas.

### **Conceitos de gerenciamento de dados e comunicações de dados**

O gerenciamento e a comunicação de dados envolvem o modo pelos quais os componentes de um sistema de computador interagem para processar, manipular e distribuir dados.

Os termos **online** e **offline** indicam, respectivamente, a presença ou ausência de conexões elétricas ou outras entre dispositivos de computação. Quando usados para descrever configurações particulares de computadores, os adjetivos *online* e *offline* são algumas vezes confundidos com outros dois termos: *real time* (tempo real) e *batch* (em lote). Estes dois referem-se ao modo de processamento de dados (relação entre transações ou outros eventos e processamento de informações sobre estes eventos pelo computador) e não especificamente ao fato de haver conexão elétrica ou não.

Sendo assim, no **processamento de dados em lote** (*batch processing*), há um intervalo de tempo entre a ocorrência de uma transação ou qualquer outro evento e o processamento do computador da informação relativa a este evento. Sistemas que desempenham este modo de processamento são tipicamente caracterizados por coleta de dados *offline* com armazenamento em intervalos pré-determinados em dispositivos de armazenamento massivo, pelo procedimento direto dos dados armazenados em fita e pela geração de relatório em papel ou outros suportes. Estes relatórios refletem a condição de um sistema em um dado momento, não incluindo as transações ou eventos ocorridos desde então. Nos sistemas de **processamento em tempo real**, as informações sobre transações ou outros eventos são introduzidas e processadas no momento em que o evento ocorre, mantendo estas informações sempre atualizadas. Nestes sistemas, a informação é armazenada em disco ou em outros dispositivos de acesso direto e recuperada por meio de terminais conectados *online*.

Atualmente, especialistas em computador estão crescentemente voltados para o desenvolvimento de **sistemas interativos**, nos quais o computador guia o usuário na execução das várias tarefas de processamento da informação. Programas interativos bem desenhados são simples de usar, prevêm erros dos usuários e permitem que estes controlem o ritmo da interação.

Enquanto programas interativos proporcionam a interface intelectual entre o usuário e o computador, o **terminal online** fornece a necessária interface física por ser um dispositivo que permite ao usuário transmitir e receber informações de um computador ou de outra máquina de processamento de informações. Os terminais para aplicações interativas podem ser classificados em função da complexidade de seu circuito eletrônico, de seus recursos de saída e da natureza de seus componentes (ou características de comunicação). Quanto ao seu circuito eletrônico, os terminais podem ser burros, espertos ou inteligentes: terminais burros utilizam todas as capacidades de processamento de informação do computador a que estão conectados por não possuírem estas funções; terminais espertos são versões melhoradas dos terminais burros ou não programáveis; e os terminais inteligentes são microcomputadores programáveis que podem também funcionar como terminais. Com referência a seus recursos de saída, os terminais podem ser impressoras ou telas. Em função de seus componentes, os terminais podem ser configurados com teclado para comunicação interativa, sem teclado para apenas receber dados ou com um dispositivo de gravação ou impressão dos dados recebidos de um computador hospedeiro.

Independente do modo de processamento ou tipo de terminal utilizado, os dados armazenados em fita ou disco devem ser organizados em formatos estruturados. Em termos de **estrutura física**, dados são armazenados em um dispositivo dado, tais como, um disco de *drive* ou uma fita, e a unidade de dados que pode ser lida ou gravada em uma única operação é chamada de registro físico. Assim, os dados físicos implicam o armazenamento físico dos dados em conjuntos de arquivos. Em termos de **estrutura lógica**, a informação (dados, textos ou imagens) é tipicamente organizada em arquivos.

Cada arquivo consiste de vários registros lógicos que, por sua vez, se subdividem em campos.

Em muitas aplicações, arquivos de dados são especificamente desenhados para satisfazer necessidades de um programa de aplicação determinado, originando dados redundantes, desperdício de espaço de armazenamento e outros problemas. Assim, especialistas em computação recomendam o uso de **bases de dados** com a integração de arquivos independentes eliminando as redundâncias, reduzindo os custos de armazenamento e melhorando a precisão e a integridade dos dados. A implementação de uma base de dados integrada é realizada a partir de um conjunto de programas denominado **sistema de gerenciamento de bases de dados (SGBD)**. Alguns SGBDs utilizam linguagens, como as de quarta geração, que simplificam o desenvolvimento da aplicação e permitem recuperação da informação sem necessidade de programação formal.

Os **sistemas de armazenamento e recuperação de textos** são uma categoria especial de SGBDs. Tais como os sistemas de gerenciamento de bases de dados convencionais, estes programas criam, mantêm e manipulam arquivos de registros processados pelo computador, mas os registros contêm informação textual não estruturada de documentos completos. Implementações experimentais de armazenamento e recuperação textual datam da metade da década de 60. Tipicamente concebidos como projetos de demonstração designados para testar a eficácia de metodologias específicas de indexação e recuperação, eles envolveram baixo número de documentos, aplicações de abrangência reduzida e grupos de usuários bem controlados. Aplicações do gênero para bibliotecas foram implementadas no início dos anos 70 quando serviços de informação *online* introduziram armazenamento e recuperação de texto integral. Tais implementações foram inicialmente limitadas a pequenos registros bibliográficos, alguns acompanhados de resumos. O serviço de informação Lexis (atual Elsevier) foi o primeiro sistema a prover acesso *online* a texto completo de grandes coleções de documentos. Pacotes de *software* pré-escritos com recursos de armazenamento e recuperação tornaram-se disponíveis durante os anos 60 para computadores de grande porte. Produtos para mini e microcomputadores cresceram significativamente durante as duas décadas seguintes. Bibliotecas, arquivos e organizações correlatas têm utilizado programas de armazenamento e recuperação de textos para gerenciar uma variedade de documentos, incluindo relatórios científicos, técnicos e gerenciais, livros de notas (*notebooks*) de laboratório, patentes, apresentação de conferências e *preprints*. A maior parte das implementações tem sido em bibliotecas industriais, técnicas, médicas e jurídicas, onde complexas operações de recuperação exigem poderosos recursos de pesquisa.

Programas de armazenamento e recuperação de textos podem pesquisar o conteúdo completo de documentos por palavras específicas ou por raízes de palavras. Operadores booleanos podem ser usados para especificações de recuperação extensas ou restritas. A maior parte destes programas também possui comandos que podem localizar documentos contendo duas ou mais palavras em uma relação de proximidade especificada. Dependendo dos programas, tais comandos podem ser capazes de recuperar duas ou mais palavras no mesmo parágrafo, na mesma sentença ou em um número especificado de palavras em um parágrafo ou sentença. Um número crescente de programas de armazenamento e recuperação de textos oferece recursos de hipertexto, os quais ligam documentos relacionados e reconhecem palavras em segmentos de textos previamente selecionados para serem usados como termos de pesquisa. Alguns programas recuperam documentos ou segmentos de documentos por

sua presumida relevância, apresentando-os segundo a ordem desta relevância; o índice de relevância costuma ser medido com base na frequência e proximidade dos termos de pesquisa que o documento contém.

Uma alternativa menos comum refere-se aos processadores associativos: em vez de consultar índices para localizar arquivos, registros com termos especificados na pesquisa examinam seqüencialmente arquivos de texto. Em alguns aspectos, os processadores associativos representam uma reversão a velhas tecnologias, no entanto, antigos sistemas de recuperação de informação que empregavam pesquisa seqüencial de arquivos de fitas magnéticas estão sendo recentemente utilizados para *softwares* de alta performance que contam com índices e dispositivos de armazenamento de acesso direto. Os processadores associativos permitem certas formas de recuperação impraticáveis pelo acesso indexado. Como um exemplo, estes programas são utilizados por corporações e agências governamentais para pesquisar registros que contenham assuntos específicos, nomes de pessoas e de companhias e outros segmentos de textos.

Atividades de computação centralizada têm sido freqüentemente criticadas por falhar para responder satisfatoriamente a requisitos de processamento de informação de usuários remotos. Por este motivo, desde os anos 70, muitas organizações adotaram sistemas de **processamento distribuído de dados (redes)** que são desenhados para dispor os recursos de computação mais próximos do usuário final. Um sistema distribuído é uma configuração integrada de computadores para efetuar uma aplicação específica e que pode ser organizada horizontal ou hierarquicamente.

Computadores interconectados, facilidades de comunicação e dispositivos correlatos em um sistema de informação distribuído constituem uma rede de computadores. Redes de computadores variam em suas tipologias, abrangência geográfica e facilidades de comunicação.

*Wide-area networks* (WANs) não têm restrição geográfica e podem utilizar combinações de linhas telefônicas analógicas ou digitais, linhas telefônicas analógicas ou digitais dedicadas ou tecnologias de comunicação de rádio tais como microondas terrestres ou transmissão de dados celulares. WANs podem ser de acesso público ou implementadas por uma empresa, agência governamental ou outra organização para seu próprio uso. A Internet é o mais importante exemplo de WAN de acesso público. É atualmente uma confederação de redes de computadores independentes que são operados por uma variedade de organizações. A World Wide Web (WWW), ou apenas Web, um subconjunto da Internet baseada na tecnologia cliente-servidor, contém páginas formatadas de informação que são desenhadas para serem visualizadas por um *software* do tipo *browser*. Intranets e Extranets são redes de computadores privadas que empregam tecnologias de Internet, sendo que, as Extranets designam um acesso não limitado à empresa mas estendido a seus interlocutores privilegiados como fornecedores e clientes.

*Local area network* (LAN) se distinguem por sua menor abrangência geográfica, entre outras características. Elas normalmente utilizam pares trançados como meio de conexão, mas cabo coaxial, fibra ótica e ligações de rádio também são possíveis. Entre suas vantagens, as LANs permitem compartilhar recursos de *hardware*, de *software* e informações. Muitos sistemas de automação de bibliotecas e produtos de informações relacionados são disponíveis em LAN.

## Sistemas automatizados de escritório e tecnologias correlatas

Por considerar que as atividades em bibliotecas têm um significativo componente de rotinas de escritório, as tecnologias adequadas para trabalhos administrativos são de grande importância. O grupo dos sistemas automatizados de escritório e outros produtos relacionados incluem:

- editores de textos;
- tecnologias de micrográfica relacionadas aos computadores como o *computer output microfilm* (COM) que usa microformas como alternativa ao papel para a produção e distribuição de registros gerados em computador e os sistemas de *computer assisted retrieval* (CAR) que usam computadores para criar, manter e pesquisar índices *online* de imagens de documentos registrados em microfilmes, combinando as vantagens do pouco espaço ocupado pelas microformas com a habilidade dos computadores na manipulação de índices de dados;
- sistemas de imagens de documentos eletrônicos que utilizam *scanners* para converter documentos em imagens digitalizadas apropriadas para armazenamento, recuperação e distribuição em computador;
- copiadoras e duplicadoras;
- mensagens eletrônicas (*e-mails*) e sistemas de mensagens como fac-símile e telex;
- tecnologia de vídeo em combinação com computadores ou não: televisão de alta definição, vídeo-discos óticos, *digital video disk* (DVD), vídeo-conferência, além dos vídeo-textos que, atualmente competem com a Web que fornece o mesmo tipo de informação em um formato mais atrativo (em alguns países da Europa e de outros locais a tecnologia de vídeo-texto recebeu subsídio governamental; nos Estados Unidos, foi implementada em base experimental mas sua viabilidade comercial nunca foi demonstrada).

## Parte II - Automação de Bibliotecas, sistemas e serviços

### Computadores e catalogação descritiva

Computadores foram aplicados pela primeira vez para a impressão de fichas catalográficas. Em seguida, as bibliotecas sentiram falta de também automatizar a produção, manutenção e uso de seus catálogos, devido aos problemas e limitações verificados com o uso de catálogos convencionais em fichas impressas, como: considerável espaço ocupado; necessidade de mobiliário de preço elevado; rotinas de manutenção que exigiam muitas pessoas e horas de trabalho, aumentadas com as mudanças de regras de catalogação; e possibilidade de conforto e de recursos de recuperação limitados sob o ponto de vista do usuário. Assim, os registros que já estavam em formato computadorizado passaram a ser processados pela técnica de processamento em lotes para a geração de catálogos em forma de livro ou armazenados em disco para acesso *online*. Nos catálogos em forma de livro, a manutenção dos arquivos era realizada automaticamente, cópias podiam ser distribuídas a usuários remotos e mais espaço físico ficava disponível. No entanto, não refletiam a atualidade

dos dados registrados, diferente dos catálogos *online* que já ofereciam recuperação mais rápida e variada que as fichas catalográficas impressas e as fichas em forma de livro.

O grande impulso para a automação das bibliotecas dos Estados Unidos ocorreu no final da década de 60, quando a Library of Congress desenvolveu o **Formato MARC**, agora conhecido como USMARC, para a comunicação de informação bibliográfica na forma legível por máquina. Tornando-se a principal instituição de catalogação de dados, passou a distribuir estes registros bibliográficos em meio magnético desde então. A acumulação de tais registros é conhecida como a base de dados LCMARC.

O escopo do Programa MARC, que foi inicialmente limitado às monografias de língua inglesa, passou a incluir monografias em outras línguas (inclusive línguas com o uso de alfabetos não romanos e línguas ideográficas), além de vários formatos MARC para os diversos tipos de materiais de bibliotecas, como livros, publicações seriadas, mapas, materiais visuais e de arquivos, manuscritos, partituras de música e arquivos de computador.

Outras bibliotecas nacionais estabeleceram formatos MARC similares e programas de distribuição para registros catalográficos legíveis por máquina. No entanto, há considerável variação nos designadores de conteúdo dos formatos MARC desenvolvidos pelas bibliotecas nacionais de países como Canadá, Grã-Bretanha, França, Alemanha, Itália, Austrália, Suécia, Japão e outros países. O Formato MARC canadense (CANMARC), por exemplo, usa mais designadores de conteúdo que aqueles permitidos pela Library of Congress. Por sua vez, os formatos MARC do Reino Unido (UKMARC) e da França (INTERMARC) são ainda mais abrangentes neste aspecto que o formato canadense. Por este motivo, a International Federation of Library Associations and Institutions (IFLA) criou o Formato UNIMARC para facilitar o intercâmbio internacional de dados bibliográficos na forma legível por máquina.

Enquanto fitas MARC que contêm registros catalográficos legíveis por máquina podem ser comprados da Library of Congress, muitas bibliotecas obtêm acesso aos dados MARC através de produtos e serviços oferecidos por editores e outros provedores de informação como os serviços bibliográficos. Entre estes produtos, catálogos em CD-ROM incluem total ou parcialmente a base de dados LCMARC, muitas vezes acrescentados de registros bibliográficos de outras fontes como a National Library of Medicine, a National Library of Canada e a British Library. Comparados com outros métodos de catalogação, sistemas em CD-ROM fornecem uma relação custo-benefício positiva para projetos de conversão retrospectiva de grande volume de dados.

Serviços bibliográficos (*bibliographic utilities*) é o nome coletivo para um grupo de organizações de serviços de computação que mantém grandes bases de dados de registros catalográficos e oferecem vários serviços de suporte à catalogação e produtos correlatos para bibliotecas e outros clientes que acessam aqueles registros *online*. Todos os serviços bibliográficos adquirem registros catalográficos na forma legível por máquina da Library of Congress e de outras fontes, como informação proprietária de bibliotecas que acrescentam itens específicos de suas coleções. Alguns serviços oferecem registros anteriores ao início do Programa MARC, portanto, não incluídos na base de dados LCMARC. Usando estações de trabalho, as bibliotecas pesquisam os registros, copiam para suas bases locais e modificam, criando seu próprio catálogo *online*. A maioria dos serviços bibliográficos estimula a correção de registros com erro pelas bibliotecas participantes.

Exemplos de serviços bibliográficos norte-americanos incluem:

- A Online Computer Library Center (OCLC), maior e mais conhecida do mundo, foi criada em 1967 pela então Ohio College Library Center com a intenção de funcionar como uma central de processamento computadorizado para as bibliotecas universitárias de Ohio, nos Estados Unidos. A OCLC possui bibliotecas clientes em aproximadamente 60 países e a assim chamada base de dados WorldCat, agrega por volta de 40 milhões de registros, sendo acrescentados 2 milhões a cada ano.

- A Research Libraries Information Network (RLIN) é um serviço bibliográfico e um sistema de recuperação de informação *online*, operado pelo Research Libraries Group (RLG) desde 1978. O RLG foi fundado em 1974 como um consórcio entre Harvard University, Columbia University, Yale University e New York Public Library, nos Estados Unidos. O programa de automação de bibliotecas que levou à formação da RLIN teve origem em um sistema de recuperação de informação *online* interativo, desenvolvido para a área de Física pela Stanford University, em 1967. É o maior serviço bibliográfico somando 85 milhões de registros, se não forem considerados registros duplicados, característica que não compõe a filosofia da OCLC: as duplicatas na RLIN são consideradas importantes como alternativa para a escolha da melhor forma de catalogação pela biblioteca usuária. O sistema de recuperação de informação *online* fornece acesso a várias bases de dados temáticas, inclusive na área de Humanidades, mas apenas para pesquisa.

- O serviço Impact/ Online CAT é decorrência do crescimento da Auto-Graphics Interactive Library Exchange (AGILE) desenvolvida no final dos anos 70 e implantada formalmente em 1981.

- O Impact/ MARCIt é operado pelo A-G Canada, uma subsidiária da Auto-Graphics, formada em 1997 quando a Auto-Graphics comprou o serviço bibliográfico e o *software* que eram comercializados pela ISM Library Information Services. A ISM, por sua vez, havia adquirido, em 1992, a UTLAS (University of Toronto Library Automation System), uma das mais importantes organizações da história da automação de bibliotecas.

- O ITS.MARC começou em 1997 como o primeiro serviço bibliográfico desenvolvido especifica e exclusivamente para a Web, e é operado pela Library Corporation conhecida pelos seus produtos catalográficos em CD-ROM.

Serviços bibliográficos de outras regiões do mundo também mantêm bases de dados mas geralmente não estão disponíveis para bibliotecas norte-americanas, como o serviço BLAISE-LINE da British Library, o Dutch Project for Integrated Catalogue Automation (PICA), SIBIL da França, Australian Bibliographic Network (ABN) e o serviço bibliográfico sul-africano SABINET.

Os serviços bibliográficos diferenciam-se pelo tamanho e composição de suas bases de dados, pela natureza e número de assinantes e pelos recursos específicos de catalogação e recuperação que suportam. Alguns serviços bibliográficos têm aquisições *online* e subsistemas de empréstimos entre bibliotecas. No entanto, a ênfase destes serviços continua na catalogação cooperativa, o que claramente se distingue de outros serviços de informação *online* que fornecem acesso por tempo partilhado a registros bibliográficos similares. A base de dados LCMARC, por exemplo, há muito tempo é disponibilizada *online* pelo serviço de informação Dialog mas esta implementação é destinada para aplicações de referência e de pesquisa mais que para processamento técnico. É uma boa fonte para verificação bibliográfica, mas o Dialog não suporta entrada

*online* de dados catalográficos originais, edição de registros, produção de fichas ou outros serviços de operação catalográfica.

### **Sistemas integrados de biblioteca**

Os sistemas integrados de bibliotecas são definidos como sistemas de informação baseados em computador que usam uma única base de dados bibliográficos e um conjunto de programas inter-relacionados para automatizar múltiplas aplicações de biblioteca. Podem ser utilizados por pequenas bibliotecas e até por consórcios de grandes bibliotecas. Desde a metade dos anos 80, os sistemas integrados têm suplantado os *softwares* customizados e os produtos de automação de bibliotecas para aplicações específicas tais como catalogação ou controle de circulação e de aquisições.

Na maior parte dos casos, o sistema inclui módulos de catalogação, catálogo *online* de acesso público (OPAC) e controle de circulação. Gerenciamento de aquisições e de publicações seriadas são populares mas, normalmente, opcionais.

Todos os sistemas integrados têm recursos para a **catalogação** como um módulo principal que permite a criação, atualização e gerenciamento de uma base de dados bibliográficos de biblioteca. Também possibilitam a adoção de registros MARC e alguns permitem o uso de registros não-MARC com campos definidos pela própria biblioteca. Registros catalográficos podem ser digitados ou transferidos de fontes legíveis por máquina, seja eletronicamente ou em fitas magnéticas ou disquetes. Dependendo do sistema, o controle de autoridade pode ser incorporado no módulo de catalogação ou oferecido como um componente separado. Em alguns, o controle de autoridade estabelece e mantém formas autorizadas e referências cruzadas para campos designados. Quando registros bibliográficos são inseridos, campos de assunto são automaticamente checados segundo as formas autorizadas. Nomes de autor, títulos uniformes, títulos em série e cabeçalhos de assunto são os campos mais comumente controlados. Alguns sistemas integrados suportam tesouros com estruturas de referências cruzadas que incluem termos genéricos, específicos e relacionados, notas de escopo, em adição às tradicionais referências “veja” e “veja também”.

A substituição do catálogo de fichas pela OPAC costuma ser o principal motivo para a implementação de um sistema integrado. Alguns sistemas possuem interfaces gráficas com recursos que simplificam a entrada de comandos de pesquisa e há OPACs especialmente construídas para crianças. Estão sendo grandemente adotadas OPACs com *Web browsers* populares como Netscape Navigator e Microsoft Internet Explorer.

Todos os módulos OPAC permitem pesquisas por autor, título e assunto, replicando a tradicional funcionalidade de recuperação das fichas catalográficas. Outros parâmetros de pesquisa incluem o número de classificação, International Standard Book Number (ISBN), editora, data de publicação, país de publicação, idioma de publicação e tipo de mídia. Para permitir máxima flexibilidade, alguns sistemas permitem que a biblioteca crie campos que podem ser indexados para fins de recuperação. Recursos especiais de recuperação incluem pesquisa por palavra-chave, pesquisa por raiz de palavras e operadores booleanos.

Para expandir suas funcionalidades, sistemas integrados de bibliotecas incorporam, nas OPACs de pesquisa, serviços de informação externa tais como bases de dados de referências e produtos de informação em CD-ROM. Em alguns sistemas, um módulo opcional de informação referencial fornece aos usuários da OPAC acesso a

informações sobre grupos comunitários e eventos locais. Desde a metade dos anos 80, sistemas integrados têm fornecido acesso a catálogos de outras bibliotecas. Isto é possível por conta do padrão Z39.50, que define uma interface para a comunicação entre sistemas de computador de fabricantes diferentes. A compatibilidade com este padrão também tem sido implementada por serviços de informação *online*, produtos de informação em CD-ROM e *sites* da Internet.

A função de **controle de circulação** foi uma das primeiras aplicações a serem automatizadas. O interesse dos bibliotecários em automatizar o controle da circulação baseia-se nos problemas decorrentes da circulação tradicional como: tempo excessivo dispensado para a anotação das transações; imprecisões e erros; dificuldade para gerar estatísticas; e ausência de interface entre os dados de circulação e outros dados da biblioteca. Historicamente, sistemas automatizados de empréstimo foram implantados desde os anos 60, mas utilizavam técnicas de processamento de dados em lote, que tinham certas limitações como a impossibilidade de se obter dados atualizados, dificultando localizar usuários infratores ou fazer reservas. A partir dos anos 70, estes obstáculos foram superados com a implantação dos sistemas em tempo real pelas grandes bibliotecas. No entanto, como muitas bibliotecas não possuíam condições para desenvolver sistemas adaptados a suas necessidades, optaram por adquirir pacotes de *software* de controle de circulação. Atualmente, módulos de controle de circulação fazem parte dos sistemas integrados e são altamente parametrizáveis, permitindo à biblioteca especificar as condições sob as quais os itens de acervo devem circular, como os arquivos podem ser consultados e quais relatórios podem ser produzidos.

Todo módulo de controle de circulação executa as operações de entrada, saída e renovação em tempo real. Muitos módulos empregam leitores de códigos de barra para simplificar a entrada dos itens e identificar os usuários que solicitam empréstimos. Muitos deles também fornecem recursos para a geração de relatórios operacionais e estatísticos. Módulos de controle de circulação são especialmente construídos para bibliotecas escolares e universitárias contando com: recursos de reserva domiciliar e cópia de documentos cujos registros estão ligados aos cursos e professores, além de módulo para multimeios (*media booking*).

As funções de **aquisição e de controle de publicações seriadas** foram automatizadas desde a década de 1960, mas até recentemente, eram as funções que menor atenção recebiam no processo de automação de bibliotecas. Entre as razões para automatizar o processo de aquisição, está o desejo de melhorar o processo de tomada de decisões com a ajuda de análises estatísticas e outros relatórios, além da facilidade decorrente, com a criação de sistemas integrados, de os registros bibliográficos serem criados ao mesmo tempo em que o exemplar é solicitado. A busca bibliográfica prévia ao pedido e a entrada de dados propriamente dita podem ser automatizadas com o uso das bases de dados oferecidas pelos serviços bibliográficos, serviços de informação *online* e editores de CD-ROM.

Os módulos de publicações seriadas dos sistemas integrados são desenhados para gerenciar revistas, periódicos, jornais, séries monográficas e outros materiais que as bibliotecas recebem continuamente, seja a intervalos regulares, seja como suplementos irregulares ou publicações especiais. Os recursos deste módulo incluem: pedidos *online* e renovação de assinaturas, entrada de números recebidos, solicitação de números perdidos ou danificados e controle de pedidos de encadernação.

## Serviço de referência automatizado

A função do serviço de referência é definida como aquela que ajuda os usuários de biblioteca a obter acesso à informação que necessitam. Assim, enquanto as atividades de controle de circulação e de catalogação refletem os aspectos bibliográficos e de custódia da Biblioteconomia, a referência é um serviço público que o bibliotecário presta como especialista e profissional da informação.

No entanto, o uso de ferramentas automatizadas para estes serviços de referência implica a disponibilidade das fontes bibliográficas legíveis por máquina. Desde o final da década de 60, muitos editores de índices impressos e de revistas de resumos (*abstracts*) têm oferecido seus produtos em versões informatizadas para bibliotecas e outras organizações. Historicamente, estas bases de dados bibliográficos têm sido encontradas, em sua maioria, cobrindo áreas técnicas e científicas, mas recentemente a atenção voltou-se também para as áreas de Ciências Sociais, Humanidades e negócios. Além das bases de dados que são publicadas no meio impresso, um crescente número de serviços de referência tem sido desenvolvido especialmente para uso em sistemas computadorizados sem contrapartida no meio impresso. Enquanto as bases de dados mais antigas trabalharam com dados bibliográficos, muitas bases estão sendo desenvolvidas a partir de informação não-bibliográfica ou factual. Em termos de sua utilidade para o serviço de referência, bases de dados não-bibliográficos podem ser divididas em dois grupos: bases de dados textuais, que contém informação equivalente àquela que é encontrada em fontes de referência impressas como enciclopédias, diretórios e outras publicações; e bases de dados numéricas que contém dados estatísticos, financeiros e outras informações quantitativas.

Já no final dos anos 60 e início dos anos 70, algumas bibliotecas compraram bases de dados bibliográficos e não bibliográficos para pesquisa em computador. No entanto, muitas bibliotecas careciam de recursos para obter *hardware* e *software* necessários. Este problema foi resolvido com a oferta de serviços de busca bibliográfica pelas mesmas empresas, ou seja, o pagamento passou a se dar pelas buscas realizadas e não pela obtenção de toda a base. A princípio, estas bases funcionavam *offline* e depois passaram a oferecer consulta *online*.

Os serviços de informação *online* trabalham com bases de dados multidisciplinares, especializadas ou de notícias e negócios, segundo a seguinte metodologia comum: as empresas compram ou obtêm dados bibliográficos e não bibliográficos em formato legível por computador de seus produtores e convertem estes dados no formato de armazenamento compatível para seus computadores, permitindo às bibliotecas e outros clientes a recuperação destas informações mediante pacote de *software* de gerenciamento de bases de dados. Alguns destes serviços oferecem arquivos privados que permitem às bibliotecas desenvolver sua própria base de dados. Com relação aos custos, depende do tempo de conexão e das taxas de cada base de dados. As duas principais áreas de redução de custos, em comparação aos sistemas de fichas catalográficas e acervo impresso são: eliminação de assinaturas de índices impressos pouco utilizados e redução do trabalho do profissional que realiza as pesquisas ou outras tarefas da referência. Além disso, melhores resultados são obtidos com a pesquisa *online*.

## Bibliotecas digitais

Uma definição restrita indica uma biblioteca digital como uma biblioteca que mantém toda coleção, ou uma parte substancial da mesma, na forma processável por computador como uma alternativa, suplemento ou complemento para materiais impressos e microfimes que correntemente têm dominado as coleções de biblioteca. Coleção é aqui entendida como o conjunto de documentos que uma biblioteca adquire ou mantém, em oposição a catálogos, índices ou outras formas encontradas para fornecer informação sobre aqueles documentos. A coleção de uma biblioteca pode incluir materiais publicados, tais como livros e periódicos, ou materiais não publicados, como relatórios de pesquisa mantidos por muitas bibliotecas técnicas, entre muitos outros. Alguns consideram as bibliotecas digitais como sistemas de computador em rede e bases de dados *online*, não diferenciando bibliotecas digitais enquanto coleções de informação processável por computador de bibliotecas digitais como organizações que mantém e fornecem tais informações. Também não distinguem bibliotecas digitais de outros fornecedores de informação processável pelo computador, tais como serviços bibliográficos, serviços de pesquisa *online* ou editores de CD-ROM.

Uma biblioteca digital pode ser acessada *online* por pessoas autorizadas através de LANs ou WANs. Uma vez acessada, materiais da coleção de uma biblioteca digital podem ser visualizados, impressos, baixados ou manipulados de outro modo para satisfazer necessidades particulares de usuários.

Contudo, historicamente, a automação de bibliotecas tem enfatizado a automação de ferramentas bibliográficas como catálogos, índices e outras formas de ajuda, mais que o armazenamento e recuperação de livros, periódicos e outros documentos contidos em coleções de bibliotecas. Enquanto uma porcentagem crescente de bases de dados legíveis por máquina incorpora o conteúdo completo de livros, artigos de revistas, relatórios técnicos ou outros trabalhos, a maior parte das bibliotecas continua mantendo tais materiais em suas coleções na forma de papel ou microforma. Os conceitos e tecnologias da biblioteca digital fornecem uma abordagem completamente computadorizada para armazenamento e recuperação de materiais de biblioteca. Sendo assim, as formas de ajuda às coleções de biblioteca são crescentemente computadorizadas pois o gerenciamento de coleções baseado em computador é o último passo lógico na automação de bibliotecas.

Enquanto as coleções de bibliotecas digitais sugerem uma abordagem revolucionária de gerenciamento de bibliotecas, conceitos e tecnologias de bibliotecas digitais são mais precisamente caracterizados como evolução natural. Bibliotecas têm criado, mantido e usado informações processadas por computador há décadas. O conceito de uma biblioteca digital tem uma longa história na produção de literatura e implementações práticas baseadas em texto ou em imagens datam das décadas de 70 e 80. Desde o início dos anos 90, um número de bibliotecas e agências de informação nos Estados Unidos iniciou projetos de bibliotecas digitais. Bibliotecas universitárias têm implementado bibliotecas digitais para obter maior número de usuários por pesquisa (para materiais de arquivos e manuscritos, por exemplo) e para facilitar o acesso a itens de alta demanda, como reserva domiciliar de materiais ou coleções que fazem parte de currículos de ensino. Projetos de biblioteca pública têm enfatizado materiais históricos locais. Implementações de bibliotecas governamentais fornecem acesso *online* a publicações governamentais e registros públicos. Bibliotecas corporativas têm focado materiais de produção e uso próprio como relatórios técnicos que exigem atividades de

pesquisa específicas. Um número de organizações profissionais tem sido formado para desenvolver estratégias e procedimentos para implementações de bibliotecas digitais.

Conceitos e tecnologias de bibliotecas digitais oferecem significativo potencial para a melhoria de serviços de biblioteca, com relação a certos problemas tradicionalmente associados ao gerenciamento de coleções. Bibliotecas digitais fornecem acesso *online* a livros, artigos de revista e outros materiais de biblioteca, transformando irrelevante a localização física destes materiais. Conceitos de biblioteca digital também implicam o acesso transparente a múltiplas coleções de biblioteca. Entre outras vantagens, bibliotecas digitais possibilitam recursos avançados de recuperação. Elas podem também simplificar o controle e a manutenção diária das coleções, reduzindo custos no processo.

A disponibilidade de livros, revistas e outros materiais na forma processável por computador é pré-condição para a implementação de bibliotecas digitais. Algumas iniciativas de bibliotecas digitais têm se concentrado sobre a informação que já nasce na forma digital. Outras fornecem ao usuário acesso *online* a versões digitais de livros, periódicos e outros materiais de biblioteca que são adquiridos na forma legível por máquina de editores ou outras fontes externas. Em muitos casos, no entanto, as bibliotecas devem converter estes materiais de sua coleção para textos codificados por caractere ou imagens eletrônicas: tais conversões envolvem esforços e custos enormes sendo, por isso, o maior obstáculo para implementações de bibliotecas digitais.

Implementações de bibliotecas digitais baseadas em textos, livros, periódicos e outros materiais podem ser convertidos por digitação ou reconhecimento de caracteres óticos. Nas implementações baseadas em imagens, documentos são digitalizados por *scanners*. Em ambas, os documentos são ligados aos registros bibliográficos em um catálogo *online* de uma biblioteca ou em outra base de dados. Quando os registros da base de dados são recuperados, usuários têm a opção de ver ou imprimir seus documentos associados.

### ***Library Automation, segundo William Saffady***

A primeira edição do livro "Introduction to automation for librarians" foi escrita por William Saffady no início da década de 80. Para esta pesquisa foram utilizadas, num primeiro momento, a tradução para o espanhol desta primeira edição e a terceira edição em inglês, publicada em 1994. Só então foi obtida a quarta edição da obra, datada de 1999. Esta casualidade possibilitou interessante comparação entre as primeiras edições e a última, cujo enfoque e organização foram alterados.

Uma das mudanças ocorreu, do foco nas operações da biblioteca e nas tecnologias envolvidas, para a abordagem sobre os conceitos relacionados aos serviços e sistemas de bibliotecas e as tecnologias que lhe são decorrentes. Outro aspecto relaciona-se à inclusão das características dos sistemas integrados para gestão de bibliotecas, causando estranheza a apresentação fragmentada das operações de biblioteca nas edições anteriores, uma vez que, sistemas com algum nível de integração já existissem no mercado. Também os termos "leitor" e "livros" foram substituídos por "usuário" e "documentos" (ou "materiais"). Esta alteração é significativa no que se refere ao conceito ampliado do termo "usuário" e à tipologia diversificada envolvida no termo "documentos". E o tópico novo, por excelência, é o das bibliotecas digitais, abordado a

partir da ampla e variada gama de definições que vem recebendo e, provavelmente, principal responsável pelas alterações acima citadas. Vale comentar que, não só os tópicos apresentados e sua forma de organização receberam verdadeiros re-arranjos, mas a linguagem e a abordagem são definitivamente alteradas: as operações automatizadas de representação e gestão de acervos são acrescidas das noções de informação, elaboradas internamente ou não, e a integração de operações enfatizam a otimização dos recursos tecnológicos para gerenciamento.

Resenha sobre esta nova edição, cujo título trata da importância da aprendizagem da automação, inicia com a seguinte afirmação: “Parece que todos os bibliotecários são bibliotecários de sistemas atualmente”. Esta frase é justificada pela percepção de que quase todos os profissionais de informação usam sofisticados recursos tecnológicos de *software* e *hardware* para realizar seu trabalho. Catalogadores usam os serviços da OCLC, da Library of Congress ou de outros vendedores para editar e fazer *download* de registros e acrescentar dados nos seus catálogos *online*, adotando o próprio sistema local da biblioteca para gerenciar os dados. Bibliotecários de referência e de pesquisa podem usar uma variedade de recursos de pesquisa *online*, desenvolver métodos sofisticados para criação de bibliografias com registros eletrônicos e desenhar interfaces públicas para sistemas de computador. Muitos dos usuários têm computadores em suas casas e em seus escritórios e esperam por eles na biblioteca. No período da primeira edição deste livro, a pesquisa *online* era utilizada apenas por aqueles bibliotecários que acessavam serviços como o Dialog; atualmente a biblioteca contemporânea depende da automação em todas as áreas e é praticamente impossível ser um bibliotecário sem usar a automação. O bibliotecário deve entender dos conceitos de automação para que esteja capacitado para planejar, desenhar e fornecer os melhores serviços possíveis. O resenhador do livro, Gregory (2000), é professor e chefe dos serviços bibliográficos da New Mexico State University Library, Las Cruces.

Relevante notar na obra de Saffady, a forma clara como são apresentadas informações sobre tecnologia, inclusive aquelas mais herméticas para os bibliotecários, como *hardware* e linguagens de programação. Além disso, apresenta informações históricas sobre o desenvolvimento dos sistemas de computadores e dos processos informatizados em bibliotecas, essenciais para proporcionar a compreensão dos tempos atuais. Dentre as obras analisadas nesta pesquisa, esta é a de organização mais didática.

Sobre os serviços de referência automatizados, o aspecto mais amplamente tratado foi o dos serviços de informação *online*, reforçando a noção de que os acervos locais e o tratamento tradicionalmente dispensado a eles, não dão conta das necessidades de usuários. Isto aponta para a necessidade de que estes serviços de informação *online*, externos às informações contidas na biblioteca local, sejam incorporados e compatibilizados às metodologias de trabalho da biblioteca local.

O tópico sobre catalogação descritiva automatizada, diferente das edições anteriores, trata da descrição dos grandes projetos e programas de catalogação cooperativa, menos que da natureza e características destes processos técnicos. São citadas as empresas e instituições que oferecem estes serviços, sempre em função do Formato MARC, fornecendo um quadro do caráter mercadológico destas operações em nível nacional, mas que, de fato, extrapolam as fronteiras do país.

Sobre as bibliotecas digitais, cita que os documentos são anexados aos registros bibliográficos em um catálogo *online* de uma biblioteca ou em outra base de dados e quando estes registros são recuperados, usuários têm a opção de ver ou imprimir seus

documentos associados. A partir desta descrição, fica evidente que não está em uso a linguagem de marcação de textos XML (eXtensible Markup Language), na qual, campos de informação podem ser recuperados a partir da representação do documento que é realizada no próprio texto e não fora dele, como tradicionalmente acontece.

Conclui-se que o autor trabalha com a apresentação das operações realizadas pelos sistemas de automação, em detrimento dos métodos para realizá-las. Além disso, parece centrar-se nos sistemas de automação já amplamente difundidos em seu país, supostamente, numa tentativa de manter o público de bibliotecários antenados com o mercado estabelecido de sistemas de automação de bibliotecas. Por isso, pode-se especular sobre a baixa incidência de desenvolvimento tecnológico em bibliotecas para os recursos e técnicas de tratamento e recuperação da informação.

## **ALGUNS CURRÍCULOS DE ENSINO SUPERIOR**

Para subsidiar a abordagem acima apresentada sobre Automação de Bibliotecas, foram levantados currículos de ensino de cursos de graduação e pós-graduação do curso de Library and Information Science, da Palmer School, da Long Island University (COURSE..., 2001), propositadamente onde Saffady leciona na área de gerenciamento da informação. A intenção é fazer conhecer o conteúdo geral para Biblioteconomia e Ciência da Informação a partir da produção de conhecimento nos Estados Unidos localizando melhor, neste contexto, o conhecimento sobre automação de bibliotecas apresentado por Saffady.

Dentre as diversas disciplinas, seguem as que foram selecionadas por tratarem das tecnologias da informação em geral e do tratamento, armazenamento, recuperação e disseminação da informação pelos recursos automatizados ou não:

- tecnologia da informação: tecnologias associadas à transferência de informação, incluindo *hardware*, *software* e sistemas de telecomunicações;
- redes: princípios de telecomunicações incluindo protocolos, padrões, meios de transmissão e todos os tipos de redes usadas na transferência da informação como LANs, WANs, Internet, comerciais e privadas;
  - Editoração;
  - Bibliometria;
  - gerenciamento de bibliotecas e centros de informação;
  - gerenciamento de recursos eletrônicos;
- organização do conhecimento: princípios e técnicas de métodos e padrões para organização física e lógica do conhecimento nos vários formatos, descrição e desenvolvimento de bases de dados relacionais e outras;
- teoria da análise de assunto, vocabulário controlado e classificação: CDD, CDU, classificação da Library of Congress, Library of Congress Subject Headings, PRECIS, Art & Architecture Thesaurus, Colon Classification, Medical Subject Headings (MeSH) e outros sistemas;

- resumo e indexação para sistemas de informação: conceitos e métodos de indexação e resumos no contexto dos sistemas de recuperação da informação manual e baseados no computador, construção de tesouro e indexação assistida por computador, avaliação de sistema de indexação e recuperação;

- armazenamento e recuperação da informação: fundamentos dos sistemas de recuperação da informação (SRI) incluindo componentes, modelos, estruturas, representação da informação, vocabulário controlado e tesouro, estratégias de pesquisa, interação homem-computador e avaliação, desenho e implementação de SRIs;

- construção de tesouro: análise da relação entre conceitos, entre termos e entre conceitos e termos e como eles são explicitados em um tesouro, o papel do tesouro nos sistemas de recuperação da informação que suportam pesquisa por vocabulário controlado.

Além destas, disciplinas genéricas ajudam a compreender a proposta destes cursos:

- bibliotecas de todos os tipos, arquivos, museus e outros centros de informação;

- tecnologias desde a escrita até a informação eletrônica;

- procedimentos documentários: indexação, elaboração de resumos, metadados, sistemas de classificação, SRIs, cabeçalhos de assunto e tesouros;

- informação em diversas áreas: Humanidades, Ciências Sociais, Ciência e Tecnologia, Ciências da Saúde;

- Biblioteconomia e Ciência da Informação;

- pesquisa e ensino;

- serviços de informação baseados em estudos de usuários de áreas especializadas do conhecimento e de determinados grupos de indivíduos como crianças e jovens, portadores de deficiências, pessoas não alfabetizadas, imigrantes etc.;

- globalização e informação: comunicação e indústria da informação (incluindo bibliotecas).

Os currículos apresentados abordam a indexação, a elaboração de resumos, o uso e construção de tesouro e os SRIs a partir dos recursos automatizados. No entanto, não é citada a cadeia documentária: há organização da informação, por um lado, e armazenamento e recuperação da informação, por outro. Assim, não há disciplinas que tratam do processo documentário mas a noção sobre cadeia documentária provavelmente se apresenta, com outro enfoque, na discussão sobre indústria da informação. As poucas disciplinas abrangentes tratam da Biblioteconomia e Ciência da Informação mas seu conteúdo não foi detalhado o suficiente para que se pudesse analisá-lo.

De uma forma geral, há disciplinas técnicas e tecnológicas e outras relacionadas a tipos de usuários e a áreas específicas de conhecimento. Há uma ênfase para as tecnologias em muitas disciplinas. Chama a atenção a inclusão das bibliotecas no quadro apresentado sobre comunicação e indústria da informação.

Sendo assim, como era de se suspeitar, a Automação de Bibliotecas de que trata Saffady está incluída em um espectro maior que aquela das operações automatizadas dos serviços de uma biblioteca. No entanto, sua obra não aponta suficientemente para isso.

### **“THE ELECTRONIC LIBRARY”, Jennifer Rowley, 1998**

Outra obra no idioma inglês pertinente para a análise proposta sobre o tema da Informática Documentária é a de Rowley (1998).

O livro de Jennifer Rowley, “The electronic library”, de 1998, é uma atualização do livro “Computer for libraries”, de 1993. Este último foi publicado no Brasil pela Editora Briquet de Lemos em 1994 com o título “Informática para bibliotecas”. Foi editado pela primeira vez em 1980 e é adotado nas escolas de Biblioteconomia e Informação do Reino Unido. Assim como na análise da obra de Saffady, também foi realizado um trabalho anterior com a edição brasileira de 1994, para então vir a obter a edição mais atualizada em inglês publicada em 1998. Desta forma, além da comparação da abordagem inglesa de Rowley com a estadunidense de Saffady, também são comentadas as duas últimas edições produzidas por Rowley.

#### **Introdução**

Parte do princípio, do qual parece não haver discordância na literatura, de que a Informática têm exercido uma influência fundamental no funcionamento das bibliotecas e serviços de informação. Sendo assim, examina as aplicações da tecnologia da informação que influem sobre as rotinas e o gerenciamento de bibliotecas, bem como outras aplicações que acarretam implicações para a maneira como as informações são tratadas, armazenadas e gerenciadas.

Com base neste entendimento, que abarca o gerenciamento de bibliotecas e outras aplicações para serviços de informações, o livro é dividido em três partes: introdução à tecnologia da informação, recuperação da informação e sistemas de gerenciamento de bibliotecas. A primeira parte introduz as plataformas nas quais as aplicações são montadas e as outras duas tratam das aplicações propriamente. Os termos usados para descrever estas aplicações evoluíram significativamente em poucos anos e os sistemas designados para satisfazer as diferentes funções tornaram-se ainda mais indistintos.

#### **Parte I – Introdução à tecnologia da informação**

Para caracterizar o próprio título do livro, uma das definições apresentadas considera a **biblioteca eletrônica** como a coleção de informações organizada e gerenciada em uma variedade de mídias (texto, imagens fixas e móveis, som ou a

combinação delas), mas todas na forma digital. A coleção é organizada e gerenciada para o benefício de uma população usuária atual ou potencial e é estruturada para acesso fácil de seu conteúdo. Tipicamente, uma biblioteca eletrônica inclui um número de recursos para pesquisa ou navegação que operam na biblioteca em particular e dão acesso a outras coleções de informações disponíveis em rede. No entanto, não são tratadas as bibliotecas eletrônicas completamente na forma digital como a definição acima, por ser pouco convincente a afirmação de que usuários de informação possam vir a ter acesso a coleções de informação completamente digitais em um futuro previsível. Argumenta-se que nem todos os documentos estarão na forma digital e nem todos serão acessíveis por rede digital: o desafio das bibliotecas não difere muito do atual que é o de gerenciar uma coleção multimídia para sua comunidade de usuários que reflete as diferentes formas e formatos que esta comunidade considera conveniente para comunicação e armazenamento. Por outro lado, as opções e ferramentas disponíveis para dar conta deste processo estão submetidas a uma grande e significativa mudança.

Em função do forte papel ocupado pela informação eletrônica no mercado, a **indústria da informação** configura-se a partir do fluxo indicado abaixo:

- produtores do conteúdo intelectual (autores, ilustradores, equipes de criação multimídia);
- controladores do conteúdo intelectual com relação a padrões de qualidade e conveniência (editores, *referees*, críticos);
- intermediadores entre produtores do conteúdo intelectual e distribuidores (editores, gráficos, produtores de bases de dados);
- distribuidores (bibliotecas providas de agentes, livreiros, *web sites*, serviços de pesquisa *online*, serviços de vídeo-texto, serviços de áudio-texto, distribuidores de CD-ROM)
- arquivadores (bibliotecas e colecionadores particulares);
- redistribuidores (bibliotecas, consultores de informação);
- usuários (individuais e corporativos).

Os **sistemas de gerenciamento de bibliotecas** focam as encomendas e as aquisições, a catalogação, as OPACs, o controle da circulação e das publicações seriadas, o gerenciamento da informação, o empréstimo entre bibliotecas e a informação comunitária. Os sistemas de computador são vistos como apropriados para estas organizações ou redes onde o número de transações é grande. Sistemas de gerenciamento de bibliotecas podem ser considerados como sistemas que gerenciam acesso a documentos em uma coleção de biblioteca, ou ainda, como documentos que podem estar temporariamente em uma coleção de biblioteca através, por exemplo, do empréstimo entre bibliotecas. A preocupação central é manter a coleção de biblioteca e monitorar o paradeiro dos documentos a fim de que a equipe de trabalho e os usuários da biblioteca estejam a par da disponibilidade e do *status* destes documentos. Cada vez mais OPACs têm melhorado seus recursos de pesquisa para incluir muitas das características originalmente encontradas em aplicações para recuperação da informação, de modo que, desde que os registros do catálogo forneçam informação suficiente, eles também fornecem recuperação de informação, que identifica a informação contida nos documentos. Além disso, muitas OPACs atuam como janelas sobre uma

extensa coleção de recursos que está disponível em uma biblioteca, incluindo recursos de Internet e coleções de outras bibliotecas.

Os **sistemas para recuperação da informação** foram tradicionalmente desenhados para fornecer acesso à informação a partir de distintos documentos. Originalmente, isto foi efetuado através de bases de dados bibliográficos que incluíam registros de artigos de revistas e outros documentos. Mais recentemente, aplicações para recuperação da informação têm fornecido acesso à informação contida nos documentos eletrônicos, especialmente documentos baseados em textos, mas cada vez mais sistemas têm incluído alguns recursos para acesso a documentos multimídia e acesso à informação e objetos com documentos multimídia (tais como fotos e *video-clips*). Aplicações para recuperação da informação incluem sistemas de gerenciamento de documentos, serviços de pesquisa *online*, Internet e CD-ROM. Serviços de notificação corrente (boletins de alerta) e índices impressos são outros exemplos de sistemas especiais de recuperação da informação. Sistemas de gerenciamento de documentos podem ser criados para gerenciar coleções de documentos corporativos em uma organização; o sistema de gerenciamento de documentos pode armazenar documentos na forma eletrônica e fornecer mecanismos de recuperação apropriados, de modo que, documentos individuais ou conjuntos de documentos sobre tópicos específicos possam ser recuperados. Em alguns sistemas, os documentos podem existir na forma de microficha ou impressos e somente o índice estar na forma eletrônica. A Internet é uma grande rede que fornece acesso a um número de computadores ou servidores, cujos recursos de pesquisa possibilitam a recuperação neste banco de informação que é, ao mesmo tempo, enorme e indefinido. Alguns desses computadores ou servidores são serviços comerciais de pesquisa *online* que fornecem acesso a bases de dados selecionadas e avaliadas. Outro modo de acessar algumas destas bases de dados é pela aquisição de CD-ROMs. O CD-ROM é abastecido com *software* para recuperação da informação, de tal modo que possibilita localizar informação específica na base de dados.

Com o advento da biblioteca eletrônica, a distinção entre sistemas de gerenciamento de bibliotecas e sistemas de recuperação da informação, está gradualmente sendo diluída. Isto teve início com aperfeiçoamentos nos recursos de recuperação da informação disponíveis em OPACs, mas é necessário progredir muito mais. Ultimamente, o modo pelo qual a informação é recuperada deve ser dependente das necessidades antecipadas de usuários e não sobre o acaso de seus tipos de publicação (jornais, livros, diretórios), suas formas (impressas ou eletrônicas) ou sua localização (localmente ou em uma base de dados remota). Em tal ambiente, também seria necessário refletir sobre a natureza do acesso. O conceito de empréstimo é anacrônico, no entanto, cópia ou propriedade parece ser alternativa viável no presente mas pode não ser ainda a melhor solução. Quando estas questões tiverem sido resolvidas, não haverá distinção entre sistemas de gerenciamento de bibliotecas e sistemas de recuperação da informação, mas nesse meio tempo ambos vão provavelmente continuar a existir, com sobreposições e funções complementares.

Quando a primeira edição deste livro foi escrita, o princípio foi o de tratar sistemas de gerenciamento de bibliotecas e de recuperação da informação como dois campos diferentes. Desde aquele período, o argumento é o de que tecnologias comuns estão subjacentes a ambos os tipos de sistemas e que desenvolvimentos similares figuram simultaneamente nas diferentes categorias de sistemas.

As bases de dados são anteriores às bibliotecas eletrônicas. No entanto, os diversos tipos de bases de dados encontrados atualmente apontam para os conceitos apontados acima para bibliotecas eletrônicas. Seguindo este princípio, as bases de dados são categorizadas a partir de dois grupos: bases de dados de referências e bases de dados de fontes.

As **bases de dados de referências** encaminham ou orientam o usuário para uma outra fonte que pode ser um documento, uma instituição ou um indivíduo, a fim de obter informações adicionais ou conseguir o texto integral de um documento. Incluem:

- bases de dados bibliográficos: citações ou referências bibliográficas, às vezes, acompanhadas de resumos;
- bases de dados catalográficos: são, de fato, um tipo de base de dados bibliográficos, porém sem informações adicionais de conteúdo dos documentos e referentes sempre ao acervo de determinada biblioteca ou rede de bibliotecas;
- bases de dados referenciais: informações ou dados como nomes e endereços de instituições e outros dados característicos de guias e cadastros.

Já as **bases de dados de fontes** contêm os dados originais, como no caso dos documentos eletrônicos, dividindo-se em:

- bases de dados numéricos: dados estatísticos e de resultados de pesquisas e outros;
- bases de dados de texto integral: notícias de jornal, especificações técnicas e outros;
- bases de dados mistas: tipos numéricos e texto integral simultaneamente, como relatórios anuais de empresas e outros;
- bases de dados multimídia: informação armazenada em diferentes tipos de mídia como som, vídeo, foto, texto e animação.

Para a formatação de algumas destas bases e o compartilhamento de registros por diversas organizações, adotam-se **formatos de registro bibliográfico**, sendo o Formato MARC o mais utilizado. O Formato UNIMARC foi desenvolvido para troca internacional de registros MARC: organizações nacionais que produzem registros MARC empregando normas locais, reformatam seus registros segundo o UNIMARC para o intercâmbio com outros países. Recentemente, contudo, um número maior de fornecedores de registros MARC tem concordado em usar o formato USMARC, sendo provável que muito rapidamente este formato torne-se o padrão internacional. Assim, a edição anterior deste livro que adotava como exemplo o UKMARC (adaptação local do Formato MARC para o Reino Unido), agora utiliza o Formato USMARC.

A descrição do Formato MARC (com base na versão USMARC) é realizada a partir da divisão em duas seções: a que contém informações descritivas dos dados bibliográficos e a que contém os dados bibliográficos propriamente ditos.

Os campos da segunda seção são todos de tamanho variável, necessitando, portanto, de sinalizadores para indicar o começo e o fim de cada campo: cada campo é precedido de um parágrafo ou etiqueta (*tag*) de três caracteres e dois indicadores numéricos e termina com um delimitador especial. Os parágrafos consistem de três algarismos entre 000-945. Cada um dos campos principais contém também dois

indicadores de campo que são algarismos de um dígito colocados em seguida aos parágrafos, empregados com a finalidade de: distinguir tipos diferentes de informações inseridas no mesmo campo, acolher entradas secundárias de título, indicar o número de caracteres a serem desprezados na alfabetação dos títulos e mostrar se certas informações, como edição e impressão, se relacionam com uma parte ou com o todo. Quanto aos campos, muitos deles possuem unidades menores conhecidas como subcampos, precedidos de um símbolo e de uma letra, sendo sempre o mesmo símbolo, e para cada subcampo de um mesmo campo são adotadas letras seqüenciais.

A primeira seção do Formato MARC é composta pelo rótulo de registro, pelo diretório e pelos campos de controle. Apenas estes últimos são preenchidos pelo catalogador, incluindo dados como ISBN, idioma do texto, código de país de publicação, além de controlar o acesso ao registro principal. Já o rótulo e o diretório são inseridos pelo programa: o rótulo (também chamado de líder) contém informações sobre o registro como tamanho, situação (alterado ou não), tipo e classe, e o diretório é uma lista de localização que relaciona a etiqueta, o comprimento e a posição inicial de cada campo.

A estrutura do registro MARC é deliberadamente complexa a fim de permitir flexibilidade. Esta complexidade é decorrente de um período de criação do formato em que os catálogos impressos e os microfimes eram a norma. Apesar do formato de registro MARC ter trazido grande contribuição à padronização e à comunicação por redes, alguns acreditam que tenha chegado a sua hora de reavaliação. Assim, um formato diferente poderia ser mais apropriado para as OPACs.

Há muitos formatos para registros bibliográficos cujas descrições diferem bastante dependendo da fonte. Serviços de indexação e resumos usam regras diferentes de descrição bibliográfica das que são adotadas na catalogação realizada nas bibliotecas. O Formato MARC que é usado como formato de intercâmbio por grandes bibliotecas utiliza a norma ISBD (International Standard Bibliographic Description) como padrão, enquanto muitos serviços de indexação e resumos adotam o manual de referência UNISIST (da UNESCO). Como estes dois importantes formatos diferem na forma de definir, organizar e identificar os elementos de dados, é praticamente impossível reunir em um único arquivo registros bibliográficos procedentes de diferentes origens. Assim, o CCF (Common Communication Format) foi projetado para facilitar a comunicação de dados bibliográficos entre os setores da comunidade de informação.

Existem ainda outros formatos de registro bibliográfico para aplicações não catalográficas, mas que não são padronizados. Essencialmente há duas categorias de sistemas que funcionam desta forma: bases de dados públicas e sistemas locais desenhados em pacotes de *softwares*. No caso das grandes bases de dados públicas, não se verifica movimentos significativos no sentido da adoção de formatos padrões, pois cada produtor geralmente escolhe um formato de registro adequado às suas bases de dados. Além destas bases de dados se diferenciarem entre si devido às características das informações que veiculam e do serviço de pesquisa *online* em que estão hospedadas, a presença dos textos integrais e, mais recentemente, das bases de dados multimídia, são fatores variáveis que têm levado à constituição e uso de formatos distintos dos bibliográficos. Quanto aos muitos sistemas locais desenhados em pacotes de *software*, alguns destes pacotes oferecem sistemas de catalogação que trabalham em registros de Formato MARC ou que produzem registros compatíveis com os registros do Formato MARC, mas outros não oferecem tais opções. Virtualmente todos os pacotes de *software* oferecem a possibilidade de se desenvolver um formato de registro adequado a

uma aplicação específica. Assim, em sistemas locais pode ser grande a variabilidade dos formatos de registros.

Atualmente, o tema dos formatos de registro bibliográfico inclui um novo termo denominado **metadado**. O metadado é definido como o dado sobre o dado, sendo os registros bibliográficos um tipo de metadado. Contudo, metadado está cada vez mais sendo usado no contexto mais especializado de dados que se referem aos recursos digitais disponíveis em uma rede. Metadados também diferem de dados bibliográficos e catalográficos pois a localização da informação está no registro permitindo seu acesso fácil pela própria rede diretamente para o texto integral ou programas de entrega de documentos (*document delivery*). Motores ou robôs de pesquisa da Internet usam metadados nos processos de indexação de recursos da própria Internet. Metadados precisam acomodar a falta de estabilidade da Internet, os dados redundantes, as perspectivas diferentes sobre a granularidade da Internet (se é um documento ou um recurso) e as diversas localizações sobre uma variedade de redes diferentes.

Assim como a obra de Saffady anteriormente analisada, toda a primeira parte do livro está centrada na caracterização de *hardware* e *software*. Seguem então apenas alguns acréscimos quanto às **estruturas de bases de dados**.

Como não é possível pesquisar registro a registro no caso das grandes bases de dados, alternativas foram sendo desenvolvidas. Os serviços de pesquisa *online* e outras aplicações que usam sistemas de gerenciamento de documentos têm usado o recurso do arquivo invertido. Este arquivo é adequado para pesquisas complexas em bases de dados baseadas em texto, onde o pesquisador não conhece a forma pela qual a chave de pesquisa entrou na base de dados e tem que supor a forma mais apropriada.

O arquivo invertido é semelhante a um índice e depende pelos menos de dois arquivos: arquivo texto que contém os registros atuais e arquivo índice que fornece acesso a esses registros. O arquivo de índice contém um registro para cada termo indexado de todos os registros da base de dados, ordenados alfabeticamente. Cada termo é acompanhado pela informação sobre sua frequência de ocorrência na base de dados, pelo arquivo no qual está localizado e pela possibilidade de informação de localização como parágrafo (ou campo) do registro. Quando um novo registro é acrescentado na base de dados, é necessário atualizar o arquivo índice.

Estes arquivos são usados na pesquisa da base de dados: um usuário entra um termo na tela de busca e o sistema busca o termo no arquivo de índice. Se o termo não está presente no arquivo índice, o sistema responde que não há registros para aquele termo. Se o termo é encontrado, o usuário é informado sobre quantas ocorrências do termo há na base de dados. Para mostrar os registros, é usada a localização do arquivo texto.

Esta é uma explicação simplificada do funcionamento de arquivos invertidos pois, na prática, as estruturas dos arquivos são mais complexas. Por exemplo, para pesquisar termos de proximidade com outros termos (ou seja, termos com duas palavras entre eles), o arquivo índice deve conter informação sobre as posições das palavras com um campo para cada termo.

Em uma base de dados, nem todos os campos são indexados devido ao espaço ocupado no disco. Índices são criados para aqueles campos que são comumente pesquisados como nomes de autor, palavras do título e termos de indexação de

assuntos. Nos registros de textos integrais longos, é necessário dividi-los em parágrafos, os quais devem ser indicados por identificadores; alternativamente, as posições de palavras individuais no arquivo podem ser usadas como identificadores. Arquivos de autoridade como listas de autor, listas de cabeçalhos de assunto ou tesouros, são usados em algumas bases de dados, por meio da ligação entre o arquivo índice (ou o arquivo do termo) e a lista de autoridade, de tal modo que termos aprovados e seus relacionamentos possam ser mostrados.

Para dar conta do gerenciamento dos diversos dados decorrentes das atividades e das necessidades das organizações, surgiu o conceito de base de dados e o *software* para gerenciar tais bases ficou conhecido como **sistema de gerenciamento de bases de dados (SGBD)**.

Historicamente, há três tipos de estruturas principais de arquivos de bases de dados e sistemas de gerenciamento de bases de dados associados: hierárquicos, em rede e relacionais. As **bases de dados hierárquicas** são estruturadas em um arranjo do tipo árvore ramificada e os dados são alocados em cada um dos níveis da árvore. As **bases de dados em rede** são baseadas em ligações explícitas ou ponteiros entre entidades relatadas; neste modelo, as ligações entre os dados dos vários níveis são mais diretos que o modelo hierárquico, por conta dos ponteiros que ligam dados de diferentes níveis, os quais, no entanto, ocupam significativo espaço em disco. As **bases de dados relacionais** usam um tipo de estrutura que tem sido bastante adotado em sistemas de bases de dados. Nos sistemas relacionais, os dados são alocados em uma série de arquivos ligados entre si: neles, os dados são mantidos num conjunto de relações ou tabelas, sendo as linhas das tabelas equivalentes aos registros e as colunas a campos, e os dados nas várias relações são interligados mediante uma série de chaves. Bases de dados relacionais são desenhadas usando uma técnica de análise de dados chamada normalização. Esta técnica é usada para quebrar o dado em tabelas de tal modo que os campos em cada tabela são dependentes somente do campo chave e não são ligados com nenhuma outra chave. Relacionamentos são representados pela replicação de itens de dados. A normalização assegura que inserções, eliminações e alterações de registros podem então ser feitas sem diminuir a consistência dos dados da base.

Mais recentemente, as bases de dados multimídia (fotos, animações, som, texto e dados em tabelas) apresentam novos desafios para estrutura de bases de dados. **Sistemas de gerenciamento de bases de dados multimídia** estão sendo desenhados para gerenciar tais bases, adotando um conjunto de tecnologias, como tecnologia relacional para tabelas, bases de dados textuais para documentos e dispositivos de armazenamento de imagens para gráficos e animações. Um problema central é a manipulação de itens não-texto tais como desenhos e imagens móveis: para acesso sofisticado destes itens, recorre-se à indexação por palavras-chave de uma forma similar aos documentos baseados em texto.

**Sistemas de gerenciamento de bases de dados orientadas a objetos** fornecem recursos para armazenar itens criados por programas orientados a objetos. Desde que um objeto seja uma coleção complexa de itens de dados que são definidos como sendo relacionados a outros objetos, sistemas de gerenciamento de bases de dados orientadas a objetos precisam fornecer recursos para objetos grandes que são organizados em hierarquias de acordo com seus relacionamentos com outros objetos. Estas estruturas têm algo em comum com o modelo hierárquico mas é uma tecnologia ainda em desenvolvimento. Não é provável que substituam a tecnologia relacional para aplicações de negócios pois não oferecem as mesmas vantagens para o gerenciamento

de dados. Contudo, quando itens de dados são complexos e heterogêneos como uma base de dados multimídia, sistemas de gerenciamento de bases de dados orientadas a objetos podem fornecer uma possibilidade interessante.

O gerenciamento dos sistemas de informação implica a análise e projeto de sistemas. Para tanto, é necessária a definição de objetivos, requisitos do sistema, projeto, implementação e avaliação. Esta atividade pode contar com um planejamento estratégico de sistemas de informação. Finalmente, não se deve prescindir do planejamento e implantação da manutenção das rotinas do sistema.

## **Parte II – Recuperação da Informação**

Os fundamentos da **Recuperação da Informação** são apresentados em função da diferenciação entre os serviços voltados para a recuperação e disseminação de informações ou documentos e os sistemas que suportam atividades de gerenciamento de bibliotecas. Assim, especialmente as OPACs são vistas como uma categoria especial de sistemas de recuperação da informação.

Os sistemas de recuperação da informação quase chegaram a ser sinônimos de computadores, apesar de os sistemas de fichas catalográficas e outros recursos em papel ainda existirem e terem surgido antes mesmo do advento dos computadores. Todos os sistemas de recuperação da informação apresentam três estágios: indexação, armazenamento e recuperação.

A **indexação** é o processo de alocar índices de termos ou chaves a partir de um registro ou documento. Este índice de termos ou chaves ajuda na recuperação posterior destes registros ou documentos. A indicação de termos de indexação pode ser intelectual (ou seja, conduzida por um homem) ou baseada em computador, o qual seleciona termos para o índice de acordo com um conjunto de instruções. Neste caso, o índice de termos pode ser indicado a partir de uma lista padrão ou tesouro de computador, com base na ocorrência de palavras em um registro ou documento. Alternativamente, computadores também podem gerar índices a partir da seleção de indexadores humanos, os quais se baseiam na avaliação subjetiva do conteúdo ou na indicação de termos provavelmente procurados pelo pesquisador, a partir de discriminações intelectuais para a escolha dos termos apropriados; os termos podem ser extraídos de uma lista controlada ou podem não ser controlados.

Selecionados os termos de indexação, estes são atribuídos a cada registro de documento e um arquivo invertido é compilado, como um índice em fichas ou um índice impresso, processo que é designado de **armazenamento**. Sistemas de recuperação da informação podem usar o próprio computador para armazenar os arquivos de documentos e os arquivos índices e manter as bases de dados.

A **recuperação** é o processo de localizar documentos e itens de informação que tenham sido objeto de armazenamento por meio dos índices. Estes índices garantirão uma melhor recuperação pelo usuário se oferecerem sugestões e orientação quanto à seleção dos termos. A maior questão é que este processo é muito dependente dos estágios de indexação e armazenamento. Computadores são ainda usados para gerar

índices impressos e a contribuição destes índices à recuperação da informação não deve ser esquecida. No entanto, as pesquisas *online* introduzem uma flexibilidade impossível a partir de índices impressos, como formas e estilos de diálogo, linguagens de indexação e lógicas de busca. Alguns dos ambientes de recuperação da informação são os serviços de pesquisa *online*, a Internet, os sistemas de edição e gerenciamento de documentos, o CD-ROM e outros. Além disso, os sistemas de recuperação da informação fazem uso da base de dados para a geração de diversos produtos como índices impressos e boletins de notificação corrente.

Os tesouros têm sido uma característica dos sistemas de gerenciamento de documentos desenhados para administrar grandes coleções e estão cada vez mais presentes em OPACs e outros ambientes de recuperação da informação. Os tesouros normalmente mostram o termo de indexação controlado com as relações de termos específicos ou genéricos e podem ser apresentados em uma janela durante a formulação da estratégia de pesquisa para ajudar o usuário na seleção dos termos. Frequentemente os termos podem ser selecionados de uma lista simplificada do tesouro clicando-se sobre o termo. *Links* de hipertexto podem ser usados para a navegação das diferentes ocorrências do mesmo termo no tesouro. Outra aplicação de tesouros é como base para a indexação automática, onde todos os termos dos documentos que aparecem no tesouro geram uma entrada no índice invertido. Aplicações relacionadas de tesouros estão na criação de redes semânticas e bases de conhecimento semântico.

Indexação baseada em linguagem natural e indexação baseada em linguagem controlada são largamente adotadas em muitas aplicações de recuperação da informação. Ambas são adotadas na recuperação por CD-ROM, serviços de pesquisa *online*, sistemas de gerenciamento de documentos e OPACs. As linguagens controladas de indexação são indicadas por serem mais coerentes e, portanto, mais eficientes e de fácil compreensão para quem faz a busca; no entanto, as experiências feitas a esse respeito não conseguiram comprovar tal alegação de modo convincente. O dilema face aos desenhos dos sistemas é que, oferecer outra coisa além da indexação pela linguagem natural no contexto dos grandes bancos de dados disponíveis na Internet, seria proibitivamente caro. Por outro lado, indexação por linguagem controlada é vista como valiosa em ambiente utilizado por usuários não experientes por não exigir a navegação por todas as variantes inerentes da linguagem natural. Esforços significativos estão sendo voltados para o desenvolvimento de interfaces de sistema que gerenciam esta variabilidade, implícita ou explicitamente, a favor do usuário. Muitas bases incluem termos extraídos de linguagens de indexação, além de suportar buscas nos textos dos registros, unindo assim as duas opções.

Além disso, a busca da informação conta com outros recursos como a lógica booleana, utilizada na maioria dos sistemas, e que é uma lógica empregada para ligar os termos que descrevem os conceitos presentes no enunciado da busca. Os operadores lógicos booleanos são E, OU e NÃO.

A outra lógica de busca, menos comum, baseia-se nos termos ponderados, que conta com a indicação de que certos termos são mais importantes que seus vizinhos. Pode ser usada como auxiliar para diminuir a quantidade de registros resultantes a partir da busca pelos operadores booleanos ou para a formulação de perfis de busca no caso de um sistema que queira implementar, por exemplo, DSI. No segundo caso, pesos diferenciados são fornecidos a termos em função dos perfis de busca. Os pesos atribuem uma classificação segundo a relevância do termo existente no documento localizado. Então os perfis de busca combinam os termos e seus respectivos pesos numa soma

simples, resultando como itens para a pesquisa realizada aqueles que apresentam pesos superiores a determinados pontos. No primeiro caso, emprega-se a lógica de busca com termos ponderados para complementar a lógica booleana. Como resultado da busca são exibidos apenas os registros considerados relevantes segundo a classificação decorrente da ponderação alcançada.

Outros recursos básicos de recuperação são:

- mensagens de ajuda;
- possibilidade de alteração de valores pré-programados (*default values*) como largura e altura da tela;
- exibição dos termos de busca para ajudar na seleção para busca;
- facilidade para entrada de termos para executar a busca depois da seleção;
- combinação dos termos de busca por uma ou ambas lógicas descritas acima;
- especificação dos campos de busca para possibilitar pesquisa mais precisa;
- truncamento de palavras para fazer buscas a partir de radicais das mesmas;
- operadores de adjacência e de proximidade que ajudam na descrição de um assunto representado por várias palavras;
- buscas de amplitude e limitação de amplitude, usadas para campos numéricos ou de datas, como indicadores de igualdade ou não igualdade e de conjunto maiores ou menores;
- flexibilidade na exibição de registros, ao permitir que o usuário escolha quais registros quer ver integralmente a partir de uma lista e com qual formato quer visualizá-los;
- gerenciamento de busca que implica a possibilidade de rever a estratégia de busca adotada e gravar um perfil de busca para utilização posterior;
- opções avançadas de exibição, adequadas, por exemplo, para textos longos em que seja possível correr o texto para frente e para trás, visualizar apenas os tópicos principais para serem selecionados, ou em outros tipos de textos ou de dados numéricos, ordenar os registros segundo a ordem que se deseja;
- buscas em múltiplos arquivos como no caso dos hospedeiros *online* que dispõem várias bases de dados, a partir das quais pode-se realizar a mesma busca simultaneamente em todas as bases, com a possibilidade de eliminar as duplicatas entre os registros resultantes;
- exibição de tesouro com os termos e suas relações auxiliando na escolha dos termos e na restrição e ampliação da busca;
- hipertexto que depende, para garantir buscas efetivas, do estabelecimento, por parte de um indexador, de ligações conceituais entre os documentos.

Também como fator envolvido na recuperação da informação estão as estratégias de busca, ou seja, o conjunto de decisões tomadas e procedimentos adotados durante uma busca.

A **produção de bases de dados** é realizada corporativamente (geradas dentro de uma organização e, quase sempre, para atender a ela mesma) ou por vendedores externos (bases de dados comercializadas). Gerenciadores de informação (*information management*) têm a responsabilidade de gerenciar o acesso a ambos os tipos de fontes de informação e, especificamente, a interface entre a informação das duas categorias de fontes (o que freqüentemente é mais problemático). Isto porque, estes recursos têm diferentes origens, *softwares* de recuperação e características. Além destas duas grandes categorias de sistemas de informação, usuários individuais também desenvolvem seus próprios sistemas de informação pessoal. Gerenciadores de informação podem precisar compreender o relacionamento entre estes sistemas e os sistemas corporativos e externos e devem estar em posição de fornecer treinamento para estes fins. Na edição anterior, afirmou-se que as bibliotecas especializadas, como as de empresas, técnicas e de órgãos governamentais, mantêm bases de dados locais, enquanto bibliotecas públicas e universitárias dispensam pouca atenção a suas próprias bases de dados locais de recuperação da informação, mas todos os tipos de bibliotecas e unidades de informação fazem uso do acesso a bases de dados externas.

Três gerações dos **sistemas de recuperação da informação** são sistematizadas:

- metadados: interfaces baseadas em comandos; usuários intermediários especialistas; um número limitado de sistemas *online* em organizações e disponíveis externamente;
- dados em texto integral: interfaces baseadas em comandos e em menus; recursos adicionais de recuperação como hipertexto e recursos de pesquisa em texto integral; interfaces baseadas em DOS; acesso pelo usuário final mas nem sempre possível; sistemas *online*; primeiros sistemas baseados em CD-ROM;
- multimídia: interfaces GUI (graphical user interface); foco no acesso ao usuário final; orientação ao mercado e ênfase sobre pacotes de produtos; armazenamento e distribuição em CD-ROM ou em redes de alta capacidade; multimídia; intermediários como treinadores; maior uso em casa e em ambientes de acesso público.

Como futuro da recuperação da informação, aponta-se que muitos sistemas de recuperação da informação foram desenvolvidos e avaliados, inclusive para oferecer as melhores vantagens comerciais. No entanto, a recuperação tem tradicionalmente se preocupado com informação razoavelmente estruturada contida em bases de dados relativamente bem definidas. Neste sentido, é necessário levar em conta que a expansão corrente de acesso a redes via Internet está oferecendo enorme quantidade de informação mal definida para grande número de usuários que são relativamente novos em recuperação de informação. Sistemas de recuperação adequados para este ambiente podem ser efetuados por meio de:

- melhores desenhos de sistemas: incluem tentativas para melhorar a eficiência e eficácia do sistema, incluindo características como requisitos de armazenamento e de recuperação rápida. Trabalhos nesta área procuram superar as limitações dos arquivos invertidos pelo desenvolvimento de métodos rápidos de explorar o conteúdo de uma base de dados. Duas linhas de desenvolvimento têm sido investigadas. A primeira delas é a de algoritmos de exploração de textos para melhorar a velocidade de pesquisas em arquivos seriados. A outra é procurar soluções baseadas em *hardware*, a maioria das quais acelera a exploração de textos com o emprego de processadores paralelos, uma vez que eles possibilitam que várias operações sejam executadas simultaneamente.

- melhoria de estratégias e recursos de recuperação: busca levar a que os métodos de descrições dos documentos sejam cada vez mais condizentes com as descrições das consultas, como alternativa à lógica de pesquisa booleana. Trabalhos nesta área têm levado a um grande uso das buscas baseadas na melhor coincidência e nos *links* de hipermídia. Ultimamente o usuário está interessado em ser hábil para interagir com o sistema usando linguagem natural e quer um mecanismo de recuperação que forneça um número razoável de registros relevantes (e não uma enormidade deles) mesmo em grandes bases de dados não estruturadas. Além disso, as respostas do sistema não devem demorar mais que dez segundos. Produtos disponíveis usam uma combinação de consultas em linguagem natural, listas de conteúdos ou seleção de conceitos e técnicas de sistemas especialistas para efetivar um ambiente de pesquisa que depende de indexação automática.

- melhores desenhos visando a qualidade da interface homem-máquina: trabalhos nesta linha têm focado interfaces GUI que são mais amigáveis e auto-explicativas, além dos sistemas de computador intermediários ou auxiliares (*front-ends*) que se interpõem entre o sistema e o usuário. Mais recentemente, trabalhos têm procurado *front-ends* que simulam pesquisas pela melhor coincidência ou modelam ações de um intermediário humano. Esta pesquisa usa técnicas baseadas em conhecimento que são do campo da Inteligência Artificial.

Estas três linhas são apresentadas como os modos pelos quais os sistemas poderão ser melhorados, mas ressalta-se que as bases de dados existentes e os serviços delas derivados representam uma força inercial importante levando a que estes avanços sejam bastante lentos.

Os **sistemas de gerenciamento de documentos** são sistemas projetados para suportar a criação, armazenamento e a conseqüente recuperação e disseminação de documentos e/ou representações de documentos (ou metadados) em formato eletrônico. Estes documentos podem ser: textos, gráficos, som, imagens fixas ou móveis, vídeo ou uma mistura deles na forma de um documento multimídia. Os sistemas de gerenciamento de documentos são aplicáveis em um grande número de programas diferentes de recuperação de informação, desde o acesso de documentos eletrônicos na Internet ou Intranet ao acesso a bases de dados em CD-ROM e arquivos institucionais e pessoais. No caso dos sistemas de gerenciamento de documentos adotados por organizações para o gerenciamento de suas coleções de documentos, gerentes de informação estão se confrontando com decisões associadas à seleção e gerenciamento destes sistemas. De uma forma geral, bibliotecas e gerentes de informação têm importante papel nestes contextos.

Muitos sistemas de gerenciamento de documentos têm relacionamento próximo com sistemas de gerenciamento de bibliotecas. Além disso, os sistemas de gerenciamento de documentos surgiram dos primeiros sistemas de recuperação de textos que foram desenvolvidos em bibliotecas especializadas e unidades de informação e que davam suporte aos serviços de indexação e resumos. Documentos baseados em textos, em geral, eram previamente gerenciados através das primeiras gerações de sistemas de gerenciamento de documentos, então descritos como sistemas de gerenciamento de textos ou sistemas de recuperação de textos. Sistemas de gerenciamento de textos eram distintos de outros *softwares* de gerenciamento de bases de dados por um conjunto de características relativas ao gerenciamento efetivo de

aplicações baseadas em textos. Estas características ainda são aplicáveis aos sistemas de gerenciamento de documentos:

- capacidade de trabalhar com muitos campos de tamanho variável;
- acesso aos registros realizado normalmente por intermédio de um arquivo invertido de palavras do texto (ou outras chaves de índice) dos registros que compõem a base de dados;
- conjunto de recursos de recuperação baseados em palavras dos registros, os quais são necessários em função do controle limitado quanto à forma como a chave de busca aparece no registro;
- ênfase no gerenciamento de uma ou mais bases de dados distintas, onde a habilidade para puxar dados de bases de dados relacionadas não é central na aplicação;
- aplicações relativamente fixas que exigem recursos limitados de programação ou desenvolvimento de sistemas.

Apesar do crescimento da abordagem multimídia, muitos documentos ainda são em sua maioria baseados em textos. O volume das informações baseadas em texto está aumentando a uma velocidade alarmante e sua diversidade de forma, desde memos relativamente desestruturados, cartas ou artigos de jornal aos relatórios, diretórios ou livros mais formalmente estruturados é continuamente ampliada. Muitos dos dados que uma organização possui são textuais por natureza, embora nem todos possam estar disponíveis no formato eletrônico. Contudo, com o uso difundido de práticas de automação de escritórios tais como processamento de textos, correio eletrônico e EDI (electronic data interchange), uma proporção crescente de registros das organizações são em formato eletrônico. Sistemas de gerenciamento de documentos são crescentemente usados para realizar armazenamento, publicação e distribuição corporativa destes documentos, embora reconhecendo também que este sistema não deve ser usado isoladamente no âmbito das atividades dos sistemas de informação da organização. Quanto aos tipos de aplicações, cada organização é diferente e suas necessidades podem ser únicas. Há também aplicações de pequena escala que são usadas, por exemplo, por pesquisadores individuais para a criação de suas próprias bases de dados bibliográficos (muitos dos primeiros programas para textos eram de natureza bibliográfica).

Os recursos de recuperação associados ao sistema de gerenciamento de documentos são centrais para a operação efetiva do sistema e servem para distinguir tais sistemas de outros sistemas de bases de dados. Estas facilidades eram existentes apenas nos sistemas de gerenciamento de textos e de pesquisa *online*. Com o crescimento da conectividade, mais usuários têm necessitado de recursos de pesquisa sofisticados para navegar a seu modo através de vastas quantidades de conhecimento, recursos estes representados agora pelas ferramentas de pesquisa da Internet e sistemas de acesso público de bibliotecas. Contudo, sistemas de gerenciamento de documentos continuam a refinar os recursos de indexação e de pesquisa. Alguns oferecem recursos de indexação únicos, incluindo indexação em qualquer campo e combinações com indexação automática, além da indexação realizada pelo homem. Recursos para construção e manutenção de tesouro são importantes e podem permitir, por exemplo, a seleção de termos preferidos e de termos genéricos. O uso de interfaces GUI tem facilitado a edição de pesquisa estratégica, a utilização de pesquisas prévias, o *browsing* e a seleção de documentos relevantes para impressão posterior ou para baixar

para uma base de dados pessoal. A pesquisa por hipertexto tem sido uma das maiores inovações nos últimos anos.

O **mercado de sistemas de gerenciamento de documentos** conta com mais de 200 pacotes disponíveis no âmbito europeu e norte-americano. Embora alguns pacotes satisfaçam a todos os critérios de um sistema de gerenciamento de documentos, existem outros que apresentam apenas algumas das características desejáveis. Alguns são desenhados para suportar o gerenciamento de grandes bases de dados partilhadas e distribuídas de uma organização internacional e possivelmente oferecer acesso a usuários de fora da organização. Outros são desenhados para uso individual. Alguns têm suas origens na criação de bases de dados bibliográficos que posteriormente incluíram textos integrais, mas foram previamente descritas como sistemas de gerenciamento de textos ou sistemas de recuperação de textos. Outros estão enraizados em sistemas de criação de documentos e *softwares* de editoração. Alguns sistemas de gerenciamento de documentos são: askSam, BASIS Plus, CAIRS, Inmagic DB/TextWorks, TINLIB e Folio.

Dentre eles, são descritos o CAIRS e o BASIS Plus:

O **CAIRS** (Computer Assisted Information Retrieval System), produzido no Reino Unido, desde há muito tempo oferece recursos de gerenciamento de documentos e de gerenciamento de bibliotecas. O conjunto de produtos inclui:

- sistema de indexação em texto livre;
- recursos para *links* hipertextuais entre os dados do sistema CAIRS e um recurso multimídia externo no mesmo computador ou em rede local;
- gerenciamento de imagens e textos para coleções de arquivos e museus;
- sistema de gerenciamento de biblioteca combinado com sistema de gerenciamento de textos destinado a grandes bibliotecas públicas, universitárias e de empresas;
- *software* de pesquisa para o usuário baseado em GUI;
- *software* de criação e pesquisa em CD-ROM;
- sistema para edição Web.

Catálogos projetados pelos bibliotecários ou outros profissionais podem ser gerados e indexados em uma variedade de formas diferentes. A indexação pode suportar qualquer campo, incluindo texto livre, frase e termos selecionados do texto de forma automática ou intelectual. Títulos e resumos podem ser indexados por palavras com uso de listas de palavras vazias.

Entre outras características possui: pesquisa booleana, edição de perfil de usuários, pesquisa por *browser* e expansão, geração de índices, boletins e relatórios, DSI, segurança, menus e comandos, edição de dados, exportação e importação de dados. Para recuperar, pode-se usar um tesouro ou um índice.

O **BASIS Plus**, produzido nos Estados Unidos, foi desenvolvido como um sistema de gerenciamento e recuperação de documentos para toda a organização. É desenhado para gerenciar documentos eletrônicos em uma variedade de diferentes formatos. O BASIS Plus é baseado em cinco pilares: localização, gerenciamento, controle, distribuição e integração.

A localização de documentos permite:

- pesquisar por palavras específicas e frases no texto;
- pesquisar por qualquer campo do registro do documento;
- encontrar informação com base em qualquer palavra, frase, tópico ou valor numérico;
- consultar índices a fim de captar idéias para possíveis termos de buscas;
- pesquisar por prefixo, sufixo ou cadeias de caracteres;
- expandir termos de pesquisa com uso de tesouro.

O gerenciamento da coleção de documentos é realizado:

- pela atualização instantânea de documentos e do conteúdo de índices após alterações realizadas;
- pelo monitoramento de quando e quem faz as alterações nos documentos;
- pela recuperação de coleções de documentos;
- segurança de que a coleção completa de documentos esteja intacta e que os documentos recuperados sejam atuais e precisos;
- exibição e exportação do texto de qualquer documento em formato legível;
- acesso ao histórico do registro do documento.

A segurança é garantida pela proteção ao acesso a grupos de documentos ou partes de documentos individuais para leitura e escrita, pela proteção flexível de documentos quando há muitas contribuições para um documento e pelo controle sobre a produção de um documento revisado ou com porções revisadas.

Uma questão importante quanto à integração do sistema é a habilidade para apresentar simultaneamente gráficos, imagens e textos, característica válida para qualquer estação de trabalho.

Como futuro dos sistemas de gerenciamento de documentos, mais e mais organizações vêm reconhecendo o valor do efetivo gerenciamento de suas coleções de documentos eletrônicos, relevando o papel dos sistemas de gerenciamento de documentos. É provável que um número significativo destas aplicações venha a ser suportado pela tecnologia de Intranet porque ela oferece o atrativo da plataforma independente. Grandes organizações sempre terão uma mistura de plataformas e gerações de estações de trabalho. A principal questão para fornecedores de sistemas de gerenciamento de documentos é se estas aplicações farão uso das antigas versões estabelecidas destes sistemas ou se os novos produtos de Intranet absorverão uma proporção significativa do mercado.

No escopo da recuperação de informação, constam o ambiente Internet e o suporte CD-ROM, os serviços como os de pesquisa *online*, a notificação corrente e a

entrega de documentos (*document delivery*), além de produtos como os índices impressos.

A **Internet** é um ambiente formado por uma série de redes ligadas que operam na base de padrões de protocolos de redes. A World Wide Web é um importante componente da Internet. Serviços e recursos disponíveis na Internet incluem: *listservs* e grupos de discussão, bases de dados temáticas, informação comunitária, recursos governamentais, catálogos de bibliotecas, recursos comerciais, transações comerciais e entrega de documentos (*document delivery*). Com tantos e vastos recursos e bases de dados, tornou-se importante desenhar interfaces que ajudem os usuários a pesquisá-los. Dois tipos de ferramentas são usados para pesquisar na Internet: os *browsers* e as ferramentas de busca. Os *browsers* possibilitam percorrer e mover-se entre *sites* e páginas com base em *hiperlinks*, mas a identificação de recursos específicos necessita de uma ferramenta de pesquisa. Aplicações de biblioteca na Internet incluem: novos modos de acesso a recursos de biblioteca, *home pages* interativas, ligações para informações remotas, aquisições, catalogação e classificação. As Intranets refletem o uso da tecnologia de Internet para o sistema de comunicação interna de uma organização. Questão que necessita atenção para dar conta do desenvolvimento das tecnologias de Internet é ironicamente chamada de “World Wide Wait”, além da segurança, da propriedade e da estrutura.

O **CD-ROM** tem se tornado cada vez mais importante como uma mídia para o armazenamento e disseminação da informação, compondo uma das áreas da indústria da informação. Bases de dados em CD-ROM podem ser categorizadas em: bibliográficas, catálogos comerciais de livros (*book-trade*), fontes, referências rápidas, bases de dados multimídia e mistas. Recursos para pesquisa em CD-ROM podem não se diferenciar significativamente dos recursos para pesquisa em outras bases de dados, mas produtos em CD-ROM têm estado na frente em desenho de interface, como as GUIs que estão se tornando comuns. O CD-ROM é um produto do usuário final e usuários que nunca pesquisaram uma base de dados com sucesso contam com o CD-ROM como um produto mais amigável para uso. Com o CD-ROM, o papel do gerenciador de informação como intermediário tornou-se menos central, com a função de treinamento ganhando maior relevância. O futuro do CD-ROM dependerá do desenvolvimento de outras mídias de armazenamento e da economia relativa do uso de redes nacionais e internacionais comparadas com o acesso ao CD-ROM localmente.

Um **serviço de pesquisa online** procura montar uma lista de bases de dados desenhadas para encontrar as necessidades de um público específico e atua como um intermediário entre o produtor de bases de dados e o usuário final. Pesquisas *online* podem ser realizadas por intermediários ou usuários finais. Os tipos de serviço de pesquisa *online* incluem o tradicional mercado de serviços, serviços de pesquisa especialista, editores e serviços bibliográficos.

Já os **serviços de notificação corrente** são considerados importantes para o estabelecimento de inteligência competitiva. Seus componentes básicos são: bases de dados, perfis de interesse de usuários, notificações, *feedback* e entrega de documentos. Serviços de notificação corrente podem ser divididos em corporativos e de fornecedores externos. Estes últimos incluem: serviços de sumários e serviços de pesquisa *online*. O futuro dos serviços de alerta é parte de uma solução de informação integrada.

No contexto dos serviços de notificação corrente, a **entrega de documentos (*document delivery*)** é a área na qual tem havido a mudança mais significativa. Entrega de documentos pelo correio ou por aparelho de fac-símile e o abastecimento da cadeia

entre o serviço de notificação corrente, o usuário e o serviço de entrega de documentos era freqüentemente pequeno, enquanto sua eficácia variou dependendo do documento particular que estava sendo requisitado. A digitalização de documentos oferece muito mais rapidez e efetiva entrega, embora correio e fac-símile permanecem como opções. Além disso, os principais fornecedores de serviços de notificação corrente têm feito ligações ou alianças estratégicas com serviços de entrega de documentos.

Por fim, **índices impressos** foram um dos primeiros produtos de sistemas baseados em computador. São usados em serviços de indexação e resumos, livros, periódicos e em um conjunto de outros contextos. Há duas categorias principais de índices impressos: aqueles baseados na manipulação de termos em títulos e aqueles baseados na manipulação de cadeias de índices assinalados intelectualmente. Tais cadeias de índices são também usadas nas ferramentas de busca da Internet. Índices de citação são um tipo especial de índice impresso.

### **Parte III – Sistemas de gerenciamento de bibliotecas**

O tópico sobre os **sistemas de gerenciamento de bibliotecas** inclui também uma visão geral dos sistemas no mercado e as atividades de entrega de documentos (*document delivery*) no contexto dos serviços e produtos da indústria da informação.

Sistemas de gerenciamento de bibliotecas são atualmente ferramentas essenciais para o suporte de um efetivo serviço ao usuário, para o gerenciamento de estoques e, em geral, para o gerenciamento dos serviços oferecidos pelas bibliotecas e outras agências engajadas na provisão de acesso a coleções de documentos. O foco de tais sistemas está na manutenção, desenvolvimento e controle de documentos de uma coleção, sendo a informação mais corrente, a mais efetiva no controle da operação. Muitos sistemas contemplam a maioria das coleções de biblioteca.

Os primeiros sistemas eram poucos, com algumas bibliotecas desenvolvendo primeiramente sistemas de catalogação e outras, ao contrário, centradas nos sistemas de controle de circulação. Os sistemas correntes são sistemas integrados baseados em arquitetura de bases de dados relacionais. Em tais sistemas, os arquivos são interligados de tal modo que eliminação, adição e outras mudanças em um arquivo, automaticamente ativam mudanças apropriadas nos arquivos relacionados, oferecendo maiores oportunidades de economia de pessoal e melhores informações gerenciais.

As principais funções presentes nos sistemas de gerenciamento de bibliotecas são as que seguem abaixo, sendo essenciais as cinco primeiras:

- encomendas e aquisições;
- catalogação;
- OPACs e outras formas de catálogo;
- controle de circulação;
- controle de publicações seriadas;
- gerenciamento de informações;

- empréstimo entre bibliotecas;
- informação comunitária.

Nos sistemas de **encomendas e aquisição**, a informatização é adequada ao processo de realização de compras, uma vez tratar-se de um processo administrativo relativamente simples.

Os sistemas de **catalogação** incluem as diversas formas de criação de catálogos: arquivos de registros MARC comercialmente disponíveis; dados do catálogo coletivo do acervo de várias bibliotecas ou outra base de dados compartilhada; registros do catálogo da própria biblioteca; sistema de aquisição da biblioteca; e catalogação local. Exceto na catalogação local, os registros dos arquivos citados são selecionados e modificados para se conformarem às necessidades locais, dados adicionais (como códigos de classificação) são incluídos e, então, estes registros são depositados no catálogo da biblioteca. As principais características de um sistema de catalogação são: entrada de dados, importação (*downloading*) e controle de autoridade.

Assim, dentre as vantagens oferecidas pela catalogação automatizada podem ser citadas:

- o registro catalográfico passou a ser o registro fundamental do sistema de gerenciamento de bibliotecas por ser utilizado em outros subsistemas como controle de circulação e aquisição;
- o intercâmbio de registros catalográficos levou a uma maior padronização dos mesmos;
- os catálogos coletivos contribuíram para empréstimos mais eficazes entre bibliotecas;
- o surgimento de políticas de aquisição cooperativa e de iniciativas de armazenamento cooperativo;
- os catalogadores não precisam mais fazer a intercalação de fichas nem qualquer outra rotina de manutenção de catálogos;
- o processo de catalogação tornou-se mais estruturado e rápido em função do trabalho cooperativo que, por sua vez, levou ao compartilhamento da experiência profissional e da redução do trabalho de catalogação e, portanto, dos recursos financeiros da biblioteca.

O módulo de **OPAC** vem normalmente com alguma OPAC *default* que pode ser usada como ponto de partida. A biblioteca pode identificar e desenhar menus com mensagens específicas para uso da equipe e do público da biblioteca. Alguns sistemas também permitem definir os campos indexados e os menus e informações disponíveis à equipe e não disponíveis ao público. Recentes inovações em OPACs incluem terminais em quiosques e *links* para a Internet.

Os sistemas de **controle de circulação** são destinados a obter a disponibilidade dos documentos da biblioteca por meio da manutenção de registros que especificam, no mínimo: o material que está disponível no acervo para empréstimo; o material que está emprestado e com quem ou onde pode ser recuperado; e quando o material que está emprestado voltará a estar disponível para outros usuários.

O **controle de publicações seriadas** apresenta três subsistemas que também compõem o sistema para monografias: sistema de aquisição, catalogação e controle de circulação. As publicações seriadas diferenciam-se das monografias por conta do seu caráter contínuo que gera alguns requisitos específicos.

Os sistemas de **gerenciamento de informações** contêm muitos dados que, se extraídos, resumidos e analisados adequadamente, servirão de apoio ao processo decisório administrativo, sendo assim nomeados de módulo de informações gerenciais.

Os sistemas de **empréstimos entre bibliotecas** realizam o processamento decorrente das solicitações de empréstimo de material pertencente a outros acervos, o que inclui: a geração de pedidos iniciais a outras bibliotecas; a notificação de usuários sobre a disponibilidade do material; a manutenção de registros do material solicitado e do material que acha emprestado; o controle de devoluções; e o monitoramento em geral dos usuários e dos pedidos.

Além dos módulos-padrão que caracterizam um sistema de gerenciamento de bibliotecas, recursos extras têm sido acrescentados ao sistema, em função de o mercado ter se tornado mais competitivo. Assim, alguns sistemas incluem um módulo de **informação comunitária** que é similar ao módulo OPAC mas com maior flexibilidade no formato dos registros, na pesquisa e na apresentação das diferentes informações. Um serviço de informação comunitária consiste basicamente de listas de nomes e endereços de instituições, além de informações como oportunidades de lazer em âmbito local, empregos, atividades para crianças e informações para os cidadãos. Permite à biblioteca desenvolver sua própria base de dados contendo todo e qualquer tipo de informação que julgue interessar a seus usuários, sendo nada mais que um módulo de gerenciamento de base de dados.

Outros sistemas, em especial, para bibliotecas especializadas oferecem recursos para a criação de boletins impressos ou eletrônicos, como os boletins de notificação corrente, para a geração de notificações de DSI e para a produção de índices impressos. Outros sistemas oferecem recursos de escritório como processadores de textos e correio eletrônico para facilitar a realização de processos básicos.

Nos últimos anos, o **mercado de sistemas de gerenciamento de bibliotecas** cresceu e amadureceu. A maior parte dos fornecedores de sistemas está no mercado há alguns anos, tendo passado a se concentrar mais recentemente no aperfeiçoamento de seus sistemas, num esforço de ampliar e consolidar sua base de consumidores. Em geral, o mercado de sistemas de grande porte (para bibliotecas universitárias, por exemplo) está estabilizado e saturado, enquanto sistemas para pequenas e médias bibliotecas surgiram mais recentemente.

Neste sentido, o Bookshelf foi desenvolvido no Reino Unido, no início da década de 80, a partir de um projeto de pesquisa financiado pela British Library, com o objetivo de preencher uma lacuna existente para sistemas de gerenciamento de bibliotecas destinados a bibliotecas de médio porte. Seu sucessor, o Heritage, possui os seguintes módulos: catalogação (contando, entre outros recursos, com tesouro), OPAC, controle de circulação, gerenciador de informação e aquisição. Outros programas são o Dynix, o Advance e o Unicorn, os quais contêm a maioria dos recursos e aperfeiçoamento que seguem abaixo:

- aquisição *online*;
- criação rápida de um catálogo a partir da captura de registros MARC;

- melhor controle de autoridade: ao mesmo tempo em que se trava o debate entre os recursos de busca em linguagem natural e a necessidade de controle sobre os termos e cabeçalhos de indexação, estes últimos vêm sendo melhorados em vários sistemas. Alguns sistemas, em especial, aqueles voltados para o setor de bibliotecas especializados, permitem que se use uma maior gama de relações no tesauro (como “ver” e “ver também”), o qual também pode ser consultado, com o emprego de janelas, durante a indexação e a pesquisa;

- controle de circulação: interfaces GUI dão acesso mais direto a funções específicas, permitindo estações de trabalho *self-service* para renovação e reserva;

- empréstimo entre bibliotecas;

- gerenciamento da informação: alguns sistemas de gerenciamento de bibliotecas passaram a se preocupar com as informações gerenciais que podem gerar. Estes recursos costumavam estar presentes no próprio módulo de geração de relatórios, mas as opções de relatórios têm se tornado mais variadas e provavelmente venham a ser oferecidos crescentemente recursos para a análise de dados, a fim de identificar tendências e correlações entre variáveis;

- OPACs;

- GUIs;

- conectividade;

- informação comunitária.

Os sistemas vêm se aperfeiçoando em diversas áreas, mas levará algum tempo para que estes aperfeiçoamentos sejam levados à prática, pois as bibliotecas melhoram seus sistemas de forma gradativa. Há também um mercado de *softwares* voltado especialmente para bibliotecas escolares. Hoje, os módulos principais de um sistema para bibliotecas como os de aquisição, catálogo, OPAC e controle de publicações seriadas e circulação são considerados essenciais, sendo a norma agora um pacote que apresente mais que estas características.

Quanto à **entrega de documentos (*document delivery*)**, esta atividade tem sido considerada o coração dos serviços oferecidos por bibliotecas, editores e outros da indústria da informação. A entrega de documentos inclui tanto os documentos impressos quanto os eletrônicos. Embora mecanismos para entrega impressa estejam bem estabelecidos, este serviço tem se tornado mais eficiente por meio da introdução de pedidos eletrônicos. Além disso, tem havido rápido desenvolvimento nos últimos anos na entrega de documentos eletrônicos, particularmente a partir das revistas eletrônicas. Há um número de diferentes agentes na entrega de documentos, entre eles:

- redes e consórcios de bibliotecas, incluindo bibliotecas nacionais e serviços de entrega de documentos, sendo o foco central destas redes o compartilhamento de recursos e a criação de bases de dados catalográficos e bibliográficos compartilhados, embora elas também ofereçam um conjunto de outros serviços de valor agregado como sistemas de gerenciamento de bibliotecas e serviços de pesquisa *online*;

- entrega de documento em CD-ROM que é particularmente apropriado no contexto de livros, enciclopédias e outros documentos longos;

- serviços comerciais de entrega de documentos que freqüentemente operam em alianças com serviços de pesquisa *online* e fazem uso de recursos de outras bibliotecas;
- fornecedores de bibliotecas e agentes de assinaturas de publicações seriadas que têm usado redes eletrônicas e intercâmbio eletrônico de documentos para melhorar o processo de encomenda do documento e processos associados tais como alerta, fatura e processamento;
- editores e fornecedores de revistas eletrônicas: há atualmente um número limitado de revistas eletrônicas mas significativo interesse na digitalização de revistas que tem sido originalmente distribuídas na forma impressa.

### **Automação de Bibliotecas, Recuperação da Informação e bibliotecas digitais, segundo Jennifer Rowley**

Quanto à comparação entre as edições analisadas do livro “The electronic library” (1993 e 1998), o principal ponto a ser ressaltado refere-se à distinção mais acentuada apresentada em 1993 entre os sistemas de gerenciamento de bibliotecas e os sistemas de gerenciamento de informações textuais. Esta divisão demonstra que, já neste momento, não havia uma abordagem unilateral. O entendimento, então, era de que a maioria das bibliotecas trabalhava tanto com o gerenciamento de seus serviços quanto com a recuperação de informações, apesar das diferentes prioridades destinadas a cada uma delas, e que havia uma tecnologia comum subjacente a estas duas áreas que costumam decorrer em desenvolvimentos similares. O argumento era de que as aplicações da Informática em bibliotecas e em serviços de informação referem-se a um campo integrado e esta perspectiva, por ser mais abrangente, poderia ajudar profissionais da informação na compreensão de sistemas específicos e na capacidade de prever tendências. Considerava-se que os sistemas de gerenciamento de bibliotecas possibilitam a automação das atividades de processamento dos seus serviços de tratamento e disponibilização de seus acervos, enquanto, nas aplicações voltadas para a recuperação de informações, a tecnologia busca concretizar a aspiração máxima do especialista da área, ou seja, dispor essas informações ao alcance de qualquer pessoa, quando e onde forem necessárias. Também se percebe, nesta edição anterior, uma postura bastante crítica quanto à atuação das bibliotecas.

Estes pontos de vista são relativizados na edição mais recente. Não há críticas acirradas à atuação das bibliotecas e a descrição dos serviços de recuperação de informação está mais próxima daquela dos serviços de biblioteca. Isto é, provavelmente, consequência do reconhecimento das profundas mudanças que vem se desenrolando no contexto das bibliotecas. Por outro lado, é recorrente a afirmação de que muitas mudanças ainda estão por vir e são necessárias. Sendo assim, aceita-se e reforça-se a observação da autora de que os sistemas existentes continuam a reproduzir os processos tradicionais e que uma abordagem mais radical dos sistemas de gerenciamento de bibliotecas implicaria uma reavaliação dos seus processos fundamentais.

Na edição atual, o termo adotado “sistemas de gerenciamento de informações textuais” foi substituído por “sistemas de gerenciamento de documentos”. No entanto, mesmo com a alteração do nome, o texto manteve-se relativamente similar: a definição,

as características técnicas, os itens para avaliação e a lista de produtos apresentada. Os sistemas de gerenciamento de documentos são descritos como originados dos sistemas de gerenciamento de informações textuais. Ao que consta, num primeiro momento, era recorrente o uso dos sistemas de gerenciamento de informações textuais, em especial, para os serviços de indexação e resumos, que necessitavam de recursos flexíveis e potentes para indexação e recuperação de informação. No entanto, algumas tendências ampliaram este tipo de uso. Uma delas foi a massificação da pesquisa e do uso de informação eletrônica por meio do ambiente de Internet que oferece ferramentas de busca menos complexas e um repositório de informações diversificadas (ou seja, não apenas as informações especializadas). Outro aspecto refere-se à importância crescente que vem sendo dada pelas organizações à própria informação que produzem e àquela que consomem e que é acessada externamente. Assim, o termo “sistemas de gerenciamento de documentos” parece ser mais abrangente e condizente com a realidade. O *software* BASIS Plus, por exemplo, que aparece nas duas edições, e outros sistemas apresentados, apontam para o gerenciamento de documentos em empresas, em oposição aos tradicionais sistemas de gerenciamento de documentos: os primeiros centram-se na produção de documentos em uma empresa que é, muitas vezes, realizada coletivamente e conta com um ciclo de produção, distribuição e armazenamento de documentos, e, os segundos, focam o tratamento do conteúdo de documentos produzidos externamente e o armazenamento para recuperação. O tratamento de documentos prontos avançou para a noção de gerenciamento as etapas de produção dos documentos.

Sobre os sistemas de recuperação da informação são tratadas as etapas de indexação, armazenamento e recuperação. Nesta última, descreve-se o papel ocupado pelos tesouros nos sistemas atuais, ressaltando sua crescente existência em OPACs e sistemas na Internet de uma forma geral, em função da facilidade que oferece para pesquisa pelo usuário. Esta afirmação não constava na edição anterior, reforçando a tendência para aproximação entre as características dos sistemas documentários (ou sistemas de recuperação da informação, segundo a forma usual na literatura em inglês) e os sistemas de gerenciamento de bibliotecas.

No entanto, entre os sistemas de gerenciamento de bibliotecas, são apontados vários recursos que são freqüentemente mais presentes nos sistemas voltados para bibliotecas especializadas, como o uso maior de relações do tesouro e consulta do mesmo para indexação e pesquisa, além dos recursos para criação de boletins impressos e eletrônicos de notificação corrente e de DSI e para produção de índices impressos.

Uma alteração entre as edições está evidente na mudança de nome do livro “Computers for libraries” para “The electronic library”. O termos “computador” e “bibliotecas” utilizados conjuntamente no título anterior, não necessariamente indicam a relação entre o tratamento e disseminação de informações e os recursos tecnológicos utilizados para tal. O termo “biblioteca eletrônica” parece pretender indicar esta interface sem que, no entanto, fosse desconsiderado o fato de que este termo apresenta algumas diversidades de entendimento, evidentes nas distintas definições encontradas.

Como o livro divide a apresentação entre recuperação da informação e sistemas de gerenciamento de bibliotecas, alguns elementos se repetem, reforçando a afirmação da autora de que há um conhecimento subjacente a estas partes. Sendo assim, são comuns a estes dois tópicos, alguns produtos e serviços de disseminação de informações e de acesso a documentos como: boletins de notificação corrente, DSI e entrega de

documentos (*document delivery*). Já quanto ao mercado, também apresentado distintamente para cada uma das partes, alguns sistemas de recuperação da informação incluem módulos de gerenciamento de bibliotecas.

Comparando-se a obra sobre bibliotecas eletrônicas de Jennifer Rowley com aquela que aborda a automação de bibliotecas, de William Saffady, vários aspectos podem ser apontados.

Rowley apresenta os recursos para gerenciadores de tesouro e aplicativos de indexação quando trata dos sistemas de gerenciamento de documentos e, no caso dos sistemas de gerenciamento de bibliotecas, são citados os recursos, em constante aperfeiçoamento, de controle de autoridade que incluem os cabeçalhos de assuntos e tesouros com as relações “ver” e “ver também”. Saffady, por sua vez, apresenta os sistemas integrados para bibliotecas incluindo estes aspectos, no entanto, não trata destes recursos no tópico sobre os sistemas de armazenamento e recuperação de textos. Neste tópico, descreve apenas o processo de recuperação pelas palavras do texto por meio de cruzamentos booleanos, proximidade, relevância e, mais recentemente, por *links* de hipertexto; ou seja, não trata dos sistemas para recuperação de documentos com recursos para o tratamento intelectual do conteúdo.

Com relação aos formatos de registro bibliográfico, Saffady aborda unicamente o Formato MARC e suas variações regionais, discorrendo longamente sobre a origem e a ampla estrutura de funcionamento cooperativo e de serviços por instituições fornecedoras de registros. Também não cita formatos para aplicações não catalográficas, apesar de referir-se brevemente, na edição antiga, aos pacotes de *softwares* voltados para as bibliotecas especializadas como o Inmagic, o Star e o MINISIS, a partir dos quais é comum adotar registros de formatos variados. Na edição atual, cita empresas de serviços bibliográficos de catalogação *online* que oferecem registros bibliográficos não usuais, como aqueles de bases de dados de áreas especializadas (ou bases de dados temáticas). O serviço bibliográfico mais importante neste aspecto, o RLIN, é o único que opera como um verdadeiro serviço de informação *online* possibilitando, no entanto, apenas a consulta destes registros bibliográficos não usuais, ou seja, não podem ser aproveitados automaticamente para catalogação. Rowley, por outro lado, ressalta a existência de outros serviços de informação que não os da biblioteca e seu acervo localmente organizado ao apresentar o Formato CCF e sua origem européia pautada nas necessidades dos serviços de indexação e resumos e dos centros de documentação. Esta visão é mais abrangente e condiz com a visão que considera bibliotecas e serviços de informação de forma integrada, por mais que, de fato, seus desenvolvimentos tenham caminhado até então em separado, deixando suas marcas de distinção ainda presentes.

Enquanto Saffady discorre pouco sobre a natureza e as características dos processos, como foi descrito no caso do capítulo sobre catalogação descritiva, Rowley avança bastante neste aspecto, por exemplo, na análise comparativa sobre os sistemas de gerenciamento de bibliotecas e os sistemas de recuperação da informação, assim como, sobre a recuperação a partir de linguagem controlada e a partir da linguagem natural. Ficou nítido, por meio da análise realizada, que a obra de Rowley é de maior amplitude quanto ao entendimento que apresenta sobre os aspectos dos serviços de informação.

Sendo assim, pode-se questionar, a partir da comparação entre Saffady, dos Estados Unidos, e Rowley, do Reino Unido, se a produção desta última, mesmo que arraigada na linha marcada pelo seu idioma, não tenha absorvido algumas influências européias, supostamente decorrentes da proximidade geográfica.

## 2.2.2 INFORMATION RETRIEVAL

A julgar, entre outros, pela abrangente obra sobre bibliotecas eletrônicas de Jennifer Rowley, a área entendida como Automação de Bibliotecas (*Library Automation*) não dá conta do conjunto de conceitos, técnicas e tecnologias envolvidas no tratamento e disseminação de informações com o uso dos recursos tecnológicos existentes, apesar das tendências atuais em sentido contrário. Assim, é na abordagem realizada pela Recuperação da Informação (*Information Retrieval*) que elementos complementares são historicamente encontrados.

### O QUE É INFORMATION RETRIEVAL

Foi Kelvin N. Moers que, em 1950, usou pela primeira vez a expressão *Information Retrieval* publicando, dois anos depois, seus trabalhos com a descrição do sistema de seleção documentária Zatocoding (SUTTER, 1997). Também Moers, no fim dos anos 50, criou a idéia da linguagem documentária que tem como unidade para indexação, não as palavras, mas os conceitos (grupos de palavras ou expressões), inaugurando a noção de descritor (*descriptor*) (CHAUMIER, 1973).

Meadow (1992) fornece alguns conceitos básicos válidos para uma introdução ao tema. Afirma que a Recuperação da Informação envolve o encontro de alguma informação desejada em um repositório de informação ou base de dados. Implicitamente, o conceito em questão é seletividade; exercer seletividade normalmente exige que um preço seja pago em esforço, tempo, dinheiro, ou os três. A informação descoberta não é o mesmo que Recuperação da Informação a não ser que tenha havido um processo de seleção. Copiar um arquivo de disco não é recuperar. Assistir um programa de televisão inteiro, durante o qual o telespectador não exercita o controle sobre o que está sendo mostrado, também não é recuperação. Uma biblioteca é o melhor exemplo de uma instituição voltada à recuperação seletiva. Ninguém vai até ela para ler uma coleção inteira mas para procurar algo seletivamente, freqüentemente algo que satisfaça um conjunto de necessidades informacionais altamente individualizadas.

A Recuperação da Informação é um processo de comunicação. Neste sentido, é um meio de comunicação dos autores ou criadores de registros com leitores, mas que ocorre indiretamente e com um espaço de tempo provavelmente grande entre a criação de uma mensagem ou texto e seu acesso pelo usuário do sistema de recuperação da informação. Os registros de uma base de dados são criados e reunidos sem conhecimento exato de quem irá lê-los ou a partir de quais circunstâncias. As linguagens e os canais de tal sistema de comunicação são completamente diferentes de outros modelos bem conhecidos como o de comunicação de massa.

A Recuperação da Informação não é necessariamente uma atividade de computador, mas o termo sistema de recuperação da informação (SRI) costuma referir-se ao sistema de computador consistindo de *hardware* e *software*, incluindo a base de dados. Contudo, isto depende do contexto no qual a expressão é usada pois todos os princípios da Recuperação da Informação aplicam-se aos processos não mecânicos também.

Freqüentemente pensa-se em catálogos de biblioteca, índices impressos ou enciclopédia como ajuda mecânica à Recuperação da Informação, mas de fato estes recursos representam antigas tecnologias. Enquanto a seleção de registros de um arquivo de inventário ou de um arquivo de contas de depósitos bancários pode ser considerada Recuperação da Informação, o termo é mais comumente aplicado para arquivos de registros de textos ou registros descritivos de textos. A questão de como os textos devem ser interrogados está subjacente à atividade da recuperação da informação. Em função desta percepção, uma linha da Recuperação da Informação adotou como problema central a representação de documentos para recuperação. Esta linha sofreu significativas alterações à medida que mais bases de dados consistiram de textos completos de documentos e não dependiam necessariamente de índices de termos que representavam os documentos.

O processo de pesquisar seletivamente por informação em uma base de dados tem início a partir de, ao menos, dois pontos. O ponto inicial é o usuário, ou a usuário potencial, considerando usuário a pessoa que tem uma necessidade de informação e usa um sistema de recuperação da informação para tentar satisfazer esta necessidade. O ponto final é a pessoa ou mediador que tomou decisões para a coleta, organização e armazenagem de informação para posteriores buscas e recuperação. Há passos complexos até que a informação seja armazenada e o usuário chegue a ela.

Embora qualquer um possa ser usuário de um sistema de recuperação da informação, há três classes principais: especialistas de informação ou bibliotecários de referência; profissionais ou usuários finais com necessidades de informação e considerável conhecimento do assunto como físicos, advogados, químicos ou professores; e outros, essencialmente novatos, sem conhecimento especializado, como estudantes e usuários casuais.

## **HISTÓRIA DA INFORMATION RETRIEVAL: “THE SEVEN AGES OF INFORMATION RETRIEVAL”, Michael Lesk, 1995**

O artigo de Michael Lesk traça um quadro histórico e prospectivo do tema. Apresenta uma sistematização pautada em períodos históricos inspirada por analogia a uma obra de Shakespeare e organizada em: infância (1945-1955), adolescência (anos 60), idade adulta (anos 70), maturidade (anos 80), crise de meia idade (anos 90), realização (*fullfillment*) (a partir de 2000) e reforma (2010).

É já no período histórico inicial da Recuperação da Informação denominado **infância (1945-1955)** que ocorre a origem da tensão entre métodos estatísticos simples e análises de informação sofisticadas. Em 1949, Warren Weaver pensou sobre o possível sucesso dos computadores para a tradução de línguas, afirmando que um livro em chinês é simplesmente um livro em inglês que foi escrito com uso do “código chinês”. Pouco antes, Vannevar Bush apresentou a noção de hipertexto. O processo analítico, que é mais próximo da abordagem de Bush, inclui a indexação realizada pelo homem (que é anterior à existência de computadores) e as experiências em programas de Inteligência Artificial. A linha de Weaver é inteiramente mecânica contando com técnicas de recuperação probabilística. Acredita-se que a Recuperação da Informação tenha dois pais (Weaver e Bush) e que esteja aí a origem da dicotomia entre análises estatísticas e Inteligência Artificial cujas técnicas constituem a história da Recuperação da Informação.

Dentre as previsões de Bush, o desenho da interface de usuário é considerado o mais importante. Ele enfatizou a existência de interfaces individuais, ou seja, personalizadas para o usuário. Até recentemente, todo o desenvolvimento da Recuperação da Informação foi dirigido para interfaces uniformes, que reagiam do mesmo modo para todo usuário e não dependiam de uma organização individual dos arquivos pessoais. Isto levou à existência de melhores ferramentas de pesquisa para a literatura impressa que para as anotações pessoais. Quando cientistas ou engenheiros querem informação, consultam suas notas pessoais, perguntam aos colegas próximos, perguntam aos seus supervisores ou a outras autoridades locais, menos freqüentemente, procuram em periódicos, e quase nunca perguntam a um bibliotecário. Bush visualizou um sistema que, antes de tudo, poderia fornecer notas individuais e memórias dos cientistas. Somente mais tarde, sugeriu que pessoas poderiam comprar publicações convencionais e inseri-las na máquina que imaginou e denominou de Memex. Isto teria sido melhor que o rumo que os serviços de informação tomaram, mas era tão difícil antes dos anos 70 obter informação em computadores que a espécie de sistema que Bush queria não foi praticada.

Nos anos 50, foram criados os índices KWIC e, ainda hoje, a busca simples por palavra é utilizada em muitos sistemas comerciais.

O período correspondente à **adolescência (anos 60)** contou com a criação e a experimentação dos primeiros sistemas de pesquisa, tanto em linguagem natural como com recurso a linguagem controlada e o desenvolvimento de tecnologias para a avaliação de sistemas de recuperação. Este período viu a separação do campo dominante da Ciência da Computação. Muitos dos sistemas comerciais correntes em bibliotecas (como Dialog e BRS) fizeram suas experiências neste período, o qual é considerado o *boom* da Recuperação da Informação.

As primeiras experiências usaram pesquisa mecânica a partir da indexação intelectual, potencializadas com o advento das bases de dados computadorizadas. Assim, a idéia da busca em texto livre surgiu com a recuperação completa de documentos usando uma palavra particular sem custo de indexação prévia. Esta idéia tornou-se e tem permanecido bastante popular, mas imediatamente também atraiu objeções daqueles que acreditam que indexação não era somente um modo de selecionar palavras mas um modo de escolher as palavras certas. Vocabulários oficiais tentaram padronizar a nomenclatura e foram aplicados em muitos sistemas de indexação com o fim de evitar que um documento fosse indexado por uma palavra enquanto outro adotasse algum sinônimo na indexação.

Estes problemas estimularam o desenvolvimento de técnicas de avaliação como a revocação (*recall*) (conjunto de documentos relevantes recuperados) e a precisão (conjunto de documentos recuperados que são relevantes). Estes conceitos foram utilizados como medida dos sistemas de recuperação da informação, por meio da construção do primeiro teste de coleções para medi-los e da troca entre as duas medidas: se um sistema simplesmente recupera mais documentos, a revocação é maior e, provavelmente, haverá mais documentos relevantes, mas a precisão diminui.

Uma série de experiências seguintes mostrou que o valor da indexação intelectual era baixo. Ao menos sobre as pequenas coleções que estavam sendo usadas para os testes, a indexação em texto livre foi efetivamente mais barata. Esforços para ver se termos poderiam ser atribuídos a categorias de tesouro automaticamente ou se melhores tesouros poderiam ser feitos, seguiram uma variedade de métodos. Uma delas foi a

recuperação multilíngüe, usando um tesouro bilíngüe e o mapeamento de palavras de ambas as línguas dentro dos mesmos conceitos.

Também teve início a pesquisa baseada na busca em linguagem natural. Pesquisadores de Inteligência Artificial perguntaram-se por que sistemas de recuperação tinham que ser limitados a recuperar documentos que o usuário ainda teria que ler para encontrar as respostas para suas questões. Assim, iniciaram a construção de sistemas nos quais as atuais perguntas eram recuperadas das bases de dados, enquanto outros trabalharam os problemas de documentos analisados sintaticamente para, ou detectar elementos de dados, ou fazer frases adequadas para recuperação. Neste estágio de desenvolvimento, estes sistemas eram frágeis e funcionavam somente a partir de poucos exemplos. A comunidade da Recuperação da Informação tinha efetivado suas próprias experiências com o uso de frases e tendia a rejeitar estes processos.

Todas estas experiências começaram a estabelecer a Recuperação da Informação como uma disciplina de *software* pouco comum, na qual experiências de avaliação, métodos estatísticos e outras parafernálias da ciência tradicional disputavam um papel maior. Houve relativamente pouca recuperação computadorizada concreta. A maioria dos trabalhos foi de pesquisa e aprendizagem e, embora alguns deles eram em princípio ambiciosos, não houve disponibilização de grandes quantidades de textos legíveis por máquina para acesso por usuários.

Na **idade adulta (anos 70)** surgem os primeiros sistemas com utilização real em diversos campos científicos. Talvez a causa mais importante tenha sido o desenvolvimento de processadores de texto, que levaram a que muitos textos se tornassem disponíveis na forma legível por máquina, configurando-se em material para sistemas de recuperação da informação. A outra tecnologia motivadora foi a disponibilidade de sistemas de tempo compartilhado (*time-sharing*): em vez de perguntas sendo processadas em alguma operação por lote, foi possível apresentá-las diretamente de um terminal e obter uma resposta imediatamente. A recuperação tornou-se mais prática e sistemas foram desenvolvidos para oferecer tais serviços para bibliotecários.

Dentre os sistemas em larga escala estavam os sistemas comerciais como Dialog, ORBIT (Online Retrieval of Bibliographical Information) e BRS que alimentavam fora dos computadores os dados de indexação e resumos. O Chemical Abstracts, por exemplo, gerenciava seus dados no meio impresso até que isto se tornou impraticável: a informatização fez este processo mais rápido e fácil e as fitas que continham os índices eram fornecidas a serviços como o Dialog para acesso por profissionais bibliotecários.

Outro sistema foi o OCLC que usou o Formato MARC da Libray of Congress para catalogação legível por máquina. Como a Libray of Congress catalogava livros, a OCLC adquiriu fitas desta catalogação e imprimiu fichas catalográficas para as bibliotecas membros, eliminando o trabalho delas na preparação destas fichas. Os membros da OCLC utilizaram 1/3 das fichas da Libray of Congress, pois os outros 2/3 eram referentes a obras estrangeiras e outros itens como dissertações, relatórios técnicos e publicações não convencionais. As bibliotecas membros da OCLC catalogaram seus materiais e estes registros ficaram acessíveis a outras bibliotecas, introduzindo a idéia de trabalho cooperativo. Considera-se que este foi o primeiro exemplo de organização intelectual de informação em larga escala, de forma cooperativa e computadorizada, como Bush imaginou.

Todos estes sistemas usaram recursos de pesquisa relativamente simples. Os sistemas *online* eram melhores, mas ainda limitados à pesquisa booleana a partir de

palavras do texto. Quando havia índices de termos realizava-se, de fato, análise intelectual de conteúdo, mas os resumos eram mais pesquisados que os índices.

Embora a maioria destes sistemas operasse sem resumos e índices, esta década viu o início dos sistemas de recuperação em texto integral, pela Reed Elsevier (então denominada Lexis) que disponibilizou o texto integral de decisões do tribunal.

Paradoxalmente, com o crescimento e massificação dos sistemas comerciais de bases de dados, houve um recuo na investigação no campo da Recuperação da Informação durante todo este período. Os anos 70 e 80 foram um período em que pesquisas de bases de dados e automação de escritório explodiram. Recursos foram investidos para pesquisa e, em adição, houve alguma desilusão com promessas que tinham sido feitas nos anos 60 e que pareciam não se concretizar de forma rápida o suficiente. Por exemplo, especulava-se, na primeira parte da década de 60, que dali a 20 anos as bibliotecas deixariam de existir.

Outras promessas, tais como aquelas relacionadas com tradução automática, talvez tenham encorajado financiadores a olhar para outro lado. O surgimento dos departamentos de Ciência da Computação também desencadeou efeitos negativos sobre a pesquisa em Recuperação da Informação, pois ela não recebeu maior importância nestes departamentos e, na anarquia instaurada, era mais fácil para pesquisadores estranhos encontrarem um lugar.

No entanto, houve algum progresso na pesquisa, sendo a mais importante o surgimento da Recuperação da Informação probabilística que media a frequência de palavras em documentos relevantes e não relevantes e usava medidas de frequência de termos para atribuir o peso dado a diferentes palavras. Embora estes algoritmos tivessem problemas computacionais sobre o equipamento de então, a variante (dada pelo peso do termo) foi uma das poucas técnicas que pareceram melhorar a performance sobre a busca simples por palavra, que era a forma então dominante.

Quanto aos pesquisadores de Inteligência Artificial que estavam tentando fazer análise intelectual automaticamente, os problemas dos resultados de tradução automática e da Linguística Computacional os afetaram também e o assunto, nos anos 70, foi o reconhecimento de voz e o início dos sistemas especialistas. Os sistemas especialistas, em particular, ocuparam o mesmo nicho que agentes inteligentes têm hoje. Programas estavam sendo desenvolvidos para recuperar dados do mesmo modo que um bibliotecário fazia. Dizia-se que os anos 80 muito provavelmente seriam considerados a era dos sistemas especialistas e, quando chegassem ao final da década, seria possível telefonar para um sistema especialista para obter avisos e informações sobre tópicos técnicos, legais e médicos.

Na prática, houve uma divisão entre a comunidade de Inteligência Artificial e a de Recuperação da Informação. Os pesquisadores de Inteligência Artificial sentiram que estavam atacando problemas mais fundamentais e complexos e que era necessária uma espécie de conhecimento, sem o que seria insolúvel o entendimento de problemas como aqueles decorrentes dos limites inerentes à abordagem da pesquisa por cadeia de caracteres realizada pela área da Recuperação da Informação. O campo da Recuperação da Informação apontava que pesquisadores de Inteligência Artificial não avaliavam suas experiências pois, de fato, construíram somente protótipos que apresentavam o risco de não serem generalizáveis.

Parte do trabalho de Inteligência Artificial foi a elaboração de linguagens projetadas para a representação precisa de informação chamadas linguagens de representação do conhecimento. Estas linguagens tentaram usar a notação lógica para representar conhecimento geral e ser processada por programas. Muitas destas linguagens foram definidas e seus desenvolvedores sugeriram que grandes quantidades de informação poderiam ser escritas nelas e usadas por sistemas especialistas.

No período considerado a **maturidade (anos 80)**, os processadores de texto cresceram e o preço do espaço em disco diminuiu, significando que mais informação estava disponível na forma legível por máquina. O uso da recuperação da informação *online* expandiu-se de dois modos principais. Um foi a disponibilidade de texto integral em vez de somente resumos e índices e o outro foi a propagação da recuperação *online* para uso por não especialistas como bibliotecas, substituindo ou suplementando suas fichas catalográficas com OPACs. A comunidade da Recuperação da Informação começou a voltar-se agora para um novo trabalho sobre a análise de termos e áreas relativas. A Inteligência Artificial prosperou e decaiu. E o desenvolvimento do CD-ROM forneceu uma opção para o fornecimento de informação e pareceu, por um tempo, ameaçar a idéia de trabalho cooperativo *online*.

Houve um enorme crescimento no número de bases de dados disponíveis em sistemas *online*, assim como, nos tipos de bases, envolvendo dados numéricos e de texto integral, como no caso de muitos jornais e revistas correntes.

OPACs também foram desenvolvidas durante os anos 80. Muitas bibliotecas, graças à OCLC e a cooperativas similares, tiveram todos os seus registros legíveis por máquina na metade dos anos 70. No final da década, vendedores comerciais forneceram *softwares* com OPACs.

Na comunidade de pesquisa, teve início novamente o esforço sobre técnicas de recuperação da informação. O novo material que estava disponível e o interesse em pequenos programas que poderiam rodar em estações de trabalho individuais produziram novas pesquisas em algoritmos de recuperação. Embora algumas delas ainda estavam centradas em modelos matemáticos, houve o crescimento do interesse de novas espécies de métodos de recuperação, como a pesquisa sobre a desambigüidade usando dicionários legíveis por máquina para ajudar na distinção de diferentes significados para a mesma palavra. De fato, depois de muitos anos de frustração, teve início a percepção da utilidade da Lingüística Computacional para a recuperação, mas ainda mais próxima de uma recuperação estatística que a abordagem da Inteligência Artificial. Houve trabalho sobre recuperação bilíngüe e sobre transmissão de voz.

Poucas destas técnicas foram usadas em sistemas comerciais, apresentando os mesmos tipos de algoritmos de recuperação usados até hoje. Nem sempre os sistemas comerciais estavam interessados na performance da recuperação: preocupavam-se primeiramente com o número de bases de dados que continham, secundariamente sobre sua interface de usuário e, eventualmente, sobre a performance de recuperação. Parte do problema foi que todo o paradigma da avaliação da recuperação foi baseado em testes de coleções pequenas e quase nunca na avaliação de sistemas de grande escala. Em adição, o trabalho de avaliação tinha mostrado que houve enorme dispersão sobre questões de performance de recuperação, a ponto de não ser possível oferecer alguma confiança. Quaisquer que sejam as razões, a comunidade de pesquisa sentiu-se frustrada por suas técnicas não estarem sendo empregadas na indústria que eles tinham iniciado.

A comunidade de Inteligência Artificial continuou desenvolvendo os sistemas especialistas e as linguagens de representação do conhecimento. Um dos mais ambiciosos planos trabalhou com a tentativa de codificação de todo o conhecimento do senso comum em uma estrutura formal, a partir da qual um computador “raciocina” em qualquer área de assunto. Isto foi parte de um esforço para responder aos computadores de quinta geração que estavam sendo estudados pelos japoneses, máquinas que poderiam especializar-se em processamento simbólico e lógico. Na primeira parte da década de 1980, houve grande entusiasmo por sistemas especialistas e projetos baseados em conhecimento. No final da década, contudo, a tendência geral, não suportando a pesquisa básica mais o fracasso dos sistemas especialistas em suas promessas, era de um movimento para longe desta área. De fato, mesmo não acreditando que qualquer conhecimento expresso em linguagem possa ser traduzido por um simples formalismo, voltou-se a respeitar a idéia de que a linguagem natural usada especificamente para expressar informação seja uma espécie de informação que pode ser descrita e seja também uma parte desta informação.

Uma mudança tecnológica foi a difusão do uso do CD-ROM. No final da década, a maior parte das bibliotecas estava usando CD-ROMs regularmente para distribuir informação. Os CD-ROMs atacaram o paradigma de publicação tradicional pois eram facilmente enviados pelo correio mas difíceis de serem copiados pelo usuário, ou seja, poderiam ser economicamente distribuídos e facilmente usados. Seu tamanho na época (650 Mbytes, com um baixo custo de produção) significou que grandes bases de dados, mesmo as apropriadas para arquivos de texto integral, poderiam ser distribuídas. Redes de computadores também continuaram a se desenvolver nesta década, mas o CD-ROM serviu tão adequada e economicamente para a publicação de informação tradicional que poderia transformar-se em uma real ameaça aos sistemas *online* que estavam crescendo rapidamente nas duas últimas décadas.

O balanço é de uma década em que a informação *online* tornou-se comum, mesmo se ainda não usada pela maioria das pessoas e a maior parte do tempo e, na qual, os seguidores de Weaver, usuários de estatística, basicamente derrotaram os seguidores de Bush, que aconselhavam análise de conteúdo detalhada.

Nos **anos 90**, com o surgimento da Internet, dá-se a **crise de meia-idade**, uma vez que a Recuperação da Informação havia surgido há 45 anos. As coisas pareciam estar progredindo bem: muitos textos estavam disponíveis *online*, eram recuperados por algoritmos de pesquisa em texto integral e usuários finais estavam usando OPACs.

Contudo, em alguns casos, o assunto não estava esgotado. Tratava-se ainda de uma área primariamente de interesse para especialistas em bibliotecas pois os esforços para vender serviços *online* (BRS e Dialog) para pessoas comuns não tiveram grande sucesso. E houve pouco uso de algoritmos avançados de recuperação em sistemas comerciais. Processadores de texto estavam em toda parte mas a maior parte dos materiais digitados neles ainda era impressa e os textos legíveis por máquina não mais usados.

Esta década, contudo, produziu outra revolução tecnológica com o aumento exponencial do uso da Internet. O que é notável não é que todos estejam acessando informação, mas que todos estejam fornecendo informação. Por décadas, a distribuição de informação tinha sido atividade de uns poucos, como os grandes editores que enviavam informação para muitos consumidores. Agora os consumidores estão gerando e classificando sua própria informação, assim como, pessoas produzem suas próprias *home pages* e *linkam* com toda espécie de recursos que queiram.

Comparando isto com a previsão de Bush, vê-se que seu modelo de armazenamento de informação está se tornando verdadeiro. Cada usuário está organizando informação de interesse pessoal e trocando isto com outros. O que é também notável é que isto acontece inteiramente em bases livres, sem nenhum suporte da indústria. O fato de que muito desta informação seja de baixa qualidade, não a faz menos atrativa. Em adição aos *hiperlinks* individuais, há também algumas classificações como o Yahoo que é uma organização hierárquica. Embora haja muitas páginas livres, há também muitos grupos que vendem informação na Internet fornecendo um padrão médio para publicação. A disponibilidade de figuras passou a ser um atrativo, possível graças à diminuição do preço de espaço de disco.

Outra tecnologia que decolou nos anos 90 foi o *scanner*. Editores tentavam tratar o conteúdo gráfico de suas publicações até o advento dos discos baratos e dos CD-ROMs. Muitos projetos de editores e de preservação de bibliotecas têm contado com esta tecnologia, muito mais barata que a digitação de textos. Hoje editores colocam texto integral em sistemas *online* e vendem imagens via CD-ROM. Algumas vezes, o mesmo periódico é produzido nas duas formas. Mas o mais interessante é o formato eletrônico em função de suas vantagens econômicas.

No mundo da pesquisa, a comunidade da Recuperação da Informação ficou subitamente estimulada por encontrar em uso as tecnologias com as quais havia realizado experiências por décadas. Passou a existir uma indústria ativa em *softwares* de recuperação de textos. Nos Estados Unidos, o governo federal iniciou um projeto de pesquisa sobre biblioteca digital e muitos outros projetos no país adotaram o mesmo tema.

Do lado da avaliação da recuperação, depois de trinta anos, houve um súbito salto para a avaliação de coleções reais, mas continuou uma grande dispersão na performance dos sistemas de recuperação. Assim, ainda é necessário pesquisar para tentar entender como as pessoas fazem as frases em suas questões e como obter questões que forneçam boas respostas a elas.

### **Realização (*fulfillment*) (a partir de 2000) e reforma (2010)**

Supõe-se que os dois períodos finais, a partir de 2000 (realização e reforma) serão marcados pela utilização universal da Internet nos meios acadêmicos e empresariais dos Estados Unidos. Existem quantidades enormes de informação disponível via Internet e a facilidade de utilização destes sistemas colocou de lado outras formas de acessar a informação. No entanto, o problema de saber recuperar informação relevante continua a justificar o papel daqueles que são profissionais da Recuperação de Informação.

Alguns dos problemas com que se depara esta área são as questões de direitos autorais, os abusos comerciais, os monopólios e o desenho de redes com larguras de banda assimétricas. No fundo, a questão é a de perceber as implicações e as conseqüências sociais da utilização e acesso à informação tal como hoje é feito. E neste campo o debate ainda está por começar.

A tensão entre análise de conteúdo estatística e intelectual pareceu ser movida essencialmente através dos métodos estatísticos. Atualmente, na Web, *links* realizados intelectualmente estão voltando. Por este motivo, enfatiza-se o papel do bibliotecário na organização da informação *online*, apesar de muitos não saberem que este profissional executa atividades de classificação. Se o computador passou a alfabetar, tornando esta

habilidade irrelevante, significa que a atividade do bibliotecário e de seus sucessores tornou-se mais importante que antes. Se pensarmos a informação como o mar, o trabalho do bibliotecário no futuro não seria fornecer a água, mas fazer navegar o navio.

## **ESTADO DA ARTE DA PESQUISA SOBRE *INFORMATION RETRIEVAL*: A ABORDAGEM COGNITIVA**

Jacob e Shaw (1998) apresentam exaustiva sistematização da produção da área, a partir da qual foram extraídos alguns pontos de interesse sobre a abordagem cognitiva da pesquisa em Recuperação da Informação (*Information Retrieval*).

Sendo assim, apesar de muitas pesquisas em Recuperação da Informação estarem pautadas no ponto de vista cognitivo, podem ser considerados dois paradigmas (ou quase paradigmas):

- o físico introduzido através do modelo de Recuperação da Informação na década de 50 e baseado na analogia entre sistemas de recuperação e sistemas mecânicos em engenharia aeronáutica;
- o cognitivo que é predominante em Ciência da Informação, tem como ponto central que o processamento da informação seja simbólico, mediado por um sistema de categorias ou conceitos.

A Ciência da Informação não realizou a mudança efetiva para a visão cognitiva pois, apesar da atenção para a perspectiva centrada no usuário, muitas pesquisas continuam a enfatizar a perspectiva mecânica. A pesquisa cognitiva tem adotado a perspectiva limitada do instrumental mecanicista, dado pelo processo tecnológico de geração, transmissão e recepção de sinal.

Entre as muitas pesquisas correntemente desenvolvidas sob o ponto de vista cognitivo, destaca-se a ênfase na interação do indivíduo com o sistema físico a partir de sistemas amigáveis e interfaces inteligentes. Mas há uma negligência do papel central da linguagem de representação para facilitar a comunicação efetiva entre autor/ indexador e pesquisador/ mediador. A ênfase na natureza idiossincrática e individual das estruturas do conhecimento tem tomado o indivíduo fora do contexto por retirá-lo das influências do meio social, cultural e histórico.

A recuperação efetiva depende da congruência entre a cognição do usuário (organização cognitiva imposta pelo usuário sobre o conhecimento) e a cognição do indexador (estrutura representacional imposta pelo indexador sobre os documentos). A recuperação da informação depende do desenvolvimento de uma estrutura teórica que englobe um entendimento da linguagem e do processo da comunicação e seu papel na transmissão do conhecimento, além de uma compreensão de padrões de pensamento humano, expressos de tal forma que, padrões para conhecimento registrado possam ser derivados deles. Sendo assim, a maior questão da Ciência da Informação não é técnica mas sociolinguística e agrupa-se em torno do problema central da comunicação: como o significado é gerado, representado e controlado pelos meios de comunicação em diferentes contextos sociais e, presumivelmente, sobre o tempo. A ciência cognitivista é a ciência da representação. A representação de documentos depende de representações linguísticas usadas para descrever documentos.

Linguagens de indexação e classificação tradicional não são considerados bons instrumentos de representação, mas antes as linguagens intermediárias disponíveis e convenientes. No entanto, a introdução do vocabulário controlado proporcionou uma indexação com uma estrutura representacional sistemática.

Já a categorização é o mecanismo cognitivo fundamental que simplifica a interação do indivíduo com o ambiente pois não somente facilita a armazenagem e recuperação eficiente de informação, mas também reduz demandas sobre a memória humana. Sem o processo cognitivo de categorização, o comportamento baseado na aprendizagem ou na generalização da informação adquirida seria impossível. Resultados de pesquisas empíricas focando o processo de categorização parecem ser de ajuda potencial para o desenvolvimento de linguagens de indexação e estruturas organizacionais.

A tentativa para definir normas de representação deve ser compreendida, por isso, de uma perspectiva sociocognitiva que não isole as atividades cognitivas do pesquisador nem do indexador especialista, mas leve em conta o escopo de dimensões sociais que está subjacente a todos os aspectos da recuperação do documento. Pode-se dizer que, tratar informação deve ser entendido como o processo que considera a estrutura cognitiva do indexador, do documento e do pesquisador-usuário. Representação não é um exercício independente mas é, antes, um componente integral do processo total de aquisição, organização, armazenagem e recuperação do documento.

A perspectiva sociocognitivista em Ciência da Informação apóia a representação de padrões de grupos (segmentos sociais e não individuais) para sistemas de informação (na entrada, armazenagem e saída das informações) em oposição à visão mecanicista. A concepção de bases de dados e de vocabulários controlados baseados em diagnósticos das características cognitivas de documentos e de usuários (segundo segmentos sociais e/ou institucionais) caminha na direção da perspectiva sociocognitivista em Ciência da Informação.

As bases de dados devem contemplar as necessidades por informação dos usuários, sejam estas necessidades intelectuais ou utilitárias. A questão tem relação, de fato, com aquisição, utilização e desenvolvimento de conceitos e procedimentos, ou ainda, teorias e metodologias. Trata-se do entendimento racional e objetivo de caracterização da clientela, segundo metodologias de diagnóstico, apreensão da realidade e tradução por meio de formalização para formatação de bases de dados. É o mesmo processo desenvolvido formalmente para a análise documentária.

Percebe-se que a divisão anteriormente apresentada sobre a história da Recuperação da Informação entre a linha que teve início com Weaver e aquela sinalizada por Bush estão aqui representadas pela orientação física e pela orientação cognitivista. A segunda é ressaltada como o rumo mais pertinente para a Ciência da Informação, o que é justificado pela possibilidade de representação mais adequada do conhecimento produzido e da realidade do usuário, apontando, de fato, para uma área de caráter comunicacional.

### **2.2.3 BREVE COMPARAÇÃO ENTRE OS TERMOS *LIBRARY AUTOMATION* E *INFORMATION RETRIEVAL***

Pode-se concluir que a área conhecida como Automação de Bibliotecas (*Library Automation*) inclui os processos informatizados de fato implementados e disseminados em bibliotecas. Já a Recuperação da Informação (*Information Retrieval*) envolve muitas experiências e testes, incluindo algumas aplicações em sistemas comerciais para pesquisa *online* e a construção e desenvolvimento, comercialmente ou não, de *softwares* para recuperação da informação.

Pode-se entender a Automação de Bibliotecas como segmentada entre a gestão de bibliotecas e os procedimentos documentários, tanto quanto a Recuperação da Informação definiu-se pelos métodos de recuperação estatística e pelos procedimentos intelectuais de tratamento de conteúdo. Ou seja, as duas áreas principais do conhecimento, em idioma inglês, que representam o conjunto de conceitos e metodologias que compõem a Informática Documentária, também são fragmentadas internamente. Afinal, áreas que compartilham a mesma origem, acabam por apresentar características comuns, ao mesmo tempo em que uma possui lacunas que são campos desenvolvidos por outra e vice-versa. Assim, a Automação de Bibliotecas envolve tanto aspectos de integração e eficiência de operações de fluxo de acervo, quanto o tratamento e recuperação de informações, sendo a primeira a linha privilegiada nos sistemas de automação. A Recuperação da Informação é dividida entre a abordagem dos métodos estatísticos e a do tratamento intelectual dos documentos, as quais têm travado uma disputa desde sua origem que, segundo os autores analisados, pendeu para a primeira abordagem mas a segunda vem ganhando terreno.

Quanto à relação de complementaridade entre elas, a Automação de Bibliotecas tem a função de organizar automaticamente os registros dos itens de acervo de bibliotecas, contribuindo para a preservação e o acesso a estes itens. Enquanto isso, a Recuperação da Informação objetiva possibilitar o acesso a todos os registros do conhecimento, estejam eles contidos em itens de acervo de biblioteca ou não. Enquanto as bibliotecas são serviços de informação, a Recuperação da informação é uma área de pesquisa e de aplicação que visa aprimorar o acesso a registros dos vários serviços de informação, em tese, incluindo as bibliotecas.

### **2.2.4 ANÁLISE DO MERCADO DE SISTEMAS DE AUTOMAÇÃO DE BIBLIOTECAS E DE CENTROS DE DOCUMENTAÇÃO DOS ESTADOS UNIDOS**

O *Library Journal* publica a cada início de ano uma análise do mercado de sistemas de automação de bibliotecas dos Estados Unidos referente ao ano anterior. Os artigos, nomeados *Automated System Marketplace*, publicados entre 1996 a 2001 foram aqui sistematizados.

O rendimento anual deste mercado é calculado pelas vendas, manutenção, treinamento e outros serviços: em 1990, o rendimento foi estimado em 178 milhões de

dólares, atingindo seu pico em 1994 com 523 milhões de dólares e mantendo-se abaixo desta cifra nos anos seguintes. O ano de 2000 rendeu 440 milhões de dólares estimados.

O mercado é apresentado segundo as bibliotecas universitárias, escolares, públicas ou especializadas.

Sendo assim, muitos programas são especialmente voltados para **bibliotecas escolares**, configurando-se em um dos segmentos mais competitivos do mercado de automação de bibliotecas. O Sistema K-12, por exemplo, é um sistema integrado especialmente desenvolvido para escolas que proporciona a interoperacionalidade entre vários sistemas da escola como programas de estudo, bibliotecas, administração, serviços de alimentação e contabilidade, e cujo modelo várias empresas estão adotando. Como exemplo, em 1996, cita-se um sistema com serviços de treinamento que ajuda bibliotecários escolares ou profissionais da midiateca a tornarem-se melhores professores de desenvolvimento de habilidades para a informação eletrônica. Também são citadas empresas que se esforçam em educar o mercado sobre seus produtos por meio de publicações, seminários e apresentações.

Além da área escolar, outro segmento desenvolvido no país é o de bibliotecas que atendem pessoas com necessidades especiais. A empresa Keystone Systems, Inc., permanece há anos com o seu produto KLAS especificamente desenvolvido para pessoas com deficiência visual e física.

No contexto das **bibliotecas especializadas**, comenta-se que o mercado é tão diverso que a popularidade de um sistema não necessariamente significa que seja adequado para a maioria das bibliotecas especializadas. Os sistemas para bibliotecas especializadas não são muito caracterizados nesta análise do mercado dos Estados Unidos, provavelmente por serem considerados um segmento muito heterogêneo (de fato, apenas algumas características de sistemas de informação corporativos são citadas). Dentre os sistemas para bibliotecas especializadas, estão o CAIRS e o TINLIB que, no entanto, são detalhados na edição de 1993 da obra de Rowley, apresentados como sistemas para gerenciamento de informações textuais e possuidores de recursos para gerenciamento de tesouro e de aplicativo de indexação. Na pesquisa sobre 1995, é citado que o programa SIRSI integrou no catálogo o tesouro e a lista de cabeçalhos de assunto de Medicina (MeSH) da National Library Medicine e que o programa Winnebago implementou recursos de referência cruzada. Nenhum, no entanto, inclui gerenciador de tesouro e não há uma ampla e sistemática caracterização destes programas, distintamente da área universitária, por exemplo.

Dentre as empresas de sistemas para bibliotecas especializadas ou corporativas que fazem parte da análise, algumas são aqui destacadas. A Inmagic, Inc. visa as bibliotecas especializadas e produz o Inmagic DB/TextWorks, *software* para gerenciamento de texto para Windows. Em 1995, por volta de 90% das vendas foram para bibliotecas especializadas mas a companhia relata crescimento entre bibliotecas universitárias e públicas usando o Inmagic para suas coleções especializadas. Aproximadamente 40% de suas vendas não são para os Estados Unidos. Enquanto muitas companhias de fora da indústria da automação de biblioteca têm fornecido *software* para desenvolvimento de Intranet, a Inmagic escolheu focar as necessidades únicas de bibliotecários especializados como seu nicho de mercado. Os produtos da empresa, desenhados para bibliotecas especializadas, também desempenham funções tradicionais como catálogos *online* e gerenciamento de publicações seriadas, assim como, gerenciamento de inteligência competitiva, relatórios técnicos e arquivos de fotos. O Inmagic.NET é uma solução de comércio eletrônico que permite a bibliotecários

procurar recursos na Internet e baixar dados bibliográficos enquanto trabalham no DB/TextWorks.

O SydneyPlus Library Management System da ILS (International Library Systems, Corp.) é desenhado para bibliotecas especializadas e centros de informação corporativos, mas 64% das vendas não são para os Estados Unidos. A empresa sempre focou bibliotecas corporativas, médicas e legais, mas a partir de 1998, iniciou expansão para sistemas de gerenciamento de informação de indústrias financeiras e farmacêuticas. Desenvolveu, especificamente para bibliotecas jurídicas, recursos para a catalogação e gerenciamento de documentos avulsos (chamados de folhas soltas).

O VUBIS, produzido pela Geac Computers Inc., é citado nos relatórios sobre os anos de 1995, 1998 e 1999, como somente vendido na Europa. A empresa tem outros produtos para automação de bibliotecas que são significativamente vendidos nos Estados Unidos. No entanto, Nieuwenhuysen (1991) cita o VUBIS juntamente com sistemas como BRS/Search, CAIRS, CDS-ISIS, Inmagic, Freebase, Polydoc, Pro-Cite, Questel, Status, Texto, TINLIB e outros, como pertencentes ao grupo dos *softwares* que permitem o gerenciamento de informação textual estruturada. Esta tipologia de *software* é descrita por este autor como aquela que inclui características que estão no meio do caminho entre os extremos representados pelas bases de dados textuais e as bases de dados relacionais.

A pesquisa sobre o ano de 1997 informa que os sistemas baseados em texto são mais vendidos em pequenas e médias bibliotecas. Produtos especializados para gerenciamento de bases de dados textuais ficam em linha de competição com empresas de sistemas de gerenciamento de documentos, as quais não são tradicionalmente associadas à automação de bibliotecas. Vale citar que estes sistemas baseados em textos, como o BASIS, o TINLIB e o Inmagic, são normalmente utilizados em estudos comparativos com o *software* CDS-ISIS (ou MicroISIS como é aqui mais conhecido), o que parece apontar para a diferença entre *softwares* especialmente voltados para o gerenciamento das atividades de bibliotecas e *softwares* que melhor executam o tratamento e a recuperação de informações (como é o caso do CDS-ISIS que necessita de outras aplicações adicionais para se tornar mais satisfatoriamente um *software* para bibliotecas). Quanto ao mercado dos Estados Unidos para sistemas de gerenciamento de bibliotecas, apenas os sistemas para bibliotecas especializadas incluem módulos de gerenciamento de tesouro e aplicativos de indexação. Como citado acima, muitos destes sistemas (VUBIS, Inmagic, SydneyPlus) são mais vendidos fora dos Estados Unidos, a maioria na Europa, talvez para dar conta da automação dos seus centros de documentação e bibliotecas especializadas. Centradas nos recursos para gerenciamento de acervos, as funções de tratamento e recuperação da informação são raramente apontadas nesta análise do mercado.

Quanto às **bibliotecas universitárias**, a maioria das características apresentadas centra-se neste grupo, as quais foram sintetizadas abaixo:

Na pesquisa sobre **1995**, os sistemas passaram a contemplar recursos para:

- interação com outros sistemas como OPACs de outras bibliotecas, bases de dados em CD-ROMs, bases de dados *online*, informação na Internet e outros recursos multimídia, fornecendo um só tipo de pesquisa aos usuários para fontes diversas;
- pesquisa remota a várias bibliotecas e aquisição de documentos por entrega (*delivery*) ou empréstimo entre bibliotecas *online*;

- arquitetura cliente-servidor;
- interface gráfica (GUI);
- padrões como Formato MARC e protocolo Z39.50 (que permite ao usuário acessar OPACs e bases de dados na Web usando uma interface comum, independente do *hardware* e do *software*);
- aplicações multilíngüe.

No ano de **1997**, a análise do mercado indica que sistemas automatizados baseados em servidor passaram a ser adotados em vez de sistemas em minicomputadores. As mesmas implementações citadas acima continuaram em aperfeiçoamento.

Já na pesquisa do ano de **1998**, o tema da integração continua na pauta, reforçado pelo mercado do programa K-12 que gera integração entre todos os sistemas de uma escola. Este sistema é visto como um modelo e um desafio para outros setores de biblioteca procurando integração. Quanto às novas implementações, além daquelas dos anos anteriores, podem ser citadas: aplicações para Web e Intranet, educação à distância, funções de gerenciamento de coleções digitais, recursos XML para desenvolvimento de *home pages* e para marcação de documentos e suporte de EDI para aquisições e controle de publicações seriadas.

Em **1999**, apareceram o comércio eletrônico na Internet e as aplicações para personalização que permitem aos usuários customizar *sites* Web para suas necessidades pessoais. O mais conhecido exemplo de personalização em comércio eletrônico é o Amazon que recomenda livros a usuários com base em suas compras prévias. Como usuários tornaram-se acostumados às características personalizadas no comércio eletrônico, eles esperarão as mesmas capacidades de bibliotecas. Sistemas automatizados devem dar a bibliotecas a habilidade de prover cada usuário com uma página Web customizada que recomende livros com base em seu interesse e outras informações personalizadas como mensagens semanais de *e-mails* sobre os novos livros adquiridos nas áreas de interesse do usuário. De qualquer forma, há uma década, bibliotecários estavam experimentando como acrescentar valor à OPAC, realçando registros bibliográficos com tabelas de conteúdo e outras informações para ajudar usuários a tomar decisões.

Na última pesquisa publicada, referente ao ano de **2000**, aparece a distinção entre empresas de automação de bibliotecas e empresas de serviços de conteúdo. A empresa de serviços de conteúdo Elsevier comprou a Endeavor, a qual liderou o mercado de automação de biblioteca na década de 90. Muitos não gostaram pois os serviços de conteúdo oferecidos pela Elsevier são o grande peso dos orçamentos das bibliotecas. No entanto, para bibliotecas, isto pode ser positivo porque um crescente número de serviços de biblioteca é desenhado para a provisão de conteúdo eletrônico. Colaborações entre vendedores de sistemas de automação de biblioteca e provedores de conteúdo habilitariam bibliotecas a oferecer a compatibilidade necessária entre coleções locais e materiais *online*. Atualmente, sistemas estão sendo desenhados para suprir a infraestrutura de compatibilização das novas sinergias entre bibliotecas e provedores de conteúdo, resultando no acesso ampliado a materiais.

Representantes das empresas de sistemas de automação acreditam que bibliotecários serão a razão do mundo digital pelo fornecimento de informação *online*. Para tanto, produtos já existentes no mercado, mas distintos dos difundidos sistemas

integrados, podem ajudar bibliotecas universitárias e especializadas no registro, manutenção e acesso a conteúdo digital de eventos, seminários e apresentações. Também afirmam que os bibliotecários especializados têm sido pioneiros em manipulação de recursos eletrônicos e, num futuro próximo, os recursos de biblioteca, em sua maioria, serão eletrônicos e o bibliotecário corporativo terá o papel de gerenciador de conteúdo de Intranet.

Comenta-se que o trabalho de um bibliotecário de sistemas mudou dramaticamente nesta década. Em função dos *sites* de Web na biblioteca que apontam para uma variedade de recursos eletrônicos, além das bases de dados eletrônicas locais, o sistema de automação não é mais o fator de integração na perspectiva do usuário. O papel do bibliotecário de sistemas tem mudado de bibliotecário responsável pelo sistema integrado para o bibliotecário responsável por toda a tecnologia. Avanços e novos modos de usar a tecnologia para realçar os serviços de biblioteca geralmente vêm dos bibliotecários. Muitas destas inovações são encontradas em bibliotecas universitárias maiores que possuem grandes equipes de sistemas. Esta demanda está crescendo mas ainda há poucos bibliotecários tecnicamente qualificados para tal. Cita-se que as tecnologias estão mudando como os bibliotecários trabalham, mas elas não mudam a essência do trabalho da biblioteca. O grande desafio é o gerenciamento de documentos digitais e a proximidade entre a gestão documentária do acervo local e a informação dos provedores de conteúdo.

No mercado de automação de bibliotecas dos Estados Unidos, a empresa Ex-Libris vem se destacando e vale ser apontada por estar também ganhando o mercado brasileiro. A entrada dos produtos da empresa Ex-Libris no mercado dos Estados Unidos é definida como uma trajetória triunfante, incluindo lucros, qualidade e inovação. O sistema ALEPH teve origem nos anos 80 na Hebrew University of Jerusalém e, atualmente, está instalado em mais de 200 bibliotecas em 24 países. É o maior no mercado internacional focado em bibliotecas universitárias, sendo também adequado para centros de informação corporativos, museus, grandes bibliotecas públicas, bibliotecas nacionais e agências governamentais. Suporta 17 línguas, diversos tipos de escritas e tem funcionalidade de leitura bidirecional exigida pelo idioma árabe, hebreu e outros.

O sistema ALEPH da Ex-Libris aparece pela primeira vez nesta pesquisa de mercado no ano de 1995. Desenvolvido por uma universidade para uma biblioteca de universidade, esta análise de mercado vem chamando a atenção das bibliotecas para o ALEPH, acreditando que a Ex-Libris tenha potencial para tornar-se a maior empresa do mercado de automação de bibliotecas norte-americano. Em 2000, penetrou significativamente no mercado norte-americano de bibliotecas universitárias vencendo expressivos contratos e emergindo como um competidor no segmento acadêmico. A equipe no país cresceu 200%. Tornou-se o primeiro sistema para automação de biblioteca da América do Norte a suportar o popular sistema operacional Linux. Em 2001, continuou construindo sua posição na América do Norte como um grande vendedor para bibliotecas universitárias, dobrando sua equipe nos Estados Unidos.

Por meio dos produtos SFX, MetaLib e DigiToolLibrary, tem ajudado as bibliotecas no gerenciamento de coleções digitais. O SFX, por exemplo, é um mecanismo para fazer *links* entre recursos *online* que permite aos usuários pesquisar uma base de dados e clicar na citação resultante remetendo automaticamente para outras bases. Ele funciona independente da OPAC, o que implica que não é necessário utilizar o ALEPH, seguindo uma estratégia para permitir que clientes potenciais se familiarizem com a empresa.

Outro produto deste mercado que também está presente nas bibliotecas brasileiras é o Virtua produzido pelo VTLS Inc. O Virtua é um sistema desenhado para ambiente multimídia e com tecnologia que permite aos usuários criar uma aparência customizada para seu catálogo Web. A empresa está trabalhando para dar conta de bibliotecas multilíngües para atingir o setor internacional. Tanto o ALEPH quanto o Virtua possuem o banco de dados Oracle como sistema de gerenciamento de bases de dados relacionais.

## **2.3 INFORMÁTICA DOCUMENTAL**

*Informática Documental* é uma expressão encontrada no idioma espanhol e que reforça a linha adotada para esta pesquisa.

### **“INTRODUCCIÓN A LA INFORMÁTICA DOCUMENTAL: FUNDAMENTOS TEÓRICOS, PRÁCTICOS Y JURÍDICOS”, Carlos Costa Carballo, 1995**

O livro de Costa Carballo foi reeditado em 1995. A maioria dos tópicos trata do desenvolvimento tecnológico em *hardware* e *software* aplicados ao tratamento e recuperação da informação, abordando aspectos tecnológicos básicos e os sistemas utilizados, por exemplo, para automação de bibliotecas e para Inteligência Artificial, finalizando com as questões envolvidas com segurança de dados, vírus e legislação.

Inicia com uma reflexão sobre **o ensino das tecnologias documentárias** centrando-se, de fato, no ato de ensino em geral. Um dos aspectos apontados sobre o ato de ensinar é o da interdisciplinaridade, sugerindo o modelo do ensino integral ao considerar a integração como a coordenação de atividades diferentes para conseguir o funcionamento harmônico do todo, resultando em maior eficácia. Assim, na área em questão, a integração levaria à apresentação sintética de um tema do currículo com uma exposição centrada nos problemas de gestão da documentação que permitiriam dar uma ordem lógica à exposição e resolução dos mesmos, eliminando os detalhes acessórios. Isto se daria com a formulação do problema documentário, a partir do qual todos os professores que de alguma forma tenham a ver com a questão (ou seja, professores da área de Documentação, Informática, Biblioteconomia, Catalogação, Classificação, Arquivologia), fariam uma exposição.

Ressalte-se o fato da preocupação com a questão da formação e, além disso, com ênfase para a integração de conteúdos. Este enfoque interessa uma vez que a motivação para esta pesquisa partiu da necessidade da sistematização da área Informática Documentária, objetivando servir como uma abordagem integrada para a formação básica e continuada em Biblioteconomia e Documentação. De fato, o autor desenvolve todo o livro com base nos métodos de ensino e nos conteúdos dos cursos da Escuela Universitaria de Biblioteconomía y Documentación, da Universidad Complutense de Madrid.

## Módulo 1 - Conceitos fundamentais de Informática

Os conceitos fundamentais de Informática incluem: *hardware*, redes de computadores, estrutura da informação (representação da informação e automação de bibliotecas), *softwares* (tipologias e linguagens de programação) e Inteligência Artificial (incluindo os sistemas especialistas).

Inicia tratando do conceito, elementos, estrutura e funcionamento do computador. Define o computador como um dispositivo que tem como função o processo, o tratamento e o armazenamento de informação, para o que, possui dois tipos de componentes essenciais que trabalham juntos: *hardware* e *software*.

Desenvolve os mesmos itens sobre **hardware** que Saffady e Rowley. No entanto, a forma de fazê-lo diferencia-se destes autores na conceituação pouco sistemática dos componentes. Por outro lado, algumas explicitações aparentemente óbvias podem contribuir para transmitir ao leitor o entendimento de tarefas automatizadas pela prática. Por exemplo, com relação ao teclado do computador, apresenta a tecla ALT que, acionada juntamente com teclas de números, realiza a ampliação da quantidade de caracteres disponíveis. De fato, a partir das teclas do teclado podem ser representados apenas 95 caracteres entre letras maiúsculas e minúsculas, números e caracteres de pontuação e outros símbolos especiais. No entanto, um computador pode representar 256 caracteres, graças à codificação decimal em que cada caractere é representado por um dígito, implementada pelo código ASCII e acionada pela combinação de teclas. A tabela ASCII é utilizada para acentuação de palavras, para representação de caracteres de alfabetos não romanos e, por estes motivos, para a conversão de formatos para migração de dados.

Quanto às **redes de comunicação entre computadores** são definidos três tipos de teleprocessamento: por lotes (*batch*), em tempo real e por tempo compartilhado. Os dois primeiros já foram descritos por Saffady. O teleprocessamento por tempo compartilhado (*time-sharing*) ocorre quando diversos usuários trabalham simultaneamente em computadores, mesmo não havendo o mesmo tempo de resposta para todos os usuários. Para que cada computador possa se comunicar diretamente com outro é necessário uma rede de comunicação que usa cabos como suporte físico para conexão, assim desenvolvidos durante o tempo: fios metálicos de cobre, cabo coaxial e fibra ótica. Já o modem é o dispositivo que transforma a representação digital da informação que sai do terminal do computador (impulsos) na representação analógica do som (ondas) que chega ao telefone e vice-versa. Recebe a denominação de modem em função desta sua capacidade moduladora (transforma a forma física da informação de impulsos para ondas moduladas) e demoduladora (extrai da onda modulada a informação digital nela contida). Também são descritos os tipos de redes de computadores, já contemplados por Rowley.

O tópico sobre **estrutura da informação** inclui a representação da informação, segundo os sistemas de numeração e de codificação, os arquivos automatizados, uma introdução à Teoria de Sistemas e a automação de bibliotecas.

Quanto à representação da informação, são tratados os sistemas de numeração (binário, octal e hexadecimal) e os sistemas de codificação (Hollerith, BCD, ASCII e EBCDIC). Os **sistemas de numeração** são definidos como conjuntos de símbolos (dígitos numéricos, sinais de pontuação e sinais de operação) relacionados por certas regras que permitem representar quantidades. Pode-se distinguir dois tipos de sistemas de numeração: posicionais e não posicionais. Nos sistemas posicionais, cada dígito tem

um valor diferente em função do lugar que ocupa no algarismo, por exemplo, o algarismo 47. Nos sistemas não posicionais, cada dígito não tem um valor diferente segundo o lugar que ocupa no algarismo como no sistema romano, por exemplo, o algarismo CCIII (que não vale, no entanto, para algarismos como IX e XI, onde a posição dos dígitos alterou o valor do algarismo). Nos sistemas de numeração posicionais, a base indica quantos algarismos podem ser utilizados, ou seja: o sistema **decimal** (base 10) permite trabalhar com 10 dígitos (do 0 ao 9), o **octal** (base 8) com 8 dígitos (0 a 7), o **binário** (base 2) com 2 dígitos (0 e 1) e o **hexadecimal** (base 16) com 16 dígitos (do dígito 0 ao 15 e da letra A até a F). O sistema octal tem a base 8 que é uma potência exata do sistema binário ( $2^3 = 8$ ) e a cada número inteiro octal correspondem 3 dígitos binários, o que torna simples as conversões entre eles. O sistema hexadecimal também é uma potência exata do sistema binário ( $2^4 = 16$ ) que pode ser convertido para o binário ou vice-versa como o octal, só que, neste caso, separa-se o binário em grupos de 4 dígitos ou a cada número hexadecimal faz-se a correspondência com 4 dígitos binários.

Já os **sistemas de codificação** são uma operação prévia do tratamento da informação por meios eletrônicos, ou seja, transformam dados em uma forma extensa para outra mais condensada que o computador deve entender para poder operar. Há os sistemas de codificação numéricos e alfanuméricos. Dentre os sistemas de codificação numéricos, o mais importante é o **BCD** (Binary Coded Decimal). Sistemas de codificação alfanuméricos são o **Hollerith**, o **ASCII** (American Standard Code for Information Interchange) e o **EBCDIC** (Extended Binary Coded Decimal Interchange Code). O código Hollerith era utilizado nos cartões perfurados como sistema de armazenamento da informação. O código EBCDIC é uma ampliação do BCD que só servia para representar números e, agora, utiliza letras maiúsculas e minúsculas e alguns símbolos. O código ASCII, como citado acima, foi criado para intercambiar informação entre computadores. O código BCD serve para realizar as transformações do binário para o decimal pois, para trabalhar com informação alfabética é melhor o sistema octal ou o hexadecimal, mas para trabalhar com números é mais adequado que os resultados sejam fornecidos em sistema decimal que é o que o ser humano compreende.

O tópico sobre os **arquivos de dados automatizados** apresenta conteúdo semelhante ao que Saffady tratou sob o nome de processamento de dados. Os arquivos automatizados são definidos como o conjunto de dados referentes a um tema comum e que são reunidos em uma mesma estrutura com um formato previamente definido. Assim são citados e definidos seus componentes como:

- caractere: menor unidade, que pode ser um número, uma letra, um sinal de pontuação ou de acentuação, um símbolo etc.;
- dado: unidade de informação indivisível imediatamente superior ao caractere, ou seja, nenhum fragmento dos elementos que constituem o dado pode ter significado em separado;
- campo: coleção de dados que pertencem à mesma categoria;
- registro: reunião de vários campos que constituem uma unidade de informação.

É usado como exemplo, um sistema de gestão bibliográfica de um centro de documentação, onde os campos podem ser o de nome do autor ou ano de edição e o registro a catalogação de uma obra. Há também arquivos distintos que se relacionam entre si dentro de um arquivo de dados.

Segundo sua função, os arquivos podem ser:

- arquivos permanentes ou mestres: que permanecem mais ou menos inalteráveis com o tempo, como o cadastro de usuários ou os dados estatísticos resultantes do processamento de dados de um centro de documentação;
- arquivos de situação: cuja informação varia de forma contínua como os arquivos de empréstimos, devolução e reservas.

Segundo sua utilização, os arquivos são classificados em arquivos de entrada de dados, arquivos de saída de dados e arquivos de atualização, que é uma mescla dos anteriores.

As operações relacionadas a um arquivo de dados incluem: criação (definição de nome e estrutura), manutenção (inclusão, eliminação e alteração de dados), consulta (busca de dados que se necessita), classificação (ordenação de chaves ou códigos para facilitar a consulta) e fusão ou divisão de arquivos. Apenas a operação de criação e a de entrada de dados seguem a ordem em que foram apresentadas.

Como forma de organização de um arquivo são elencados:

- as listas: referem-se a uma série de elementos em que cada um está relacionado com o anterior e/ou com o seguinte, podendo ser lineares (um elemento é precedido de outro e seguido de um outro, salvo o primeiro e o último), circulares (um elemento final se une a um elemento inicial formando um círculo), ou matrizes (listas lineares nas quais um elemento tem algum tipo de chave de localização);
- as árvores: são estruturas não lineares em que cada elemento está relacionado com outros dois ou mais, guardando entre eles relação de ramificação e seguindo uma forma hierárquica;
- os anéis: são estruturas em árvores, com a particularidade de que alguns elementos formam listas circulares existindo elementos relacionados entre os componentes de algumas das possíveis listas.

Os modos como os arquivos de dados podem ser construídos seguem abaixo, incluindo os arquivos invertidos já tratados por Rowley:

- direto: também chamados aleatórios, os dados ocupam uma posição no disco segundo um código que é acionado quando uma informação é solicitada. É um método rápido de busca mas ocupa muita memória. Só pode ser utilizado para suportes dirigidos (ou endereçados);
- seqüencial: os dados vão sendo gravados um em continuação ao outro na ordem em que são inseridos no sistema. O acesso aos dados é lento porque, para tal, o sistema lê todos os dados do arquivo. Serve tanto para suportes dirigidos (ou endereçados) como não. Há três variações possíveis neste modo: seqüencial indexada, que consiste em associar um índice e assim convertê-lo em arquivo de acesso direto; seqüencial-encadeada, que consiste em incluir um código no início ou no fim evitando a duplicação do arquivo na introdução de elementos novos; e seqüencial-indexada-encadeada que, por utilizar índices e códigos, é uma mescla dos anteriores e de acesso mais rápido;
- invertido: armazena todos os códigos em cada um dos campos do registro para indicar a localização de cada informação que é solicitada; assim, não é necessário

percorrer o arquivo todo a cada busca nem classificar todo o arquivo segundo o campo (ou chave) que se deseja buscar.

Em seguida, a **Teoria de Sistemas** é definida indicando que um sistema é uma série de elementos que se relacionam entre si por características comuns e que trabalham juntos por um objetivo. São compostos por: elementos (pessoas, serviços, materiais, organizações que interagem para atingir um resultado); objetivos (ou resultados finais que se quer atingir); ambiente (tudo que rodeia o sistema podendo influir nele de um modo ou outro); recursos (meios que o sistema se utiliza para atingir seus objetivos); componentes ou subsistemas (partes em que o sistema pode estar subdividido); *inputs* e *outputs* (fluxos de entrada e saída, respectivamente); e uma ação (gestão do sistema). Enfatiza-se o estabelecimento de metas (ou objetivos) como primeiro aspecto a ser considerado em um sistema e os elementos como essenciais para a boa gestão de um sistema. Finaliza discorrendo sobre as causas do mau funcionamento de um sistema.

Diferenciando-se das outras obras analisadas, o tópico sobre automação de bibliotecas é apresentado antes das linguagens de programação e das tipologias de *softwares*. Nas outras obras, estes itens servem de introdução ao primeiro. Além disso, dentre os sistemas de armazenamento e recuperação da informação, são tratados os sistemas de automação de bibliotecas e as bases de dados pessoais. Isto nos faz concluir que o tópico não trata de sistemas de automação de bibliotecas, como seu nome indica, mas de sistemas de armazenamento e recuperação da informação e seus tipos.

Dentre os **sistemas de automação de bibliotecas**, é descrito o SABINI, sistema criado em 1982, integrado, nacional (ou seja, produzido na Espanha), com entrada de dados segundo o Formato IBERMARC e constituído pelos seguintes subsistemas:

- catalogação: segue a norma internacional ISBD, atribui automaticamente os códigos MARC e possui um arquivo para tesouro;
- aquisição: agrega os registros dos livros adquiridos, os registros dos fornecedores etc.;
- usuários: contempla os dados dos usuários do centro e a identificação de todos os centros que formam a rede de bibliotecas, quando é o caso;
- empréstimos: relaciona os livros que podem ser retirados e os que têm depósito permanente, os pedidos de empréstimo etc.;
- recuperação: onde estão os arquivos de documentos e as diversas formas de acessá-los, incluindo também os arquivos de autoridades e o serviço de DSI;
- OPAC: interface usuário-máquina que permite que o usuário acesse a informação armazenada e encontre o que deseja, sem que tenha conhecimento de Informática.

Alguns dos outros sistemas citados são DOBIS-LIBIS, ALEPH e ABSYS/LOIS II. Este último combina uma base de dados documentários (BRS) e um gestor de bases de dados relacionais (Informix), sendo por isso considerado o mais completo.

Já as **bases de dados pessoais** são adotadas como forma de armazenamento e recuperação de informação para centros não muito grandes ou bibliotecas pessoais. O KNOSYS é um pacote de *software* que combina um tratamento de textos com uma base

de dados, possui registros e campos de tamanho variável, o que possibilita que somente a informação introduzida ocupe espaço, e foi desenvolvido no idioma espanhol (o que supostamente motivou sua inclusão neste livro). O dBASE armazena informação em arquivos compostos de linhas (ou registros) e colunas (ou campos). Os registros são de tamanhos previamente definidos, o que faz com que o espaço por eles ocupado seja sempre o mesmo, independente do espaço ocupado pelos caracteres. Já o PAPERBASE é destinado a pesquisadores e bibliotecários e tem como especialidade a referência bibliográfica pois os campos comportam as informações neste formato, além de um campo para resumo.

Como a obra trata inicialmente dos aspectos relativos ao equipamento físico (ou *hardware*), parte posteriormente para a definição e caracterização do equipamento lógico (ou **software**). Entende-se por *softwares*, os programas necessários para poder trabalhar com o *hardware*, concebidos a partir de um conjunto de instruções escritas e codificadas (algoritmos) que serão seguidas de uma forma precisa e lógica para executar as tarefas solicitadas. Um algoritmo é uma união finita de uma série de regras as quais originam uma sucessão lógica de operações dirigidas à resolução de um problema específico.

Os programas devem estar escritos em um suporte para poder acessar a informação. Este suporte é chamado de **linguagem de programação** que é o modo de expressão que permite dar uma série de instruções ao computador em um formato mais parecido ao humano. Em uma linguagem de programação, há a sintaxe e a semântica. Por sintaxe, entende-se o conjunto de construções da linguagem consideradas corretas quanto a sua forma, enquanto a semântica é o mesmo conjunto de construções consideradas corretas quanto ao significado. A sintaxe em uma linguagem de programação se dá deduzindo-se propriedades a partir de uma hipótese, ou seja, utilizando a lógica. Quanto à semântica, se o significado do programa é uma seqüência de cálculo trata-se da semântica operacional, se a construção denota um objeto matemático, trata-se de uma semântica denotacional e, quando os modelos utilizados para definir o significado do programa baseiam-se em construções algébricas, então se trata de semântica algébrica.

As primeiras linguagens criadas são os códigos ou linguagens de máquina, baseadas na combinação de um código binário. Estas linguagens só podiam rodar nas máquinas para as quais foram criadas, além de serem conhecidas por poucas pessoas. Em busca de uma linguagem que fosse mais simples para a maioria das pessoas foram criadas as linguagens simbólicas, cujo programa resultante é compreensível para os humanos mas a máquina não o entende, sendo necessário a tradução para a linguagem de máquina, surgindo assim as linguagens de montagem (*assembler*). As linguagens de montagem substituem cada ordem da linguagem de máquina por um símbolo ou uma palavra abreviada em inglês dando lugar a outro programa com o mesmo número de instruções. No entanto, percebeu-se que com este tipo de linguagem era necessária uma quantidade enorme de operações que se repetiam para fazer um programa, para o que, pensou-se em uma linguagem que englobasse todas estas operações repetitivas. Nasce assim as linguagens compiladas (ou os compiladores), a partir das quais também se traduz instrução por instrução mas uma instrução simbólica pode representar várias ordens no código de máquina. Para que não seja preciso compilar o programa novamente a cada vez que um erro de sintaxe seja corrigido, uma linguagem vai lendo linha por linha e fornecendo as sinonímias ou equivalências das instruções em linguagem simbólica para instruções de máquina, o que se faz com as linguagens interpretadas (ou interpretadores). Portanto, as linguagens de programação são de três classes: linguagem de máquina (código binário), linguagem de baixo nível (linguagens de montagem ou

*assembler*) e linguagem de alto nível, que mais se assemelham à linguagem humana (compiladores e interpretadores). Estas linguagens são consideradas, respectivamente, linguagens de primeira, segunda e terceira geração. As linguagens de quarta geração permitem o desenvolvimento rápido de aplicativos e algumas são orientadas a usuários.

Algumas linguagens de alto nível são: BASIC, PASCAL, COBOL, FORTRAN, APL e LISP. O FORTRAN (contração de *formula translation*) foi desenvolvido em 1954 e introduzido no mercado em um modelo IBM três anos depois, sendo voltado para o cálculo simples por adotar a simbologia matemática conhecida de todos. O COBOL (Common Business Oriented Language) foi desenvolvido pelo Departamento de Defesa dos EUA entre 1960 e 1964 para o gerenciamento tanto de dados numéricos como alfabéticos. O LISP (List Processor) trabalha com dados não numéricos e foi desenvolvido em 1960 pelo Massachusetts Institute of Technology (MIT) para investigar o cérebro infantil, o que decorreu na facilidade de processar listas e levou esta linguagem a ser uma das utilizadas para Inteligência Artificial e sistemas especialistas. O BASIC (Beginners' All-Purpose Symbolic Instruction Code) foi desenvolvido no Dartmouth College, New Hampshire, entre 1963 e 1965. Há mais de 200 linguagens, mas também podem ser citadas: PASCAL, RPG (Report Program Generator), FORTH, SMALLTALK, PILOT, C, MODULA-2, SIMSCRIPT, PROLOG (Programming in Logic – muito utilizado em Inteligência Artificial), GPSS (General Purpose System Simulator) e ADA (que foi o último projeto de criação de uma linguagem universal, desenvolvido pelo Departamento de Defesa dos EUA e assim denominado, em homenagem a Ada Augusta Byron, a primeira programadora da história).

Não são citadas as linguagens de programação orientadas a objetos, conhecidas como linguagens de quinta geração, as quais descrevem cálculos aritméticos, comparações lógicas e outras operações computacionais. Exemplos de linguagens de programação orientadas a objetos são C++, OBJECT PASCAL e IRIS. Java é uma linguagem orientada a objeto do início da década de 90 amplamente utilizada e que ganhou destaque por suas aplicações na Internet.

Quanto à natureza dos *softwares*, há os *softwares* de sistema ou básicos e os *softwares* de aplicação. O grupo dos *softwares* de sistema é formado por programas relacionados ao sistema do computador que permitem ou facilitam a comunicação entre o usuário e a máquina. Este grupo é composto pelos sistemas operacionais e pelos programas utilitários. Os sistemas operacionais definem como o *hardware* deve funcionar e gerenciam todas as operações que o computador deve realizar, sendo mais conhecidos e utilizados depois de terem sido amplamente provados. Alguns sistemas operacionais são: MS-DOS (Microsoft Disk Operating System), UNIX, Windows e Macintosh. Já os programas utilitários são os que permitem fazer melhor uso do computador e dos periféricos, como os programas de cópias de arquivos e de detectores de vírus de computador. O grupo dos *softwares* de aplicação é muito maior e se constitui dos programas ou pacotes que o usuário compra e que cobrem todas as aplicações informáticas, como os programas de tratamento de textos, planilhas eletrônicas, sistemas de geração de bases de dados, gráficos e outros.

Há também *softwares* desenvolvidos para algumas profissões especificamente como os de Medicina, que incluem programas para gestão de clínicas, laboratórios e consultórios médicos, para gestão da história clínica e para prevenção da medicina do trabalho. Estes programas para aplicações específicas são constituídos de grandes sistemas de gestão de bases de dados relacionais e integram imagens, configurando-se em sistemas complexos. Para a área de Biblioteconomia e Documentação, são

apontados: sistemas de gestão bibliográfica computadorizada, bases de dados em CD-ROM, gestão de bibliotecas e bases de dados.

Sobre o tema da **Inteligência Artificial**, comenta-se que, apesar de ser assunto envolto em certo ar de mistério pois é e foi muito explorado pela ficção científica, vem desde alguns anos desenvolvendo pesquisas e transformando-se em realidade tangível, como foi brevemente apontado no histórico sobre Recuperação da Informação (*Information Retrieval*). Na Medicina, por exemplo, fala-se em “máquinas que diagnosticam”; mesmo acreditando-se que nada substitui a relação humana estabelecida entre médico e paciente, as máquinas auxiliam a Medicina, área em que mais se desenvolveu esta parte da Informática conhecida como Inteligência Artificial. Inteligência Artificial é definida como o oferecimento, por parte da máquina, de um comportamento parecido com o humano (compreensão da linguagem, raciocínio, aprendizagem, resolução de problemas etc.), capaz de se acomodar ou de se ajustar a uma situação real ou fictícia e de poder escolher de acordo com uma série de particularidades para dar uma resposta rápida e o mais correta possível. Se a Inteligência Artificial é parte da Informática, os **sistemas especialistas** são parte da Inteligência Artificial e são entendidos como os sistemas informáticos que possuem os conhecimentos e comportamentos de um especialista em determinado tema, sendo por isso, passíveis de contribuir para fixar, precisar, concretizar e determinar melhor nosso pensamento.

Em qualquer época da história da humanidade, houve tentativas de criação de maquinarias dotadas de alguma forma de inteligência e que estivessem a favor de seu criador. No entanto, a Inteligência Artificial, tal como se conhece hoje, teve seu início em 1956 quando os investigadores Newell, Shaw e Simon criaram o Logist Theorist, um programa de computador que resolvia e/ou demonstrava uma série de teoremas lógicos. Depois, como já citado, em 1964 foi inventado no Darmouth College, o primeiro programa BASIC e, a partir deste, o LISP, considerado a melhor linguagem de programação para Inteligência Artificial.

No início dos anos 70, a área parecia estar desaparecendo por não terem sido atingidas as expectativas, quando programas para aplicações médicas deram novo impulso à Inteligência Artificial. Muitos modelos adotados pela Inteligência Artificial estavam baseados no método estatístico, atualmente considerado defasado por não imitar o comportamento inteligente de um profissional. Assim, com o fim de simular a linguagem humana, experiências recentes adotam linguagens baseadas na representação do conhecimento, cujos principais enfoques são:

- sistemas de reprodução: baseados na dedução, ou seja, em uma série de conclusões a partir de uma proposição prévia, utilizando para isso um número de regras e premissas do tipo: “se..., então...”;
- redes semânticas: representações gráficas de palavras ou objetos que identificam conceitos, relacionados entre si por meio de redes;
- quadros semânticos: esquemas pré-elaborados na mente representando idéias determinadas, que podem conter outros dentro de si que, por sua vez, podem relacionar-se entre si formando representações distintas, ou seja, trata-se de uma divisão dos objetos ou situações em seus componentes mais elementares.

Existem poucas experiências de programas de Inteligência Artificial para Biblioteconomia e Documentação. Assim, com uma base de dados de texto integral e um programa inteligente realizaríamos uma consulta e a máquina responderia com um texto

que seria uma síntese de todos os documentos introduzidos na base de dados que respondem às características da consulta. Para esta área de atuação, o problema está na quantidade de material que deve ser tratada. Apesar de parecer ficção científica, é aí que está a Inteligência Artificial para Biblioteconomia e Documentação. O melhor exemplo é o programa KNOESPHERE, uma enciclopédia eletrônica sobre Ciência da Informação, Documentação, recursos audiovisuais e formação em cada uma destas áreas, desenvolvida na linguagem LISP pela empresa Atari.

## Módulo 2 - Produção técnica de documentos gráficos e audiovisuais

Este módulo inclui duas partes: uma sobre as tecnologias modernas de armazenamento, distribuição e recuperação de documentos e outra sobre a microfilmagem.

A primeira parte trata das aplicações da eletrônica no acesso à informação e cita outras tecnologias que, além do computador, realizam a geração, o armazenamento, o tratamento e a recuperação da informação. Entre elas, estão as tecnologias para **reprodução de documentos**, ou seja, aquelas que possibilitam a obtenção de uma ou várias cópias iguais em todos os aspectos ao documento original, ainda que possa haver alguma diferença em função da qualidade da cópia obtida. São classificadas em dois tipos: duplicação (como no caso do *offset*) e reprodução (como a fotocópia). Outras tecnologias são o **fac-símile** (ou telecópia), que é fruto da evolução da máquina fotocopadora clássica, acrescida de características transmissoras e receptoras. Já o **teletexto e o vídeo-texto** são outras modalidades de transmissão da informação que realizam a elaboração de um documento legível por máquina, onde a citação da informação é codificada em chave e pode ser recuperada mediante um receptor de televisão com escassas variações: o teletexto transmite por ondas hertzianas de rádio de forma seqüencial e para receber os documentos solicitados deve-se esperar por um breve período de tempo e o vídeo-texto transmite por via telefônica permitindo acesso de forma interativa a uma quantidade ilimitada de informação. O **vídeo-disco e o vídeo-cassete** são dispositivos de armazenamento para informação previamente digitalizada. O vídeo-disco é um suporte analógico de forma circular gravado por agulha ou *laser* capaz de armazenar grande quantidade de informação fônica ou gráfica que pode ser codificada e recuperada. O vídeo-cassete costuma ser utilizado para gravar em fita magnética e mostrar programas televisivos, cuja técnica de grande facilidade propicia sua ampla utilização. O **vídeo digital** é uma variação do CD-ROM e é conhecido como DVI (Digital Video Interactive) ou CD-I Video (Compact Disc-Interactive Vídeo), tratando-se de um suporte de vídeo animado com qualidade de imagem televisiva que pode conter textos, gráficos, som, imagem animada e *software* para trabalhar com toda a informação a partir de um computador. Assim, os sistemas **multimídia** são uma combinação de elementos físicos e lógicos (*hardware* e *software*) que suportam um ambiente de informação multi-sensorial.

Por fim, o **CD-ROM** é um disco de plástico que armazena informação textual digitalizada. O disco óptico numérico é um tipo de CD-ROM com duas características principais: somente um feixe luminoso procedente de um *laser* pode decodificar a informação contida no disco (denominando-se por isso disco óptico) e a informação está codificada segundo o código binário de 0 e 1 que representa uma série de pequenas proeminências que formam sulcos em toda a superfície do disco que o feixe de luz analisa (sendo por isso, denominado disco numérico). Entre as vantagens do CD-ROM, estão o tamanho reduzido do suporte, a grande capacidade de armazenamento, a

facilidade de busca dos dados, a resistência à deterioração ou desgaste do disco, o baixo preço e a independência das telecomunicações para acessar as informações. Entre as desvantagens constam a necessidade de um equipamento de informática próprio para recuperar a informação e a informação estar menos atualizada que as bases de dados *online*. Conclui-se que, devido a aplicações determinadas do CD-ROM, este vai conviver de forma complementar com as bases de dados *online*: para estar em dia com as informações serão utilizadas as bases de dados *online*, enquanto que, para consultar obras de referência, os CD-ROMs são mais adequados.

O tópico sobre **microfilmagem**, discorre sobre os microfilmes e microfichas, suas formas de produção e de uso.

### **Módulo 3 - Sistemas operacionais**

São traçadas a origem e a evolução dos sistemas operacionais em geral e as características dos mais importantes. É descrito tecnicamente o MS-DOS e apresentada a forma de funcionamento do MS-DOS SHELL versão 5.0, o qual diferencia-se pela apresentação em janelas, o idioma dos menus em espanhol e a possibilidade de uso do *mouse*.

### **Módulo 4 - Edição eletrônica**

O módulo sobre edição eletrônica aborda os processadores de texto e as bases de dados. A primeira parte apresenta as características gerais dos **processadores de texto** e de alguns deles em especial como o Word Perfect.

A segunda parte trata das **bases de dados** ou sistemas de armazenamento e recuperação da informação. Justifica a existência necessária das bases de dados em função do crescente aumento da informação e da possibilidade de consulta por argumentos de busca. Considera que, em um **sistema de gestão de bases de dados convencional** pode-se consultar por meio de chaves criadas previamente ou desenhadas no momento de se realizar a busca. Ao contrário, em um **sistema de gestão de bases de dados documentários**, pode-se acessar informação a partir de qualquer termo que esteja incluído no sistema devido a sua maior flexibilidade. Atualmente, os sistemas empregam técnicas mistas combinando estas duas possibilidades. Entre os grandes sistemas estão: BASIS, BRS/ Search, MISTRAL, STAIRS (Storage and Information Retrieval System), Texto e outros.

Dentre outras bases de dados também consideradas de grande utilidade para os centros de documentação ou para necessidades individuais, são apresentados os *softwares* da família dBASE. Os primeiros membros desta família romperam com o que já se havia criado em bases de dados até o momento e se converteram em uma das melhores ferramentas do gênero. O último desta família, o dBASE IV, exige mais memória que seu antecessor, possui menus simples, inclui a linguagem de programação SQL (Structured Query Language) e novas e ampliadas funções estatísticas, além do QBE (Query By Example) ou consulta por exemplo, que consiste em fazer buscas em vários arquivos preenchendo espaços em branco de um formulário.

Já o KNOSYS foi desenvolvido no idioma espanhol e é classificado como uma base de dados documentários ou um sistema de armazenamento e recuperação de

informação, segundo seus fabricantes. Todas as bases de dados são sistemas de armazenamento e recuperação da informação, mas este programa é considerado diferente, supostamente por incluir a habilidade de tratamento e recuperação de textos. Além disso, este sistema foi apresentado acima como um sistema para bibliotecas pessoais em oposição aos sistemas para médias e grandes bibliotecas.

O sistema KNOSYS trabalha com janelas que permanecem na tela e podem ficar abertas simultaneamente. Possui grande capacidade de recuperação de informações a partir de recursos como: busca alfanumérica, busca numérica (incluindo números positivos ou negativos, inteiros ou decimais), datas e intervalos de tempo, busca truncada a direita e a esquerda, busca por adjacência de palavras, busca de termos por campos determinados do registro, busca por grupos de documentos segundo a ordem de entrada no sistema, busca por operadores booleanos (E, OU, NÃO), busca a partir de buscas anteriores, busca por sinônimos determinados por meio de glossário previamente construído, hipertexto (busca os documentos que contenham um termo determinado e selecionado em um documento visualizado) e hipotexto (realiza o mesmo que o hipertexto mas somente nos documentos selecionados na última pesquisa). O sistema trabalha com índices para busca que são gerados na construção da base de dados, cujos termos podem ser livremente determinados. A opção de importação e exportação contempla arquivos em dBASE III, ASCII e outros arquivos KNOSYS. Também possui facilidades para a geração de diversos relatórios, incluindo recursos para formatação de texto.

O livro finaliza com um apêndice sobre as questões éticas e jurídicas envolvidas na documentação automatizada, englobando aspectos como a confidencialidade e a privacidade, as tipologias dos delitos de Informática como fraude, pirataria e sabotagem, os vírus de computador, a proteção e segurança dos dados e a legislação existente, concluindo com a caracterização do caso espanhol.

### **Conteúdo complementar: organização curricular universitária**

Quanto à apresentação das idéias, muitas vezes, assuntos mais complexos que exigem explanação conceitual cuidadosa, resultam em um texto confuso que dificulta o entendimento sistemático dos temas, a propósito das reiteradas afirmações do autor sobre a necessidade de conceituação.

Pode-se dizer que o livro discorre sobre as tecnologias envolvidas no processamento de informação em geral, mas não trata dos aspectos que se localizam na interface entre a Informática e o tratamento e recuperação da informação documentária, ou seja, da Informática Documentária. São tratadas as tecnologias que fazem parte do trabalho dos profissionais na elaboração de textos ou alimentação de bases de dados, com algum detalhamento sobre a estrutura de organização desta informação computadorizada, os suportes tecnológicos para o registro e reprodução da informação e as questões implicadas na segurança dos dados. Há um salto para o campo da representação da informação apenas no capítulo sobre Inteligência Artificial.

O livro interessa aos profissionais da informação na medida em que propicia o entendimento das ferramentas tecnológicas que envolvem seu trabalho. O conhecimento

de noções básicas de estruturação de informação em computador também é essencial como base para a participação destes profissionais no desenvolvimento de sistemas junto a equipes de analistas e programadores ou na avaliação e seleção de pacotes de programas para aquisição.

O autor cita diversas vezes o conteúdo dos cursos da Escuela Universitaria de Biblioteconomía y Documentación, da Universidad Complutense de Madrid, para explicar o conteúdo do livro. Assim, foi necessário recorrer à organização curricular destes cursos (PLANES..., 2002). Com isto, constatou-se que o livro corresponde à disciplina Fundamentos das Tecnologias Documentárias. No entanto, esta disciplina acrescenta alguns tópicos com relação ao livro: história e futuro da Informática; tecnologia informática e qualidade de vida; histórico dos suportes de informação e características; cadeia documentária (produção, armazenamento e distribuição da informação); recuperação da informação e introdução às bases de dados. São as disciplinas Documentação Automatizada I e II que complementam a noção sobre Informática Documentária.

A disciplina Documentação Automatizada I inclui os temas: sistemas de informação (tipologias segundo a estrutura dos dados); bases de dados (conceitos, características e objetivos); sistemas de gestão de bases de dados (conceitos, funções e tipos: SGBD e SRI); bases de dados documentários (diferenças entre bases de dados documentários e bases de dados relacionais, sistemas de recuperação da informação, estrutura da informação em bases de dados documentários, linguagens de recuperação da informação, indexação automática); sistemas de recuperação de informação comerciais; teoria da recuperação da informação (controle do vocabulário, estrutura e apresentação do vocabulário, organização dos termos, sistemas de recuperação da informação e buscas em bases de dados); indústria da informação (indústria de bases de dados, produtores e distribuidores); serviços e recursos de informação na Internet (ferramentas de busca e de classificação e indexação de documentos eletrônicos); e difusão da informação (intermediação e elaboração de produtos por demanda).

A disciplina Documentação Automatizada II, por sua vez, é composta do módulo Teledocumentação e Comunicações e o módulo Automação de Bibliotecas. O módulo Teledocumentação e Comunicações define estes dois termos e aborda as redes de computadores, o modelo cliente-servidor, os protocolos de comunicação, a rede Internet e suas formas de navegação e a constituição de páginas na Web. O módulo sobre Automação de Bibliotecas engloba: história da automação de bibliotecas; *softwares* para bibliotecas não integrados (como KNOSYS, CAIRS, CDS-ISIS e Inmagic), integrados (DOBIS/LIBIS, SABINI e ALEPH) e de comunicação; planejamento e implantação de um sistema de automação de bibliotecas; normalização da descrição bibliográfica automatizada (formatos bibliográficos); conversão retrospectiva; e situação da automação na Espanha. Ou seja, foca só as bibliotecas apesar de apresentar os *softwares* para gestão documentária que não são centrados no gerenciamento de acervo de uma biblioteca.

Importante citar que, em busca de atualização, os nomes das disciplinas acima já sofreram transformações: Documentação Automatizada I e Documentação Automatizada II foram alteradas, respectivamente, para Desenho e Gestão de Bases de Dados e Automação de Unidades de Informação (que inclui os sistemas de automação de bibliotecas e de gestão de arquivos eletrônicos). De forma pertinente, estas disciplinas de denominação genérica receberam nome representativo de seu conteúdo.

Sendo assim, a noção de Informática Documentária é amplamente contemplada por estes cursos e aponta para um desenho desta área que se assemelha à linha francesa adotada como referencial, descrita brevemente no início do capítulo. No entanto, questiona-se se o nome *Informática Documental* que compõe o título do livro é adequado, já que seu conteúdo inclui apenas as tecnologias em geral e algumas tecnologias documentárias. A Informática Documentária é, de fato, contemplada pelo conteúdo deste livro (ou da disciplina Fundamentos das Tecnologias Documentárias) mais o conteúdo das duas disciplinas que eram nomeadas como Documentação Automatizada.

Com relação às tipologias de *softwares*, o sistema KNOSYS é apresentado no livro e nas grades curriculares como: base de dados documentários, sistema de armazenamento e recuperação de informação, sistemas para bibliotecas pessoais em oposição aos sistemas para médias e grandes bibliotecas e, finalmente, como um sistema não integrado de automação de bibliotecas. Nesta última categoria, também constam os programas CAIRS, o Inmagic e o CDS-ISIS. Pode-se assim exercitar uma comparação para melhor entender a tipologia descrita como *software* documentário. A princípio, deve-se constatar a carência de sistemas do gênero no Brasil, levando a que apenas o programa CDS-ISIS possa ser comparado ao KNOSYS. Segundo a descrição deste último, ambos podem ser equiparados, em especial, na alta capacidade de recuperação e nas opções para construção e manutenção de índices. O KNOSYS, no entanto, oferece algumas vantagens como as janelas para os menus, o glossário com sinônimos, os vários formatos de importação e exportação, aplicações para relatórios e outros que, podem ser aplicados em CDS-ISIS mas não de forma direta, ou seja, sem programação adicional. Chama a atenção o fato de desconhecermos, no país, outros sistemas com estas características, apesar de amplamente citados na literatura.

Nos currículos das disciplinas do curso levantado, são dignas de nota algumas abordagens realizadas. Entre elas, a história da Informática e dos suportes tecnológicos eletrônicos, uma vez que, se a história do livro faz parte da história da Biblioteconomia, então as origens e as características dos novos e inquestionáveis suportes também devem ser incluídas. Os tópicos que tratam da indústria da informação e da cadeia documentária apontam para a dupla influência recebida: a indústria da informação foi tratada na produção em língua inglesa e a cadeia documentária é oriunda da concepção francesa. A primeira noção carrega um teor mercadológico e a segunda acentua os procedimentos técnicos lingüísticos envolvidos no ciclo documentário. Além da produção e distribuição, o item sobre a difusão da informação é relevante por citar a intermediação e a elaboração de produtos por demanda, aspectos nem sempre contemplados devidamente pelos serviços de informação, mal acostumados ao atendimento de demandas pontuais, resolvidas de forma imediata e sem maiores reflexões. Também o enfoque no tratamento e recuperação de informações a partir de ferramentas tecnológicas sofisticadas aparece nos tópicos sobre classificação e indexação de documentos eletrônicos e sobre as características informáticas do controle de vocabulário e da recuperação de informação.

Três aspectos importantes foram acima citados: os produtores e distribuidores de bases de dados; a intermediação para o uso da informação e a elaboração de produtos por demanda; e a indexação de documentos eletrônicos e aplicação de vocabulários controlados por meio dos recursos tecnológicos. Estes três aspectos apontam para a possibilidade de caracterização da área de Biblioteconomia e Documentação como um campo da Comunicação. O aspecto comunicacional fica evidente nas estratégias envolvidas na elaboração de produtos por demanda e na intermediação, atividades impossíveis de serem executadas sem que se esteja “em comum” com seu público.

Também baseadas em princípios comunicacionais, a organização de vocabulários e a indexação são processos idealmente pautados na contextualização cognitiva que caracteriza determinado grupo de indivíduos. A construção da Biblioteconomia e Documentação como campo da Comunicação contribui para a criação de parâmetros científicos para a sedimentação da Ciência da Informação. Acredita-se que, currículos que tenham organizado o conhecimento nesta direção, estejam promovendo esta difícil tarefa de cientificação.

## **“INFORMÁTICA JURÍDICA DOCUMENTAL”, Mario Saquel, 2001**

A obra de Saquel (2001) refere-se à elaboração da Informática Documentária no domínio jurídico e foi encontrada disponível em *site* da Internet. Mario Saquel foi professor de Informática Jurídica, na Faculdade de Direito, da Universidade do Chile, atualmente ministrando a disciplina Metodologia da Pesquisa, na Pontifícia Universidade Católica de Santiago de Guayaquil. A utilização de elementos documentários e tecnológicos na área jurídica, a ponto da criação de um curso em universidade desta área e elaboração de um livro a respeito, mostra a relevância e a importância dada à questão da informação em uma atividade profissional como a do Direito.

Trata-se de obra que discorre, na primeira parte, sobre o aspecto técnico-documentário, abordando conceitualmente os termos: Ciência da Informação, informação, documento, análise documentária, linguagem documentária, bases de dados e recuperação da informação. A segunda parte trata dos aspectos técnico-jurídicos que não serão amplamente abordados aqui, como: informação e Direito, informática jurídica, informação e documento jurídico, linguagem jurídica.

### **Parte I - Aspectos técnico-documentários**

#### **A Ciência da Informação**

O problema do acesso à informação conduziu à necessidade de um método ou sistema que permitisse recompilar os avanços do saber em qualquer campo, de tal forma que se pudesse recuperá-los de forma imediata e precisa. Este intrincado problema deu origem a uma ciência autônoma, a Ciência da Informação, que pretende que nenhum dado relevante se perca logo seja dado a conhecer.

A tentativa de construção de uma definição para Ciência da Informação, admitindo as ambigüidades recorrentes, leva a concluir que o objeto desta área refere-se a:

- natureza e propriedades da informação;
- aplicações concretas de seus princípios, como auxílio às outras ciências;
- suportes ou documentos que as contém.

É questionado o aporte prático designado à área, uma vez tratar-se de uma ciência. Acredita-se, no entanto, que a ciência não pode ser exclusivamente teórica. Neste caso, se o seu objeto limitar-se a analisar o conceito, as propriedades e o fluxo da

informação, as possibilidades de comprovação de seus acertos seriam muito limitados. Seria muito difícil, neste caso, orientá-la unicamente no sentido filosófico, por que a sua questão estaria restrita à busca da essência da disciplina.

Esta ciência possui uma finalidade utilitária ou auxiliar, pois é a utilidade prática que dá razão a um trabalho desta espécie: não se concebe uma base de dados ou uma biblioteca que não tenha usuários, ou que tenham usuários que não consigam respostas medianamente satisfatórias. No entanto, está claro que, a atividade utilitária nunca constituirá ciência alguma.

Ciência da Informação é, por fim, definida como uma ciência multidisciplinar que estuda a transmissão do conhecimento, sua natureza e processamento, armazenamento, recuperação e difusão. Algumas das disciplinas citadas como elementos desta ciência são: matemática, lingüística, psicologia, informática, artes gráficas, comunicação, biblioteconomia, administração.

Com relação a estas disciplinas, faz-se importante ressaltar o questionamento do autor quanto à inclusão da disciplina Informática para a constituição da Ciência da Informação. Isto porque, se é possível tratar informação à margem de um processo computacional, pergunta-se se conceitualmente existe razão para incluí-la. Define Informática como o tratamento automático da informação, tendo o termo derivação do francês *informatique*, que é composto de *information* e *automatique*. Conclui dizendo que a abordagem teórica é extravasada pela realidade dos fatos: com o acúmulo de informação em qualquer disciplina e a capacidade de tratamento da mesma pelo computador, o tratamento informacional alcança níveis mais altos de qualidade com o uso da tecnologia computacional. Sendo assim, acredita que a Ciência da Informação é o verdadeiro pontal de todo desenvolvimento científico, o que deve em grande medida aos recursos da Informática. Este ponto faz reforçar ainda mais o pragmatismo da área em função do próprio desenvolvimento científico e tecnológico.

### **O documento**

O termo “documento” é conceituado como o suporte material, de caráter probatório e que possui a finalidade de informar sobre seu meio histórico, ou seja, sobre algo externo e distinto a ele.

Os documentos são classificados como primários, secundários e terciários. O documento primário tem como fonte original de informação, o próprio documento original elaborado pelo autor. O secundário é uma criação derivada da informação original, constituindo um dos resultados do armazenamento da informação, como os índices. Os documentos terciários apresentam definição mais difusa: trata-se de um tipo de documento em elaboração e, portanto, não goza de definição unânime. Opta-se pela definição de García Gutiérrez (que é citado várias vezes neste trabalho) como um tipo de documento secundário que utiliza informação original sem, no entanto, fazer referência direta a esta informação (dicionários, por exemplo). Uma definição simples para esta classe (não citada na obra analisada) é a que abarca todos aqueles documentos elaborados a partir dos primários e secundários: bibliografias de bibliografias, estados da arte etc.

## **Análise documentária**

A análise documentária é uma operação essencial para o tratamento documentário pois, possuir documentos, não implica em ter informação e, para tanto, é necessário um procedimento idôneo. Sendo assim, a análise documentária é a operação que consiste em selecionar as idéias informativamente relevantes de um documento a fim de expressar seu conteúdo sem ambigüidades para recuperar a informação nele contida.

A análise documentária pode ser dividida em: análise formal (elementos objetivos do documento: tipo, autor, título, editora, data, paginação, idioma etc.) e análise de conteúdo (significado geral do documento). Estes são os tipos de análise que conduzem, respectivamente, à representação descritiva e à representação temática de documentos.

Sob outro ponto de vista, a análise documentária pode ser dividida em duas fases: análise externa e interna. A análise externa se ocupa da descrição exterior do documento e dos dados que o distinguem dos demais, servindo de envoltório à mensagem (assemelha-se à análise formal acima). A análise interna se ocupa da mensagem, identificando-a e informando sobre ela. Na análise interna podem ser identificadas duas subfases: descrição bibliográfica e descrição substancial. A descrição bibliográfica possui fins de identificação e localização e seus elementos básicos são: autor, editora, título. A descrição substancial possibilita um acesso ao coletivo dos documentos, devido ao tratamento temático realizado neste processo (pode ser considerado equivalente ao que se chamou de análise de conteúdo acima). De fato, o autor não torna clara a distinção entre a análise externa e a descrição bibliográfica da análise interna.

A análise substancial (ou análise de conteúdo) tem por objetivo recuperar um número pertinente de documentos de acordo com as demandas do pesquisador, e é justamente onde demonstra sua eficácia. A análise de conteúdo tem como processos a indexação e a condensação. No primeiro caso, obtém-se como produto o índice, que traz o(s) assunto(s) do documento e, no segundo, o resumo, o qual apresenta o conteúdo do documento ou o que ele diz.

## **Linguagens documentárias**

A **linguagem natural** escrita ou falada é um meio de comunicação humana, mas que apresenta deficiências por ser particularizada por cada ser humano segundo sua cultura, condição social ou nível intelectual. Sendo assim, as **linguagens documentárias** são instrumentos criados com o objetivo de traduzir a informação contida nos documentos de forma acessível a homens de distintos modos e estilos de falar ou escrever. Com a explosão da informação, a utilização desta forma de linguagem é ainda mais justificada para a garantia da comunicação.

Comparando-se linguagem natural e linguagem documentária, pode-se dizer que a primeira se estabelece e evolui por impulsos antropológicos e psicossociais, sendo essenciais as dimensões de espaço e tempo. Já a linguagem documentária surge como resultado do trabalho de documentalistas que objetivam criar um sistema de comunicação entre o pesquisador e a informação que demandam, devendo expressar-se no menor espaço possível e sem ambigüidades, utilizando somente uma parte condensada e simplificada da linguagem natural.

A **indexação** é definida como o processo de extrair termos que apontam o conteúdo de um documento, adequando-os a uma linguagem natural ou documentária

previamente escolhida pelo analista. A análise documentária é considerada mais adequada para este fim pois busca estruturar uma linguagem cujos elementos tenham a virtude, em razão de seu significado preciso, de fornecer uma base cognitiva ao usuário que lhe possibilite encontrar a informação que procura.

A indexação inclui as etapas de extração dos conceitos significativos, conversão desses conceitos para a linguagem documentária com o mínimo possível de perdas e busca de outros conceitos vinculados por relações semânticas. Esta última etapa é considerada a mais complexa pois ajudará a configurar uma rede de relações entre os conceitos, que levará à criação de um novo documento com estrutura distinta do primeiro, ou seja, um documento secundário. O desafio está no fato que esta nova estrutura deve conter o mesmo conteúdo do documento original; apesar de a completa fidelidade ao documento original ser impossível, métodos adequados dão algumas garantias razoáveis.

Os tipos de indexação podem ser apresentados segundo: o nível ou grau de profundidade (indexação exaustiva, genérica, seletiva) ou o sistema utilizado (manual ou automática). A indexação automática é definida como a identificação dos termos significativos pelo computador a partir do texto integral ou do resumo do documento. O sistema indexa todos os termos do texto que coincidem com a linguagem documentária utilizada: no entanto, para isso seria necessário que a linguagem documentária estivesse muito mais próxima da linguagem natural. Outros métodos, como os estatísticos, determinam a indexação dos termos a partir da frequência com que aparecem nos documentos: este método não considera o valor informativo de cada termo pois este é considerado independente do todo. Já os métodos sintáticos analisam as frases de acordo com regras gramaticais previamente introduzidas no sistema considerando, para a indexação, a posição relativa das palavras na frase: continua o problema do todo pois analisar a frase é distinto de analisar o texto. Sendo assim, ainda se faz necessária a operação intelectual do documentalista para manipular as numerosas e variáveis expressões da linguagem natural.

É a partir destes procedimentos intelectuais que o documentalista, por meio de uma linguagem documentária, representa os conceitos detectados no texto, atuando sobre o tratamento dos termos (entrada no sistema) e sobre a consulta ou busca da informação (saída do sistema). A linguagem documentária é, então, um conjunto de termos utilizados por uma unidade de informação declarados porta-vozes preferenciais no momento de descrever as mensagens informativas existentes nos documentos armazenados, assim como, as perguntas feitas ao sistema, a fim de obter uma efetiva recuperação da informação.

Os **tipos de linguagens documentárias** segundo temas, palavras ou conceitos são:

#### 1 Temas

Sistemas de classificação e listas de cabeçalho de assuntos. Os sistemas de classificação são baseados na sistematização de temas a partir de ordenações hierárquicas, dispondo próximos os documentos com os mesmos temas. Os sistemas de classificação mais utilizados são a CDD e a CDU, criadas, respectivamente, por Melvin Dewey, em 1876 e por Paul Otlet e Henri La Fontaine, em 1895. Nas listas de cabeçalhos de assunto, os temas são mais especificados, mas isto sobrecarrega o sistema, pois necessita de muitos cabeçalhos, além de dificultar a previsão de busca pelo usuário.

## 2 Palavras

Unitermos: unidades documentárias do sistema de indexação criado, em 1955, por Mortimer Taube, através da representação dos documentos por muitos vocábulos ou palavras existentes neles. Deu origem à possibilidade de combinação de palavras no momento da busca que, no entanto, estavam isoladas do todo perdendo sua significação e produzindo falsas recuperações.

## 3 Conceitos

Tesauros: possuem como unidade o descritor que representa o conceito essencial do conteúdo dos documentos. As vantagens do tesauro são: os descritores têm significação própria sendo possível estabelecer relações semânticas entre eles, permitem encontrar o termo para indexação de um conceito a partir da linguagem natural e oferecem menor probabilidade de falsa combinação dos termos. O tesauro é definido como um conjunto de descritores relacionados semanticamente. O primeiro tesauro foi produzido, em 1960, pela ASTIA (Armed Service Technical Information Agency), atual DDC (Defense Documentation Centre), em Washington. A elaboração de um tesauro conta com as fases de seleção e composição de termos e construção das relações entre eles. O autor discorre sobre estas fases e os critérios que elas envolvem.

As primeiras linguagens baseadas em temas tinham a função de organizar a informação carecendo de formas mais adequadas para a busca. Quando a busca passou a ser prioritária, estas linguagens tenderam a desaparecer (no entanto, questiona-se afirmação do autor pois os sistemas de classificação tradicionais não parecem estar em processo de extinção nas bibliotecas). Apareceram então os unitermos que foram um passo decisivo para o surgimento de linguagens documentárias baseadas em conceitos, especificamente os descritores, alcançando resultados mais adequados quando aplicados nos centros de documentação.

## **Bases de dados**

Iniciando a área propriamente da Informática, onde o sistema de bases de dados é entendido como aquele que, mediante a utilização de registros em computadores, pretende guardar e recuperar informação de uma ou várias bases de dados em forma conveniente e eficiente. São considerados elementos de um sistema de base de dados: dados, *hardware*, *software* e usuários.

1 Dado: representação de uma certa entidade do mundo real em alguma forma simbólica. Um registro é o conjunto de dados de um mesmo fato, um campo representa cada tipo de dado dentro de um registro e um arquivo é o conjunto de registros que agrupa um conjunto de campos. Por sua vez, as bases de dados são o conjunto de dados inter-relacionados resultante da unificação de vários arquivos independentes e que procura eliminar, em termos razoáveis, a redundância e a inconsistência dos dados, a fim de permitir sua utilização em diversas aplicações de acordo com as necessidades de seus usuários.

2 *Hardware*: composto dos volumes de armazenamento, onde residem a base de dados e os dispositivos associados como as unidades de controle, canais etc.

3 *Software*: Sistema de Gerenciamento de Bases de Dados (SGBD) que atua como intermediário ou interface entre a base de dados física (armazenamento real dos

dados) e os usuários do sistema, tendo como objetivo final gerenciar as solicitações de acesso à base de dados formuladas pelos usuários.

4 Usuários: programador de aplicações, usuário final e administrador da base de dados. Este último é responsável pela estrutura original dos dados, pelo armazenamento e pelo método de acesso aos dados, visando a segurança, a cópia de arquivos e a recuperação de dados em consequência de danos causados ao sistema.

Segundo a abordagem do autor, o profissional da informação (documentalista ou bibliotecário) parece estar basicamente no grupo dos usuários finais. De fato, não foi contemplada a função do responsável pela integridade do conteúdo da informação, em oposição ao responsável pela integridade física dos dados, que concerne aos profissionais de Informática.

As **bases de dados documentários** são tratadas com aprofundamento, sendo definidas como aquelas que tem por objetivo armazenar informação textual contida em documentos. Este documento pode conter distintas unidades de informação, ou seja, partes independentes com valor informativo autônomo que podem ser recuperados. As características principais das bases de dados documentários referem-se ao conceito de registros e arquivos de diferentes tamanhos, distintamente dos registros fixos dos arquivos tradicionais. A implementação deste tipo de arquivo requer que o SGBD marque a divisão dos campos e registros. Essas bases são mais adequadas para fins de custo de armazenamento e facilidade de recuperação do que a definição do tamanho e da capacidade dos respectivos campos e registros de acordo com o maior texto que se pretenda armazenar, com a consequente perda de espaço e excessivo ônus para a manutenção do sistema. As bases de dados documentários podem ser organizadas a partir de: arquivos independentes e a criação de um arquivo índice correspondente com as informações daqueles arquivos que se queira recuperar; ou de um grande arquivo único delimitando-se o início e o fim de cada registro e campo com caracteres especiais, e criando arquivos índices ou tabelas de recuperação com chaves que indiquem a informação relacionada ou relacionável àquela inicialmente acessada. Esta última forma de organização apresenta semelhanças com as bases de dados relacionais por utilizar arquivos índices ou tabelas de recuperação mas difere no fato de que, em uma base de dados documentários todos os campos criados podem estar aptos para servir como chave de acesso. Desse modo, a base documentária assemelha-se à estrutura de rede, pois toda a informação está relacionada ou é relacionável e pode ser acessada de qualquer ponto.

### **Recuperação Documentária ou *Information Retrieval***

Por fim, discorre sobre o conceito e as técnicas envolvidas na Recuperação Documentária, segundo a tradução literal do termo espanhol, ou *Information Retrieval*. É conceituada como a atividade realizada por um sistema na resposta às questões colocadas pelos usuários, selecionando determinados documentos que representam temas de interesse para eles. Esta busca pode dar acesso a um ou mais documentos ou apenas às fontes dos mesmos. Dentre os sistemas de recuperação documentária, estão aqueles baseados no texto completo (*full text*) do documento armazenado, em descritores estabelecidos por documentalistas ou na utilização de ambos os sistemas simultaneamente. Entre as vantagens e desvantagens destes sistemas, estão o trabalho e a presença de diversos profissionais para indexação no segundo caso e o alto índice de

resultados falsos no primeiro, apontando o sistema misto como mais adequado por integrar ambos os métodos.

São considerados dois pontos quanto à recuperação documentária. O primeiro refere-se ao papel da técnica da indexação enquanto garantia de um efetivo e eficaz retorno de informação, por realizar a extração das noções informativas fundamentais contidas no documento. O segundo ponto está baseado na recuperação documentária como uma das premissas básicas sob as quais se assenta o desenvolvimento e o aperfeiçoamento das modernas técnicas de armazenamento da informação e, a partir das quais, estabelece-se que entre os objetivos fundamentais na organização das bases de dados se prioriza a disponibilidade imediata de dados.

Dentre os procedimentos da recuperação documentária, é tratada a aplicação da teoria de conjuntos. Estes procedimentos são definidos como as operações centradas na escolha da maneira em que se fará uma consulta determinada, de acordo com os meios disponíveis e as necessidades expressas pelo usuário, estabelecendo critérios e prioridades da busca, a fim de determinar a estratégia a usar. Assim, a indexação é o processo pelo qual se representa o conteúdo do documento através de etiquetas ou rótulos descritivos e a recuperação é o acesso a conjuntos de documentos descritos através de propriedades que são comuns aos documentos da coleção: cada documento é individualizado podendo pertencer a vários conjuntos diferentes. Para a recuperação documentária a teoria dos conjuntos ganha importância ao selecionar, a partir de um “conjunto universal” de documentos armazenados, um conjunto particular explicitado por uma consulta. Esta forma de busca obedece à lógica booleana que é a mais utilizada nos sistemas de recuperação documentária, seja por fichas impressas ou por recursos automatizados. Origina-se da aplicação da álgebra de Boole e utiliza-se dos operadores booleanos E, OU e NÃO para realizar as relações de intersecção, união e exclusão, respectivamente, que podem ser representadas graficamente pelo diagrama de Venn. As estratégias de busca podem contar com mais de um operador booleano e conjunto de dados. Também podem ser aperfeiçoadas pelo uso de ajudas sintáticas como os operadores de distância (ou proximidade) entre palavras e de truncamento das palavras.

As buscas no sistema podem ser:

- diretas (pelo próprio usuário) ou delegadas (por intermédio de especialista da informação);
- referencial (obtem-se a identificação do documento ou a fonte da informação desejada) ou informativa (concede a informação desejada no mesmo instante, sem recorrer aos documentos primários);
- pré ou pós-coordenada (segundo a coordenação entre os termos para recuperação for feita no momento da indexação ou da busca, respectivamente).

São elencadas as etapas da busca documentária: surgimento da necessidade de informação no usuário e definição da mesma; comunicação da consulta ao sistema de informação e discussão desta demanda entre usuário e documentalista do sistema; formulação da pergunta em linguagem documentária com uso de estratégias de busca; filtragem ou seleção das referências mais pertinentes pelo documentalista; comunicação ao usuário dos resultados da busca; verificação pelo usuário da validade da resposta e, se for preciso, formulação de nova demanda; apreciação pelo usuário da pertinência da resposta e eficácia do serviço; e registro pela unidade de informação da demanda e da apreciação do usuário.

Vale ressaltar que o autor aponta o retorno do usuário sobre o sistema de informação traçando um caminho para a interação entre sistema e usuário.

São citados também alguns problemas da busca documentária como: desconhecimento das necessidades de informação pelo próprio usuário, falta de entendimento adequado entre documentalista e usuário, baixo uso dos recursos do sistema de bases de dados, uso deficiente dos operadores lógicos etc.

Como tentativa de contextualização dos aspectos acima propostos, é apresentada a **teoria documentária aplicada a organizações complexas**. A motivação para este item está calcada no problema decorrente da conservação, administração, uso e recuperação de documentos em uma organização em função da geração descentralizada, abundante e heterogênea destes documentos e da diversidade de suportes físicos em que estão contidos. Este problema é ampliado quando estes documentos são emitidos por diversos autores com localização física e estrutural distintas dentro da organização e são de natureza e conteúdo diverso ou têm correspondência com um processo específico da organização. Neste sentido, são caracterizados três estados em que se podem encontrar os documentos de uma organização no que se refere ao seu ciclo de vida:

- em formação: estão em processo de criação em computadores ou máquinas que ajudam a gerar o suporte e o conteúdo e, cuja versão final, ainda não se encontra terminada ou autorizada;
- emitidos: são aqueles liberados pelo(s) autor(es) com o fim de comunicar a mensagem que contém;
- arquivados: aqueles cuja mensagem já foi comunicada e deve ser armazenado para efeito de prova e história. Em geral, a maioria dos documentos de uma organização está neste estágio.

A função principal da administração documentária associada a estes documentos consiste em sua distribuição, arquivamento e recuperação. Vale notar que a distribuição foi citada pelo autor antes das outras atividades, forma talvez mais adequada à natureza de funcionamento de uma organização complexa, mas com certeza, uma visão bastante voltada para o usuário da informação. Para o controle eletrônico dos documentos emitidos e arquivados se utilizam as ferramentas denominadas bases de dados documentários, as quais armazenam todo tipo de documentos eletrônicos. O controle dos documentos em formação corresponde a sistemas de administração da gestão documentária também conhecidos como controle de fluxo, controle de atividade e *workflow*.

Os requisitos para administrar os documentos de uma organização ultrapassam as soluções que uma biblioteca tradicional ou convencional pode oferecer. Estes requisitos passam a contar com a integração de tecnologias que potencializam a administração documentária moderna. As necessidades de algumas organizações complexas quanto à sua documentação enfatizam a obtenção de itens documentários que tenham relação com o estudo ou desenvolvimento de um projeto específico. Por outro lado, a documentação da informação gerada de um projeto específico pode servir como referência para um projeto diferente.

Isto indica que o mais apropriado para este tipo de instituição é a criação de um centro de documentação, definido como a “memória documental” de uma organização e com a missão principal de registrar, conservar e administrar a documentação gerada em

cada projeto. Este centro de documentação deveria ser administrado por um profissional com amplos conhecimentos em Documentação e Informática. Estabelecer as relações entre os documentos físicos e os documentos digitais deveria ser uma das atividades principais da administração de um centro de documentação.

Em resumo, todo documento definido previamente pela instituição como relevante e já emitido formalmente (ou seja, aqueles finalizados e autorizados), deverá ser armazenado no centro de documentação, cujo processo implica o arquivamento físico e o armazenamento ou criação dos arquivos digitais correspondentes, devendo existir entre ambos subsistemas uma perfeita relação que garanta a exaustividade e uma fiel correspondência. Por outro lado, o centro de documentação deverá administrar um sistema gestor de bases de dados documentários que permita a seus usuários interrogar as bases de dados e obter como resultado o documento digital, seja texto, gráfico ou objeto.

O sistema deverá basear-se em uma clara definição prévia do documento administrado, o que exigirá um procedimento acabado que fixe as normas de emissão destes documentos. Isto significará a criação de um sistema ou guia de normalização para a emissão de documentos quanto a sua forma e elementos essenciais de conteúdo (formato, elementos ou menções essenciais, uso de referências ou abreviaturas etc.). Fazendo uso da teoria documentária e da Informática, o primeiro passo para estruturar um sistema de administração documentária orientado à recuperação consistirá em definir e hierarquizar as unidades documentárias que se vai administrar.

Interessa comentar que o autor não utiliza o termo Recuperação da Informação, mesmo tendo acrescentado o referencial inglês *Information Retrieval*, mas adota o termo Recuperação Documentária ou, no original, *Recuperación Documental*, supostamente reforçando suas origens européias calcadas nas técnicas documentárias.

## **Parte II - Aspectos técnico-jurídicos**

Na segunda parte do livro, encontra-se a justificativa para o desenvolvimento da área **Informática Jurídica** (ou Informática Jurídica Documentária):

- surgiu pelo impacto da Ciência da Informação sobre os procedimentos tradicionais de armazenamento e recuperação de informação na área do Direito;
- a complexidade crescente da organização jurídica obriga o jurista a dedicar cada vez mais tempo para sua atualização;
- grande parte de sua atualidade e perspectivas deve-se à irrupção da computação no mundo do Direito, com toda a potencialidade que oferece sua assombrosa velocidade de cálculo e capacidade de armazenamento de dados.

No entanto, a Informática Jurídica não é um ramo do Direito mas trata-se de um aspecto da Ciência da Informação focalizado em um objeto particular, o fenômeno jurídico. O objeto de estudo da Informática Jurídica é a preocupação com o Direito, sendo que suas conclusões devem auxiliar quem atua no âmbito jurídico e quem aborda este tema deve ter conhecimentos ou assessorar-se por alguém que os tenha. Sendo assim, quem trabalha nesta disciplina não está realizando um trabalho jurídico, mas um trabalho

de informação jurídica. Ao Direito interessa o estudo da norma jurídica, sua abrangência e seu sentido. O que ocorre com a Informática Jurídica, assim como com muitas outras disciplinas, é que esta analisa a norma jurídica de outra perspectiva, tomando o Direito apenas como objeto de análise.

O autor questiona-se quanto à adoção do termo Informática Jurídica, uma vez que considera a Informática apenas como um aspecto da Ciência da Informação, ou seja, o processamento automático da informação. A questão é, por que a aplicação da Ciência da Informação ao Direito é chamada de Informática Jurídica, ao invés da adoção de um termo menos restritivo? Talvez o termo mais adequado seja informação jurídica. O fato de o próprio autor adotar este nome, mesmo sem concordar completamente com ele, deve-se a que, praticamente todos os autores tratam este tema a partir do termo Informática Jurídica. Além disso, acredita que o gigantesco salto qualitativo que o processamento eletrônico da informação proporcionou não oferece um critério correto. Ou seja, podendo a Ciência da Informação fornecer soluções efetivas aos problemas informacionais, limitar o tema à recuperação automática, implica deixar em uma espécie de “limbo científico” importantes temas da informação e sua recuperação. Além do fato de que, muitos processos da informação jurídica continuam adotando métodos tradicionais e alguns dificilmente deixarão de fazê-lo. Mesmo os autores que utilizam o termo Informática Jurídica ressaltam a necessidade de não esquecer a recuperação manual da informação. Sendo assim, o autor adota este termo para não conflitar com toda a literatura sobre o tema, desvirtuando seu sentido. Mas ressalta que, não se deve perder de vista que a Informática Jurídica é uma parte da informação jurídica.

Sendo assim, a Informática Jurídica é definida como a disciplina que se ocupa do armazenamento e recuperação dos dados jurídicos pelos meios computacionais. A Informática Jurídica é dividida em duas partes: de gestão e de ajuda à tomada de decisão. A Informática Jurídica de gestão refere-se aos trabalhos administrativos que implicam no desempenho do profissional de Direito: questiona-se se, de fato, ela pode ser considerada Informática Jurídica, ou mais propriamente a Ciência da Administração aplicada ao Direito, com uso dos computadores. A Informática Jurídica de ajuda à decisão ou Informática Jurídica Documentária trata do processamento da informação jurídica (legislação, doutrina e jurisprudência) em suportes computacionais para sua posterior recuperação. Chama-se “de ajuda à tomada de decisão” pois somente com base em uma informação adequada, o profissional de Direito pode optar de forma válida entre as distintas alternativas que um problema jurídico lhe oferece.

Deve-se enfatizar a afirmação do autor no que tange à importância de um tratamento adequado à informação jurídica e, portanto, da elaboração de uma área do conhecimento e de atuação que contemple este aspecto. Ele acredita que, o Direito unido à Ciência e à tecnologia através da informação jurídica pode atingir um aperfeiçoamento da legislação e da técnica para legislar, de modo a eliminar as contradições e lacunas da legislação vigente e as revogações genéricas ou vagas, que trazem dúvida ao jurista sobre a aplicação ou vigência de uma norma determinada. Mais que isso, a importância da informação jurídica se dá em todos os níveis, do cidadão comum até o pesquisador da área jurídica, do legislador até o juiz.

## **COMPARAÇÃO ENTRE AS ABORDAGENS PARA O TERMO *INFORMÁTICA DOCUMENTAL***

Na primeira obra em espanhol, de Costa Carballo, nota-se a experiência acumulada pelo autor na área de informação em Medicina, a julgar pelas referências sobre esta área no próprio texto e na produção bibliográfica pessoal. Apesar disto, não é enfatizada a relação entre Informática e tratamento e recuperação da informação médica. Ou seja, não foi apresentada uma “*Informática Médica Documentária*”, a exemplo do modelo de *Informática Documentária Jurídica* proposto e analisado por Saquel. Isto é decorrente da opção do autor, que comenta os princípios documentários para fins de pesquisa na área da Medicina mas não os desenvolve, em favor dos experimentos em Inteligência Artificial para esta área de atuação.

Mas o essencial a ser comentado quanto a estas duas obras é o fato de que Costa Carballo, apesar de privilegiar a apresentação de conceitos para as questões tecnológicas de que trata, não proporciona um texto claro pois as diversas noções não são alinhavadas e se confundem. Quanto a Saquel, trabalha com conceitos fundamentais da Ciência da Informação, da informação para a área jurídica e da tecnologia no uso da informação, sem entrar especificamente na caracterização das ferramentas tecnológicas. Antes de tudo, desenvolve os tópicos a que se propõe com rigor e profundidade, sobrepujando, neste aspecto, todas as obras analisadas nesta pesquisa.

## **OUTRAS REFERÊNCIAS SOBRE O TERMO *INFORMÁTICA DOCUMENTAL***

Foram encontrados outros currículos de cursos universitários com a denominação *Informática Documental*. Um deles é oferecido pela Universidad Carlos III de Madrid, conjuntamente pelos Departamentos de Biblioteconomia e Documentação e o de Informática. O curso tem por objetivo iniciar o aluno na problemática das bases de dados, apresentar a teoria em que se fundamentam e desenvolver o domínio dos aspectos básicos do funcionamento de um SGBD e de um SRI, além de sua aplicação nas bibliotecas e centros de documentação. A disciplina é oferecida para vários cursos, inclusive para o de Biblioteconomia e Documentação (DOCENCIA..., 2002).

Já na Universidade de Granada, a disciplina denominada *Informática Documental* faz parte do conteúdo do curso de Ingeniería Técnica em Informática de Gestión, do Departamento de Ciencias de la Computación e Inteligencia Artificial, da Escuela Técnica Superior de Ingeniería Informática. A disciplina inclui tópicos de sistemas de informação como organização de projetos de Informática, elaboração e avaliação de produtos de Informática, fatores de viabilidade e documentação informática, além de um item sobre técnicas documentárias que abrange sistemas documentários, problemas da automação documentária, linguagens documentárias, indexação e tesouros. Já para os cursos de Biblioteconomia e Documentação, o Departamento oferece as disciplinas Administración de Recursos em Unidades de Información, Fundamentos de Informática e Bases de Datos, Sistemas Especialistas e Recuperación de Información (DEPARTAMENTO..., 2002).

Há ainda o programa de doutorado em Informática Documental e Lingüística Computacional, do Departamento de Informática e Sistema, da Universidad de Las

Palmas de Gran Canaria (INFORMÁTICA..., 2002). O programa foi desenvolvido pelo Grupo de Investigação em Estruturas de Dados que pesquisa sobre desenvolvimento de pesquisas práticas em áreas relacionadas à recuperação da informação textual, lingüística computacional e processamento da linguagem natural. O grupo justifica que o curso parte das mesmas preocupações que os bibliotecários e documentalistas possuem quanto à recuperação da informação. Argumenta, no entanto, que as técnicas de recuperação da informação não variaram em praticamente 200 anos até que a disponibilidade de computadores e de redes exigiu novas técnicas de armazenamento, acesso, interrogação e manipulação. A julgar pelas disciplinas oferecidas e seu conteúdo, configura-se como um curso sobre recuperação da informação de textos, sem o tratamento prévio da informação neles contida, a não ser pelo uso de dicionários (de sinônimos, ideológicos e tesouros) para a recuperação automática de textos, nas linguagens de marcação de documentos e nos sistemas hipermídia.

Fora do âmbito universitário, também é recorrente a presença de páginas na Internet de empresas de países de idioma espanhol (como Espanha e Argentina) que adotam o termo *Informática Documental* para descrever a prestação de serviços que envolvem o tratamento de documentação arquivística, incluindo a transformação da informação impressa em eletrônica e seu gerenciamento. Pode-se considerar estas atividades como uma parte da Informática Documentária de que estamos tratando, haja vista a proximidade entre Documentação e Arquivística e os novos desenvolvimentos em Gerenciamento Eletrônico de Documentos (GED) a serem descritos no próximo tópico sobre a abordagem francesa.

Assim, entre as empresas e os currículos universitários citados, verifica-se que o termo *Informática Documental* é adotado tanto na área da Informática quanto na área da Biblioteconomia e Documentação, incluindo o tratamento intelectual do conteúdo da informação ou não, dependendo do contexto.

Voltando à produção de literatura, García Gutiérrez, em 1987, apresentou as novas condições e metodologias da Documentação a partir dos processos da automação daquele momento. Não utilizou a forma espanhola *Informática Documental* mas a definiu, em alguns aspectos, de forma semelhante aos autores estudados, a partir da expressão “documentação automatizada” adotada no título do seu livro. Este termo, a propósito, também aparece nos antigos nomes das disciplinas da Universidad Complutense de Madrid, onde Costa Carballo leciona.

Muitas das obras aqui analisadas produzidas nos idiomas francês e inglês foram traduzidas para o espanhol. Assim, neste idioma, é possível encontrar literatura sobre o tema da Informática aplicada à Biblioteconomia e Documentação, tanto pela produção no próprio país quanto pelas diversas traduções da produção de outros países de referência em idiomas inglês e francês; ambas representam, de certa forma, a produção no próprio país.

Neste sentido, a obra francesa usada como referência, “Informatique documentaire”, de André Deweze, foi traduzida literalmente para “Informática documental”, e muitas das obras em espanhol seguem a linha francesa. Já o livro publicado em inglês por Saffady e intitulado “Introduction to automation for librarians”, também foi traduzido para “Informática documental para bibliotecas”. No entanto, por ser a primeira edição do livro de Saffady, ainda tratava, exclusivamente e de forma tópica, das operações automatizadas de bibliotecas. Por este motivo, esta tradução supõe uma redução do conceito de Informática Documentária ou uma tentativa pouco rigorosa de indicar a automação de bibliotecas como parte daquela. Acredita-se que melhor tenha

sido a tradução do livro de Reynolds, de 1985, "Library automation", literalmente para "Automatización de bibliotecas".

## **2.4 INFORMATIQUE DOCUMENTAIRE**

O termo encontrado na literatura e que parece contemplar o objetivo deste trabalho é *Informatique Documentaire*. Além de ser recorrente na literatura, o termo faz parte da realidade do profissional e do mercado de serviços documentários e de biblioteca na França.

### **"INFORMATIQUE DOCUMENTAIRE", André Deweze, 1994**

A obra considerada como a que representa de forma mais completa e atualizada esta linha francesa recebe o mesmo nome e foi escrita por Deweze, em 1994, em sua quarta edição. O livro foi publicado pela primeira vez em 1985.

O autor possui extensa experiência profissional como responsável pelo centro de documentação da empresa industrial Merlin-Gerin, na produção de duas bases de dados acessíveis por servidores nacionais e na coordenação do grupo de trabalho que concebeu o tesauro multilíngüe THÉSÉE sobre eletricidade e eletrônica. Produz literatura sobre os sistemas de informação documentária e defendeu uma tese de doutorado da Université Claude Bernard (Lyon I) sobre as redes semânticas na indexação e pesquisa documentária, além de ensinar as aplicações da Informática à Documentação em várias universidades e escolas especializadas. O engenheiro Yves Daniel, que colaborou com esta quarta edição, é responsável por importante base de dados numéricos francesa (Thermodata) e por empresa responsável pelo *software* documentário Superdoc.

A obra é apresentada apontando a evolução e o barateamento da oferta de produtos da automação de bibliotecas e centros de documentação e os conhecimentos requeridos pelos seus responsáveis na definição de suas necessidades para a escolha de *softwares* adequados. Sendo assim, objetiva facilitar o diálogo com vendedores e a avaliação de suas proposições. Todas estas tarefas exigem capacidade de análise, um certo domínio da ferramenta informática e conhecimentos fundamentais como aqueles que são úteis para a concepção de um tesauro ou de uma linguagem de representação, operações essenciais para a gestão do material documentário. Sendo o computador incapaz de qualquer interpretação, faz-se necessário um grande rigor de expressão dos elementos do tesauro. Importante notar que, diferente de edições anteriores, não é citada na apresentação a necessidade de conhecimentos para a concepção de sistemas, uma vez que se pode contar com uma variedade de produtos a preços acessíveis; o que não exclui da obra, parâmetros para uma avaliação de custos de construção de aplicações.

Segundo o autor, é um livro direcionado para informáticos, documentalistas profissionais, estudantes de Ciência da Informação e responsáveis por serviços de informação interessados em se aprofundar nos mecanismos do tratamento informático. Já Richard Bouché, autor do prefácio, acredita que o livro é capaz de atender ao engenheiro, ao documentalista, ao bibliotecário e ao estudante que queiram compreender

os sistemas documentários. Bouché estende o público deste livro aos setores de documentação e bibliotecas em suas bases de dados e catálogos, enquanto Deweze, refere-se apenas aos serviços de documentação e aos documentalistas, com um capítulo sobre a automação em bibliotecas como parte de um serviço de documentação, como se verá a seguir.

A definição adotada para o tema do livro *Informatique Documentaire* é o conjunto de aplicações da Informática à Documentação, técnica que se refere às intervenções da informática nas diversas fases de produção e utilização de documentos: produção de textos, difusão pelo editor, gestão pela biblioteca, análise e indexação para a constituição de bases de dados bibliográficos e para a difusão seletiva e *softwares* para a interrogação dessas bases de dados.

O autor afirma a intenção de apontar e analisar corretamente os problemas de Documentação em via de sua resolução informática. Sobre a questão de como atender a informáticos e a documentalistas ao mesmo tempo, explica que o leitor não informatizado poderá se iniciar a partir dos capítulos que tratam brevemente das funções e componentes do computador. Em seguida, os métodos de análise e a definição de arquivos (de interesse aos informáticos e aos documentalistas) são aprofundados. Além disso, a apresentação de aplicações especificamente documentárias pretende possibilitar que o estudante de Ciência da Informação realize a ligação entre o ensino da Documentação e daquelas técnicas informáticas.

O livro discorre primeiramente sobre a cadeia documentária, os serviços de informação nesta cadeia e o papel do computador nos serviços. Descreve as funções e componentes do computador, os suportes de informação, as redes de computadores e a telemática. No capítulo sobre a estruturação dos dados, descreve e caracteriza arquivos de dados e apresenta os SGBDs relacionais e documentários, finalizando com os sistemas de hipertexto. Em item à parte, apresenta a forma de entrada dos dados de três SGBDs documentários, e em outro, aponta as formas de produção de documentos pelo computador, que inclui os procedimentos para impressão, tratamento de textos e de imagens e outros. Trata também de arquivos e gerenciamento eletrônico de documentos. Levanta os pontos essenciais para a informatização de uma biblioteca, relacionados aos processos necessários para a gestão de acervo. A análise de conteúdo e a concepção e elaboração de um tesouro são tratadas cada uma em um capítulo, sempre sob o ponto de vista de sua concretização informática. O capítulo sobre a interrogação de bases de dados descreve os tipos de bases de dados documentários e os diversos modos de interrogação às mesmas. Propõe a criação de um guia de especificações para uma aplicação informática. Finaliza com a análise sobre o custo de uma aplicação informática e com a apresentação de critérios para escolha de soluções.

São apresentados a seguir os tópicos que acrescentam conteúdo quanto aos demais livros e artigos analisados, quais sejam, a cadeia documentária, as características e os produtos de SGBDs documentários, a análise de conteúdo dos documentos e a concepção e elaboração de tesouros e o GED.

## **A cadeia documentária**

### **1 Produção, transferência e uso da informação**

O processo de aquisição, criação e transferência do conhecimento no âmbito da economia nacional ocorre no curso de diversas fases de formação: escola primária e secundária, escola técnica e universidade. Estes ciclos de formação levam ao uso do conhecimento na administração, na pesquisa e na produção de bens e serviços.

A transferência de conhecimento se efetua com a ajuda de diversas mídias: documentos impressos (livros, manuais escolares, revistas); documentos sonoros (discos, cassetes); imagens fixas (fotografias, diapositivos, mapas); imagens animadas (cinema, televisão); transferência oral de *know-how* (eventualmente registrados em vídeo-cassetes).

A intervenção da Informática se dá tanto na produção de documentos impressos como na de documentos audiovisuais.

## 2 O lugar do computador na cadeia documentária

Alguns modos de uso do computador na cadeia documentária constam, por exemplo, na difusão realizada pelo livro produzido por fotocomposição gerando caracteres alfanuméricos e gráficos e no armazenamento de informação alfanumérica em suportes magnéticos (fitas, discos) e em vídeo-discos ou discos óticos numéricos (que suportam informação alfanumérica e gráfica e permitem acesso ao documento integral).

O computador também possibilita o acesso seletivo à informação contida nas bases de dados bibliográficos ou referotecas (aquela que é composta por registros que fazem referência aos documentos primários como livros, revistas etc.) e nas bases de dados factuais (onde se encontra diretamente a informação pesquisada, como os anuários de empresas com suas características e produtos, os repertórios de textos jurídicos, os anuários telefônicos etc.).

Há ainda o acesso aos bancos de dados estritos (séries econômicas, constantes físico-químicas, resistência de materiais, estruturas químicas etc.) por meio da consulta aos dados brutos ou à associação destes dados aos *softwares* de exploração (modelos econômicos para os estudos de mercado, cálculo de estruturas etc.).

## 3 Os fluxos, as características e os suportes de informação textual

Para conhecer, conservar e comunicar a informação, é preciso dominar seus três componentes:

- seu conteúdo semântico (a significação da mensagem);
- seu suporte (livro, filme, fita magnética etc.);
- o meio de transmissão do suporte.

Como todo corpo social, para se manter e atingir sua finalidade, a empresa é envolvida por um fluxo de informação, para o que deve ser capaz de tratar:

- a informação vinda do exterior (fluxo externo-interno);
- a informação destinada ao exterior (fluxo interno-externo);

- a informação para consumo interno (fluxo interno-interno).

E, como decorrência deste último fluxo, deve tratar a distribuição interna da informação externa e a preparação, a partir de elementos internos, da informação destinada ao exterior.

Uma grande parte do conhecimento humano é transmitido pela linguagem falada ou escrita. As informações textuais podem estar contidas em documentos ou recursos primários (revistas, livros, relatórios, anuários) e em recursos secundários, que fazem referência aos documentos primários (boletins ou bases de dados bibliográficos).

#### 4 A informação multimídia

A informação não se expressa unicamente por palavras: existem significantes gráficos extralexicais como desenhos em patentes, curvas representando funções numéricas ou traços de medidas, gráficos representando relações hierárquicas e plantas utilizadas em arquitetura e em cartografia.

Estas informações podem ser estocadas sobre os suportes tradicionais em papel (livros, revistas etc.), sobre os suportes informáticos (fitas e discos magnéticos) ou sobre microfichas e microfilmes, vídeo-discos ou discos óticos numéricos. Estes últimos podem ser acessados por computador.

As imagens podem ser analisadas por um *scanner* e transformadas em dados numéricos (ou seja, representados por dígitos). Um caso particular de utilização destas imagens digitalizadas por *scanner* é o reconhecimento ótico de caracteres para a exploração do conteúdo dos documentos.

Os documentos sonoros (discos de 45 e 33 rotações, cassetes etc.) constituem um recurso de informação. As palavras e as músicas são registradas sob a forma analógica e representadas pelas variações de frequência e amplitude. Estes registros podem ser submetidos a uma operação análoga àquela da digitalização de imagens: as variações de frequência e amplitude são codificadas sob a forma numérica (ou digital). Esta técnica, utilizada inicialmente sob os *compact-discs audio* (CDA) e agora também sob os CD-ROM, permite uma compactação significativa dos dados.

Como o reconhecimento ótico de caracteres manuscritos, estão em estudo as técnicas de reconhecimento de voz. A voz deverá permitir os comandos vocais do computador. A combinação de uma sucessão de imagens digitalizadas com uma representação numérica de sons permite o registro de filmes sonoros.

#### 5 Papel do serviço de documentação na cadeia documentária

A finalidade do serviço de documentação é trazer ao leitor a informação útil em tempo útil. Para atender este objetivo em boas condições de rentabilidade, o serviço de documentação deve se esforçar em integrar e harmonizar as três funções abaixo:

- **Gestão da biblioteca:** execução das operações de aquisição, empréstimo, circulação e difusão de documentos ou distribuição sob a forma de cópias.

- **Análise de conteúdo:** análise, elaboração de resumos e indexação de documentos. O resultado destas operações visa informar o leitor sobre as publicações recentes por meio de boletins analíticos, DSI com ajuda de perfis documentários e acesso *online* a uma base de dados bibliográficos.

- **Pesquisa retrospectiva:** esta função consiste em encontrar, com a ajuda de palavras-chave ou de descritores, os documentos úteis correspondentes a um tema de pesquisa. Para isto, podem ser utilizados fichários tradicionais; boletins de índices (KWIC, índices estruturados, índice de citações), boletins analíticos com índices (estruturados ou não), base de dados para a produção do boletim analítico ou da DSI e bases de dados disseminadas pelos servidores.

Estas três funções são interdependentes. A análise de conteúdo e a pesquisa retrospectiva são simétricas. A etapa de análise e de entrada para a realização (com os controles necessários) de um arquivo automatizado deve permitir: a produção de boletins analíticos, índices, catálogos etc.; a difusão seletiva sobre perfil e a pesquisa retrospectiva em relação mútua; e o fornecimento dos documentos primários assinalados ou detectados. Esta última operação supõe uma boa gestão de biblioteca.

## Funções e componentes do computador

### 1 Unidades funcionais do computador

As funções essenciais de um computador consistem em receber as informações ou dados de **entrada**, aplicar a estes dados um **tratamento** definido por um programa ou série de instruções e produzir resultados de **saída**.

Para realizar estas funções, o computador compreende quatro tipos de unidades funcionais: os componentes de entrada, os componentes de saída, as memórias e a unidade de tratamento (ou unidade central).

Os componentes de entrada são dispositivos que permitem a introdução de dados sob a forma de caracteres (letras, números, signos de pontuação, signos aritméticos) representados por um código. Os principais componentes de entrada são: o teclado associado a uma tela; o disquete magnético e seu leitor; as fitas e os discos magnéticos; o leitor de cartões por *chips* e por pistas magnéticas (como os cartões bancários); as telas táteis; o leitor de caracteres magnéticos; o leitor de caracteres óticos; e o leitor de códigos de barras.

Os principais componentes de saída são: as diversas impressoras (por impacto, a *laser*, a jato de tinta); os fotocompositores; as saídas sobre microfichas ou microfilmes (COM); as fitas e os discos magnéticos; e as telas de visualização.

### 2 Representação da informação

Do ideograma ao caractere alfabético e a sua representação binária:

Os primeiros caracteres utilizados pelo homem para representar seu pensamento foram os ideogramas chineses e os hieróglifos egípcios. Inicialmente, estas escritas

utilizavam o desenho para dar uma representação esquemática figurativa das noções. Mais tarde, o homem inventou a escrita alfabética na qual cada letra representa, aproximadamente, um fonema ou seu elemento de uma língua falada.

O número de unidades de informações suplementares que se utiliza para descrever e transmitir o conteúdo de uma mensagem é considerado redundante. Assim, o código binário utilizado nos computadores é um código sem redundância. Seu “alfabeto” é reduzido a uma unidade de informação (0 ou 1) ou bit (*binary digit*). Uma configuração de 8 bits ou octeto permite representar 256 combinações diferentes. Mas uma sucessão de caracteres binários não é cômodo para a leitura: a redundância é o preço que se paga para estruturar a informação e apreendê-la de modo mais global. Ela constitui também uma proteção contra as deteriorações que podem sofrer a mensagem no curso de sua transmissão.

Código de barras:

O código denominado 2 entre 5 entrelaçado é numérico. Ele substitui o 2 entre 5 que foi o primeiro código deste gênero, mas foi abandonado por ocupar muito espaço. No código 2 entre 5 padrão, cada caractere é representado por 5 barras pretas (3 estreitas e 2 largas) separadas por espaços entre elementos. O código 2 entre 5 entrelaçado imbrica os caracteres codificados nas barras com os caracteres codificados nos intervalos. É utilizado na área industrial.

O CODABAR é um código numérico (parcialmente alfanumérico) e o padrão mundial para a codificação de transfusões sangüíneas. O Código 39 é um código de barras alfanumérico e é o mais disseminado na área industrial. Os códigos EAN8 e EAN13 (European Article Numbering) utilizam combinações para codificar produtos de grande distribuição. São utilizados no ISBN.

Códigos binários:

O sistema de numeração binária é adequado para os elementos utilizados em Informática e podem representar dois estados estáveis, como a presença ou ausência de perfuração em um cartão ou presença ou ausência de um *spot* magnetizado sobre a fita ou disco.

Códigos alfanuméricos a 8 bits:

O EBCDIC é utilizado nos computadores da IBM e de outras empresas. Compreende 8 bits mais uma posição para o controle de paridade (bit de controle para conferir a este código sem redundância alguma imunidade). Todas as 256 combinações possíveis não são utilizadas e as combinações vagas autorizadas permitem representar os caracteres particulares de uma língua (é, è, à, î, ü, ç etc.).

O ASCII é um código de 7 bits realizado para os usuários de computadores e equipamentos de transmissão para padronizar as comunicações entre diversas máquinas e entre computadores. No entanto, alguns computadores trabalham com caracteres de 8 bits, como os da IBM. Assim, estendeu-se o código para uma representação de 8 bits, chamada ASCII-8. Em alguns equipamentos pode-se utilizar programas utilitários de conversão ASCII-EBCDIC e vice-versa.

DCB:

Os caracteres são representados com a ajuda de 6 posições binárias (2 posições de zona A e B e 4 posições numéricas), mais uma posição C para o controle de paridade.

Sistema hexadecimal:

O hexadecimal é um sistema de numeração de base 16 (o binário é de base 2 e o decimal de base 10) representado pelos 10 números decimais (de 0 a 9) e pelas letras A até F (para os números 10 a 15).

A informação codificada com a ajuda dos octetos EBCDIC ou ASCII é utilizada para os tratamentos internos do computador. Mas uma sucessão de oitos 0 ou 1 não é fácil de manipular para introduzir os dados em um computador ou para ler em saída. A redundância nula do octeto lhe confere uma estrutura elementar mais difícil de apreender visualmente que a estrutura gráfica mais complexa de caracteres alfanuméricos. Além disso, certos octetos não correspondem a combinações de caracteres alfanuméricos visualizáveis (o espaço, por exemplo, ou os caracteres minúsculos ausentes da impressora e substituídos pelos pontos), sendo desejável identificá-los com mais facilidade que pela representação de sucessões de 0 e 1. A utilização do sistema hexadecimal é mais cômoda pois as 256 configurações possíveis de 0 e de 1 constituem um octeto podendo ser representados com a ajuda de dois caracteres somente (os números 0 a 9 e as letras A a F).

Imagens digitalizadas:

O **scanner** (ou analista digital de imagens) é um equipamento conectado em um computador que permite a digitalização de imagens a partir de um documento em papel, fotografia (preto e branco ou colorida) negativo, diapositivo etc. O documento é transformado em um série de pontos com um valor que depende da técnica de digitalização. As imagens ponto por ponto ocupam significativo volume de memória, para o que, os equipamentos de digitalização são associados às técnicas de compressão e descompressão de informações. A qualidade de análise da imagem depende: da resolução expressa em dpi (*dots per inch* ou pontos por polegada); do contraste ou luminosidade expresso em nível de cinza; e da análise das cores por meio de 256 níveis.

A aparição de *scanners* de mão a preços competitivos favoreceu o desenvolvimento de *softwares* para **OCR**. Estes *softwares* permitem transformar um documento em papel, datilografado ou impresso, em um arquivo texto legível por um computador. Esta transformação é efetuada pela análise de uma imagem ponto a ponto por um *scanner* e pela identificação das imagens de caracteres.

O **aparelho fotomagnético** permite a digitalização direta de imagens em cores sobre disquetes. Estas imagens podem ser, em seguida, exploradas por um microcomputador dispondo de uma interface adaptada. Este aparelho tem a vantagem de ser transportável e utilizável em condições adversas, apesar da qualidade das imagens ser sofrível para uso de ilustração.

A **câmera digital** permite a digitalização de imagens fixas ou animadas associadas ou não a som. Estas seqüências podem ser exploradas por um microcomputador dispondo de equipamentos adequados.

A tecnologia de **reconhecimento de voz** permite a todo ser humano comandar pela voz um computador. O estado de técnicas atuais permite de maneira operacional apenas um número limitado de funções.

Dentre as **normas de intercâmbio de dados**, as duas mais importantes são: SGML (Standard Generalized Markup Language) ou norma ISO 8879, de 1986, de origem nos Estados Unidos, visa indicar uma linguagem de descrição dos documentos para sua edição, independente do suporte; e ODA (Office Document Architecture) ou norma ISO 8613, de 1989 é de origem europeia e permite definir simultaneamente a estrutura lógica e a estrutura física de um documento, permitindo a troca de dado textuais, gráficos e fotográficos. (A norma XML não consta aqui por ter sido criada em 1996, depois do livro ser editado).

Compressão e descompressão de dados:

As técnicas de compressão e descompressão permitem reduzir o espaço das informações estocadas.

### 3 Suportes de registro da informação

As **fitas magnéticas** permitem o registro de arquivos sob um espaço confiável. Mas é um sistema de acesso seqüencial: lê-se sucessivamente todos os registros até encontrar o desejado. Assim, pode-se considerar as fitas magnéticas como suportes para as cópias de segurança e para intercâmbio de dados por via postal (neste caso, especialmente, antes do advento da Internet).

Os **disquetes** são memórias dirigidas ao acesso seletivo (ou endereçadas) e foram produzidos com 8, 5 ¼ e 3 ½ polegadas, utilizados em microcomputador.

Os computadores são equipados de **discos “duros”** que são unidades imóveis. Um disco duro externo pode constituir a indispensável ferramenta de salvaguarda (*backup* de segurança).

Os **discos óticos** distinguem-se em duas categorias: gravados uma vez e lidos várias vezes sem limitação (CD-ROM e WORM) e regraváveis.

A **memória da unidade central** é organizada em um certo número de células compreendendo cada uma um local e os dados como as instruções do programa e as informações a tratar. Há memórias vivas ou memórias de acesso aleatório, com as quais se pode escrever, modificar e ver as informações (como a RAM - Random Access Memory) e memórias mortas (ROM - Read Only Memory) que estocam informações que serão utilizadas de forma imutável.

### 4 Acesso à informação

Em um arquivo, cada registro pode ser recuperado por uma ou mais chaves de acesso, como os nomes de autores em um arquivo bibliográfico e o código de um fornecedor em um arquivo de encomendas. Para efetuar esta pesquisa, o sistema compara a cadeia de caracteres pesquisada com as palavras-chave dos diversos registros. Os métodos de acesso dependem das memórias utilizadas: aquelas sobre suportes não dirigidos ou endereçados (fitas magnéticas) e aquelas sobre suportes dirigidos ou endereçados (disquetes, discos, memórias RAM).

No **acesso seqüencial** ocorre obrigatoriamente a leitura sucessiva da chave de todos os registros. Mesmo efetuando uma seleção prévia (alfabética por autores, por

exemplo), é necessário explorar a metade do arquivo antes de acessar o registro pesquisado.

O **acesso seletivo** consiste na pesquisa dos registros a partir de sua localização sobre o suporte, ou seja, na busca do seu endereço em um suporte. Distingue-se dois modos de acesso seletivo: a organização aleatória e a organização seqüencial indexada. As duas repousam sobre a organização de tabelas de correspondência entre a palavra de acesso a um registro lógico e seu endereço físico sobre o suporte.

Na organização aleatória ou seletiva, os registros de um arquivo não são classificados segundo uma seqüência correspondente a sua chave de acesso, mas nos locais nos quais correspondem um endereço. A tabela de correspondência entre estes endereços e a chave de acesso resulta de um cálculo mais ou menos complexo assegurado pelo sistema de exploração.

A organização seqüencial indexada vem se expandindo. No aspecto físico, os registros são dispostos seqüencialmente sobre a unidade de disco. A transformação de chaves de acesso de registros em endereços sobre os discos é inteiramente assegurada pelo sistema de exploração. O sistema fornece muitas tabelas de índice (de onde o nome deste tipo de organização) que permitem acelerar a pesquisa de registros. As tabelas de índice compreendem, em geral, muitos níveis.

As tabelas de índice compreendem dois elementos: argumento de pesquisa (chave de acesso a um registro) e valor (endereço do registro sobre o suporte). Os métodos de pesquisa mais utilizados são: varredura seqüencial simples (comparação do argumento de pesquisa com todos os argumentos da tabela na ordem de aparecimento); varredura seqüencial tendo em conta uma freqüência de aparecimento; e pesquisa dicotômica (a consulta da tabela não é seqüencial mas dirigida ou endereçada e os argumentos são classificados pelo valor crescente). Este método consiste em comparar o argumento de pesquisa com o argumento central da tabela, pois em função do resultado obtido ao pesquisar por refinamentos sucessivos, se alcança o argumento pesquisado.

## 5 Tratamento da informação

Toda operação de entrada e saída (ou escrita e leitura) é efetuada pela unidade central. A unidade central permite a um programa, representado pelas seqüências de instruções, tratar a informação de entrada em via de um resultado de saída.

Memória central:

A memória central é composta de um grande número de células idênticas encontradas por seu endereço. Cada célula pode estocar uma mesma quantidade de informação ou palavra, cujo tamanho varia segundo o computador.

Na memória do computador se encontram os dados a tratar e os dados obtidos como resultados, além do programa constituído de uma série de instruções que o computador executa sucessivamente.

A unidade aritmética e lógica:

A parte aritmética calcula, desloca os números, coloca o sinal do resultado etc. A parte lógica executa as operações de tomada de decisão que modificam eventualmente a seqüência de execução de instruções.

A unidade aritmética e lógica opera com três elementos chamados registros: o acumulador (no qual se encontra o resultado das operações efetuadas); o registro de instrução (que contém as instruções do programa durante sua execução); e o registro de endereço de instrução (chamado também contador ordinal, contém o endereço da próxima instrução a executar).

A execução de instruções:

Um programa é uma sucessão de instruções permitindo o tratamento de informação. Podem ser distinguidas cinco classes de instruções:

- Instruções aritméticas: trata-se da adição, da subtração e de outras operações que não são freqüentemente utilizadas em Informática Documentária (a não ser, por exemplo, para números de inventário, gerenciamento de orçamento e outros).
- Instruções de movimento: permitem trocar os dados de lugar na memória da unidade central. Os mais usados em Informática Documentária são para movimentar as zonas de escrita em modelo de impressão.
- Instruções lógicas: uma das mais utilizadas é a de comparação (=, <, >). Por exemplo, para incorporar uma operação de seleção sobre os nomes de autores. A operação será, neste caso, e segundo o resultado, seguido de uma instrução de conexão e de movimento: se o caractere  $x < y$ , colocá-lo a frente de  $y$ .
- Instruções de conexão: permitem saltar de uma instrução a outra. Podem ser incondicionais (de uma seqüência de instruções a uma outra) ou condicionais (a instrução depende do resultado de um teste como uma comparação, por exemplo).

Tratamento de uma cadeia de caracteres:

É um dos principais tratamentos em Informática Documentária. Chama-se de cadeia de caractere toda seqüência de caracteres (alfanuméricos, caracteres especiais e outros) precedidos e seguidos de um espaço.

No caso da inversão de pré-nomes e nomes, por exemplo, no caso de um anuário telefônico que tem os nomes em ordem alfabética, sendo que os pré-nomes estão na frente dos sobrenomes. Deseja-se colocar os sobrenomes na frente, ou seja, efetuar a inversão de pré-nomes e sobrenomes. Existem diversos algoritmos de seleção mais ou menos eficientes. Com um algoritmo simples, pode-se inverter as posições começando a leitura pelo último caractere da direita.

Outra questão refere-se aos problemas característicos do alfabeto “rico”, definido como aquele que distingue maiúsculas e minúsculas e eventualmente os caracteres acentuados, enquanto o alfabeto “pobre” não distingue maiúsculas. Segundo o exemplo em língua francesa, a alfabeto “rico” permite distinguir: *Il exerçait la profession de maçon dans la ville de Mâcon* (no alfabeto pobre, *maçon* e *Mâcon* são representados indistintamente por MACON).

## 6 Modos de operação

Distinguem-se muitos modos de operação como:

- Tratamento por lotes (*batch processing*): adequado quando o volume de dados justifica este processo ou nas horas ocias da noite;
- Tratamento em tempo real: a informação é tratada no momento de sua emissão a partir de um terminal, cujo resultado é enviado imediatamente;
- Multiprogramação: permite tratar seqüências de um ou mais programas;
- Multitratamento (*multiprocessing*): utilizando várias unidades aritméticas e lógicas com uma só memória central, permite efetuar simultaneamente várias instruções pertencentes a um ou mais programas;
- Tratamento à distância (*teleprocessing*): utiliza linhas de comunicação para trabalhar segundo um dos três modos anteriores;
- Tempo compartilhado (*time-sharing*): permite a um grande número de usuários utilizar, em tempo real, a unidade central.

Também são citados os sistemas monotarefa e mono-usuário, que vem sendo substituídos pelos sistemas multitarefas e multi-usuários.

## 7 Sistemas operacionais

O sistema operacional ou *Operating System* (OS) é um conjunto de programas, estocados sobre disco e/ou em memória ROM que completa as funções do *hardware* e garante a ligação entre os programas e os dados de usuários, de uma parte, e os recursos físicos do computador, de outra parte.

Os sistemas operacionais mais utilizados para microcomputadores são MS-DOS, Mac-OS (Macintosh Operating System), Unix e OS/2 (IBM).

Atualmente, ao sistema operacional se associa uma interface gráfica que permite o uso de ícones, de janelas e de menus, oferecendo maior amigabilidade ao usuário.

## 8 Redes de computadores

O sucesso dos microcomputadores se deu em boa parte devido à independência que forneceram a seus usuários com relação aos minicomputadores e computadores de grande porte. No entanto, esta independência foi paga, durante algum tempo, pela impossibilidade de compartilhar dados e recursos (como impressora, por exemplo) com outra máquina. A demanda por conexão e compartilhamento foi contemplada pelas redes de computadores que ofereceram uma forma de trabalhar em grupo, de compartilhar recursos e dados e até de se conectar a minicomputadores e computadores de grande porte.

As redes de computadores são geralmente compostas de um microcomputador especial chamado **servidor**, cujo papel é servir os outros, ou seja, as **estações de trabalho**, pela alimentação com recursos e dados. Diferente dos terminais, as estações de trabalho possuem recursos para a efetivação de tarefas.

Considerando as diferentes possibilidades de se ligar computadores em rede, a instalação, a administração, a gestão e a manutenção de uma rede são do domínio de especialistas (dispensando assim a atuação dos documentalistas).

## 9 Telemática

Envolve a conexão de computadores no domínio privado e público. Os serviços desta natureza existentes na França são divididos em duas categorias: as conexões especializadas e as redes de comutação. Uma das redes de comutação é o Minitel, utilizado como um terminal de computador e que foi distribuído gratuitamente pela França tendo, como primeira função, a interrogação de bases de dados, como o anuário telefônico eletrônico.

### **Estruturação dos dados**

#### 1 Propriedades de um arquivo de dados

A noção de arquivo é bem anterior à Informática, como demonstram as agendas de endereços de uma secretária onde constam nomes, endereços e números de telefones, ou as bibliografias resultantes do trabalho de pesquisadores, escritas em fichas intercaladas alfabeticamente segundo, autores, títulos, assuntos e datas de publicação.

Um arquivo automatizado é constituído de um certo número de registros e cada registro compreende um certo número de campos distintos nos quais as informações de mesmo tipo podem figurar de forma obrigatória ou facultativa. Um arquivo assim constituído ganha em rigor e em homogeneidade, além de permitir uma operação capital em Informática: a seleção de registros a partir de um ou mais critérios (nome, cidade, data de nascimento etc.).

Um arquivo do tipo agenda de endereços que está ordenado alfabeticamente, por exemplo, impossibilita a recuperação por um segundo critério, sendo necessário percorrer toda a agenda seqüencialmente. Isto pode ser um pouco mais fácil com um arquivo de fichas móveis, como no caso das bibliografias de pesquisas, mas ao custo da duplicação de fichas para cada critério de busca.

#### 2 Estrutura e representação de um arquivo

Um arquivo é um conjunto de registros, os quais compreendem vários campos. Os campos podem ser: alfanuméricos, numéricos inteiros, numéricos com decimais, data, texto, imagem e outros. Os quatro primeiros tipos podem ser indexados, ou seja, uma tabela de índice permite recuperá-los rapidamente.

Quanto aos tipos de campos, em alguns sistemas os campos de tamanho fixo são preferidos com relação aos de tamanho variável em função da simplicidade de programação, organização do disco e limitação do tempo de acesso. Exemplifica-se a partir da criação de um sistema com um arquivo com campos fixos e outro arquivo com o campo resumo de tamanho variável acessível pela mesma chave. Também foram

adotadas listas de autoridade, por meio de sistemas relacionais. Assim, a partir de arquivos de formato fixo, pode-se fornecer mais portabilidade ao sistema com outras aplicações.

Para a criação de um arquivo invertido são necessários:

- um argumento de pesquisa, como nomes de autores ou palavras-chaves;
- para cada autor ou palavra-chave, os números de acesso às citações dos documentos no arquivo de referências e, eventualmente, o número de documentos (ou ocorrências).

Sendo assim, na organização informática das relações entre o arquivo invertido e o arquivo de referência, consta um arquivo intermediário com os números que ligam o número de acesso de cada citação com o número da palavra-chave correspondente. Em alguns sistemas, cada número de acesso é seguido da indicação de lugar no registro do arquivo de referências, o que permite pesquisar as palavras em uma ordem e distância dadas (operador E acrescido de condições de ordem e de distância entre os termos).

### 3 Sistemas de gerenciamento de bases de dados (SGBD)

O SGBD permite gerar um conjunto de arquivos nos quais as informações são associadas por *links*.

Propriedades de uma base de dados relacionais:

O autor justifica sua insistência sobre as vantagens das bases de dados relacionais, em função de a experiência ter demonstrado que estas vantagens não são imediatamente percebidas pelos documentalistas, por estarem habituados à utilização de arquivos únicos e possuírem conhecimentos rudimentares de Informática.

Considerando os SGBDs relacionais mais difundidos (dBase IV, Omnis 7 4D) cita que, em outros sistemas, o pesquisador deve escrever com precisão, o nome de um autor ou editor, o título de uma revista ou o assunto de um artigo, cada vez que os utilizar. Para evitar esta dificuldade e os erros possíveis, os SGBDs relacionais podem colocar à disposição do pesquisador, arquivos como os de autores, editores, léxico (ou vocabulário), a partir dos quais o pesquisador poderá buscar os valores convenientes. Trata-se assim de uma base de dados com vários arquivos. Nestas bases, há uma ligação entre os campos autor, editor, revista e assunto do arquivo de documentos e os campos correspondentes nos quatro arquivos de busca.

O SGBD dBase IV:

As bases de dados documentários apresentam caráter específico por necessitarem de um controle mais restrito do vocabulário utilizado para indexação. Sendo assim, uma aplicação dBase IV possui como vantagens o recurso de especificar um campo particular para ser modificado e a possibilidade de ligar muitas bases de dados para as funções que necessitam de uma base única, sem necessitar unir os dados.

Como grande inconveniente de uma aplicação dBase IV, está o espaço do disco em função dos campos fixos: independente de possuírem caracteres preenchidos,

ocupam o espaço definido como tamanho dos campos. Em especial, para as aplicações documentárias, observa-se:

- ausência de controle de palavras-chave na entrada
- ausência de gestão de tesouro
- ausência de campos multivalorados

Como o dBase IV é uma ferramenta simples, de bom desempenho, oferece uma grande facilidade de aprendizagem e grande capacidade de arquivos, muitos profissionais de Informática envolvidos da realização de bases de dados documentários adotam este SGBD. No entanto, para que este projeto tenha sucesso, é necessário que sejam considerados o controle de palavras-chave na entrada, a gestão de tesouro e os campos multivalorados. É comum constatar projetos do gênero que, depois de absorver horas de trabalho, não contemplam as necessidades de gestão de centros de documentação e devem ser convertidos para *softwares* de gestão documentária especificamente concebidos para este fim. O mesmo tipo de inadequação é descrito adiante nas realizações documentárias com Hypercard.

#### O SGBD 4D (Quarta Dimensão):

Trata-se de um SGBD relacional que foi utilizado para desenvolver o SGBD documentário Basethèque. Possui como características gerais:

- roda em computadores Macintosh e possui limite de 16 milhões de registros por arquivo, 100 arquivos aberto simultaneamente, número ilimitado de relações entre os arquivos, 511 campos por arquivo e 511 subcampos por sub-arquivo;
- tipos de campos: alfabéticos (até 80 caracteres, tamanho variável); texto (até 32.767 caracteres, armazenados em tamanho variável para impressão); data (4 formatos); numérico (reais, inteiros, inteiros longos); imagem (todas as imagens Macintosh); e raiz (sub-estrutura à qual corresponde um sub-arquivo podendo compreender até 511 campos);
- pesquisa e seleção: pesquisas relacionais multicritérios; 8 operações de comparação (=, ≠, <, ≤, >, ≥, contém, não contém); truncamento; seleção visual de registros; salvamento de procedimentos de pesquisa; e pesquisa com a ajuda de menus criados pelo programador;
- recursos para 8 estilos de formatos que podem ser criados automaticamente;
- recursos para 8 modelos de gráficos (linhas, pontos, barras simples e outros) permitindo visualizar diversos tipos de distribuição de dados;
- procedimentos globais ou associados aos formatos de entrada: procedimentos globais (série de instruções associadas a toda a base); fórmula de formato (série de instruções ativas no nível dos formatos); e *script* (série de instruções associada a um campo ou a uma variável).

Desenvolvimento do *software* documentário Basethèque com o SGBD 4D (Quarta Dimensão):

Para ilustrar a utilização do SGBD relacional 4D na estruturação de uma base de dados documentários, são apresentadas apenas as funções de análise de conteúdo e de pesquisa retrospectiva, sendo as operações de gestão de biblioteca examinadas mais à frente.

Constam na Basethèque, os seguintes arquivos interconectados:

- monografias (livros, relatórios, normas etc.);
- pessoas (autores, leitores de uma biblioteca, pessoal de uma empresa);
- editoras;
- seriados e publicações periódicas;
- tabelas de diálogo (para a entrada automática de números de livros, criação de formatos de diálogo utilizados nos menus e criação de textos explicativos utilizados nas informações dos menus);
- tesouro (contém os descritores, suas características e relações);
- hierarquia (estruturação de relações hierárquicas, associativas e de sinonímia do arquivo tesouro).

Um conjunto de procedimentos globais e de fórmulas de formato pode ser utilizado, por exemplo, para distinguir entre os livros propriamente ditos e as outras monografias no arquivo que leva este nome, e então calcular o número de livros por temas/ autor ou por coleção/ autor.

O SGBD Superdoc:

Uma base de dados Superdoc é dividida em diversos arquivos físicos sob MS-DOS:

- o arquivo de registros contendo o texto integral da informação introduzida pelo usuário;
- o arquivo de léxicos contendo tanto os arquivos lógicos quanto os arquivos de tipo índice da base;
- o arquivo de referências dos registros chamado arquivo invertido contendo tanto os arquivos lógicos quanto o de descritores ou palavras-chave;
- o arquivo dos endereços físicos dos registros contendo, para cada registro da base, um ponteiro sobre o início do registro com relação à origem dos dados (primeiro registro da base). Ele contém também as datas de criação, de modificação ou de supressão de cada registro;
- o arquivo de títulos de campos contendo a descrição da base (lista e características de cada campo).

A estes arquivos são associadas duas tabelas primárias: uma de pesquisa e uma de acesso. O papel destas tabelas é melhorar a busca dos léxicos pelos usuários, a partir

das informações fornecidas por ele, como proximidade alfabética, truncamento, valor numérico etc.

A noção de registro físico estruturado de tamanho fixo não existe no Superdoc. Em conseqüência, o tamanho da informação é totalmente livre. Isto traz duas restrições:

- falta assegurar a manutenção das tabelas de ponteiros permitindo encontrar o início de cada informação e a necessidade de manter, no interior das informações, marcadores ou separadores especificando a natureza de cada campo;
- falta dispor de um processo de compactação e de recuperação, necessário para as alterações dos conteúdos de registros.

Estes limites são grandemente compensados pela densidade dos arquivos residentes sobre o disco que permitem otimizar o espaço ocupado pelas informações. A plena aplicação destas técnicas se encontra, em particular, no suporte pelo Superdoc, de imagens digitalizadas por um *scanner* depois compactadas. As listas de autoridade e os tesouros do Superdoc apresentam as mesmas técnicas de organização.

Hipertexto:

O hipertexto gera informação graças a uma rede de ligações permitindo estabelecer um sistema de relações entre os dados. Nas bases de dados documentários, a técnica de hipertexto permite pesquisar as informações navegando através da rede. Os *links*, que são em geral fixados no momento da realização do hipertexto, são ativados por meio da seleção de valores.

Para ilustrar o conceito de hipertexto, pode-se recorrer à enciclopédia, a qual não se lê do começo ao fim. Consulta-se um artigo que nos sugere um outro, e assim por diante. Esta peregrinação ou navegação representa um caminho heurístico. Descobrimos aquilo que, decerto, existe mas ao qual não tínhamos pensado inicialmente. Este mecanismo de descobrir é enriquecedor e gerador de novas idéias.

O **Hypercard** é um *software* que associa gráficos e textos. Os campos permitem a entrada de valores alfanuméricos ou numéricos e os gráficos podem ser colados, não importa onde. Os botões constituem a ferramenta de navegação.

Um certo número de aplicações é realizado pelo Hypercard mas esta facilidade é fruto de programação realizada por usuários experientes. Além disso, a memória ocupada pelo Hypercard é significativa.

Pode-se perguntar se esta ferramenta é a mais apropriada para a descrição de documentos com listas de autoridade indispensáveis para a coerência de informações e a indexação com a ajuda de um tesouro e para a pesquisa e a visualização de resultados e a sua impressão, assim como, se a resolução de problemas correntes em Informática Documentária não é preferência dos SGBDs, mesmo os relacionais. Sua aparente facilidade conduziu documentalistas a realizar aplicações documentárias com Hypercard, para então perceber que ele não permite a realização de uma gestão documentária.

Por não atordoar o usuário nas navegações guiadas imaginadas pelo programador e permitir a ele a descoberta a partir de uma navegação pela tela, o *software Eureka* possibilita explorar bases de dados documentários tradicionais. Este *software* considera dois tipos de navegação: uma de ordem técnica e outra de ordem

semântica. Do ponto de vista técnico, o usuário pode clicar diretamente nos valores com o *mouse* ou então nos termos indicados pelos botões de seleção. No aspecto semântico, as bases de dados documentários possuem a solução para o problema, por meio de duas possibilidades:

- o usuário clica no léxico sobre uma ou mais noções que lhe interessam. O sistema apresenta o número de documentos recuperados e propõe a visualização. Simultaneamente, o sistema pesquisa no arquivo invertido (constituído a partir da linguagem natural contida no título e no resumo e organizado para este fim) os outros termos que apresentam mais ocorrências. Estes são visualizados nos botões a partir dos quais o usuário pode estender sua questão;
- o sistema pode fazer o mesmo com os descritores da zona de indexação. Isto pode ser também uma forma de objetivar a relação associativa “ver também” de um tesouro.

### **Entrada dos dados**

Com o fim de proceder à interrogação *online* de bases de dados bibliográficos, depois da revisão pelo analista-indexador, pode-se dispor dos seguintes elementos:

- sistema de entrada, de controle e de manutenção do arquivo de referências (ou arquivo direto);
- sistema de organização de arquivos invertidos ou de chaves de acesso indexadas no arquivo direto;
- sistema de interrogação da base a partir de arquivos invertidos, compreendendo os procedimentos de acesso e uma linguagem de comandos, em particular para a edição de resultados de interrogação.

São apresentados alguns sistemas existentes, a partir dos quais, os tipos de entrada possíveis e as formas de chegar a elas incluem, entre outros:

- especificação dos controles de entrada: determinação dos campos e de suas características para a entrada dos dados;
- controle de palavras-chave: a partir da visualização da hierarquia, procede à eliminação de variações ortográficas com a ajuda de recursos do sistema e impede a entrada de palavras-chave sem esta validação;
- telas de entrada com validação dos dados;
- entrada com ajuda de listas de autoridade: uma lista de autoridade é associada a um campo índice contendo os termos pré-registrados que são selecionados no momento da entrada e automaticamente recopiados no campo correspondente, assegurando-se assim, uma entrada homogênea, rápida e sem falhas de ortografia nem variantes de escrita;
- controle de unicidade: efetuado sobre um campo do tipo índice da base de dados, permite verificar a unicidade de um valor, ou seja, conferir se este valor não existe no mesmo campo em um outro registro da base;

- entrada com ajuda de um tesouro: entrada de termos e de seus atributos, a partir de um módulo de gestão de tesouro. Inclui a hierarquia e relações de sinonímia e de “ver também”, além da possibilidade de associar uma nota de aplicação a um termo dado. Toda entrada de um descritor, neste caso, é impossível se ele não existir no tesouro. Assim, utiliza-se um campo suplementar de tipo índice na estrutura da base de dados que pode ser intitulado, por exemplo, “descritores candidatos”.

## Computador e produção de documentos

Para produzir um documento, deve-se proceder à redação, à composição e à impressão. A composição, efetuada a partir de um teclado, é mais ou menos automatizada e oferece mais ou menos possibilidades de intervenção no tratamento do texto e da imagem segundo o procedimento e o tipo de equipamento utilizado.

As possibilidades de intervenção sobre o texto (supressão, inserção ou substituição de uma letra, palavra ou parágrafo) são nulas com uma máquina de escrever clássica, são trabalhosas para o linotipista que compõe seu texto linha por linha e um pouco melhores em máquinas de escrever dotadas de memória. No entanto, máquinas dotadas de memórias voláteis não permitem a continuidade de um trabalho após terem sido desligadas.

Para amenizar este inconveniente foram criadas máquinas com memórias auxiliares (como os disquetes), a partir das quais, o texto pode ser gerado como um arquivo informatizado (realizando as operações de criação, exclusão e substituição). As máquinas de tratamento do texto são as que, em especial, permitem o acesso a palavras, linhas e parágrafos. Estas máquinas podem ser munidas de programas permitindo efetuar operações de composição automática como na escolha de estilo e tamanho de caracteres e de enquadramento de textos.

### 1 Técnicas de impressão

Impressoras de impacto:

Nas impressoras de impacto, um caractere móvel bate uma fita grafite em um papel, como as impressoras de esfera ou margarida, a agulhas e em cadeia. Estas últimas são impressoras rápidas utilizadas em saídas de computadores.

Técnicas de impressão sem impacto:

Dentre estas técnicas estão aquelas que utilizam **jato de tinta**, cuja qualidade gráfica é boa mas relativamente cara, portanto, não adequada para grande volume de textos.

A **fotocomposição com tubo catódico** não apresenta os limites dos anteriores para estilos, tamanhos e tipos de caracteres: os caracteres desenhados são digitalizados e sua “imagem” binária é registrada no computador com um código tipográfico apropriado.

Na **fotocomposição a laser**, o texto a ser composto é digitalizado em imagens ponto por ponto (*bit-map*) ou em imagens vetorizadas (*Postscript*). Neste último, o caractere é considerado, não como um conjunto de pontos mas como uma série de traços geométricos sobre o qual se memoriza a formulação matemática. Os valores binários servem para modular um raio *laser* que varre a superfície de um tambor revestido de um suporte com uma carga elétrica estática. A luz do raio *laser* descarrega a superfície do tambor no lugar correspondente aos caracteres a imprimir e atrai assim partículas de tinta carregadas eletricamente. A tinta depositada é transferida ao papel (este é o princípio utilizado por todas as fotocopiadoras eletrostáticas).

As técnicas de fotocomposição por tubo catódico ou a *laser* são também utilizadas para a produção de listagens de computador sobre microficha ou microfilme, ou seja, para as **saídas de computador sobre microforma (COM)**.

## 2 Escolhas de modos de operação

Distingue-se dois modos de entrada em função do volume tratado, da necessidade de dispor mais ou menos rapidamente do resultado, da repetitividade e da complexidade do tratamento:

- entrada em conexão direta (*online*) por um teclado ligado à fotocompositora;
- produção de um arquivo intermediário (sobre fita magnética) a ser tratado *offline* por um computador gerando os programas de fotocomposição. É adotada, por exemplo, para a produção de boletins analíticos.

## 3 Fotocomposição programada por computador

A fotocomposição compreende quatro fases:

- Fase de interpretação: o arquivo compreende uma informação estruturada (campos recuperáveis pela posição do primeiro caractere) e uma informação codificada (dando acesso às tabelas), mas nenhuma informação de ordem tipográfica;
- Fase de composição: em função dos comandos tipográficos que descrevem os estilos e tamanhos de caracteres a utilizar, o programa de composição separa as palavras na linha, seu enquadramento e o corte da palavra, se for o caso;
- Fase de paginação: formatação segundo indicações introduzidas na fase de interpretação para as colunas e títulos com a divisão adequada dos blocos de informação em colunas e páginas e enquadramento do texto nas colunas;
- Fase de visualização (*display*): o arquivo resultante é transformado em código de comando para a fotocomposição.

## 4 Tratamento do texto e fotocomposição assistida por computador

Os *softwares* de tratamento de texto funcionam com microcomputadores munidos de uma impressora e de um disquete ou disco “duro” para a memorização de dados e programas. Têm as funções de tratamento de texto (*word processing*) propriamente dito,

gerenciamento de arquivos e cálculos (muitas vezes em associação com uma planilha de cálculo).

O **tratamento do texto** compreende os seguintes recursos:

- com a inserção de cadeias de caracteres a partir de uma posição dada, o texto que segue é deslocado automaticamente;
- substituição de uma cadeia de caracteres por outra, de tamanho inferior ou superior e o deslocamento do texto que segue;
- supressão de uma cadeia de caracteres, como um caso particular de substituição;
- realocamento de uma cadeia de caracteres que é uma combinação de inserção em um novo local e de supressão do antigo;
- repetição automática de constantes (como títulos ou datas, no cabeçalho ou pé das páginas);
- paginação automática com a possibilidade de determinar o número da primeira página;
- enquadramento de títulos no texto;
- escolha de caracteres, além do seu tamanho e estilo;
- corte de palavras pela divisão de palavras longas a partir de suas primeiras sílabas, seguidas de um traço e espaço e com a parte restante seguindo na linha seguinte.

O **gerenciamento de arquivos** deve permitir pesquisar certos registros com a ajuda de chaves de acesso. Para produzir, por exemplo, um boletim analítico, deve-se exportar os dados do SGBD documentário e tratar os diferentes campos em função dos caracteres escolhidos.

## 5 Tratamento de imagens e fotocomposição assistida por computador

A maior parte dos *softwares* com estes recursos comporta:

- uma caixa de ferramentas de desenho nas quais pode-se escolher ferramentas de traços retos, de figuras geométricas, de desenho à mão livre etc.;
- uma coleção de “motivos” permitindo forrar o fundo de uma figura, escolher a espessura de traços etc.;
- elementos de tratamento de texto permitindo utilizar o teclado para inserir no texto desenho em modo *bit-map* (representação ponto por ponto dos elementos constitutivos dos caracteres). Para a geração de figuras geométricas se utiliza o modo vetorizado, onde o desenho é considerado uma série de traços geométricos no qual se conserva a formulação matemática permitindo, por exemplo, efetuar rotações dos desenhos.

## 6 Os SGBDs e a produção de documentos

Os SGBDs permitem a produção de documentos variados, como catálogos e listas, possibilitando a escolha de caracteres, estilos, normas e outros.

### **Arquivamento e gerenciamento eletrônico de documentos (GED)**

O arquivamento de informações e de documentos com vistas a sua utilização posterior pode ser realizado sobre diversos suportes:

- o papel, que é geralmente o suporte original;
- as microformas (microfilmes, microfichas) que constituem uma redução do volume do suporte anterior;
- os arquivos magnéticos (fitas e discos com registros em código binário como EBCDIC ou ASCII);
- os discos óticos ou magneto-óticos com registros analógicos ou digitais.

#### 1 Indexação de microformas e COM

Uma aplicação informática para gestão de microformas pode contar com a cópia de documentos primários, os quais podem ser munidos de códigos de acesso ou endereços (inscritos no alto das microfichas, por exemplo) e refletir o sistema de documentos secundários que constitui a base de dados bibliográficos.

As microfichas podem ser produzidas pelos meios fotográficos clássicos, mas também por impressora a *laser*. Sobre as microfichas produzidas por computador ou COM, os endereços podem ser gerados automaticamente por computador. A informação dos COM é uma informação que se produz desde a entrada ou por uma transformação com a ajuda de um programa. Não é o mesmo caso para os documentos primários que compõem um arquivo criado por fotocomposição ou em bancos de dados em texto integral.

Então se procede ao armazenamento em fitas magnéticas, mas não para acesso direto, o que torna custosa esta operação. O armazenamento é mais adequado com os discos óticos que podem conter os documentos primários, sejam os produzidos por fotocomposição, seja sob a forma gráfica do documento primário restituído sobre papel.

#### 2 O disco ótico digital

Representação e leitura da informação:

A informação é representada sobre o disco sob a forma binária pelas cavidades de tamanho variável de uma espiral, sendo que a presença ou a ausência de cavidade representa 0 ou 1. A formatação da espiral em setores dá uma organização do suporte e pistas endereçáveis (e, por isso, de acesso direto). A leitura se faz graças a um raio *laser*

se deslocando de pista em pista ou ao longo da espiral. O nível do sinal refletido difere com a presença ou ausência de cavidades caracterizando os valores 0 ou 1 (considerando que os registros de vídeo-discos são analógicos). Os discos podem ser registrados sobre duas faces.

Existem discos digitais de dimensões mais reduzidas (5¼ polegadas) que possuem dimensões análogas às das dos discos compactos de música. O registro sobre um disco compacto áudio resulta da transformação do sinal analógico correspondente ao som musical em valores numéricos (ou digitais). Estas novas memórias óticas são conhecidas sob o nome de CD-ROM e podem ser lidas sobre leitores, associados aos microcomputadores, de dimensões ligeiramente superiores aos leitores dos disquetes.

O disco é gravado a partir de uma matriz, com os mesmos procedimentos e ferramentas utilizados para os *compact-discs audio* (CDA), mas é uma memória “morta” (ROM - Read Only Memory).

Análise de imagens e compressão dos dados:

O raio *laser* permite uma maior densidade de registro. O disco permite registrar documentos com a ajuda de um *scanner* (ou analista de imagens) que digitaliza a informação sobre uma página (que pode ser um texto datilografado ou manuscrito). Assim, na saída do *scanner* utiliza-se uma técnica de compressão de dados.

### 3 Sistema de exploração de discos óticos

Compõe-se necessariamente de três partes:

- leitor de vídeo-disco ou CD-ROM: pode funcionar associado a um computador;
- *software* documentário: tal qual os utilizados para pesquisa em bancos de dados, permitem efetuar pesquisas segundo diversos critérios. No entanto, qualquer que seja a técnica de registro, ou seja, a partir de um arquivo magnético ou de um *scanner*, os documentos sobre o suporte ótico devem ser indexados de tal forma que o sistema possa recuperar seu endereço físico sobre o disco;
- *software* de navegação do vídeo-disco ou do CD-ROM: este *software* constitui a interface indispensável entre o *software* documentário e o acesso à informação sobre o disco ótico numérico que permite a visualização dos dados recuperados, sejam bancos de referências ou de dados tipo textual ou factual.

### 4 Aplicações de discos óticos ou magneto-óticos:

Pode-se classificar estas aplicações em cinco grandes categorias:

- difusão de bancos de dados;
- suporte para a telecópia (fac-símile) de documentos primários;
- difusão de obras pelos editores;
- arquivamento eletrônico de dados de uma instituição;

- gerenciamento eletrônico de documentos (GED) de uma instituição.

Difusão de bancos de dados:

Trata-se de bancos de dados difundidos por servidores, tais como, Chemical Abstracts, Medline, ERIC, cuja realização de CD-ROMs é simplificada pelo fato dos dados existirem já em arquivo automatizado assim como os índices de pesquisa. Os CD-ROMs são reeditados duas ou três vezes por ano. A vantagem para o usuário que utiliza o banco de dados com frequência é não ter que pagar pelo tempo de conexão por acesso. O inconveniente é a lacuna de atualização que pode ser amenizada por eventuais consultas *online*. (Supõe-se que, até a elaboração desta obra, as facilidades de acesso proporcionadas pela Internet ainda não estavam disseminadas).

Suporte para a telecópia (fac-símile) de documentos primários:

A maior parte dos servidores oferece os serviços de telecópia (fac-símile) de documentos primários detectados por uma pesquisa. Isto supõe um sistema de acesso fácil aos documentos arquivados, seja sobre microformas ou discos óticos digitais. Estes últimos, já gerados pelos meios informáticos para acesso aos dados, são particularmente adaptados ao transferir estes dados a um sistema de telecópia.

Difusão de obras pelos editores:

Trata-se sobretudo de obras do tipo enciclopédico, cujo interesse se evidencia pelas seguintes razões:

- a entrada do texto pode resultar em um arquivo magnético ASCII;
- os arquivos assim constituídos podem ser munidos de um índice que facilita a consulta e a navegação na enciclopédia;
- uma enciclopédia de dezenas de milhares de páginas é custosa sob a forma do papel em função do gasto com impressão e da defasagem de atualização dos dados com o passar do tempo;
- o editor poderá difundir sobre assinatura edições inteiramente atualizadas.

Arquivamento eletrônico:

Por arquivamento se entende a conservação de informações que respondem a critérios de patrimônio, de história ou de precauções jurídicas. Por definição, trata-se de informações não modificáveis e acessíveis somente para a consulta (o mesmo que as bases de dados bibliográficos). Em uma empresa ou instituição, arquivar implica em conservar, classificar e gerar os documentos a fim de recuperar a informação em caso de necessidade.

O arquivamento pode incluir:

- documentos datilografados registrados sobre suporte magnético ou de arquivos antigos digitalizados por um *scanner*;
- esquemas ou plantas constituídos de caracteres alfabéticos e de traços digitalizados por um *scanner*;
- fototecas (em preto e branco ou em cores) digitalizados por um *scanner* a partir de fotografias ou de dispositivos existentes.

A questão técnica é a mesma das três aplicações anteriormente analisadas; o problema neste caso é de ordem econômica. Considera-se o custo de produção do primeiro exemplar (*software* + análise e registro) e das cópias do disco em exemplares a partir de uma matriz, segundo a importância da tiragem. A rentabilidade da operação é avaliada em função do número de consultas anuais e do número de pontos de consulta. Por outro lado, se a capacidade de distribuição do produto é grande, a amortização do disco inicial é mais rápida.

#### Gerenciamento eletrônico de documentos (GED):

O GED é sensivelmente diferente do simples arquivamento. A diferença está na perenidade no caso dos arquivos e na obsolescência rápida no caso das numerosas informações administrativas, técnicas ou de gestão.

O GED se situa em um ambiente dinâmico, continuamente em movimento, de gerenciamento de documentos (notas, relatórios, anotações de reuniões etc.) referentes à vida da empresa ou da organização. A vocação do GED é a de saber difundir a informação, tanto aquela que é ferramenta de trabalho (produção e ajuda à decisão) quanto aquela de assistência à formação. Esta gestão necessita não somente da utilização de suportes técnicos apropriados, mas sobretudo, de um vasto sistema de informação e de comunicação global, gerado por programas de computador.

Deve-se considerar que uma tecnologia não implica, ela mesma, nenhuma vantagem na gestão de uma organização complexa. É a capacidade de uma organização em utilizar os recursos de uma tecnologia que decorre em vantagem. No entanto, o autor acredita que raramente uma nova tecnologia do setor informático suscitará tantas esperanças como o GED tem feito.

A gestão de informação em uma empresa tende a se complicar dia a dia em razão da multiplicação incessante do número de documentos. Trata-se de instaurar o trabalho cooperativo propício, tanto ao aumento da produtividade individual quanto ao de uma equipe, por meio de intercâmbios intensivos de informações no quadro de projetos comuns bem definidos. Um sistema de GED deve fazer interfaces com as aplicações já existentes, sejam de gestão, de produção ou de rotinas de escritório, para realizar consultas e extrações de informações ou indexações e adições de novos documentos na base de dados.

A implantação de um sistema deve ter uma justificativa econômica e organizacional. A questão do volume não é uma delas. A justificativa deve ser ligada à produtividade, ou seja, às melhoras evidentes sobre o tempo de pesquisa e de acesso, mais que ao ganho de espaço.

Dedicado à entrada, ao armazenamento, à consulta e à circulação de documentos de trabalho, um sistema de GED assegura o compartilhamento de dados entre todas as pessoas da empresa. São a eficácia da pesquisa de informações e a riqueza de suas possibilidades que vão garantir a adesão dos usuários ao sistema. A interface amigável também faz parte deste rol, mas cores, ícones e mensagens não devem encobrir uma carência funcional.

Apesar de os investimentos em GED serem seriamente considerados em empresas, meios materiais mais modestos podem gerar resultados surpreendentes. Mais que outra aplicação, o GED necessita de um estudo prévio aprofundado e um guia de especificações rigoroso e preciso.

## Informatização de uma biblioteca

Diversos casos podem se configurar a partir do funcionamento de uma biblioteca. Alguns deles seriam:

- biblioteca integrada a um **serviço de documentação** em uma empresa pública ou privada, um centro de pesquisa etc.;
- biblioteca de **leitura pública** com consulta no local e, muitas vezes, com empréstimo em domicílio;
- biblioteca ligada a um **estabelecimento de ensino** e funcionando aproximadamente como uma biblioteca de leitura, em que a comunidade de leitores é conhecida e pode ter sido registrada ao mesmo tempo em que se matriculou.

As funções destas bibliotecas podem ser diferentes segundo os imperativos do leitor: a atividade da biblioteca integrada a um serviço de documentação pode diferir daquela da biblioteca de leitura pública pois, além do empréstimo e da leitura no local deve, por exemplo, providenciar os documentos que ela não possui a fim de satisfazer as necessidades expressas pelos leitores. A exploração da Informática por estas bibliotecas pode também diferir pela natureza das informações contidas nos arquivos que elas utilizam, pela proveniência e controle destas informações e pela sua obtenção enquanto sub-produto de arquivos já existentes.

Como exemplo é apresentado o módulo de gestão de biblioteca do *software* documentário Basethèque (que foi desenvolvido a partir do SGBD relacional 4D). Neste módulo, são realizadas as funções:

- gestão de encomendas e de assinaturas;
- controle de chegada e circulação de periódicos;
- gestão de empréstimos;
- gestão orçamentária.

### 1 Operações de abastecimento (provimento) de documentos

Qualquer que seja o tipo de biblioteca, toda a sua atividade deve ser orientada para sua tarefa maior: satisfazer as necessidades do leitor nos melhores prazos e nas melhores condições de rentabilidade.

As necessidades do leitor podem ser estimuladas pelas informações que lhes são transmitidas ou indicadas:

- pelos editores (catálogos, publicidades diretas ou em revistas);
- pelas citações bibliográficas (em boletins, em bases de dados, pelas citações de autores);
- pelo circuito informal (congressos, colegas, superiores hierárquicos).

Isto pode e deve ser realizado por pesquisas sistemáticas efetuadas pelo serviço de documentação e pela indicação espontânea de documentos suscetíveis de interesse.

A biblioteca situa-se no cruzamento da relação existente entre os fornecedores do documento (editores, organismos especializados, outras bibliotecas) e o usuário final, o leitor. As diversas operações efetuadas pela biblioteca a partir disto são:

- criação dos meios de expressão das necessidades;
- verificação das demandas e, se for o caso, encaminhamento de encomendas;
- efetivação da encomenda de acordo com o tipo de material (periódico ou não periódico);
- recepção e controle das mercadorias chegadas: pontuais (livros) e repetitivas (revistas);
- comunicação ao setor de análise (quando se tratar de uma biblioteca integrada a um serviço de documentação);
- arquivamento de documentos primários (eventualmente sob forma de microfichas);
- contabilidade (pagamento de fornecedores, imputação aos usuários).

Estas operações de gestão clássica são de fácil informatização pois são formalizáveis por um sistema de relações entre fornecedores, documentos, leitores etc.

## 2 Definição de operações

A gestão de encomendas e de assinaturas em um serviço de documentação é um caso particular de gestão de estoques. A gestão de um estoque é realizada por um sistema de transações colocando em relação:

- objetos definidos por suas características próprias (produtos industriais, produtos alimentares, livros, revistas);
- atores que podem ser pessoas físicas (consumidores, leitores) ou jurídicas (empresas, serviços de uma empresa, editores).

Entre os atores de uma transação, pode-se notar:

- o produtor do objeto (fabricante, editor);
- o fornecedor do objeto (centrais de assinaturas, livrarias);
- o usuário final (consumidor particular, pessoas ou serviços destinatários de documentos em uma empresa).

A gestão do estoque implica também:

- o controle do estado do estoque;
- a gestão do abastecimento (provimento) em função do consumo anterior e dos desenvolvimentos atendidos;

- os aspectos financeiros (pagamentos de fornecedores, fatura aos clientes, imputação aos serviços usuários de uma empresa).

No caso particular de gestão de um serviço de documentação, distingue-se:

- as encomendas pontuais de documentos (livros, relatórios, anuários, números isolados de uma revista etc.): cada operação tem fim desde que o usuário final esteja de posse do documento (definitivamente ou por empréstimo) e que o fornecedor esteja pago;

- as assinaturas de periódicos: trata-se de um pedido reconduzido a cada ano.

O fornecedor de um serviço de documentação pode ser o próprio editor, um grossista (central de difusão de assinaturas), um detalhista (livraria local para o fornecimento de livros), uma associação ou uma sociedade científica (que difunde uma revista juntamente ou não a um editor comercial).

São definidos da seguinte forma os diversos atores:

- fornecedor: organismo (editor, livraria, sociedade científica, central de assinaturas etc.) que comercializa documentos (periódicos ou não);

- solicitante (serviço): serviço solicitante da aquisição de um documento ou da subscrição de uma assinatura;

- solicitante (pessoa): responsável (pessoa a parte ao serviço solicitante) de um serviço que estabelece a demanda de compra ou de assinatura, podendo ser ele mesmo o destinatário;

- dono da encomenda: serviço que estabelece a encomenda e a envia para a biblioteca ou para o editor;

- destinatário (serviço): serviço ao qual a livraria ou o editor deve enviar o documento solicitado ou a assinatura;

- destinatário (pessoa): pessoa à parte do serviço destinatário para a qual o fornecedor envia diretamente o documento;

- organismo pagador: serviço em nome do qual a livraria ou o editor deve liberar a fatura.

### 3 Serviço de documentação: central de compras

Considera-se o serviço de documentação como uma central de compras dentro da empresa. Para cumprir esta função, o centro de documentação deve realizar certas operações:

- passar as encomendas aos fornecedores (editores ou livrarias);
- efetuar ou fazer efetuar o pagamento por um serviço central de contabilidade;
- registrar os documentos em sua chegada (com um módulo de entrada e um sistema de controle de chegada de periódicos);

- colocar os documentos à disposição de usuários por um sistema de empréstimo ou de circulação de revistas (com os módulos existentes ou a criar).

#### 4 Análise orgânica: definição de arquivos

A aplicação realizada com o SGBD relacional 4D, utiliza-se dos seguintes arquivos:

- arquivo mestre: monografias;
- arquivos servidores: pessoas, coleção de periódicos e editores;
- arquivos de união dos dados para aplicação de: encomendas e assinaturas e, numa etapa posterior, para a circulação.

Estes arquivos são essenciais para que se estabeleça um sistema de relações entre objetos e atores. Estas relações constituem a essência de suas próprias informações como:

- data da encomenda;
- datas de vencimento e de renovação de uma assinatura;
- quantidades;
- montantes em função das quantidades e do preço unitário (calculados pela relação entre os arquivos de monografias e de coleção de periódicos);
- modo de pagamento;
- data de pagamento;
- número de cheque;
- banco de pagamento.

No arquivo mestre e nos arquivos servidores, existem campos que permitem assegurar os aspectos contábeis da gestão. No arquivo de revistas, podem ser calculados, por exemplo, os preços de assinatura anteriores e os correntes. O arquivo de organismos conta com os recursos: previsão orçamentária para o ano em curso, despesas do ano em curso e do ano anterior e orçamento corrente.

#### 5 Entrada de encomendas

Trata-se da solicitação de monografias (livros, normas, relatórios etc.) menos os periódicos que são tratados com as assinaturas (a não ser, um número de um periódico do qual não se tem assinatura). Pode-se escolher um ou mais exemplares do mesmo documento que não se encontra na biblioteca. Pode-se também solicitar um ou mais exemplares suplementares de um documento existente na biblioteca e já descrito no sistema.

No primeiro caso, deve-se registrar o tipo de documento, tipo de arranjo que se deseja, fornecedor, pessoa que demanda, serviço que demanda (que é preenchido

automaticamente se o campo anterior for utilizado), documentos solicitados e número de exemplares. Todos estes campos são preenchidos por conteúdos previamente existentes, não sendo, assim, necessário digitar todo o termo, mas apenas as primeiras letras. Os campos “número do pedido”, “data do pedido” e “número do registro” do livro são fornecidos automaticamente pelo sistema.

## 6 Entrada de assinaturas

Supõe-se que os dados necessários já existam nos arquivos de pessoas, editores e periódicos. O tratamento da assinatura necessita da entrada de características da assinatura, do envio de uma carta ao editor ou ao fornecedor e da renovação anual das assinaturas.

Para uma revista, é criado um registro no arquivo de assinaturas que inclui uma relação entre a revista e a pessoa ou serviço que a solicitou, qualquer que seja o número de assinaturas. Para novas pessoas ou serviços solicitantes, novos registros são criados.

Na tela de entrada das assinaturas, constam os campos:

- revista: que traz automaticamente, a partir do arquivo de coleções de periódicos, o editor comercial, o ISSN (International Standard Serial Number), preços e taxas;
- assinatura: cujo valor é calculado automaticamente pela concatenação da revista e do solicitante;
- preço unitário da revista: não é utilizado se o preço varia no decorrer do ano;
- fornecedor: é criado automaticamente no arquivo de assinaturas a partir do campo “editor comercial”, carregado junto com o endereço apenas quando o fornecedor for um editor comercial;
- número da assinatura corrente: número de assinaturas do ano em curso para este registro;
- número da assinatura anterior: número de assinaturas do ano anterior para este registro. Se a assinatura for anual ou mensal, este valor pode vir automaticamente. Para este campo e o anterior, o programa utiliza a comparação entre os valores dos campos para produzir automaticamente um texto apropriado nas cartas de renovação;
- data da operação: atribuída automaticamente pelo programa;
- data de início e de vencimento da assinatura: campos de preenchimento obrigatório;
- solicitante (pessoa): vem automaticamente do arquivo de pessoas com a digitação das primeiras letras;
- solicitante (serviço): é preenchido automaticamente quando o campo anterior é preenchido.

## 7 Controle de chegada de revistas

Em bibliotecas não informatizadas, utiliza-se freqüentemente fichas chamadas kardex (do nome da marca do móvel no qual elas são dispostas em gavetas) para anotar a chegada de revistas. Esta designação é mantida aqui para descrever a tela de controle de chegada. Este controle deve permitir perceber números ou exemplares faltantes e revistas das quais não se recebe exemplares depois de um certo tempo.

No programa de automação usado como exemplo, o arquivo kardex faz parte do arquivo de coleção de periódicos e cujo formato de entrada permite:

- visualizar as 12 últimas operações efetuadas sobre uma revista;
- listar todas as revistas em que faltam um ou mais exemplares depois de determinada data;
- listar todas as revistas sobre as quais não foram feitas operações depois de determinada data.

A tela do kardex possui os campos: volume, número, data da publicação, data de entrada, número de exemplares recebidos, exemplares faltantes (o programa calcula o número de exemplares a receber de acordo com a assinatura) e data da última operação.

## 8 Circulação de revistas

A circulação é a transferência de documentos a  $n$  leitores sucessivos.

Na **circulação restrita**, um documento parte do ponto de origem (a biblioteca), vai sucessivamente aos leitores e volta à biblioteca, onde o arquivamento é o último ponto da circulação. O documento pode também ser classificado como perdido, cujo leitor é considerado desconhecido.

Na **circulação em margarida**, a biblioteca transfere um documento a um leitor, recupera o documento e o transfere a outro leitor, e assim por diante. O **empréstimo** é uma operação em margarida. Sua duração pode ser fixa (com um sistema automático de lançamento), indefinida (não se recupera o documento até que haja outra demanda) ou sob a forma de depósito (o leitor faz uso permanente do documento). Um segundo exemplar da mesma obra pode ser necessário e seu empréstimo será representado por mais uma pétala na margarida. O mesmo procedimento automatizado pode ser utilizado para o empréstimo e para a circulação. A consulta no local é um empréstimo efêmero.

Para fazer circular as revistas junto a diversas pessoas ou serviços de um organismo, utiliza-se uma papeleta de circulação. Com a papeleta de circulação, pode-se adotar tanto o método restrito como o de margarida.

A circulação controlada em margarida oferece as seguintes vantagens:

- assegura-se a cada leitor receber a revista em um tempo razoável;
- a todo momento, pode-se saber onde se encontra os diversos números de dada revista (para uma consulta à biblioteca, por exemplo);
- a todo momento, pode-se encontrar o conjunto de revistas que um leitor tem em mãos.

O sistema produz papeletas de controle de circulação a partir dos campos:

- revista;
- número de circulação: gerado automaticamente, é mais recomendado para circulação de vários exemplares de uma revista para vários usuários ao mesmo tempo;
- data ou número da publicação: anteriormente introduzidos no kardex;
- data de circulação: data automática do dia da operação;
- destinatários: registro de cada destinatário já existente no sistema.

## 9 Gestão de empréstimos

A gestão de empréstimos é realizada por um sistema de transações, portanto, sobre objetos documentários (livros, relatórios, teses, revistas, suportes audiovisuais) entre o serviço de documentação ou a biblioteca, de uma parte, e os usuários, de outra.

Os objetos documentários podem ser (no caso exemplificado):

- monografias (livros, relatórios, teses, normas etc.);
- catálogos de fornecedores;
- números de uma revista;
- suportes audiovisuais (vídeo-cassetes, discos compactos, filmes, dispositivos, fitas cassetes etc.).

Os usuários podem fazer parte da instituição de forma permanente (funcionários de uma empresa) ou de forma temporária (alunos de uma escola) ou pertencer a instituições externas. No arquivo de pessoas, o pertencimento de um leitor a um serviço interno ou a uma instituição externa, pode ser definida pela ligação com o arquivo de organismos. Este método permite responsabilizar pessoalmente o beneficiário de um empréstimo pela restituição do documento.

O empréstimo pode ser: efêmero (leitor que faz consulta local e não há registro no sistema), de duração determinada e de longa duração ou permanente. A atribuição permanente de documentos permite gerar pontos de documentação ou de bibliotecas locais em grandes empresas.

A natureza das transações pode ser de saída ou de restituição de documentos. A estas transações, pode-se incluir a noção de “espera” para documentos indisponíveis que tenham sido solicitados por um leitor que, então, entra em uma lista com este nome.

O objetivo da aplicação é saber:

- o estado permanente dos empréstimos em curso;
- o estado permanente dos leitores;
- o estado permanente dos leitores com documentos emprestados;
- o registro individual do leitor com documentos emprestados;
- a estatística de frequência de empréstimo de documentos;

- a possibilidade de encontrar rapidamente quem emprestou um documento;
- a possibilidade de enviar cartas de cobrança aos usuários atrasados na duração do empréstimo;
- a possibilidade de informar um leitor sobre a disponibilidade da obra que ele reservou.

No sistema Basethèque constam os arquivos de monografias, revistas, organismos, suportes audiovisuais e pessoas, que foram descritos anteriormente. Ao menos um dos quatro arquivos é necessário para realizar a operação de empréstimo, além do arquivo de pessoas que é pivô desta transação.

## 10 Gestão orçamentária

O objetivo desta aplicação é o controle de despesas de diversos serviços de uma instituição em matéria de encomendas de obras ou de assinaturas, além da comparação com orçamentos anuais anteriores e de previsão. Nas grandes empresas ou em uma rede de bibliotecas em que as compras são processos cooperativos, é indispensável atribuir um orçamento anual a cada serviço ou biblioteca e, deduzir do montante, estes orçamentos de despesas decorrentes da demanda por encomendas de documentos ou assinaturas.

A previsão de orçamentos destes serviços ou bibliotecas é realizada por meio do arquivo de editores, o qual possui um campo para despesa realizada que é ligado aos arquivos de encomendas e de assinaturas.

A repartição de despesas entre os diversos serviços ou bibliotecas pode ser visualizada ou impressa sob a forma de listas ou sob a forma de gráficos, entre muitos outros relatórios.

## **Análise de conteúdo dos documentos**

### 1 Constituição de bases de dados bibliográficos (referoteca)

A base de dados bibliográficos (referoteca) é constituída por um conjunto de registros que representam documentos secundários, por fazerem referências aos documentos primários. Cada registro compreende um certo número de campos (autores, palavras-chave, resumo etc.).

O conteúdo dos campos “autores” e “palavras-chave” de cada registro é unido ao conteúdo dos outros registros e separados alfabeticamente, constituindo um novo arquivo. Este é chamado arquivo invertido, por oposição ao arquivo direto que constitui o arquivo de referências.

Em resposta a uma questão colocada com a ajuda de palavras-chave ou nomes de autores (algumas vezes, combinados entre eles com a ajuda dos operadores booleanos E, OU, NÃO), o sistema pesquisa no arquivo invertido os números de acesso dos documentos secundários que tratam da questão colocada. O sistema utiliza os números de acesso para apresentar ao usuário as referências (título do artigo, revista em

que foi publicado, data, página etc.) do(s) documento(s) que respondeu(ram) à questão. Estas referências podem ser acompanhadas de uma indicação sobre o conteúdo dos documentos com a ajuda de um resumo e/ou de palavras-chave.

## 2 Objetivos das bases de dados bibliográficos (referotecas)

As bases de dados bibliográficos (referotecas) têm como objetivos essenciais informar com rapidez ao especialista aquilo que se publica de novo em sua área e permitir-lhe pesquisar retrospectivamente todos os documentos úteis à resolução de um problema particular.

O primeiro dos objetivos pode ser efetivado por produtos tradicionais clássicos como a publicação de fichas bibliográficas ou a difusão de boletins analíticos ou pelo método de DSI. O segundo objetivo pode ocorrer pelos meios antigos (acesso por assunto em fichários manuais ou por índices de classificação) ou pelo método mais atual de pesquisa retrospectiva em uma base de dados informatizada.

A difusão seletiva consiste em fazer chegar periodicamente ao usuário os registros correspondentes a sua área de interesse, com ajuda de um conjunto de palavras-chave ou descritores que constituam seu perfil documentário. É efetuada no conjunto de registros integrados à base em um dado período de tempo.

Já a pesquisa retrospectiva consiste em interrogar a referoteca com ajuda de um conjunto de palavras-chave ou descritores relativos ao tema de pesquisa do usuário e visualizar os registros dos documentos correspondentes a este perfil.

Os registros integrados à base podem servir para a edição de um boletim bibliográfico ou analítico (por um procedimento de fotocomposição, por exemplo) munidos ou não de um índice de autores e palavras-chave.

O resumo e a indexação que resultam da análise de conteúdo constituem um meio de expressão reduzido do texto integral em linguagem documentária. As palavras-chave ou descritores podem ser expressos pela linguagem natural ou por um léxico controlado ou um tesouro. O resumo pode ser redigido em linguagem natural ou em formas de proposições representadas por palavras-chave e reunidas por indicadores de sintaxe.

A operação de pesquisa retrospectiva ou de DSI representam uma filtragem da questão colocada pelo usuário em linguagem natural a ser transformada em linguagem documentária.

Os procedimentos de análise e de representação reduzida do conteúdo dos documentos são examinados e analisados quais deles são informatizáveis.

## 3 Procedimentos para elaboração de um resumo

A significação de um documento escrito é detectada por uma série de integrações sucessivas sobre a significação de unidades lingüísticas e semânticas de dimensões crescentes e sobre a significação das associações destas unidades em um nível mais complexo. Nessa ordem têm-se: as letras, as palavras, as frases e os conjuntos de frases. A significação global do documento depende da significação dos elementos

considerados isoladamente em um certo nível e a significação que eles tomam quando são associados a elementos do mesmo nível. A detecção da significação de um texto passa, entre outros, pela apreensão dos pólos semânticos. Mas a significação de um texto não pode ser detectada apenas pelo uso de seus constituintes lingüísticos. De fato, para “compreender” um texto, o leitor utiliza um grande número de subconjuntos de fora do texto que fazem parte de seu espaço cognitivo.

O espaço cognitivo é constituído pelo conjunto de informações do texto e fora dele, dispostos ao leitor para que ele apreenda a significação de um texto: experiências de citações bibliográficas, substrato sociocultural do leitor, conhecimento do autor (biografia) e de sua obra (bibliografia), imagens induzidas pelo leitor-receptor (a uma descrição do autor, o leitor associa as imagens mentais, que é em particular o que faz o cineasta quando transpõe uma obra literária).

A elaboração de um resumo resulta de um conjunto de operações intelectuais conduzidas por um analista, quais sejam: leitura do texto integral, filtragem do texto integral pela apreensão dos pólos semânticos, integração sucessiva dos pólos semânticos e apreensão da significação global. Ele redige um relatório dessa significação: em linguagem natural, em linguagem documentária com ajuda de frases canônicas ou em linguagem natural com sintaxe reduzida.

Os obstáculos maiores à realização automática de resumos são de ordem prática e teórica: necessidade de compreender o texto integral e de detectar a significação de um texto, produzindo um resumo significativo e legível sobre o conteúdo de um documento.

Um bom resumo deve permitir saber de que fala o documento com a ajuda de palavras de caracterização e que estão ditas com a ajuda de palavras operatórias.

No “texto” de uma referoteca (título ou resumo de um documento), encontram-se muitas categorias de palavras conforme o interesse que elas apresentam para a indexação desse documento:

- palavras vazias (artigos, preposições, locuções prepositivas): não são utilizadas para constituir o arquivo invertido de uma referoteca. O conjunto de palavras vazias constitui o antídicionário;
- palavras atemáticas ou operatórias: expressam condições operatórias como: estudo, descrição etc. Indicam o assunto das coisas mais que as coisas elas mesmas. Não apontam o tema tratado no documento, mas precisam o que é que se diz, a propósito das palavras de caracterização (citadas em seguida) que definem o documento. São termos mais ou menos comuns a todos os assuntos e situações e, por isso, apresentam pouco interesse para a indexação e a pesquisa retrospectiva;
- palavras de caracterização: palavras temáticas genéricas ou palavras-chave específicas que definem o assunto ou tema tratado no documento. Em uma pesquisa retrospectiva, a sua utilização é primordial pois permite isolar um subconjunto de documentos a partir dos quais se pode pesquisar em condições operatórias particulares.

#### 4 Métodos de indexação

Entrada automática no arquivo invertido em linguagem natural:

Cada registro inserido no sistema contém palavras (compreendidas como as cadeias de caracteres incluídos entre dois espaços) em certas partes como no título e no resumo, por exemplo. Depois de eliminados os caracteres não alfabéticos (signos de pontuação, apóstrofes etc.), estas palavras servem para a entrada no arquivo invertido que é acrescido do número do documento se a palavra já existia, ou da palavra nova e do número do documento caso contrário. Vários dos sistemas de indexação são munidos de um antídicionário ou dicionário de palavras vazias, as quais não são utilizadas para a entrada no arquivo invertido.

Esse tipo de indexação utilizado, entre outros, pela IBM no *software* STAIRS, tem por vantagem necessitar de recursos informáticos simples, mas apresenta como desvantagens:

- proliferação do número de entradas no léxico, em que se podem encontrar várias formas nominais, adjetivais, verbais referentes à mesma noção (a morfologia do inglês é mais restrita que a do francês, o que explica que este tipo de arquivo invertido é mais adequado para bases de dados bibliográficos anglo-americanas), sem contar as formas saídas de erros de ortografia;
- separação alfabética de palavras do mesmo campo semântico, expressando outras noções como, por exemplo, *oeil* (olho) e *yeux* (olhos);
- palavras compostas que ficam separadas e dispersas devido, em alguns casos, ao dicionário de palavras vazias, como por exemplo, as palavras ferro de passar roupa e caneta-tinteiro, que geram no arquivo invertido a seguinte seqüência aproximada: caneta, ferro, passar, roupa, tinteiro.

Indexação manual com um léxico controlado:

Para evitar aqueles inconvenientes, a maior parte das bases de referências reserva em cada registro referente a um documento, um campo de indexação onde são introduzidas os descritores extraídos antecipadamente pelo analista/ indexador do léxico controlado. O número de entradas no arquivo invertido é assim consideravelmente reduzido, obtendo-se a regularidade de escrita dos descritores, o que rende maior confiabilidade às buscas e a possibilidade de expressar uma noção com a ajuda de uma palavra-chave correspondente a diversas formas lexicais.

O léxico pode ser monolíngüe compreendendo a lista de unidades lexicais utilizadas por uma técnica ou uma ciência em uma língua, ou bilingüe ou multilingüe compreendendo uma lista de unidades lexicais com seus equivalentes em uma ou mais línguas. Para a indexação de documentos, utiliza-se um léxico controlado que é, com relação ao léxico geral da língua, um léxico referente a um vocabulário de especialidade.

Indexação com ajuda de um tesouro:

A relação de ordem alfabética que preside a organização de um léxico é tal que ela não permite a aproximação semântica entre as partes de um todo (como os nomes dos países de um continente), entre os hipônimos (palavras que se referem a uma noção específica, como gato, vaca) e entre hiperônimos (palavras que se referem a uma noção genérica como, animal doméstico, mamífero) ou reciprocamente.

Tais relações devem ser organizadas em rede semântica, cuja aplicação particular é o tesouro, utilizado para melhorar as performances dos sistemas de pesquisa

documentária, em particular, porque permite a extensão da formulação da pergunta com a ajuda de descritores mais genéricos ou mais específicos.

Indexação com analista automático de textos:

A entrada automática em um arquivo invertido apresenta inconvenientes. Ela tem, entretanto, a vantagem de não necessitar da intervenção humana, como no caso da indexação com léxico controlado ou com um tesouro. Mas a indexação manual pode contribuir a uma melhor definição do conteúdo de documentos e permitir a organização de arquivos invertidos mais econômicos. A utilização de um analista automático de texto deverá permitir combinar as vantagens dos dois métodos.

O primeiro objetivo do analista automático consiste em obter as cadeias de caracteres ou palavras do texto (título e resumo) do registro documentário e compará-las com um léxico, a fim de reduzir as variantes morfológicas (singular/plural, masculino/feminino, desinência dos verbos etc.) a uma forma lexical canônica (a entrada de um dicionário, por exemplo). Para isso, pode-se utilizar um analista morfo-gráfico.

O segundo objetivo do analista automático é converter as diversas categorias lexicais (substantivos, adjetivos, verbos) referindo-se a uma mesma noção (como as palavras “detecção”, “detector”, “detectar”) a uma forma semântica única (representada, por exemplo, pelo substantivo) com saída para o analista morfo-gráfico.

O arquivo invertido assim criado não contém praticamente mais que substantivos no singular, o que constitui uma redução considerável do número de entradas do arquivo invertido com relação à entrada automática em linguagem natural. Mas esse arquivo contém apenas unitermos. Para constituir um arquivo invertido, devem ser utilizados sintagmas. Para tanto, faz-se necessário um analista morfo-sintático capaz de reconhecer sintagmas.

O analista morfo-sintático explora não somente as categorias lexicais, mas também o gênero, o número etc. para efetuar uma análise morfológica mais completa de palavras que apareçam no texto. Graças ao resultado da análise morfológica, efetua-se uma análise sintática do papel respectivo das palavras na frase que pode ser recortado em sintagmas.

Um dos problemas colocados para os sintagmas é o reconhecimento de seus limites e, para os sintagmas longos, o problema de sua distribuição em sintagmas mais elementares. Assim, pode ser mais simples utilizar os resultados do analista morfo-gráfico e comparar cada segmento de duas palavras com os sintagmas possíveis no léxico controlado ou um tesouro.

## 5 Funcionamento de um analista morfo-gráfico

As funções do analista morfo-gráfico são:

- selecionar no texto cadeias de caracteres limitados por um espaço (e por certos caracteres tais como apóstrofe e os signos de pontuação);
- comparar as cadeias de caracteres selecionadas com as diversas formas de palavras (singular/plural, masculino/feminino, pessoas e tempos verbais) que podem ser geradas a partir das bases de palavras às quais se ajunta o conjunto de desinências associáveis a uma base (algoritmo de reconhecimento de uma palavra);

- selecionar no dicionário das bases uma forma canônica única para as diversas formas de palavras (o singular para os substantivos, o masculino singular para os adjetivos, sendo que os verbos não são normalmente considerados relevantes para indexação) que será a única a figurar no arquivo invertido;
- selecionar e assinalar os homógrafos eventuais;
- permitir restabelecer diversas formas canônicas a uma forma considerada como semanticamente equivalente: ao verbo “detectar” corresponde a forma semântica “detecção”;
- utilizar a forma semântica para tratar os sinônimos documentários: à forma “Suíça” pode corresponder a forma “Confederação Helvética”;
- utilizar a forma semântica para o antidicionário: às palavras não significativas (artigos, preposições etc.) será associada uma forma semântica vazia e seu reconhecimento não cria a entrada no arquivo invertido. O reconhecimento de locuções prepositivas é efetuado conforme o mesmo princípio dos descritores sintagmáticos.

As restrições deste analista são menores que as de um analista morfológico que deve fornecer as informações gramaticais (gênero, número, pessoa etc.) necessárias à análise morfo-sintática (seleção de grupos de palavras ou sintagmas e de seu papel dentro da frase).

Os resultados do analista morfográfico permitem reduzir consideravelmente o tamanho do arquivo invertido da referoteca pois as palavras não significativas não entram no arquivo e o número de formas de palavras significativas é reduzido (para 2 para os substantivos, 4 para os adjetivos e 10 para os tempos e modos mais usuais dos verbos).

Isto leva a uma economia do espaço em disco, além da diminuição do tempo para consulta à referoteca pelo usuário.

## **Concepção e elaboração de um tesouro**

### **1 Apresentação e finalidade do tesouro**

O tesouro permite estruturar as noções representadas pelas palavras do léxico e dar evidência às relações semânticas que são ausentes da relação de ordem alfabética de um léxico. Interessa a utilização de um tesouro em oposição a um simples léxico alfabético: em especial, em sociedades onde o vocabulário é mais rico, a necessidade de falar a mesma linguagem leva à utilização de um tesouro para indexar documentos a fim de que cada usuário possa recuperá-los com mais facilidade.

A maior parte dos tesouros é editada sob forma de documento em papel ou consultáveis em tela. O tesouro pode compreender diversos modos de acesso: pela lista alfabética, por campos semânticos, por árvores hierárquicas.

No **acesso por lista alfabética**, os descritores são organizados alfabeticamente e cada um é seguido de seus termos genéricos (TG), específicos (TE) e associados (TA). Eventualmente pode apresentar também uma indicação de um sinônimo (ou um quase-sinônimo) preferencial, numa relação de substituição. Além disso, o descritor pode ter um

equivalente em uma língua estrangeira. A entrada alfabética pode ser seguida de uma nota explicativa (NE) ou nota de aplicação (NA), precisando a aplicação de um descritor. É interessante apresentar os descritores sintagmáticos em duas entradas, permutando as palavras significativas (princípio do índice KWIC).

No **acesso pelos campos semânticos**, o descritor de entrada alfabética pode ser seguido de um número de campo semântico no qual encontram-se todos os descritores designando as noções que possuam entre eles uma ou mais características comuns. Os campos semânticos são apresentados sob forma de lista alfabética de descritores fazendo parte deste campo semântico ou sob forma de um esquema sinalizado. No esquema por flechas (ou setas), os descritores são localizáveis por suas coordenadas verticais e horizontais. As duas representações permitem ao indexador situar um descritor no seu desenvolvimento semântico.

Para o **acesso pela estrutura hierárquica**, a relação genérico-específica ou a relação todo-parte pode ser representada por uma estrutura em árvore, que é representada sob forma de indentação, cujas linhas sucessivas desenham o nível do descritor sobre a árvore. Esta apresentação permite ao analista-indexador situar um termo no seu desenvolvimento hierárquico e utilizar eventualmente outros TGs ou TEs para indexar seu documento. Mas, neste caso, a indexação pode ser feita automaticamente, se o arquivo em que foram registradas as relações hierárquicas estão organizados convenientemente.

## 2 Representação informática do tesouro

Para gerar um tesouro são necessários: um arquivo dicionário, arquivos de relações (hierárquicas, associativas, substitutivas) e o arquivo-tabela dos campos semânticos.

O **arquivo dicionário** é um dicionário simples apresentado sob a forma de código e nome do termo correspondente.

O **arquivo de relações hierárquicas** permite levar em conta a relação genérico-específica e a relação todo-parte. A poli-hierarquia é admitida.

Uma vez conhecidas as hierarquias do tesouro e descritas para um sistema de indentação, aplica-se a cada termo, segundo seu nível na árvore, um código correspondente. Por ocasião da criação ou manutenção das árvores, os seguintes controles são exercidos:

- trata-se de um descritor autorizado;
- existência de uma árvore de nível imediatamente superior;
- não duplicação de um vértice (ou validação por uma pluri-hierarquia interna à estrutura em árvore);
- indicação dos níveis inferiores e validação à frente da supressão de um vértice.

O tesouro pode ser utilizado para a interrogação da referoteca. Ele permite atender à interrogação de todos os documentos indexados com ajuda de TG e TE com relação à questão colocada inicialmente. Por isso, o sistema utiliza o código registrado na

árvore seguido do código do descritor e efetua uma pesquisa sobre todos os descritores situados sob essa árvore: o número de níveis explorados pode ser limitado ou estendido efetuando-se o truncamento sobre o código da árvore.

Nos **arquivos de relações associativas**, a relação entre noções descritas pelos descritores associados pode ser transitiva (similitude) ou intransitiva (semelhança). Na relação transitiva (similitude), a relação entre três termos A, B e C é reflexiva e simétrica, ou seja, é uma transitividade fechada. Na relação intransitiva (semelhança), a intersecção entre os termos é mais fraca e a diferença simétrica entre subconjuntos tomados dois a dois (AB, BC, CA) é mais significativa (a diferença é maior entre A e C que entre AB e BC). Neste caso, há uma transição aberta em função do afastamento semântico.

Já no **arquivo de substituições**, a relação de substituição é utilizada em um certo número de tesouros para indicar o emprego preferencial (TP= termo preferido) de um descritor, com relação a um outro termo (NP= não preferido) que não deve ser utilizado para indexação. Trata-se de sinônimos documentários que não são necessariamente sinônimos como são entendidos geralmente (sobretudo certos termos preferidos são polissêmicos). Por exemplo, a substituição “frequência” por “período” é admissível no caso de um fenômeno vibratório, mas não quando se trata de um período geológico. A relação de substituição não é, do ponto de vista semântico, uma relação de equivalência.

A **tabela de campos semânticos** é um arquivo em que os registros compreendem dois campos: código e conteúdo (teor do campo semântico). Em um sistema multilíngüe, o campo de conteúdo pode aparecer também em outros idiomas.

### 3 Coleta e estruturação de descritores para a elaboração de um tesouro

A coleta do vocabulário em um domínio de especialidade pode ser feita utilizando-se os seguintes recursos:

- consulta a tesouros existentes na área de especialidade e recuperação das noções correspondentes aos níveis mais altos das estruturas em árvore;
- mesma operação com um macrotesouro ou um tesouro de caráter enciclopédico;
  - tabela de assuntos de obras recentes (ou de cursos universitários);
  - transposição de sistemas de classificação (CDU, por exemplo);
  - recuperação das palavras importantes nos dicionários de especialidade;
  - recuperação dos termos freqüentes em uma referoteca.

Os termos utilizados poderão ser:

- unitermos ou palavras simples: agricultura;
- grupos de palavras ou sintagmas: agricultura tropical, estrada de ferro;
- palavras polissêmicas seguidas de um termo entre parênteses: coluna ( pilar cilíndrico) e coluna (vertebral/ anatomia humana).

Para a estruturação deste vocabulário, os termos coletados serão divididos em subconjuntos. Um descritor deve pertencer a um subconjunto e a um somente. Caso pareça necessário incluir a noção que ele representa em dois subconjuntos, significa que ele constitui uma polissemia, que poderá ser discriminada utilizando um termo entre parênteses. Os subconjuntos constituídos representam os campos semânticos que serão designados com a ajuda de um descritor. Cada descritor correspondente aos campos semânticos será descrito sobre uma ficha separada com indicação do domínio e número do campo semântico. A interclassificação alfabética do conjunto de fichas permitirá detectar, em função dos termos entre parênteses, as polissemias eventuais de descritores utilizados nos diversos domínios.

#### 4 Analista morfológico - utilização e construção de tesouros

Pode-se associar o dicionário de um tesouro ao arquivo das bases:

- pelo tratamento da relação de substituição: ao termo não preferido reconhecido a partir da base cria-se uma correspondência à forma semântica do descritor preferido (se é um unitermo) ou o código do termo no campo;
- para o reconhecimento dos descritores sintagmáticos consecutivamente à análise: para isso, cria-se um campo suplementário indicando o código do termo no dicionário do tesouro. Para permitir, a uma forma semântica entrar na composição de diversos sintagmas, pode-se criar segmentos no campo suplementário, ou ainda, utilizar o analista para verificar o dicionário do tesouro e criar a informação necessária no campo suplementário. Esta técnica de reconhecimento de sintagmas, mais simples que aquela que necessita o emprego de um analista morfo-sintático, pode também ser utilizada para o reconhecimento de locuções propositivas (forma semântica vazia).

Se no curso da análise de um arquivo, todas as formas semânticas apontam um mesmo descritor, o resultado de análise pode ser utilizado para introduzir o código do termo ou o próprio termo (segundo a técnica utilizada para a constituição do arquivo invertido) no campo de indexação. O analista pode assim ser utilizado para a indexação automática em linguagem controlada e estruturada (tesouro).

A coleta inicial e a entrada dos termos do tesouro podem ser efetuadas explorando os resultados do analista. Por conta das ocorrências, os unitermos mais freqüentes podem dar uma indicação sobre a estrutura do tesouro. Mas, em um *corpus* dado, os unitermos mais freqüentes não são necessariamente os portadores significativos de informação. Por outro lado, os sintagmas constituem descritores mais característicos. Pode-se colocá-los em evidência no estabelecimento de correlações entre duplas (duas palavras) ou trios (três palavras) aparecendo em uma ordem dada no texto de um certo número de registros de documentos. Os resultados de análise com os termos menos freqüentes podem ser considerados como descritores candidatos.

#### 5 Polissemia e tesouro multilíngüe

A concepção de um tesouro multilíngüe (ou a tradução de um tesouro em outra língua) consiste, em um primeiro momento, em entrar os termos da(s) língua(s) suplementária(s) no(s) campo(s) conveniente(s) do arquivo dicionário. Mas esta operação faz aparecer imediatamente o problema dos homógrafos por ocasião da duplicação de

um termo. Deve-se precisar o sentido dos homógrafos por uma palavra entre parênteses e por um código do campo semântico. Pode-se ainda observar que a terminologia utilizada continua sendo coerente nas diversas relações. Estabelecer um tesauro multilíngüe não consiste unicamente em traduzir os termos do dicionário. Deve ser examinado se cada dois termos utilizados não representam algum dos casos abaixo:

- sinônimo para utilização de duas palavras equivalentes em um contexto dado;
- polissemia-homografia;
- quase-equivalência em uma relação genérico-específica em outra língua;
- uma polissemia apresenta cada um dos termos nas três línguas mas com variações de significação, apresentando orientações diferentes em cada língua. A significação de cada termo deve ser precisada entre parênteses. A multiplicidade de polissemias nas três línguas conduz a relações não transitivas entre os termos de cada uma delas;
- homografia (cadeias de caracteres idênticos em duas línguas) que introduz um cruzamento semântico entre duas noções em duas línguas diferentes.

## 6 Introdução de um tesauro no sistema Basethèque

A estrutura do tesauro implementada pelo sistema Basethèque com o SGBD 4D conta com três arquivos que se relacionam entre si: hierarquia, tesauro e de ligações.

O **arquivo tesauro** é o principal e possui os campos:

- DESC1, DESC2 e DESC3: permitem a entrada de descritores sintagmáticos e o cálculo automático da chave de acesso do descritor. A chave do descritor é calculada automaticamente pela concatenação das 5 primeiras letras de DESC1, as duas primeiras letras de DESC2 e a primeira letra de DESC3 e, eventualmente, a chave especificante. O programa controla a unicidade deste campo;
- chave especificante: duas palavras (*Japon, japonais*) podem conduzir ao cálculo de um mesmo valor de chave que será recusado. Pode-se então acrescentar uma ou duas letras (*ai* para japonês, por exemplo) vindo devolver a chave mais específica;
- descritor: este campo é inteiramente calculado pela concatenação de DESC1 + DESC2 + DESC3. O valor é declarado único a fim de obrigar o tratamento de ambigüidades presentes para tais termos;
- preferência: campo bi-valorado que adota as siglas U (utilizar) ou UP (utilizar para);
- nota explicativa ou sinônimo: permite explicitar uma sigla ou acrônimo, visualizar um sinônimo ou um termo associado;
- chave sinônimo: se, para o termo de entrada, existe um sinônimo, o valor do sinônimo é carregado por seleção automática a partir do **arquivo de ligações**. Basta bater as duas ou três primeiras letras seguidas do símbolo @ do sinônimo para fazer aparecer uma janela de seleção de sinônimos começando por estas letras e clicar sobre o valor conveniente. Os dois sinônimos são então afetados ao mesmo vértice hierárquico.

Ao efetuar uma pesquisa pelo vértice hierárquico, os documentos indexados com um ou outro sinônimo serão recuperados;

- nível genérico (ou nível hierárquico): permite situar a noção a um nível mais ou menos específico sobre uma árvore hierárquica. O valor deste nível, imediatamente superior àquele da noção em curso de entrada, é controlado no **arquivo hierarquia**. Os valores dos vértices sobre uma árvore hierárquica são obtidos e selecionados pelo mesmo método que descrito para os sinônimos;

- nível descritor: cada nível hierárquico é representado por uma fatia de quatro letras separada da precedente por um ponto. A posição sobre uma árvore é calculada automaticamente por concatenação de fatias de quatro letras representando o cimo imediatamente superior e as quatro primeiras letras do descritor (duas primeiras letras de DESC1 e DESC2 se o descritor não é um unitermo). O nível descritor será reportado automaticamente no **arquivo hierarquia**, quando se introduzirá este nível para permitir a entrada de um nível imediatamente inferior no tesauro e iterativamente (repetidamente) para os níveis seguintes. Para os descritores candidatos, atribui-se o nível de valor X;

- posição nível: criação automática de um número de \* (asterisco) correspondente ao nível sobre a árvore hierárquica. Este símbolo não serve para a extensão ou a diminuição automática da pesquisa, mas é utilizado para produzir listas hierárquicas com indentação;

- classificação: permite estabelecer uma equivalência entre a noção registrada no tesauro e seu índice em um outro sistema de representação (CDU, CDD, esquemas de flechas etc.). Isto permite às grandes bibliotecas, cujo código de arranjo é um sistema de classificação, estabelecer uma ligação entre a representação de noções no tesauro e a posição física dos documentos nas prateleiras. O campo nomenclatura desempenha um papel análogo para a equivalência com os códigos de produtos em uma nomenclatura (INSEE, Bottin, Kompass etc.);

- termo associado - poli-hierarquia: sendo duas hierarquias em dois campos semânticos distintos A e B e duas noções pertencentes, uma ao campo semântico A e outra ao campo semântico B, e possuindo entre elas uma semelhança suficiente para associá-las, pode-se estender automaticamente a pesquisa de um lado a outro, qualquer que sejam as duas hierarquias para se efetuar a pesquisa.

## 7 Gestão do tesauro no Superdoc

Adotado para a entrada ou interrogação de bases de dados, trata-se de um vocabulário de dez níveis de hierarquia, gerado independentemente da base ou das bases de dados. Ele gera dois outros tipos de relação: a sinonímia e a associação (tipo “ver também”). Cada termo pode igualmente ser acompanhado de uma nota de aplicação.

Possui quatro arquivos de tamanho variável, cada um contendo:

- o vocabulário e as relações de filiação;
- os ponteiros e o sistema de acesso rápido à informação;
- as notas de aplicação cujo papel é acrescentar uma definição curta ou um breve comentário (limitado a 125 caracteres) a um descritor. A nota de aplicação não é

parte integrante do descritor mas será apresentado durante a interrogação de uma base ou a impressão do tesouro;

- a cadeia alfabética dos termos independente de sua posição hierárquica.

Quanto aos limites do tesouro no Superdoc, têm-se que:

- o número total de termos (descritores e não descritores) do tesouro é limitado a 32.000;
- o tamanho máximo de termos é de 127 caracteres e os espaços são autorizados;
- um termo de nível 1 não dependente de qualquer outro termo é chamado “termo genérico” de nível 1 e é único;
- um termo não genérico de nível 1 é chamado “termo específico” de termos de nível superior;
- um termo, com exceção de termos de nível 1, não pode estar presente em cinco filiações diferentes ou mais;
- os termos do tesouro obedecem às mesmas regras que aquelas utilizadas para as palavras-chave dos léxicos: os espaços não significativos são ignorados e os termos são editados conservando a grafia da entrada;
- um termo genérico pode ter tantos termos específicos sob sua dependência quanto restar de lugar disponível no tesouro.

A sinonímia liga obrigatoriamente um termo descritor introduzido previamente a um ou mais termos não descritores introduzidos automaticamente no tesouro durante o estabelecimento da relação ND (não descritor). Um termo não descritor pode ser ligado somente a um descritor. Um descritor pode ter até 10 relações de sinonímia.

A relação “ver também” é muito útil pois aconselha o usuário do sistema na interrogação da base de dados por um termo. O usuário não entra automaticamente a interrogação da base de dados com o termo “ver também”. Trata-se de um conselho que o usuário aceita ou não. Esta relação pode ser estabelecida de forma recíproca.

## **Interrogação de bases de dados**

### **1 Tipos de bases de dados documentários**

A **base de referências bibliográficas** é uma base de dados tal que, seguida a uma questão expressa com ajuda de palavras-chave, palavras de linguagem natural, nomes de autor, índices de classificação ou códigos de uma nomenclatura, o sistema responde indicando as referências (título do artigo, revista na qual foi publicado, data, páginas, resumo etc.) do(s) documento(s) primário(s) no(s) qual(is) a questão é tratada. A base de referências bibliográficas ou referoteca compreende um conjunto de documentos secundários que fazem referência aos documentos primários nos quais se encontram o texto integral.

Já as **bases de dados factuais** (ou cadastrais: diretórios de instituições e pessoas) fornecem diretamente a informação pesquisada. Tal é, por exemplo, o caso de certas bases de dados nas quais se pode encontrar as características de uma empresa da qual se conhece o nome (endereço, nome dos dirigentes, produtos fabricados etc.) ou da qual se obtém o nome pesquisando todas as empresas que fabricam um produto definido. Este também é o caso de certas bases de dados jurídicas ou de jornais que dão acesso ao texto integral.

Por definição, as bases de referências bibliográficas são textuais e as bases de dados factuais podem compreender texto e dados numéricos, muitas vezes, sob a forma de tabelas.

Enfim, existem os **bancos de dados estritos** (séries econômicas, constantes físico-químicas, estruturas químicas etc.) que podem eventualmente associar *softwares* de cálculo permitindo, por exemplo, testar diversas hipóteses para fazer funcionar um modelo, simular um ensaio, produzir histogramas etc.

A obra trata essencialmente das bases de dados bibliográficas e factuais por acreditar que a utilização de bancos de dados estritos é reservada a uma categoria de usuários cuja competência altamente técnica e especializada ultrapassa o quadro de atividade tradicional do documentalista.

## 2 Bases de dados difundidas pelos servidores

Os servidores são organismos públicos ou privados, dispendo de computadores potentes que difundem, mediante taxa, bases de dados bibliográficos e factuais produzidas por diversos organismos públicos ou privados. O servidor reverte aos produtores *royalties* em função do número e da duração de interrogações efetuadas pelos organismos terceiros, os clientes.

Para tornar possível a interrogação destas bases, o servidor deve dispor de um *software* munido de uma linguagem de comandos que permite as operações de:

- formulação de noções pesquisadas com ajuda de operandos que são as palavras-chave em linguagem natural ou os descritores (unitermos ou sintagmas) em linguagem controlada (algumas vezes, a partir de um léxico visualizável);
- combinação booleana de operandos com ajuda de operadores lógicos (E, OU, NÃO);
- combinação eventual de operandos com os operadores de distância (palavras de adjacência na ordem dada e palavras presentes na mesma frase ou parágrafo de uma referência);
- pesquisa de uma cadeia de caracteres no texto (título ou resumo), por exemplo para a pesquisa de compostos químicos, muitas vezes com ajuda de sinais para truncamento;
- visualização de resultados *online*: pode-se obter a visualização de todas as referências ou somente de um número definido e sob diversas formas de edição;
- impressão de resultados *offline*;

- salvamento de pesquisas prévias e possibilidade de executá-las posteriormente, a partir da mesma base ou outra. A re-execução periódica de pesquisas prévias corresponde a um perfil documentário que permite realizar a função de DSI.

### 3 Interrogação guiada por menus

Os *softwares* de interrogação de bases difundidas por servidores apresentam como característica comum linguagens de comandos relativamente complexas. Assim, seus usuários necessitam de treinamento prévio e de uma utilização freqüente. Estes *softwares* são pouco propícios à banalização do acesso à informação por um grande número de usuários que não tenham recebido um treinamento especializado. Em geral, trata-se de engenheiros, pesquisadores e técnicos. Estes usuários não são especialistas das técnicas documentárias, não praticam o uso dos *softwares* de interrogação e não conhecem a estrutura da base de referências que vão interrogar.

Levando em conta estas hipóteses sobre os usuários, o *software* deve, entre outros, ser auto-explicativo e propor escolhas simples com ajuda de menus e utilizar um mínimo de teclas bem localizadas em uma zona do teclado.

Pesquisa guiada por menus – *software* Eureka:

Trata-se do *software* de interrogação adotado pelo serviço de documentação da empresa Merlin Gerin, desde 1982, para acesso a várias bases de dados.

O usuário pode conduzir sua pesquisa de três maneiras diferentes: acesso hierarquizado (das noções genéricas às noções mais específicas), acesso por consulta ao léxico e acesso por consulta à lista de autores.

Como resultado de sua pesquisa, o usuário se vê diante de dois modos de visualização:

- rápida: título e referências do documento;
- máxima: autores, título, referências, termos genéricos de indexação e resumo;

Em seguida a esta primeira pesquisa, o usuário pode estender a pesquisa a noções vizinhas daquelas tratadas no primeiro subconjunto de documentos isolados ou salvar a pesquisa para re-executá-la periodicamente na base.

O **acesso hierarquizado** permite ao usuário ter conhecimento do conteúdo da base percorrendo as noções mais genéricas até as noções mais específicas, que é a própria estrutura do tesauro que serviu para indexar os documentos. Este acesso é necessariamente mais longo, mas é mais sinótico que o acesso pelo léxico alfabético. Quando o usuário termina de percorrer as telas de acesso hierárquico, o sistema registra os temas selecionados (12 no máximo) e propõe de os combinar com ajuda de operadores booleanos E ou da seguinte forma:

- carrega sobre uma mesma linha todos os números de temas a serem combinados segundo o operador OU (reunião);

- carrega sobre linhas diferentes todos os números (ou uma reunião de números sobre uma mesma linha) de temas a serem combinados segundo o operador E (intersecção).

O **acesso alfabético** pelo léxico utiliza uma lista constituída alfabeticamente por termos genéricos, palavras-chave específicas e unitermos vindo do analista.

Os termos genéricos TG ou palavras-chave específicas TE são os termos controlados do tesouro para indexar os documentos. Eles podem ser constituídos de palavras simples (Brasil), palavras complexas ou sintagmas (América do Norte) e palavras compostas (caneta-tinteiro).

Os unitermos são constituídos por cadeias constituídas de caracteres. São obtidos pela recuperação, com ajuda do analista morfográfico, de todas as palavras significativas (exceto as palavras vazias como artigos, preposições etc.). A partir dos termos apresentados acima, se obtém os unitermos: Brasil, América, Norte, caneta, tinteiro. Graças ao analista morfográfico, os substantivos e os adjetivos são transformados na forma lexical masculino-singular, permitindo reduzir consideravelmente as entradas no arquivo invertido.

A pesquisa pelo léxico é mais rápida que a pesquisa hierárquica e permite uma pesquisa mais extensa por meio da ajuda de palavras significativas do título e do resumo. Por outro lado, ela não oferece o mesmo acesso sinótico pelas noções suficientemente gerais. Para amenizar este inconveniente, o usuário pode, depois do resultado da combinação booleana de termos selecionados a partir do léxico, estender sua pesquisa aos temas (TGs) correspondentes às palavras-chave específicas (TEs) selecionadas.

No **acesso por um autor**, o princípio da pesquisa é o mesmo que aquele do acesso alfabético pelo léxico. O usuário pode pesquisar um só autor, combinar dois ou mais autores ou combinar o autor e o nome de sua empresa ou instituição. Depois do resultado desta combinação, ele é conduzido à tela na qual se escolhe um modo de visualização.

Qualquer que seja o modo de visualização escolhida, o usuário pode limitar a seleção de documentos a uma data igual ou posterior àquela da entrada na base. Esta possibilidade de limitação no tempo é interessante para a re-execução periódica de um questionário salvaguardado. Realiza-se assim uma difusão sobre perfil adaptado a cada usuário, permitindo-lhe conhecer a cada mês as novidades em seu domínio.

O subconjunto de documentos selecionado pelo usuário, seja pela via hierárquica lexical, compreende um certo número de termos de indexação TGs ou TEs, a partir dos quais ele pode estender sua pesquisa através de noções associadas tratadas nos documentos detectados por sua questão inicial. Ele pode também partir para a relação de vizinhança para descobrir documentos tratando de noções que ele não havia pensado de início.

Pesquisa guiada por menus – *software* Basethèque:

Realizado com o SGBD 4D, o *software* Basethèque utiliza mais amplamente as propriedades de menus rolantes da Macintosh por permitir ao usuário expressar facilmente seus critérios de pesquisa com ajuda de telas de diálogo.

O usuário dispõe de seis menus de pesquisa: temas genéricos, palavras-chave (descritores específicos), autores, cadeias de caracteres de títulos ou resumos, artigos publicados em uma revista, pessoas (autores de documentos diversos e de audiovisuais) e catálogos de fornecedores.

Na pesquisa por **temas genéricos**, é possível ter uma idéia rápida da natureza da coleção: saber se há documentos que tratam do tema pesquisado e, se há, quais são. Isto porque, cada documento pode ser registrado com até três grandes temas. Este tipo de pesquisa deve ser executado antes da pesquisa de uma cadeia de caracteres do resumo, para fazer uma seleção prévia de documentos.

A pesquisa por **descritores específicos** pode optar pela forma hierárquica ou alfabética e por operadores E e/ou OU. Nela, também é possível realizar a pesquisa extensiva a qual inclui a busca de documentos indexados por descritores mais específicos que aquele expresso, além daqueles indexados com ajuda de descritores sinônimos.

A pesquisa hierárquica permite aumentar ou precisar uma pesquisa. É possível, por exemplo, encontrar todos os documentos referentes à América do Norte sem ser obrigado a lembrar todos os países deste continente.

Na **pesquisa por cadeias de caracteres de títulos ou resumos**, a tela de diálogo permite selecionar a cadeia de caracteres por chave de acesso a todos os documentos indexados por determinado tema. No caso do resumo, por não se tratar de um campo indexável, o programa percorre seqüencialmente todos eles para localizar a cadeia de caracteres procurada, tornando-se assim, uma operação mais longa. Por isso, sugere-se que esta pesquisa seja adotada em documentos pré-selecionados. Esta forma é particularmente útil em domínios onde não é fácil constituir um tesouro completo (como a química orgânica, por exemplo).

Na **pesquisa de artigos publicados em uma revista**, pode-se buscar por aquelas de um ano ou de um mês preciso.

Na **pesquisa de pessoas** (autores, músicos, compositores etc.), pode-se localizá-los pelo índice de autores.

Na pesquisa de produtos e catálogos (**catálogos de fornecedores**), é possível encontrar o conjunto de fornecedores fabricantes de um produto ou prestadores de um serviço, assim como aqueles de uma família de produtos ou serviços. Também permite listar os catálogos que se possui sobre tal produto e imprimir uma seleção de fornecedores.

A **pesquisa direta nos arquivos** permite maior variedade de opções de busca pois, apesar de o programa Basethèque cobrir de forma exaustiva as necessidades de usuários, não pode prever todas elas. Para tanto, o usuário deve escolher o arquivo desejado na janela de seleção que lhe é apresentado em forma de lista. A partir desta forma de pesquisa, é possível, por exemplo, buscar por idioma da publicação que não é um campo indexado.

Os modos de interrogação para a versão do Superdoc em estudo são: modo guiado, por máscara de entrada de critérios nos diversos campos, por perfil pré-registrado, em linguagem documentária (descritores e operadores booleanos) e em linguagem natural.

Assim como na entrada de dados, listas de autoridade e o tesouro são disponíveis no Superdoc como ferramentas de ajuda á interrogação.

Além disso, o usuário dispõe dos seguintes recursos:

- uma tecla de função que permite conhecer os campos interrogáveis da base de dados, ou seja, a lista de campos indexados e visíveis, assim como, sua ligação com as listas de autoridade ou com o tesouro;
- uma função de edição de palavras-chave dos léxicos organizados alfabeticamente com o número de ocorrências e o número de registros onde elas aparecem;
- os operadores lógicos E, OU e NÃO;
- o operador de truncamento à direita;
- os operadores de comparação aritmética < e > utilizados quando os critérios são de tipo numérico (data, código, preço, número de referência etc.);
- a expressão lógica obtida poderá incluir, entre outros, cinco níveis de parênteses.

A interrogação por palavras-chave truncadas à direita é necessária quando o usuário tem uma dúvida sobre a finalização de uma palavra-chave. É evidente que o tempo de resposta de toda pesquisa documentária utilizando as possibilidades de interrogação por meio de descritores truncados à direita será diretamente proporcional ao número de descritores do léxico que responde ao critério de truncamento. Em consequência, o operador de truncamento não deve ser utilizado sistematicamente pois a consulta do léxico é assim uma forma de levar à ambigüidade e os melhores tempos de resposta e pertinência são obtidos a partir de descritores não truncados.

A sintaxe das questões obedece às regras da lógica formal clássica. Assim, por exemplo:

(arte **E** clássico) **OU** (amor **E** moderno): para as obras clássicas tratando da arte assim como as obras modernas tratando de amor, enquanto que, para:

(arte **E** amor) **E** (clássico **OU** moderno) **NÃO** Verlaine: para todas as obras tratando por sua vez de arte e de amor que sejam clássicas ou modernas, salvo aquelas que se referem ao poeta Verlaine.

Sobre toda seleção obtida depois de uma interrogação, pode-se novamente aplicar critérios lógicos para refinar a pesquisa, mas também utilizar um critério de pesquisa no texto integral. Com a pesquisa final, uma seleção manual é colocada à disposição do usuário.

As saídas sobre impressora são possíveis em modo padrão ou com o uso dos formatos de impressão, por meio de um módulo de criação de formatos inteiramente

parametrizáveis. As saídas em arquivos são do tipo ASCII e de fácil importação em todos os tratamentos de textos modernos funcionando em MS-DOS ou Windows.

## 5 Critérios de performance

As performances de um *software* de interrogação podem ser avaliadas segundo os seguintes critérios:

- **pertinência:** qualidade que apresenta as referências extraídas ao responder exatamente ao questionário estabelecido com ajuda de palavras-chave, descritores ou unitermos;
- **taxa de revocação:** relação entre o número de referências pertinentes extraídas e o número total de referências pertinentes existentes na base de referências interrogada;
- **ruído:** proporção de referências não pertinentes extraídas com ajuda de um questionário;
- **silêncio:** número de referências pertinentes perdidas (ou faltantes), apesar de existirem na base de referências.

O ruído e o silêncio podem resultar de uma indexação ruim, e então, não é mais o *software* de interrogação que está em causa. Daí vem o interesse do usuário nos unitermos da linguagem natural, com o fim de evitar os silêncios. A experiência mostra que não é raro recuperar, pela combinação de unitermos (extraídos de títulos e resumos), o descritor que deveria estar presente no campo de indexação.

Para combinar a saída de unitermos, efetua-se a operação booleana com o conector E. Mas, certos *softwares* (como STAIRS, BRS, Questel, Dialog) oferecem a possibilidade de pesquisar dois termos não somente no mesmo documento, mas a uma certa distância entre eles (expressa em números de palavras que os separam) ou na mesma frase ou parágrafo e em uma ordem dada. Isto permite adicionar precisão à pesquisa evitando recuperar documentos nos quais os dois termos estão presentes sem uma ligação semântica real. Possibilita a distinção de, por exemplo, “carros a gás” ou “gás liberado por carros”.

Os *softwares* diferem entre si pela linguagem de comandos (conjunto de palavras de comando disponíveis ao usuário) e pelas mensagens colocadas pelo sistema. Mas a maior diferença entre eles está nas possibilidades oferecidas no quadro da função pesquisa, em particular com relação à possibilidade de poder selecionar termos a partir da visualização do léxico e/ou de utilizar operadores de distância. Os *softwares* que oferecem estes recursos são os que possibilitam melhores resultados. A performance do sistema de interrogação está estreitamente ligada às modalidades de interrogação autorizadas pelo *software* em função da organização dos arquivos invertidos.

## 6 Teletransferência (*downloading*) e importação de dados

O *downloading* permite salvar o resultado de uma sessão de interrogação de um banco de dados difundido por um servidor.

O resultado de uma pesquisa visualizada em tela ou impressa com ajuda de uma impressora exige um tempo de conexão proibitivo. Assim, é preferível que o resultado seja registrado em um disquete ou um disco “duro” de microcomputador.

O resultado da pesquisa pode ser:

- **impresso** a partir de uma impressora;
- **modificado** com ajuda de um *software* de tratamento de texto para melhorar a apresentação e eliminar as linhas inúteis;
- **importado** para um SGBD documentário a fim de que as referências possam ser encontradas com os mesmos procedimentos de pesquisa deste SGBD para as referências introduzidas diretamente no local.

## 7 Importação para um SGBD documentário

Um problema para a importação de dados refere-se à formatação dos arquivos originais ser distinta da formatação do arquivo destinatário, o que acontece na maior parte dos casos.

A outra dificuldade é aquela ligada à presença de campos que podem conter mais valores. Este é particularmente o caso do campo “autor” e do campo “descritores”, por conta das formas distintas utilizadas para separação entre as ocorrências (por vírgula, espaço, barra, ponto e vírgula). No campo fonte do artigo de periódico, por exemplo, estas diferenças são ainda maiores entre os servidores. Estas diferenças levam à impossibilidade de organização alfabética do arquivo invertido.

Importação de dados para o SGBD relacional Basethèque:

Em um SGBD relacional, as dificuldades de reformatação são ainda maiores em função dos diferentes valores (escritos de forma distinta) de cada campo nos diversos arquivos como: autores, editores, periódicos, tesouro etc.

Para evitar todo este trabalho de reformatação, o Basethèque permite criar uma ficha para interrogação, onde os campos são utilizados da seguinte forma:

- título: objeto da pesquisa podendo ser expresso de forma mais explícita e detalhada;
- autor 1: o demandador da pesquisa;
- autor 2: o documentalista que efetuou a pesquisa;
- editor: o nome do servidor;
- coleção: o nome do banco de dados interrogado;
- data: da interrogação;
- descritores: utilizados para efetuar a interrogação.

O resultado da pesquisa pode ser introduzido em um ou mais campos de tipo texto (32.000 caracteres), tal qual o do resumo, depois da eliminação das referências não pertinentes.

Dada a simplicidade de entrada, esta solução apresenta as seguintes vantagens:

- permite ao usuário as facilidades oferecidas pelos procedimentos padrão de pesquisa pelos menus rolantes do Basethèque;
- autoriza a pesquisa, na zona de texto utilizada para importar as referências do servidor com ajuda do operador “contém”, que retorna ao pesquisador uma cadeia de caracteres;
- representa economia de espaço de disco pois não utiliza mais que um registro podendo armazenar muitas dezenas de referências, o que minimiza o problema das duplicatas;
- potencializa a adequação do tesouro com os descritores utilizados nas bases mais freqüentemente utilizadas.

O conteúdo do campo assim criado pode ser tratado com um programa de tratamento de texto para permitir: o enriquecimento da cadeia de caracteres (minúsculos, maiúsculos, caracteres acentuados); a diferenciação de diversas zonas do texto pela determinação de estilos e de fontes de letras; e a eliminação das mensagens de serviço que se repetem.

Importação de dados no SGBD Superdoc:

Para proceder ao carregamento de um arquivo texto em uma base Superdoc, deve-se dispor:

- de uma estrutura memorizada de base Superdoc;
- de um arquivo texto ASCII obtido, seja por meio de um editor ou de um tratamento de texto, seja por meio de um *software* de comunicação entre um microcomputador e um servidor de informações;
- de um arquivo de parâmetros fixando as regras de reconhecimento de informações no arquivo texto fonte e a estrutura de recepção Superdoc.

Um utilitário é utilizado para introduzir em uma base Superdoc as informações contidas em um arquivo de tipo texto ASCII respondendo às especificações mínimas de organização e de estrutura.

O arquivo texto deve conter cadeias de caracteres de no máximo 252 caracteres por linha, sendo todos imprimíveis (alfabeto maiúsculo e minúsculo, pontuação, números e espaços); os caracteres não imprimíveis são filtrados pelo Superdoc.

Para se comunicar com uma base Superdoc, deve-se poder identificar no texto aquele que corresponderá a um registro Superdoc e identificar, no interior deste registro, aquele que corresponderá a cada um dos campos Superdoc.

O delimitador de registros é uma cadeia de caracteres com até 37 caracteres e sempre enquadrada à esquerda no início de uma linha sobre a tela. Quanto este

delimitador estiver ausente no arquivo, o primeiro nome de campo citado no arquivo de parâmetros servirá, não somente para designar um campo, mas também para assinalar o início do registro. Isto implica que todos os registros devem obrigatoriamente começar por este campo. Por outro lado, se há um delimitador de registro, a ordem dos campos no interior de um registro é independente. Um campo pode ocupar tantas linhas quantas necessárias, as quais devem ser consecutivas e não conter o identificador de campos no seu início.

Cada campo é identificado pelos 16 primeiros caracteres de seu nome, no máximo. O conteúdo do campo é precedido por um delimitador de nome de campo. Se este delimitador estiver ausente, deve-se verificar se não há ambigüidade no texto para encontrar os títulos dos campos. Normalmente, o conteúdo de cada campo do arquivo texto é introduzido em um campo Superdoc designado por um número. O tratamento do conteúdo é diferente segundo o campo Superdoc ser indexado ou não, pois há delimitadores de palavras-chave para o primeiro caso.

## 8 Telecópia (fac-símile) de documentos primários

Trata-se do fornecimento *online* de documentos primários aos quais fazem referência diversos bancos de dados. É o caso, por exemplo, do objeto do projeto TRANSDOC realizado no âmbito da Comunidade Européia e que agrupa cinco fornecedores (CNRS, EDF, INPI, Télésystèmes-Questel, CEE). Os usuários consultam os bancos de dados implantados no servidor Télésystèmes podendo encomendar *online* os documentos selecionados que lhe serão enviados por telecópia (fac-símile).

Já a digitalização de milhões de páginas de diferentes formatos possui o problema de ordem jurídica (*copyright*).

## **Guia de especificações para uma aplicação informática**

Uma definição rigorosa da aplicação constitui uma prévia a toda implantação de arquivo, atividade de programação e escolha de equipamento. Pois o caminho lógico consiste em, primeiramente, definir corretamente os objetivos e não em examinar depois os meios (inclusive as necessidades de equipamento).

Uma análise correta do problema compreende, em particular, uma análise da existência e uma descrição precisa dos objetivos e resultados que se procura obter. Os objetivos e resultados pesquisados devem ser definidos em um dossiê compreendendo o conjunto de especificações da aplicação. O dossiê deve ser preparado entre o demandador da aplicação e o programador que fará a aplicação; este diálogo é indispensável para a elaboração correta do programa segundo as necessidades dos usuários.

Em princípio, um dossiê deve ser estabelecido independente da limitação tecnológica, sem preocupação com equipamentos disponíveis e suas condições de exploração. Em seguida, examinam-se as condições de viabilidade da aplicação em função destas limitações tecnológicas.

As condições de exploração da tecnologia podem variar em função do domínio do usuário sobre os meios utilizados. Alguns aspectos a serem considerados são:

- serviço informático com um ou mais computadores e pessoal da área (como programadores e outros);
- computadores em tempo compartilhado onde cada usuário coloca suas próprias aplicações e a gestão e a manutenção do sistema são asseguradas por uma equipe central reduzida;
- microcomputador onde o usuário coloca suas aplicações e gera o sistema.

A implicação do usuário no processo informático é inversamente proporcional aos meios colocados a sua disposição. No limite, ele mesmo deve efetuar a análise funcional, a análise orgânica e a programação.

As conclusões do estudo do dossiê devem levar à escolha do *software* mais adequado já existente no mercado. A utilização de um estudo deste tipo permite não esquecer o essencial em um catálogo de especificações.

### **Custos de uma aplicação informática**

A fim de efetuar uma escolha racional do equipamento e dos *softwares* que serão utilizados, convém fazer uma previsão de custos para a aplicação e suas condições de rentabilidade. Devem ser examinados os custos de realização da aplicação, de exploração da aplicação e de preparação dos dados de entrada.

Para os **custos de realização** que representam um investimento, devem ser consideradas as regras de amortização do investimento de *softwares*, além da questão sobre a elaboração ou aquisição de um *software*.

Os **custos de exploração** podem ser considerados como uma amortização de custos fixos ou de custos diretos variáveis segundo as soluções adotadas.

Os **custos de preparação** dos dados de entrada são essencialmente os custos diretos variáveis.

O exame destas três linhas de despesas deve conduzir à justificativa da rentabilidade do projeto.

### **Escolha de soluções**

Critérios de escolha de um equipamento:

O critério principal é a pesquisa de um equipamento capaz de satisfazer as especificações do dossiê da aplicação informática. Uma vez determinado este equipamento, o segundo critério é aquele das condições de exploração (*softwares*

implantados, assistência técnica, manutenção). Enfim, determina-se a melhor solução possível em função do orçamento que se dispõe.

Critérios de escolha de um *software*:

Antes de se lançar na programação da aplicação, convém pesquisar se certos *softwares* não respondem no todo ou em parte às especificações do dossiê da aplicação informática. Em caso afirmativo, deve-se examinar em que medida eles são implantáveis sobre o material que se dispõe. Os principais critérios de portabilidade de um *software* sobre um novo computador hospedeiro são:

- o computador hospedeiro aceita a linguagem na qual o *software* foi criado;
- o *software* é compatível com o sistema de exploração do computador hospedeiro;
- o *software* não utiliza elementos ausentes da configuração do computador hospedeiro;
- a capacidade de memória da unidade central é suficiente para a exploração do *software*.

Para realizar esta escolha, existe um certo número de *softwares* que foram desenvolvidos especificamente para as aplicações documentárias e de gestão de biblioteca.

Entre os mais conhecidos *softwares* documentários podem ser citados: BASIS, CDS-ISIS, MISTRAL, STAIRS, MINISIS e Texto. Há também aqueles utilizados pelos servidores BRS, Questel, Dialog e ORBIT. Os *softwares* compreendem, em geral, um certo número de módulos funcionais.

O programa BASIS é provavelmente o mais completo, apesar de ser também o mais caro e o que mais utiliza a capacidade da unidade central. Inclui os módulos: entrada de dados em tratamento por lotes; entrada de dados *online*; seleção; gestão e controle do vocabulário (tesauro); DSI; cálculo; estatísticas de uso da base de dados; definição e edição de relatórios. Sobre o MISTRAL, cita-se que contém um módulo para a gestão de tesauro.

Para os *softwares* de gestão de biblioteca, mais particularmente orientados para a gestão de aquisições, a catalogação e o empréstimo, existe um certo número adotando o formato bibliográfico MARC. Trata-se de sistemas de grande porte gerando milhões de documentos e exigindo meios informáticos potentes, em particular para a exploração em rede com dezenas ou centenas de terminais. Entre eles:

- o sistema OCLC nos Estados Unidos;
- o sistema SIBIL (Système Intégré pour les Bibliothèques Universitaires de Lausanne) usado pela maior parte das bibliotecas na Suíça e na França;
- o sistema BLAISE (British Library Automated Information Service);
- o sistema MONOCLE adotado pela Bibliothèque National de Paris, possui o mesmo nome do formato que utiliza, o qual é baseado no Formato MARC. O Formato MONOCLE foi desenvolvido por Marc Chauveinc, responsável pela Bibliothèque National de Paris. Este formato também é adotado no sistema SIBIL.

Para o desenvolvimento de uma aplicação, pode-se examinar se certos *softwares* e sistemas de gestão de arquivos, não especialmente destinados a aplicações documentárias ou de biblioteca, não dão conta destas necessidades, além de resolver outros problemas da empresa ou da instituição. Tais *softwares* ou módulos permitem evitar longas e custosas jornadas de programação (notadamente pela possibilidade de desenvolver aplicações com ajuda de procedimentos catalogados). Eles permitem também utilizar um produto cujas potencialidades estão claras e a evolução é assegurada pelo produtor.

Considerações de rentabilidade:

Coloca-se a questão sobre fazer um programa adequado à instituição ou adquirir um pronto. Tratando-se de Informática Documentária, o problema pode ser visto da seguinte forma:

- a própria instituição desenvolver a aplicação (ou solicitar a uma empresa de Informática que o faça) e esperar por meses ou anos;
- adquirir um *software* documentário funcionando já em dezenas (às vezes centenas ou milhares) de instituições e projetos, livre de demandar qualquer adaptação.

A partir dos *softwares* apresentados nesta obra, ou seja, a base de dados de Merlin Gerin, desenvolvido em parte com o SGBD Power, o *software* Basethèque e o Superdoc, o autor avalia tempo, mão-de-obra e recursos financeiros, demonstrando a viabilidade da segunda opção, ou seja, a aquisição de um *software*. Segundo os dados apresentados, dentre os bons *softwares* disponíveis no mercado (especialmente, o francês) e amplamente utilizados, os custos são, no mínimo 20 vezes menores, sem contar a relação custo-benefício envolvido no tempo gasto com programação e elaboração da documentação técnica.

## **Perspectivas**

Depois de 30 anos de serviço direto com usuários finais, o autor constata que interessa não apenas obter as referências de documentos pertinentes, mas os próprios documentos e ainda, a informação precisa que os usuários necessitam.

Este objetivo será atendido pelas técnicas de armazenamento de documentos primários. O acesso à informação contida nestes documentos poderá ser possibilitado com a utilização de técnicas de marcação de textos empregado por um analista especialista.

Entre esta categoria de usuários, há também aqueles que pesquisam dados factuais ou dados estritos (estatísticos, constantes físico-químicas, resistência de materiais etc.), muitas vezes associados a modelos matemáticos. Estes sistemas, adequados para pesquisadores de alto nível e usados somente por especialistas, saem do quadro das aplicações gerais da Documentação. Mas estes especialistas têm também recorrido à Documentação para saber o que se faz de novo em seu domínio, conhecer as comunicações presentes nos diversos congressos etc.

Para tanto, os sistemas documentários devem ser acessíveis e de fácil utilização. Isto supõe:

- o aperfeiçoamento de *softwares* de pesquisa guiados e de navegação por listas de conteúdo;
- a utilização de equipamentos reduzindo ao máximo o emprego do teclado pela utilização do *mouse*, telas táteis, recursos interativos etc.

Estes aperfeiçoamentos necessitarão de pesquisas e de investimentos consideráveis, sendo difícil mensurar a rentabilidade com relação aos sistemas atuais.

Mas, além da restrita rentabilidade da aplicação informática, convém não negligenciar seus efeitos induzidos sobre os desenvolvimentos futuros graças à abertura e à familiarização de usuários com a Informática.

E, por fim, com relação à rentabilidade da aplicação dos problemas documentários e da informação científica e técnica, será preciso argumentar a favor da rentabilidade da própria informação científica e técnica. Vasto debate sobre a finalidade da informação científica e técnica não foi contemplado nesta obra, mas a Informática Documentária constitui elemento poderoso para contribuir para esta finalidade. Este foi o propósito do autor.

## **COMPARAÇÃO ENTRE O TERMO *INFORMATIQUE DOCUMENTAIRE* E OS TERMOS *LIBRARY AUTOMATION* E *INFORMATION RETRIEVAL***

Assim como nas obras anteriormente analisadas, também foi possível comparar diferentes edições do livro de Deweze. Foi apresentada aqui, a quarta edição de 1994 que, comparada com a segunda edição publicada em 1986, acrescentou: os SGBDs, as bases de dados relacionais (antes apenas citada), o hipertexto, as redes de computadores e a telemática, além de alterações no capítulo sobre informatização de biblioteca. Também agregou outros temas como discos óticos digitais, CD-ROMs, GED, norma SGML e multimídia. Ao que consta, a estrutura básica do livro permaneceu a mesma nas diversas edições, em especial, na sua essência baseada na noção de cadeia documentária, apenas acrescida de novas tecnologias de suporte e de tratamento dos dados.

E é justamente o aspecto citado acima que aponta para o caráter diferencial deste livro que inicia por tópico que discorre sobre o ciclo documentário e não pela descrição de computadores e aspectos de tecnologias de informação em geral. Assim, os recursos informáticos são definidos em função do ciclo documentário, incluindo a gestão do acervo realizada pela biblioteca. Ou seja, quando trata das funções e partes do computador, discorre desde as características técnicas da máquina como a representação da informação (códigos binários, sistema hexadecimal), suportes de registro da informação (fita magnética, cartão perfurado) até os tipos de acesso à informação, tratamento e modos de recuperação. Da mesma forma, o item sobre produção de textos pelo computador inclui a impressão e procedimentos de composição até a produção de boletins de índices para pesquisa documentária. Por outro lado, as obras anteriormente analisadas, focam a gestão automatizada do acervo e o atendimento aos usuários baseado no acesso a bases de dados por acesso remoto ou em CD-ROM (como em Saffady). Ou ainda, centram-se na gestão do acervo e nos sistemas de recuperação da informação que é, no mínimo, uma abordagem mais abrangente mas também fragmentada (como em Rowley). De fato, o principal da obra de Deweze com relação às

de origem no idioma inglês é que, diferente destas, não trata em separado dos fundamentos de Informática para depois abarcar a gestão e recuperação de documentos e de informação, mas antes, focado na cadeia documentária de produção de documentos, tratamento, armazenamento, recuperação, disseminação e uso, inclui os recursos tecnológicos que tornam possíveis estes processos.

Pode-se concluir que, segundo o modelo francês apresentado, a existência e o uso de bases de dados externas a um centro de documentação não prescindem do tratamento de conteúdo dos documentos eletrônicos ou disponíveis localmente pelo centro. As atividades destes centros denotam a institucionalização dos serviços, o que, por sua vez, compreendem uma aproximação do processo comunicacional da informação, já que promovem a contextualização da mesma. O modelo francês, calcado na noção de cadeia documentária, também reforça a abordagem comunicacional, como se disse, até na forma de discorrer sobre as tecnologias envolvidas.

No entanto, é fato que os modelos analisados de origem inglesa e espanhola (a exceção de Saquel), discorrem sobre a indústria da informação, entendida como o ciclo do mercado composto pelos produtores, armazenadores e consumidores da informação. Esta visão sobre o mercado da informação não inclui os elementos teóricos e metodológicos envolvidos na produção, tratamento e disseminação da informação que contemplam a noção de cadeia documentária de Deweze. A análise em idioma inglês apresentou a idéia da indústria da informação, mas no idioma francês, o ciclo da informação é tratado sob outro prisma, qual seja, o da cadeia documentária ou fluxo documentário. Esta cadeia inclui as noções de produção, aquisição, tratamento, armazenamento, recuperação e disseminação da informação e assimilação pelo usuário. Este conjunto de noções não está centrado na capacidade mercadológica de produção e consumo de informação, mas inclui estes aspectos mais as questões envolvidas com a natureza do conteúdo e com o perfil e necessidades do usuário. Uma das abordagens em espanhol aqui apresentada apresenta os dois modelos.

No entanto, deve-se considerar a existência de distinções sutis antes de partir para delimitações rígidas. De fato, há preocupações com a natureza da informação e com as características comuns de grupos de usuários no modelo em idioma inglês, mas a discussão destes aspectos é realizada de forma segmentada, funcionalista e pragmática (assim como com outros aspectos analisados na distinção destas culturas). Desta forma, a essência fica encoberta pela eficiência, pois aquela se justifica principalmente pelo aprimoramento de processos para a realização efetiva de operações.

Em outra demarcação entre países e idiomas, Deweze aponta para diferenças na concepção de *softwares* segundo as peculiaridades idiomáticas, ao afirmar que os *softwares* de tratamento documentário construídos para suportar idiomas anglo-saxões são mais pobres de recursos que aqueles que tratam textos em idiomas latinos. Isto porque, os idiomas de origem latina possuem alfabetos mais sofisticados e então exigem recursos tecnológicos complexos para o tratamento documentário. Isto é visível nos itens em que o autor descreve as formas de alfabetação e de criação de dicionários de palavras vazias.

Nos tópicos sobre indexação, elaboração de resumos e concepção e elaboração de tesauro, são examinados o que são estes processos para então avaliar qual deles e de que forma são informatizáveis. A história da Automação de Bibliotecas apresentada também cita esta etapa (que supostamente não foi rápida) de definição de operações e investigação sobre sua automação. No entanto, apenas a obra de Deweze discorre, sob este ponto de vista, sobre as operações documentárias por excelência, em oposição às

operações unicamente de cunho administrativo. Como estas operações não estão relacionadas diretamente à gestão de suportes físicos mas à representação do conteúdo informacional neles contido, a explicitação destas atividades exige maior esforço intelectual, tanto quanto sua transposição para o ambiente informático.

Importa ressaltar que, outros trabalhos de Deweze tratam dos aspectos semânticos da análise automática de documentos, ou seja, a produção deste autor parece estar baseada na interface entre Informática Documentária e Linguística Documentária. Os bancos de dados reforçaram a necessidade do estudo e uso das linguagens documentárias, o que supostamente contribuiu para a constituição da Informática Documentária.

Quando o autor trata da importação de dados para bases de dados locais, diferente do enfoque da Automação de Bibliotecas, não resalta a cooperação catalográfica nem a padronização de procedimentos para tal. Foca as atividades de importação de dados a partir de bases de servidores e as dificuldades relacionadas à reformatação e tratamento dos dados para a base local. Apenas cita o formato bibliográfico MARC ao tratar dos *softwares* para gestão de bibliotecas do mercado e das redes de bibliotecas na Europa e nos Estados Unidos. Para o autor, a falta de padronização parece não ser o maior problema das bases de dados de centros de documentação ou de serviços de informação *online*, uma vez que, neste campo, interessa mais a adequação da representação de informações segundo o segmento de público a que se destina que, neste caso, é bem mais diversificado que o público de bibliotecas universitárias de um país, por exemplo. No entanto, não se discute se alguma compatibilidade mínima entre estas bases de dados documentários é possível e desejada. Assim também, na produção de literatura sobre Automação de Bibliotecas, as bases de dados documentários, sua produção e suas características não são discutidas, sendo apontados somente seu uso nos serviços de referência. Este fato indica a distinção de dois mundos em que cada um cita o outro mas não se aprofunda, gerando lacunas que só vem sendo superadas com o desenvolvimento tecnológico que deflagra a informação como elemento subjacente a estas áreas.

Neste sentido, o autor trabalha com a noção de bibliotecas como espaços de leitura pública e independente de um centro de documentação, ou como suporte na gestão de acervos deste. Neste último caso, a gestão da biblioteca é considerada a parte dos serviços do setor de documentação relacionada a acervo (controle de inventário, empréstimo de documentos etc.). Assim, o serviço de documentação é composto pela gestão de biblioteca, pela análise de conteúdo e pela pesquisa retrospectiva. Também se encontra no livro a expressão “setor de documentação e de bibliotecas”.

Sendo assim, é interessante notar que, diferente do uso no Brasil, o termo documentação aparece antes de biblioteca, que é a uma das formas recorrentemente empregada na língua francesa. Além disso, o livro explicita o que entende por um e por outro termo, enquanto é comum na literatura brasileira a referência a “bibliotecas, centros de documentação e outros serviços de informação” para caracterizar os espaços de atuação da Biblioteconomia. Acredita-se que esta expressão é objetiva e clara no contexto brasileiro apenas para o primeiro termo (bibliotecas), sendo que no segundo (centros de documentação) não fornece uma idéia concreta para a maioria das pessoas da área, enquanto que, no terceiro (outros serviços de informação) imagina-se qualquer serviço de informação, desde os de ordem utilitária até os de cunho jornalístico, com menor alusão às bibliotecas. Assim, mais vale o uso culturalmente compreensível dos franceses que as expressões nebulosas adotadas no Brasil que, parecem pretender

ampliar a dimensão desta área de atuação e, no entanto, confundem mais que esclarecem.

O Dictionnaire Encyclopédique de l'Information et de la Documentation, coordenado por Serge Cacaly e editado em 1997, além de ser um instrumento abrangente, rigoroso e atualizado para a pesquisa na área, proporciona uma visão sobre a abordagem francesa, entre outros, porque a maioria dos redatores é do país ou é versado no idioma e cultura franceses.

Neste sentido, dois verbetes merecem ser destacados. O termo “formato” (*format*) é definido como o modelo de organização de dados destinados a representar uma classe de objetos, sendo que a representação associa, em geral, um aspecto conceitual e uma ou mais expressões destinadas ao tratamento informático. São tratados os formatos para: registros bibliográficos e catalográficos, textos, gráficos, imagens, sons e multimídia. No caso dos formatos bibliográficos e catalográficos, as normas do ISBD realizam as especificações conceituais que são precisadas por regras de catalogação como o AACR2, enquanto os formatos MARC definem os modos de organização e de representação dos dados para o tratamento informatizado. O Formato UNIMARC é um subconjunto criado para facilitar a troca entre os formatos MARC mas, na prática, a conversão de formatos encontra dificuldades técnicas que inviabilizam a conversão automática completa em todos os casos, em razão das diferenças conceituais relacionadas às regras de catalogação. Já nos formatos para texto, é considerado, de uma parte, o conteúdo textual que tem um sentido e que é organizado em parágrafos, frases, palavras e expresso por cadeias de caracteres e sinais de pontuação, e, de outra parte, a apresentação física do texto, como as fontes dos caracteres e a forma do texto na página. Três tipos de formatos para textos são descritos: o texto com caracteres codificados em ASCII ou outros sistemas; a marcação lógica de cadeias de caracteres no corpo do texto selecionando categorias de objetos (títulos, citações) como os campos de uma base de dados referenciais (como exemplo, cita-se a norma SGML); e os formatos para tratamento de texto e impressão como o PDF (Portable Document Format) da Adobe e o RTF (Rich Text Format) da Microsoft (MARTIN, 1997).

O verbete também se estende sobre os outros formatos mas, o que mais interessa aqui, é a abstração realizada para a conceituação do termo “formato”, a abrangência sobre os tipos de formatos em função do reconhecimento da diversidade de tipos de informações e, em decorrência dos anteriores, é digno de nota o entendimento sobre o formato como uma metodologia de representação da informação (ou seja, o formato não é definido como um padrão para intercâmbio de informações como é recorrente na literatura dos Estados Unidos sobre bibliotecas, efetivando uma inversão ao transformar um dos fins do formato em sua própria definição). Além disso, é registrada a percepção de que regras catalográficas estão sujeitas a interpretações distintas, fenômeno nem sempre admitido no universo bibliotecário, inclusive na produção literária e no ensino.

O termo “pesquisa documentária” (*recherche documentaire*) tem a sua tradução para o inglês como *information retrieval*. No entanto, diferente da literatura da língua inglesa levantada sobre este termo, não é apontada a disputa sobre a adequação entre os recursos de análise estatística e os de análise linguística para a recuperação de informações. Ao contrário, cita que uma pesquisa documentária eficaz implica a organização da informação e a existência de ferramentas de recuperação disponíveis ao usuário como índices, árvores hierárquicas ou listas, ou seja, documentos secundários

que sinalizam a existência ou a localização de documentos que respondam a uma demanda. Esta operação de indexação deve se efetuar, de preferência, por meio de uma linguagem documentária de forma a permitir a coincidência entre as palavras-chave da questão e aquelas que caracterizam o documento. No entanto, a recuperação probabilística não é dispensada, mas agregada como uma forma de diminuir a rigidez da busca booleana e acrescentar mais possibilidades para o acesso à informação (SUTTER, 1997). Sendo assim, além de não ser possível uma equivalência na tradução do termo entre os idiomas francês e inglês, também o significado (ou a abordagem) é relativamente distinta. A linha francesa (*recherche documentaire*) parece apresentar uma sistematização bem realizada, ou seja, consistente e alinhavada, enquanto a inglesa (*information retrieval*) aponta para uma área essencialmente de pesquisa, de enfoque quantitativo e com baixa aplicação. É fato que, também na França estas experiências foram realizadas, no entanto, distinguem-se pela sistematização abrangente do conhecimento, pela ênfase na abordagem lingüística e pela aplicação conjunta de técnicas de recuperação automática e de análise intelectual do conteúdo dos documentos nos sistemas de automação (como se verá mais a frente na análise do mercado francês de sistemas de automação de bibliotecas e centros de documentação).

Por fim, este dicionário sobre informação e documentação divide seus diversos termos em grandes tópicos. O tópico sobre tratamento da informação e da documentação inclui os termos acima apresentados, formato e pesquisa documentária, respectivamente, enquanto ferramenta e operação documentária. Mas é no tópico das técnicas eletrônicas que se encontra o termo Informática Documentária juntamente com as Telecomunicações. Sendo assim, a partir desta organização, o conjunto que compreende o conceito de Informática Documentária inclui: banco de dados, de sons e de imagens; codificação; compressão; Cibernética; edição eletrônica; GED; HTML (Hypertext Markup Language); hipertexto; HyTime (Hypermedia time-based structuring language); Inteligência Artificial; interface; *software* documentário; microinformática; digitalização; OCR; segurança de dados; SGML; sistema especialista; engenharia documentária (conjunto de procedimentos e metodologias para a concepção da produção e gestão de documentos e dos serviços de informação); e engenharia lingüística (conjunto de procedimentos e metodologias para a concepção e a realização de *softwares* que permitem o tratamento automático da linguagem natural).

## **OUTRAS REFERÊNCIAS SOBRE O TERMO *INFORMATIQUE DOCUMENTAIRE***

O termo *Informatique Documentaire* é encontrado em currículos profissionais e cartões de visita de páginas eletrônicas francesas. Também é usado para nomear diversas disciplinas de cursos da área de Ciência da Informação.

No Canadá, na Université de Montreal, a École de Bibliothéconomie et des Sciences de l'Information (EBSI), possui ao menos duas disciplinas, ambas oferecidas por Yves Marcoux, especialista em metodologias de implantação de documentos estruturados, bases de dados textuais, normas SGML e XML e hipertexto/ hipermídia (YVES MARCOUX, 2002). Este professor desenvolve, além das disciplinas Documentos Estruturados e Criação de Bases de Dados Documentários, dois módulos denominados Informática Documentária. O primeiro módulo inclui: culturas informáticas e microinformáticas, representação da informação textual, conectividade, HTML, pesquisa em

texto integral, SGBDs textuais, redes locais e metodologias de informatização. O segundo módulo desenvolve noções de abordagens metodológicas documentárias, introdução ao modelo relacional e aos documentos estruturados e a interoperacionalidade de sistemas documentários. O programa parece contemplar questões conceituais e de aplicação de *softwares* documentários, formando um conjunto abrangente sob o ponto de vista das demandas do bibliotecário ou documentalista.

Já na França, o Département des Sciences de l'Information et de la Communication, da Université de Paris X, oferece um curso de caráter profissionalizante que inclui a disciplina Informática Documentária (BIBLIOGRAPHIE..., 2001). A bibliografia para esta disciplina trata dos seguintes temas: sistemas especialistas, Recuperação da Informação (*Information Retrieval*), técnicas de hipertexto, GED, pesquisa documentária, informatização de bibliotecas e bibliotecas digitais. São contemplados os aspectos da automação da informação em ambientes virtuais ou locais como bibliotecas e centros de documentação, incluindo a análise comparativa entre estes dois últimos.

Também na França, em Lyon, a École Nationale Supérieure des Sciences de l'Information et des Bibliothèques (ENSSIB), oferece um Diplôme d'Études Supérieures Spécialisées (DESS) denominado Informática Documentária (LA DESS..., 2002). Este DESS é dividido em quatro áreas: Sistemas de Informações, Biblioteconomia, Documentação e Informática. O curso busca formar profissionais em informação científica, técnica ou econômica, para que se tornem aptos para analisar um problema, conceber uma solução, executá-la e evoluir a partir das técnicas documentárias e das ferramentas informáticas.

Segundo Cunha (1999), a distinção entre as carreiras e a formação de bibliotecários e documentalistas é uma especificidade francesa, sendo que o primeiro curso técnico de documentalista se realizou em Paris, em 1945. A partir dos anos 60, foram realizadas tentativas de criação de um tronco comum relativo às profissões de informação (Biblioteconomia, Documentação e Arquivística) sem resultados concretos. Atualmente, embora existam formações conjuntas de bibliotecários-documentalistas, no geral, as duas profissões desenvolvem cada uma seu próprio sistema de ensino. No Brasil, esta tentativa de união das profissões de informação, recomendada pela UNESCO, também não teve sucesso, provavelmente reforçada pela inexistência de uma área marcadamente constituída pela Documentação.

Voltando à literatura sobre o tema, pode ser citado também o livro francês editado pela UNESCO e traduzido para o português como "Introdução geral às ciências e técnicas da informação e documentação" (GUINCHAT, MENU, BLANQUET, 1994). No capítulo sobre a Informática nas unidades de informação, consta um tópico que discorre sobre os programas de Informática Documentária, os quais são entendidos como programas, procedimentos e regras concebidos especificamente para as tarefas documentárias. É considerada a existência de dois tipos de programas: os programas integrados são desenvolvidos para tratar um conjunto de aplicações documentárias e de gestão de bibliotecas, e, os programas dedicados a uma tarefa específica são concebidos para uma aplicação precisa da cadeia documentária como, por exemplo, o empréstimo, o tesouro e o acesso a bases de dados. Como campos de intervenção de um programa documentário constam:

- aquisição (encomendas, registros);
- gestão das entradas (catálogo, análise, indexação);

- gestão dos instrumentos lingüísticos (tesauros, léxicos);
- gestão dos arquivos (criação e atualização, restauração, salvamento);
- pesquisa documentária (pergunta e resposta, pesquisa retrospectiva, DSI);
- edição gráfica (índices, boletins);
- administração (estatísticas, gestão financeira, gestão de empréstimos).

Os itens acima apontam para processos que evidenciam a atenção com o controle do registro dos documentos, o tratamento da informação, os recursos para uso da informação pelo usuário e a gestão destes processos, em detrimento do cadastramento e controle de documentos como são característicos nos sistemas tradicionais de automação de bibliotecas.

Já o livro intitulado “L'accès automatisé à l'information: informatique documentaire et bases de données”, de Jacques Chaumier, foi publicado na França, em 1982. É um dos trabalhos que fundamenta a produção de Deweze, o qual é tão citado por Chaumier quanto este por aquele. Chaumier aborda:

- o tratamento da informação: *softwares* para pesquisa documentária ou de gestão de bases de dados textuais em minicomputadores e em computadores de grande porte (os microcomputadores ainda estavam em desenvolvimento) e *softwares* para gestão documentária que inclui as aquisições, a catalogação, o empréstimo e a gestão de periódicos, considerado de maior interesse à bibliotecas (inclusive, formatos MARC, formatos locais e redes de cooperação catalográfica da França);
- o tratamento lingüístico: sistemas informatizados que realizam a gestão de tesouro, a pesquisa em linguagem natural e a tradução automática;
- o acesso à informação: redes, servidores, linguagens de interrogação e acesso aos documentos;
- os novos meios de acesso à informação: considerados, a esta altura, como o vídeo-texto.

Este trabalho não é aqui amplamente explorado em função de as tecnologias nele tratadas estarem defasadas. No entanto, o que deve ser ressaltado, é justamente o fato de que, mesmo no caso de uma obra do início dos anos 80, os princípios básicos não se alteraram e ainda servem como norte para a literatura mais recente da mesma linha de abordagem. Isto porque, entre outros, a Documentação nunca preteriu a tecnologia, mesmo tendo surgido antes da Informática; ao contrário, fez da tecnologia, unida às linguagens documentárias, seu grande instrumento de trabalho e de operacionalização de seu escopo teórico e metodológico.

A Informática Documentária é assim concebida por Chaumier:

- tratamento da informação
  - pesquisa documentária
  - gestão documentária
    - aquisição

- catalogação
- empréstimos
- acesso à informação
  - sistemas *online*
    - bases de dados
    - bancos de dados
    - vídeo-texto
  - acesso aos documentos

A estrutura de conteúdos apresentada por Chaumier, no início da década de 80, é semelhante àquela do trabalho de Deweze, produzida mais de uma década depois, até porque, ambos parecem pertencer à mesma escola. Sendo assim, enquanto a produção francesa, em função de sua essência, tem permanecido atual, as obras de Rowley e Saffady, especialmente este, tornaram-se abrangentes, apenas nas últimas edições publicadas no final da década de 90. Saffady trabalhava com a idéia das operações automatizadas de uma biblioteca e de seu caráter histórico-evolutivo: na edição mais recente (1999) mudou sua abordagem para a noção de conceitos de sistemas e serviços que geram as operações automatizadas e atualizou as ferramentas tecnológicas envolvidas. Na obra de Deweze, não houve alteração de enfoque entre as primeiras edições até a mais atualizada que, a propósito, nem é tão recente (1994). Assim mesmo, parece não haver prejuízo sobre o entendimento do cerne do livro que é a noção de cadeia ou fluxo documentário, apesar de esta edição não apresentar as mudanças decorrentes da disseminação de documentos eletrônicos pela Internet. Enquanto isso, o foco de Saffady que estava na gestão de acervos é mais fortemente abalado com as bibliotecas digitais. Em outras palavras, um livro sobre Informática Documentária ainda estará relativamente atualizado se seu foco for o fluxo da informação documentária e não as tecnologias (*hardware*, *software*, linguagens de programação e demais aplicativos e produtos tecnológicos) para a gestão de estoques de informação e de documentos.

#### **2.4.1 ANÁLISE DO MERCADO FRANCÊS DE SISTEMAS DE AUTOMAÇÃO DE BIBLIOTECAS E CENTROS DE DOCUMENTAÇÃO**

A Association des Professionnels de l'Information et de la Documentation (ADBS) tem publicado trabalhos sobre o mercado francês de sistemas para automação de bibliotecas e centros de documentação, na revista *Documentaliste* e na *Collection Sciences de l'Information, Série Études et techniques*. São relatadas abaixo, informações extraídas de alguns artigos publicados na revista citada, a partir de 1994.

Chaumier, em **1994**, apresentou os *softwares* para pesquisa documentária em função do aparecimento no mercado de produtos novos e de novas versões de produtos já implantados. Descarta os *softwares* especialmente dedicados a microcomputadores e

aqueles para a gestão de bibliotecas (que caracteriza como produtos centrados sobre a Biblioteconomia), pois outros artigos da revista já os apresentaram.

Apresenta as principais características da oferta naquele período na França, a partir da análise de 8 *softwares* de pesquisa documentária: AIRS (sigla de ASTRA Information and Retrieval System), BASIS Plus, Doris, Doxis, Ful/Tex, Status, Taurus e Topic. Dentre os outros programas do mesmo gênero, opta por não analisar os *softwares*: Texto, JLB-DOC, BRS, MINISIS, Spirit, Search Manager (desenvolvido pela IBM para substituir o STAIRS, mas até então, não instalado na França), DAP Text e Folio Views (estes dois últimos não são conhecidos na França).

Os *softwares* para pesquisa documentária analisados apresentam as seguintes características:

- *Softwares* integrados são raros;
- As funções biblioteconômicas não são levadas em conta na maioria das vezes. Apenas dois dos sistemas analisados (BASIS Plus e Doris), em função de módulos específicos, oferecem ao usuário o conjunto de funcionalidades de aquisição, catalogação, empréstimos, OPAC, assim como a possibilidade de utilizar o Formato MARC;
- Capacidade para arquivamento sobre disco ótico digital;
- *Softwares* para GED;
- Todos possuem capacidade para tesouro: ao lado das funções de pesquisa sobre texto integral, numerosos *softwares* oferecem gestão de tesouro com gráficos semânticos utilizados para a interrogação (as pesquisas em linguagem controlada e em linguagem natural são dois métodos complementares e não métodos que se opõem);
- A utilização da norma SGML aparece pouco e, não por acaso, é mais freqüente naqueles de origem norte-americana (o autor não explica porque a não casualidade deste fato);
- A pesquisa por hipertexto é quase ausente;
- A lacuna considerada mais grave é a ausência quase total de integração de ferramentas lingüísticas à exceção de tesouros: apenas um *software* (Topic) integra uma implementação do tipo indexação automática a partir de “tópicos” constituídos de gráficos semânticos;
- Também não são muito presentes as funções de gerenciamento de informação digital;
- Todos têm arquitetura cliente-servidor;
- Todos correm em ambiente UNIX.

Seguem abaixo as características do programa BASIS Plus, por ser sucessivamente citado nas outras obras e por fornecer informações relacionadas a esta pesquisa:

- Foi desenvolvido pelo Institut Batelle Collumbus Laboratories, dos Estados Unidos, mas outra empresa comercializa o produto a partir de filiais em diversos países;

- É usado em mais de cem instituições ou empresas da França, sendo o *software* documentário mais utilizado neste país, fora os *softwares* para microcomputadores;
- Sua portabilidade permitiu uma larga difusão;
- É o que reúne mais funcionalidades somando 14 módulos;
- Foi um dos primeiros *softwares* documentários (junto com o MINISIS) a ser desenvolvido a partir de um SGBD relacional;
- Inclui todos os índices de pesquisa (arquivos invertidos) sob a forma de árvores balanceadas (*B-trees*), as quais permitem um acesso rápido aos termos mesmo em índices com muitas chaves;
- Possui excelente gestão de tesouro;
- Permite a pesquisa em texto integral com operadores de proximidade;
- Foi o primeiro a aceitar documentos estruturados sob SGML;
- As funções de cálculo e estatística são desenvolvidas;
- Possui módulo com todas as funções de gestão de biblioteca e interfaces com o Formato MARC, sendo considerado um *software* integrado;
- Não possui qualquer ferramenta lingüística como analista, fonetizador e lematizador (que consiste em conduzir termos para sua raiz eliminando as marcas de conjugação ou de declinação).

Mais tarde, em **1996**, Lénart também realiza pesquisa sobre a evolução dos *softwares* de pesquisa e de gestão documentária. Aponta que as principais evoluções funcionais dos *softwares* documentários incluem dois aspectos: as funções de ajuda à pesquisa e o acesso Internet/Intranet (ou módulos Internet).

As funções de ajuda à pesquisa podem ser classificadas em duas grandes categorias: aquelas que relevam os tratamentos lingüísticos e aquelas que respondem a análises estatísticas. No caso das análises lingüísticas, os produtores de *softwares* não têm se lançado em desenvolvimentos complexos e custosos, priorizando as funções clássicas (como sinonímia de tesouros e truncamentos) para tratar as dificuldades da língua. Foram detectados poucos *softwares* com tratamento do singular e plural e nenhum com processos de lematização. Os tratamentos estatísticos, usuais em pesquisas sobre a Internet, são também raros para esta categoria de *softwares*.

Os *softwares* documentários com módulos para Internet estão sendo desenvolvidos pelos produtores em função do impulso que esta tecnologia tem causado. No entanto, a maioria dos projetos está baseada sobretudo em páginas HTML de apresentação (com organograma, anuário etc.) e menos sobre o acesso à base documentária. Os módulos Internet fornecem recursos, de fato, para o acesso Intranet/Internet. As tecnologias utilizadas para as aplicações Internet e Intranet são basicamente as mesmas, sendo que a diferença está no grau de abertura do sistema e da garantia de segurança para tal. Em um projeto baseado na Internet, a base é acessada via rede pública para qualquer usuário que é cliente Web. Uma aplicação

Intranet restringe esta possibilidade somente aos usuários que tem acesso à rede privada da empresa. A noção intermediária de Extranet designa um acesso não mais limitado à empresa mas estendido a seus interlocutores privilegiados como fornecedores e clientes.

Dentre os *softwares* analisados, são apresentadas aqui as características do Spirit-Biblio e do AIRS, respectivamente, em função da existência de ferramentas lingüísticas e da origem a partir de *software* documentário pioneiro (MISTRAL).

Spirit-Biblio:

*Software* bem conhecido dos documentalistas, é indicado aos grandes centros de documentação interessados em investir em funções elaboradas de tratamento lingüístico, pois as funções de análise semântica são sua grande originalidade. A pesquisa conta com:

- Análise morfológica a partir da lematização e do reconhecimento de expressões idiomáticas (a partir de um dicionário em francês, alemão e inglês que pode ser enriquecido pelos documentalistas levando em conta, por exemplo, os termos próprios do domínio de uma empresa);
- Análise sintática com a supressão de ambigüidades semânticas por atribuição de uma categoria gramatical (artigo masculino, nome feminino etc.);
- Cálculo das dependências lingüísticas como aquelas introduzidas pelo “d” (em francês, por exemplo, *contrat d’assurance*);
- Análise estatística destinada a definir o peso informacional dos termos em função de sua freqüência;
- A análise da questão em linguagem natural inclui uma reformulação da questão baseada sobre uma extensão aos sinônimos e aos termos da mesma família (genéricos, específicos e termos associados). A transcrição de uma questão em uma outra língua leva em conta os idiomatismos, evitando a tradução palavra a palavra.

Apesar de estas funções estarem longe de serem correntes nos centros de documentação, estão baseadas nas pesquisas em Lingüística e em Inteligência Artificial aplicadas à Documentação iniciadas nos anos 60, que levaram à elaboração de protótipos e depois a aplicações, nos anos 80.

A pesquisa por idéias, a ser implementada futuramente, visa a compreensão de idéias e questões do texto e não mais de conceitos tais como são representados em uma abordagem do tipo tesouro. Esta análise dinâmica de idéias está baseada em três eixos:

- Uma análise da retórica graças à modelização de estilos ou gêneros;
- O tratamento de raízes indo-européias de palavras que são associadas a 8.000 sentidos primitivos. A modelização das regras de evolução de línguas permite uma aproximação multilíngüe e enciclopédica da linguagem;
- A modelização dos 873 conceitos de base da lógica aristotélica.

Menos novos mas também necessários ao trabalho documentário, as funções do módulo de gestão de biblioteca também estão presentes no Spirit Biblio. A saída é feita em formato de tipo documentário mas a exportação de dados em Formato MARC é possível desde que haja uma boa estruturação da base. O campo “resumo” e a ficha de

descrição bibliográfica permitem, respectivamente, sumários e fotografias digitalizadas. Dispõem de uma OPAC em linguagem natural para os campos textuais. Os resultados são classificados por ordem de pertinência a partir da análise de termos do registro. A base pode ser interrogada pela Web e os resultados são convertidos em formato HTML.

O Spirit também permite a preparação de discos compactos para a difusão de informações em CD-ROM.

#### AIRS:

Criado por uma empresa formada por antigos membros da equipe de desenvolvimento do MISTRAL, da Bull, um dos mais conhecidos *softwares* documentários, o AIRS foi desenvolvido a partir deste conhecimento da problemática documentária. O AIRS oferece as possibilidades de pesquisa multicritérios, pesquisa em texto integral, utilização do tesouro e *links* de hipertexto e de acesso às imagens associadas. É constituído dos módulos:

- gestão de bases de dados editoriais;
- gestão de biblioteca;
- gerenciamento eletrônico de documentos: gera diferentes tipos de documentos como imagens, arquivos Word, documentos SGML e outros, que são acoplados às aplicações de gestão de documentos;
- serviço de Internet: possibilita acesso e interrogação à base documentária em uma configuração Internet ou Intranet e a recuperação de mensagens de *e-mail* e sua integração em bases AIRS;
- concepção e elaboração de CD-ROM: com todos os recursos de pesquisa, tesouro, hipertexto e imagens associadas do próprio sistema AIRS.

Sendo assim, até 1994, a diferença maior na evolução dos *softwares* estava na oferta de GED. Nesta análise de 1996, a autora considera que estas disparidades foram atenuadas, a exceção de alguns *softwares* do mundo de bibliotecas que, não estando ainda confrontados com uma demanda por parte dos usuários, não têm desenvolvido esta função. Afirmo que, o que está em jogo neste momento é a Internet, mas acredita que a pressão baseada na moda é tal que contrastes devem ficar à mostra rapidamente. Suspeita que o aspecto diferencial está em torno de funções complexas como o tratamento lingüístico, mas alguns produtores também podem vir a desenvolver seus sistemas sobre domínios que estão no limite do tratamento documentário clássico, como o gerenciamento da revisão de documentos e de arquivos.

Em geral, estes movimentos seguem a evolução da profissão dos documentalistas. A autora acredita que no próximo estudo será possível ver a geração de *softwares* integrando todos os aspectos da gestão documentária, do tratamento de documentos primários para a difusão sobre todos os tipos de suportes, passando pela pesquisa documentária sobre todas estas formas. No entanto, os produtores não poderão fazer evoluir seus produtos se o mercado representado pelos centros de documentação não apresentar um número suficiente de reformatizações.

Pesquisa de 1994, realizada por Lénart e Bourdin, foi baseada na pergunta sobre a existência e sobre a necessidade de um *software* único para a pesquisa documentária

e para a gestão de biblioteca. Os autores afirmam que esta análise aponta para uma situação bem conhecida dos profissionais de informação e de documentação: informatizar o centro de documentação tanto em seus aspectos especificamente documentários como nas equações de pesquisa complexas, difusão seletiva da informação ou gestão de tesouro, quanto em suas funções de gestão de acervo, tais como o empréstimo de documentos, registro de periódicos, controle de aquisições, além de importação e de exportação de dados bibliográficos em Formato MARC.

Na pesquisa de 1998, Lénart retoma a questão da ambigüidade entre os *softwares* de pesquisa documentária e os de gestão de bibliotecas, argumentando que os sistemas informatizados necessários à gestão de centros de informação e de documentação devem, na maior parte dos casos, cumprir dupla função. Assim, busca identificar as evoluções destes grupos de *softwares* e analisar a oferta de módulos Web, sem tratar dos *softwares* mistos mas, prioritariamente, daqueles destinados aos centros de documentação e daqueles voltados às bibliotecas especializadas, mesmo que os limites entre os diferentes tipos de *softwares* sejam, às vezes, muito imprecisos.

No trabalho de 1994, os principais critérios de seleção dos grupos de *softwares* estavam na presença de funções mínimas de gestão de bibliotecas, de uma parte, e de gestão de tesouros de outra. A gestão de tesouros é um dos pontos cruciais de diferenciação com os sistemas de gerenciamento de biblioteca tendo em conta a relação de hierarquia e o recurso de *autopostage* (indexação complementar de um documento, que consiste em atribuir automaticamente descritores pertencentes às mesmas cadeias hierárquicas que os descritores atribuídos diretamente pelo indexador). A distinção entre *softwares* destinados ao mundo das bibliotecas e os mais próximos ao meio documentário (por sutil que esta diferença seja para um não especialista), estava já patente neste ano.

Em 1998, a diferenciação se acentuou: dois *softwares* que a princípio respondiam às necessidades de bibliotecas especializadas, orientam seus esforços para as bibliotecas de leitura pública ou bibliotecas de leitura e empréstimos de livros de empresas privadas (*bibliothèque de comité d'entreprise*), abandonando as funções de gestão de tesouro e de *autopostage*. A autora supõe que o mercado aponta para a concentração no segmento onde se é mais eficaz.

Por outro lado, dois outros *softwares*, o ALEPH, *software* israelense comercializado na França, e o Bestseller, que é canadense, podem se encontrar na categoria de *softwares* mistos: a versão documentária do ALEPH, difundida especialmente na Alemanha, equipa centros de documentação, e o Bestseller desenvolve atualmente uma gestão de tesouros.

Existe uma continuidade entre os *softwares* documentários orientados ao GED e os sistemas de gerenciamento de biblioteca puros. Para os primeiros, uma interface com um *software* de gestão de biblioteca será uma solução que lhes permite cobrir o conjunto de necessidades. Os segundos têm optado por aplicações de GED, em especial, nas bibliotecas públicas na gestão de dossiês temáticos digitalizados, bases iconográficas etc. O que se aponta é o problema da integração com suas vantagens e seus limites. Para os centros de documentação e, mais ainda, para as bibliotecas especializadas, a definição precisa de necessidades é, por isso, uma condição prévia indispensável para a escolha de um *software*.

A pesquisa indica que alguns *softwares* originários dos Estados Unidos e Reino Unido estão ganhando o mercado francês e vice-versa.

Dentre os novos *softwares* do mercado, seguem:

Ex Libris:

Produzido pela empresa Cadic, é um SGBD relacional que possui todas as funções clássicas: a utilização de operadores booleanos, de truncamentos e de máscaras, a ativação e visualização de tesouros e a *autopostage* parametrizável em até cinco níveis do tesouro. Conta também com as funções de pesquisa em texto integral: pesquisa de proximidade por adjacência e distância (parametrizável de 2 a 50 palavras), ponderação de critérios, pesquisa por similaridade e por extrapolação da questão (por exemplo, substituindo a questão já realizada em operador E pelo operador OU).

Integra um módulo lingüístico permitindo os tratamentos morfológicos para as formas conjugadas de verbos, singular/plural, masculino/feminino, além de formas fonéticas, sintáticas e semânticas (com possibilidade de definição de suas próprias regras semânticas em inglês ou francês). Oferece funções de tratamento do multilingüismo sobre 8 línguas pelas funções semânticas do módulo lingüístico.

As seleções por ordem de pertinência são particularmente sofisticadas permitindo:

- Número total de contextos: se a pesquisa foi realizada sobre diversos critérios, os primeiros documentos resultantes respondem a todos estes critérios;
- Número total de contextos diferentes: os documentos que comportam os termos mais freqüentes com relação ao índice serão fornecidos prioritariamente como resultado;
- Termos ordenados: os documentos serão selecionados com relação à ordem de termos escolhidos pelo usuário;
- Termos críticos: os documentos que comportam os termos menos freqüentes com relação ao índice serão fornecidos prioritariamente como resultado.

MARCo Server:

A empresa inglesa que comercializa o sistema de gerenciamento de biblioteca TINLIB, criou o MARCo Server para investir no mercado francês, prioritariamente nos centros de documentação e bibliotecas especializadas. O *software* aproxima-se do mundo das bibliotecas e, como seu nome indica, fundamenta-se no Formato MARC. Mas é também um sistema integrado destinado aos centros de documentação e às bibliotecas especializadas: é conhecido por servir como ponto de entrada comum a todos os tipos de informações, sejam bibliográficas ou não, de fonte interna ou externa (em especial, a informação disponível na Internet).

Para a pesquisa de texto integral, possui os recursos de:

- Pesquisa por proximidade (adjacência e distância);
- Pesquisa por similaridade com os métodos de reconhecimento de formas;
- Tratamentos lingüísticos: morfológicos, fonéticos, sintáticos e semânticos graças aos dicionários (400.000 significações de palavras, 50.000 expressões idiomáticas e 1,6 milhões de relações de associação para o inglês);
- Cálculos de seleção por pertinência.

No ano seguinte (1999), o *software* integrará a norma Unicode para a gestão de línguas e de escritas não latinas, visando a difusão internacional.

Cindoc:

É o novo *software* de 32 *bits* da mesma empresa que comercializa o Texto (16 *bits*). Dispõe de :

- Sistema de GED integrado com funcionalidades como anotações do tipo *post-it*, gestão de revisão de documentos, além de suportar mais de 200 formatos;
- Pesquisa em texto integral e OCR;
- Funções de DSI para *push* (envio, encaminhamento) via serviço de mensagens eletrônicas;
- Integração com *software* de *workflow* para fazer face ao desenvolvimento do trabalho cooperativo na empresa.

Quanto aos módulos Web de *softwares* documentários, a pesquisa aponta que todos os *softwares* possuem um módulo Web e suas vendas são significativas. No entanto, nos centros de documentação, a demanda pauta-se, em especial, sobre projetos para Intranet, sendo as opções por Internet mais comuns em bibliotecas.

Atualmente, no ambiente Web, as distinções correntes entre produtos documentários (constituídos pela iniciativa do documentalista) e serviços (prestações acionadas pela demanda do usuário) vêm com força com as tecnologias denominadas *push* e *pull*. O *pull* corresponde, para o usuário, à ação de obter (puxar) a informação pela consulta, impressão ou teletransferência (*downloading*). O *push*, ao contrário, consiste em encaminhar (empurrar) as informações ao usuário e, se possível, em função de suas necessidades próprias. Recupera-se assim, a antiga noção de difusão seletiva da informação, no entanto, rejuvenescida e complementada com a utilização apropriada de listas de difusão e de serviços de mensagens eletrônicas.

Outro aspecto quanto aos módulos Web é que sua utilização tende a se estender pelo conjunto de funções documentárias. O esquema tradicional da cadeia de tratamento está dividido em três tipos (a coleta, o tratamento documentário e a difusão) e a situação evoluiu diferentemente sobre estes três grupos de tarefas. A Web foi percebida inicialmente como uma ferramenta de consulta e de difusão da informação, levando a que seus primeiros desenvolvimentos estivessem centrados nestas atividades: recupera-se informações a partir de todo o leque de possibilidades, dos mais clássicos aos mais sofisticados, incluindo tratamentos estatísticos e lingüísticos. Se o servidor Web é antes uma ferramenta de consulta e de difusão da informação, a pressão dos documentalistas é tal que passará a ser também uma ferramenta de entrada e tratamento. Já as funções de coleta aparecem pouco, mas poderão ser tratadas a partir dos agentes inteligentes encarregados de interrogar automaticamente e de remeter em forma de fontes de informação.

Sendo assim, a autora conclui que a oferta de *softwares* com módulos Web está longe de ser estável e é normal que a atenção maior seja para a difusão pois esta é a razão de ser da Web. No entanto, os outros aspectos da cadeia documentária deverão se desenvolver. Assim, uma das tendências marcantes é a geração de funções de pesquisa sofisticadas baseadas em tratamento lingüístico.

Com isto, surge o problema da capacidade dos usuários finais em utilizar as potencialidades do sistema, dominando a lógica booleana e construindo uma equação de pesquisa. Para tanto, os sistemas devem caminhar no sentido da construção de telas de diálogo parametrizáveis e com mensagens de qualidade. Por outro lado, se a Internet entra na vida cotidiana das pessoas, seja profissional ou privada, o domínio de técnicas de pesquisa próprias dos motores de busca deste ambiente (Yahoo, Altavista e outros) difundem-se progressivamente. Com isto, os produtores de *softwares* documentários têm interesse em usar como base estas formas de pesquisa. Além disso, os dois navegadores predominantes (Internet Explorer e Netscape) promovem, de certa forma, uma normalização efetiva de interfaces para os usuários.

O argumento final é o de que, apesar do amadurecimento dos usuários no uso das ferramentas de pesquisa da Internet, os documentalistas e profissionais de informação devem trabalhar no aperfeiçoamento dos *softwares* e no reforço do papel de profissão baseada na mediação.

### **Apontamentos sobre os mercados francês e estadunidense**

A responsável pela análise do mercado francês, Michèle Lénart, é consultora em Engenharia Documentária pela TOSCA Consultores, localizada em Paris. Ela induz à conclusão de que os usuários intermediários dos *softwares* documentários (ou seja, os documentalistas) são exigentes e possuem necessidades bastante particulares. Enquanto isso, os usuários intermediários dos *softwares* de bibliotecas, representados pelos bibliotecários, exploram menos as ferramentas tecnológicas de que dispõem, ao mesmo tempo em que possuem demandas menos específicas (por exemplo, quanto à evolução de sistemas para GED que não tem ocorrido no universo das bibliotecas).

Isto é reforçado pela afirmação de que os sistemas informatizados necessários à gestão de centros de documentação devem, quase sempre, cumprir a dupla função de tratamento e pesquisa documentária e de gestão do acervo. Como decorrência, os centros de documentação têm mais encargos em função da gestão do acervo, enquanto as bibliotecas (e seus usuários, ao que se supõe) não têm se incomodado na mesma medida em aprimorar os aspectos estritamente documentários. Parece ser mais óbvia a um centro de documentação a necessidade de tratar e armazenar o acervo, mas isto não justifica a desconsideração histórica pelas bibliotecas sobre o tratamento e recuperação da informação documentária, atualmente amparada pela ampla oferta de bases de dados documentárias externas e outras informações acessadas remotamente.

São apresentados elementos que convergem para a idéia de que os sistemas documentários já apontavam para uma concepção mais ampla e complexa, em oposição aos sistemas de biblioteca marcados, ao menos tradicionalmente, pelo controle dos acervos locais. O caráter avançado dos sistemas documentários está calcado na própria origem dos princípios documentários que se fundamenta no reconhecimento da linguagem e na exploração das tecnologias para sua manipulação.

Também é realizada uma distinção entre centros de documentação e bibliotecas especializadas fazendo supor que, nos primeiros há mais procedimentos documentários e nos segundos, ocorre a gestão de acervo especializado. Além disso, é interessante mas não conclusiva, a afirmação de que é justamente o segundo grupo que merece maior atenção na definição de necessidades.

Com relação à segmentação de *softwares* para bibliotecas e para centros de documentação, pode-se concluir que, a tendência é a de que os sistemas para gerenciamento de bibliotecas venham agregando os recursos dos *softwares* documentários, ao mesmo tempo em que se verifica um movimento no sentido de que estes dois grupos satisfaçam as necessidades de tratamento e acesso pela Internet. No entanto, alguns produtos de características documentárias caminharam para sistemas puros de gerenciamento de biblioteca, perdendo suas características de gestão de tesouros e de *autopostage*. Isto demonstra a força deste mercado representado pelas bibliotecas, que é maior e mais homogêneo que o dos centros de documentação e de outros serviços de informação. Ao que parece, é a Internet que está mudando este quadro, aproximando estes dois mundos que, há décadas foram se constituindo em separado. Resta saber se esta confluência instituirá características documentárias aos sistemas de automação de bibliotecas, ou haverá apenas esforços para a compatibilização entre as bases de dados locais e as externas, promovendo assim, um “casamento de fachada”. Ou seja, mesmo com recursos refinados de pesquisa, os sistemas de automação de biblioteca, ao acessarem catálogos de biblioteca onde não houve tratamento documentário elaborado, obterão como resultado uma recuperação restrita.

Com relação à cadeia documentária no ambiente Web que aponta para a concentração na difusão de informações em detrimento da coleta e tratamento, vale notar que são os documentalistas a reivindicarem mais e melhores ferramentas de tratamento documentário. Além disso, muito provavelmente, esta não seja uma demanda, em geral, de bibliotecários, que estão essencialmente centrados nas operações de acervo e público locais e, cujas informações externas a serem fornecidas a seus usuários constam, freqüentemente, apenas de seus serviços de difusão e não de organização. Na análise de mercado dos Estados Unidos, comentário semelhante é realizado, apontando os bibliotecários corporativos (de empresas) como os impulsionadores do desenvolvimento e uso de *softwares* e outras tecnologias para estes fins.

Neste aspecto, importa considerar que, no Brasil, também há pouca demanda para organização das informações no ambiente Web em comparação com a oferta já existente para consulta. No entanto, com a difusão acelerada de disseminação de informações pela Internet, o mercado tem sentido necessidade de profissionais que atuem na organização de informações, não nos requisitos unicamente tecnológicos, mas indexação do conteúdo de documentos eletrônicos e na organização geral de *sites* e portais, entre outros. De fato, não há clareza, entre os bibliotecários e entre os difusores de informação na Internet, sobre a relação de continuidade entre o ambiente da biblioteca e o ambiente Web, o qual, por este motivo, exige os conhecimentos e métodos da Biblioteconomia e Documentação.

Também cabe ressaltar que a análise do mercado francês sobre a noção de cadeia documentária na Web não foi encontrada na análise do mercado estadunidense. Esta análise pauta-se essencialmente na descrição das características dos *softwares*, nas tendências do mercado e no quanto isto têm afetado o trabalho dos bibliotecários, por exemplo, possibilitando a eles a realização da gestão da informação institucional. Esta discussão, no entanto, não decorre de um princípio norteador como o da noção de cadeia documentária. Sendo assim, a constatação da maior inserção do bibliotecário na instituição em que atua em função do potencial tecnológico e de seu uso estratégico, pode implicar em maior *status* profissional, mas o conhecimento que está por trás desta estratégia não é evidenciado. No caso da análise francesa, este conhecimento é apresentado como composto por um conjunto de teorias e metodologias documentárias

baseadas na noção de fluxo ou cadeia e concretizadas por meio de recursos tecnológicos eletrônicos.

Recorrendo a uma comparação com a automação de bibliotecas apresentada essencialmente por Saffady, um aspecto que chama atenção pela abordagem diferenciada é o do Formato MARC. Enquanto nas bibliotecas estadunidenses, a implementação do Formato MARC é determinante para possibilitar o trabalho cooperativo de catalogação, experiências distintas são relatadas na França.

Uma delas refere-se à compatibilização dos registros da base de dados bibliográficos da biblioteca municipal francesa de Fresnes com os registros da base da Bibliothèque National de France, visando o controle bibliográfico do país. Aponta-se que a França é o único entre os países do G7 (grupos dos 7 países mais ricos) que não realiza o controle bibliográfico nacional (GIAPPICONI, 1998). Após estudos, foi adotado o Formato INTERMARC (Formato MARC francês) pois o UNIMARC e o LCMARC são formatos mais limitados. Como não se desejava diminuir a riqueza do catálogo daquela biblioteca, assim o INTERMARC foi adotado seguindo a idéia de que “quem pode mais, pode menos”.

Este relato aponta para dois aspectos. Primeiro, demonstra a natureza das diferenças entre os formatos MARC: o Formato LCMARC é menos abrangente quanto aos campos designadores de conteúdo que os formatos MARC canadense, do Reino Unido e francês, sendo que, os dois últimos ainda são mais extensos que o primeiro. Esta diferença de formatos aponta para a tendência já detectada de que o tratamento documentário realizado nos países europeus, mesmo nas grandes bibliotecas, caracteriza-se por mais detalhamento e abrangência, supostamente pela influência dos princípios documentários. Segundo, a experiência citada em particular, é digna de nota por deflagrar uma biblioteca pública municipal que realiza tratamento documentário acurado e, além disso, apresenta a preocupação, simultaneamente, com o compartilhamento de dados e com o desejo de que este tratamento elaborado não se perca.

Foram apresentados vários levantamentos do mercado de *softwares* com base em análises de determinado período de tempo, em especial, sobre os dois países em que esta pesquisa está centrada. Assim, foi sistematizada a análise do mercado estadunidense de sistemas de automação de bibliotecas publicada na coluna anual do Library Journal chamada Automated System Marketplace, de 1995 a 2000. Na França, os artigos produzidos por Chaumier e, especialmente, por Lénart tratam de *softwares* para gestão de bibliotecas e de pesquisa documentária e as tendências destes dois grupos, cuja análise abarcou o período entre 1994 e 1998. Também foram encontradas análises realizadas pelo grupo holandês VOGIN, de 1991, centrados nos então conhecidos *softwares* para recuperação textual. Há o material em espanhol de García Gutierrez e, no mesmo idioma, a produção da Argentina apresenta significativos avanços, entre outros, em função de sua reconhecida maturidade com o *software* CDS-ISIS.

Apesar de esta pesquisa pretender analisar modelos constituídos sobre Informática Documentária, quais sejam, o norte-americano (centrado nos Estados Unidos) e o europeu (em especial, o francês), o foco tem recaído na Automação de Bibliotecas (*Library Automation*) em detrimento da Recuperação da Informação (*Information Retrieval*), nos Estados Unidos, e na Documentação em detrimento da Biblioteconomia, na França. Mesmo assim, fica claro que a constituição de sistemas de recuperação de informação dos franceses, denominados *softwares* documentários, caracterizam-se pelo tratamento lingüístico das informações, em oposição aos mesmos

sistemas estadunidenses, mais enfaticamente voltados para a análise estatística (tratamento quantitativo). Da mesma forma, os sistemas para gerenciamento de bibliotecas dos Estados Unidos atingiram grande desenvolvimento por diversos motivos:

- Há aproximadamente um século teve início o projeto de popularização e disseminação de bibliotecas públicas no país, além de altos investimentos na construção e ampliação de bibliotecas universitárias;

- Os sistemas de automação de bibliotecas são significativamente padronizáveis pois atendem a instituições relativamente semelhantes, quais sejam, grandes e médias bibliotecas públicas, escolares e universitárias, sendo por isso, mais adequados à constituição e desenvolvimento de produtos para o mercado;

- Em uma nação neoliberal e, portanto, altamente mercadológica, ou seja, com fortes segmentos empreendedores e consumidores, o terreno é adequado à exploração das facilidades apontadas no item anterior.

Todos estes aspectos, a orientação ao mercado, a popularização das bibliotecas por todo o país e seu caráter homogêneo, promoveram a formação e consolidação de um mercado de sistemas de automação de bibliotecas forte e competitivo nos Estados Unidos.

Em função do caráter mercadológico de organização e funcionamento deste país, pode-se perceber que muitas inovações são impulsionadas essencialmente pela competitividade entre as empresas produtoras dos sistemas. Isto não necessariamente implica em sistemas mais adequados pois os fins a que se propõem podem estar baseados em resultados imediatistas, não comprometidos com propostas revolucionárias na área de informação, mas ao contrário, para garantir a clientela, dependem do que é demandado. Isto fica evidente nos sistemas voltados para a catalogação cooperativa em Formato MARC e nos módulos para informação comunitária. Em especial quanto a estes últimos, são sistemas que incluem bases de dados abertas para a formatação mais adequada e inserção de informações úteis à comunidade local, como dados de eventos e de serviços de interesse público. Estes sistemas, no entanto, parecem indicar alguma semelhança com os *softwares* documentários tão amplamente abordados pelos franceses.

Já os *softwares* para bibliotecas especializadas dos Estados Unidos são considerados, segundo os artigos sobre este mercado, difíceis de serem apresentados a partir de alguns modelos comuns, por serem muito distintos uma vez que atendem necessidades específicas. Acreditam, por este motivo, que os produtos mais populares não representam as necessidades de todas as bibliotecas, mas não há uma investigação sobre os motivos da popularidade destes produtos. No entanto, a literatura européia, incluindo a do Reino Unido, mas contando especialmente com a francesa, descreve os *softwares* documentários ou de gerenciamento de documentos, distintamente daqueles mais preocupados com a automação de bibliotecas. Estes *softwares* são apresentados com suas características e tendências, fornecendo modelos bastante claros e sedimentados. Sendo assim, ou não existe acúmulo de conhecimento gerando difusão destes modelos nos Estados Unidos ou não foram pesquisadas as fontes corretas. De fato, na análise de mercado dos Estados Unidos, muitas vezes, os *softwares* classificados como adequados para bibliotecas especializadas, são vendidos principalmente na Europa. A impressão que fica é a de que, quanto ao acúmulo metodológico e tecnológico, os Estados Unidos estão para os sistemas de automação de

bibliotecas como a Europa está para os *softwares* documentários (ou sistemas de recuperação da informação).

O mercado europeu em geral não atingiu esta evolução de produtos para bibliotecas, reforçado pela cultura, nem tão mercadológica e tecnológica como a dos Estados Unidos, e por um certo traço de erudição nas bibliotecas, advindo dos séculos medievais (em oposição à forte presença do projeto da modernidade nos Estados Unidos e Reino Unido que deflagrou a implementação massiva das bibliotecas públicas). De qualquer forma, as bibliotecas europeias estão na linha de desenvolvimento das bibliotecas dos Estados Unidos, partindo para a digitalização dos documentos para fazer uso das potencialidades da Internet.

## **2.5 TIPOLOGIAS DE SISTEMAS PARA AUTOMAÇÃO DE BIBLIOTECAS, CENTROS DE DOCUMENTAÇÃO E OUTROS SERVIÇOS DE INFORMAÇÃO**

À exceção dos textos que tratam exclusivamente dos sistemas de automação de bibliotecas, todos os outros aqui analisados fizeram menção a uma distinção, nem sempre relativa, entre sistemas para bibliotecas e sistemas para centros de documentação ou documentários. Várias outras denominações são encontradas nestes textos:

- Bases de dados relacionais e bases de dados documentários;
- SGBDs relacionais e SGBDs documentários;
- SGBDs e SRIs;
- Sistemas de gerenciamento de bibliotecas e sistemas de gerenciamento de documentos (ou sistemas de gerenciamento de informações textuais);
- Programas de gerenciamento de bases de dados convencionais e sistemas de armazenamento e recuperação de textos (ou sistemas de gerenciamento de dados baseados em textos, sistemas de informações textuais e sistemas de recuperação de texto integral ou *full-text*);
- Sistemas para automação de médias e grandes bibliotecas (bibliotecas públicas, escolares e universitárias) e sistemas para pequenas e médias bibliotecas ou bibliotecas especializadas;
- Sistemas de automação de bibliotecas e sistemas documentários (ou sistema de armazenamento e recuperação de informação);
- *Softwares* para gestão de bibliotecas e *softwares* para gestão e pesquisa documentária.

Deve-se considerar que, mesmo que um grupo esteja citado em oposição a outro, nem sempre representam *softwares* distintos, mas configuram-se como sistemas heterogêneos, mistos ou complementares, eventualmente com mais características de um tipo ou de outro. No entanto, houve períodos e contextos em que esta distinção foi mais marcante, levando por isso, à recorrente referência a estas tipologias nesta

pesquisa. Neste sentido, julga-se necessário caracterizar cada um dos grupos, mesmo que não correspondam, atualmente, a uma realidade tão delimitada.

Sistematização desenvolvida em publicação espanhola por García Gutierrez e Lucas Fernandez (1987) com base no autor francês Faure (1982) é bastante pertinente. Neste estudo, os dois grupos são denominados como sistemas de gestão de bases de dados (SGBD) e sistemas de gestão documentária (SGD). Tratando-se de uma abordagem francesa do início da década de 80 e que foi, na mesma década, adotada para obra sobre o assunto da documentação automatizada da informação jornalística (periodística) de produção espanhola, indica também uma tendência neste idioma. De fato, García Gutierrez se aprofunda na caracterização destes dois tipos de sistemas (distinção válida para a época), mas centra-se nos sistemas de gestão documentária, descrevendo alguns produtos como BASIS, CAIRS, MINISIS, MISTRAL, Status, Texto e outros.

Os tópicos principais do quadro são os seguintes:

<b>SISTEMAS DE GESTÃO DOCUMENTÁRIA (SGD)</b>	<b>SISTEMAS DE GESTÃO DE BASES DE DADOS (SGBD)</b>
programas concebidos para armazenar e recuperar informação de tipo textual	orientados para o processamento convencional de dados, ou seja, para aplicações nas quais seja necessário acessar simultaneamente grupos pré-definidos de dados
capazes de manipular com facilidade informações textuais, inclusive as de tamanho grande	tem sido otimizado para o tratamento de valores numéricos ou de identificadores alfabéticos (termo simples ou expressão formada por alguns termos), manipulando com dificuldade informações textuais de tamanho grande
no momento de criação de uma base de dados, é preciso definir as zonas ou campos que conterão os dados suscetíveis de serem utilizados na interrogação (critérios de acesso), permitindo que uma pergunta possa ser uma combinação (em lógica booleana) de qualquer um destes critérios de acesso	a obtenção de respostas implica o conhecimento de todos os tipos possíveis de informações recuperáveis a partir de perguntas cuja estrutura foi pré-definida (alguns SGBD oferecem uma resposta análoga a dos SGD porém sem a mesma rapidez, sendo que tal limitação dificulta muito a realização de consultas verdadeiramente interativas)
a disposição dos dados não permite que as repetições sejam reduzidas mas otimiza o aproveitamento da memória do computador (como os dados são organizados seqüencialmente, e não sob a forma de tabelas, o espaço que eles ocupam não é previamente definido, mas depende dos caracteres necessários)	a disposição dos dados em forma de tabelas se efetua de tal maneira que as repetições (e, portanto, o mau aproveitamento da memória) são reduzidas ao máximo; em muitos casos, no entanto, a busca desta otimização aumenta o custo do sistema
muitos podem administrar e utilizar um tesouro	só podem manipular dicionários (nomenclaturas não-estruturadas)
apesar de ser capaz de assegurar a administração de muitos tipos de documentos, não é concebido para uma gestão coerente de suas relações eventuais, ou seja, cada documento é independente dos demais	é capaz de gerenciar as relações entre os objetos diferentes: por exemplo, uma base de dados factual sobre laboratórios de pesquisa, ações de pesquisa e pesquisadores e todas as relações existentes entre estes três tipos de objetos

Pode-se perceber, a partir dos textos analisados, a afirmação predominante de que os sistemas mais adequados para a Informática Documentária são, como o próprio nome indica, os *softwares* documentários. Estes programas são entendidos em termos globais como aqueles que tratam do documento e, nomeados no idioma inglês, como *softwares* de gerenciamento ou de recuperação de documentos. Os *softwares* documentários, além do tratamento de textos, podem atender ao OCR, à gestão de mídiotecas ou ao GED. Foram desenvolvidos nos Estados Unidos e na França entre 1955 e 1960, em grande parte por organizações e, comercialmente a partir dos anos 70, em especial, pela IBM (com o STAIRS) e pela Bull (com o MISTRAL). Neste período, o aperfeiçoamento destes sistemas já contava com a descrição do conteúdo e da busca a partir de palavras-chave de uma linguagem documentária controlada (mais tarde, de um tesouro). Antes dependentes dos programadores, com a chegada dos microcomputadores, os documentalistas passaram a ser formados em Informática Documentária, possuindo mais domínio na escolha e operação destes sistemas.

Para a gestão de acervos, os *softwares* documentários são preferencialmente acrescidos de algumas funcionalidades dos sistemas relacionais. No entanto, os sistemas relacionais foram amplamente adotados no mercado para a maioria das soluções chamadas comerciais. Como a área que desenvolve o tratamento e a pesquisa documentária é muito particular e pouco disseminada, também foram adotadas bases de dados relacionais, gerando baixa adequação de programas para fins documentários. Sendo assim, muitos autores, já na década de 90, escreveram sobre as questões envolvidas na alta disseminação dos programas comerciais e na falta de entendimento sobre a natureza dos sistemas adequados ao tratamento documentário.

Estes autores (NIEUWENHUYSEN, 1991, ROWLEY, 1998, MARCONDES, 1994) tratam dos sistemas para gerenciamento da informação textual estruturada, como as descrições bibliográficas. Para eles, este tipo de informação exige demandas específicas, para as quais é preferível trabalhar com registros, divididos em campos, por sua vez em subcampos.

Consideram que isto requer especificações distintas das comumente contempladas para as bases de dados não-bibliográficos comerciais, como Access, Foxpro, Oracle, Paradox e a geração dBase (II, III, III Plus e IV). Nestes programas, os campos são definidos por algum tamanho fixo, sem subcampos ou campos repetitivos. Assim, todos os registros devem ser idênticos, possuindo forma e tamanho fixos e sempre as mesmas ocorrências por campos. Essas características adequam-se às necessidades das aplicações comerciais, onde a informação é geralmente uniforme, como os registros de estoque ou de cadastro de funcionários. Como a informação não muda de tamanho, o *lay-out* fixo dispensa a necessidade de atualização constante e de tamanhos variáveis de campo. Estes programas são desenhados segundo o modelo relacional, onde as informações são organizadas em tabelas nas quais as linhas são os registros individuais de informação e as colunas, os diferentes campos. Por isso, são definidos como bases de dados relacionais.

Já os sistemas de gerenciamento de informações textuais (ou sistemas de gerenciamento de documentos) são considerados programas destinados à criação, manutenção e utilização de bases de dados formadas de textos. A fim de garantir a recuperação rápida de bases de dados textuais, esses sistemas caracterizam-se pela capacidade de manipulação de dados de tamanho variável, pelos índices baseados em arquivos invertidos e por recursos de recuperação bastante aperfeiçoados. Vantagens importantes desses sistemas são: preço, suporte, recursos de entrada de dados,

recursos de saída dos dados em tela ou impressos, recursos de recuperação da informação e facilidades com que esses sistemas podem ser integrados ou interfaceados com outros programas. Em oposição às bases de dados relacionais, formadas por tabelas, os sistemas de gerenciamento de informações textuais são compostos de dados seqüenciais, ou seja, cada registro é disposto seqüencialmente, independente do espaço que ocupem.

Entre os extremos traçados pelas bases de dados textuais e as relacionais, estão os programas de gerenciamento de informação textual estruturada, procurando oferecer o melhor dos dois grupos citados, como BRS/ Search, CAIRS, Cardbox-plus, CDS-ISIS, Inmagic, Polydoc, Pro-Cite, Questel, Texto, TINLIB/ TINMAN. Muitos deles são adequados para informação bibliográfica mas nem todos oferecem facilidades para as tarefas de uma biblioteca.

Já os sistemas integrados baseiam-se em estruturas de bases de dados relacionais e são considerados os mais adequados para gerenciar os serviços de uma biblioteca por possuírem módulos como: aquisição, catalogação, circulação, controle de publicações seriadas e pesquisa e recuperação de informações. Esses sistemas permitem processar informações textuais (como as bibliográficas) ao mesmo tempo em que integram os diversos módulos que compõem o processo biblioteconômico. Para tanto, há programas de gerenciamento de biblioteca que incluem um módulo para informação textual estruturada em uma ou mais bases. Além disso, são chamados de sistemas integrados baseados em estruturas de bases de dados verdadeiramente relacionais aqueles que eliminam a necessidade de duplicação de registros, operando com uma única base de dados bibliográficos, que pode ser utilizada por todo o sistema independente do serviço que o usuário necessite realizar.

Há também os programas para gerenciamento de textos completos e relativamente longos como o Folio Views, caracterizados por serem rigidamente estruturados.

Sendo assim, devido principalmente à dificuldade dos bibliotecários na explicitação das especificidades dos programas para bibliotecas, muitas vezes, programadores ficaram encarregados de criar sistemas, sem possuir conhecimentos específicos. Este tipo de confusão não existiu entre as grandes empresas de programas e equipamentos: a IBM desenvolveu o DB2 para aplicações comerciais e o STAIRS para aplicações bibliográficas, e a Bull, o IDSII para aplicações comerciais e o MISTRAL para aplicações bibliográficas. Da mesma forma, a UNESCO desenvolveu um gerenciador de informações textuais, o CDS-ISIS, para bases de dados bibliográficos, e o IDAMS (Internationally Developed Data Analysis and Management Software), para bases de dados numéricos.

Segundo Hopkinson (2001), a habilidade para relacionar dá mais escopo para aplicações de bibliotecas, como circulação, aquisição e controle de autoridades. Mas sistemas de recuperação textual possibilitam um nível enorme de sofisticação na pesquisa (booleana, combinações de termos no mesmo campo ou em campos distintos, adjacência, truncamento e manipulação de conjuntos). Onde há um grande número de relatórios para ser pesquisado por modos imprevisíveis, então uma ferramenta de recuperação textual é requerida em função de sua eficiência e flexibilidade de indexação. Uma ferramenta de recuperação de texto integral está à beira da análise textual e pode ser combinada com um sistema de gerenciamento de bases de dados estruturados para produzir o melhor de ambos.

Já Manzanos (1999) desenvolve uma comparação entre sistemas de recuperação de informação (SRI) e sistemas de bases de dados relacionais, com o fim de caracterizar o tipo de bases de dados que o programa CDS-ISIS representa, mesmo acreditando que esta distinção possa parecer um pouco anacrônica, dada a diversidade de programas existentes atualmente. Conclui que os sistemas de bases de dados relacionais estão em crise por não responderem a necessidades atuais como arquivos de imagens, ferramentas CAD e outros. Para estes casos, acredita em duas soluções: sistemas híbridos (aqueles que apresentam as vantagens dos dois modelos citados) ou bases de dados orientadas a objetos. No entanto, a vantagem da orientação a objetos está em pretender modelar a realidade de tal forma que o objeto seja a representação de uma entidade real, aproximando-se da visão intuitiva do usuário. A representação do documento como um objeto, com seus atributos e comportamentos próprios, está muito mais próxima da realidade e dos usuários que uma enormidade de tabelas sincronizadas que só pode ser modificada pela pessoa que as desenhou. No entanto, ainda não existem muitas aplicações comerciais que contemplem uma implementação completa do modelo orientado a objetos. Curiosamente, a estrutura física da orientação a objetos assemelha-se a do CDS-ISIS, usualmente categorizado como um gerenciador de informações textuais. Isto porque a forma de representar objetos em um suporte físico por um programa do tipo CDS-ISIS se dá a partir de estruturas de tamanho variável que permitem relações e trabalha partindo do conceito de que um registro representa a entidade que mais interessa, por exemplo, um documento.

Os sistemas mistos ou orientados a objetos são uma tendência impulsionada pelo uso crescente dos documentos eletrônicos (que unificam os tipos documentários, antes distintos) e pela recuperação remota destes documentos.

## **2.6 COMPARAÇÃO ENTRE OS TERMOS *INFORMATIQUE DOCUMENTAIRE* E *INFORMÁTICA DOCUMENTAL***

Segue uma análise comparativa sobre os dois autores selecionados (André Deweze e Mario Saquel) de cada uma das duas linhas adotadas para esta pesquisa, representadas respectivamente pelos termos *Informatique Documentaire* e *Informática Documental*. Estas linhas refletem o conteúdo considerado mais condizente com a proposta de pesquisa em função de dois aspectos principais presentes nas obras destes autores:

- A relativa abrangência dos aspectos conceituais e tecnológicos sobre os diversos serviços de informação documentária, quais sejam, as bibliotecas, os centros de documentação e os serviços de informação eletrônicos;
- A profundidade de abordagem calcada na busca da essência do uso da informação e das formas de possibilitar este uso, por meio de embasamento teórico e metodológico.

No entanto, como se observa abaixo, as duas obras têm características bastante distintas, podendo ser consideradas complementares quanto ao conteúdo que apresentam e à abordagem que desenvolvem.

Sendo assim, quanto à obra “*Informática Jurídica Documental*” de Saquel, pode-se inferir que o curso que leva este nome ministrado pelo próprio autor, no Chile, seja

voltado ao público dos profissionais de informação (bibliotecários, documentalistas e outros equivalentes). O fato de o curso ser oferecido em uma Faculdade de Direito fornece elementos para a discussão sobre a especialização em determinada área do conhecimento para os profissionais de informação. Neste sentido, o autor justifica a existência do profissional da informação jurídica como um especialista em Ciência da Informação e não como um profissional do Direito. Aponta para a necessidade da existência de um curso específico, com o objetivo de capacitar profissionais que dêem apoio a todos aqueles que precisam da informação jurídica para suas vidas, desde cidadãos comuns até os diferentes profissionais da área jurídica.

O autor também discorre sobre a escolha do termo que dá título ao livro, acreditando que seria mais adequado chamá-lo informação jurídica ou Ciência da Informação para a área jurídica, mas opta por *Informática Jurídica Documental* devido a sua utilização já majoritária na literatura.

Em comparação à obra de Deweze, “Informatique Documentaire”, o texto do chileno é muito mais conceitual, claro e objetivo, por certo por ser um texto acadêmico. Por sua vez, o texto francês apresenta-se repetitivo em muitos momentos.

Esta dispersão também pode ser explicada pelo fato de o livro francês ser mais voltado à explicação dos processos informáticos. Ao invés de fornecer uma explicação conceitual para algo, a explicação se dá pelos seus elementos, suas características e formas de funcionamento. E a questão é que a Informática é mais difícil de ser apreendida no meio impresso sendo mais compreensível no momento da operação.

Das duas obras, interessa essencialmente para este trabalho o fato de que ambas integram, à definição de Informática Documentária, o conhecimento e a atividade do tratamento temático da informação, ou seja, a análise do conteúdo dos documentos. A inclusão da análise de conteúdo nesta conceituação deve-se à informatização que veio influenciando este processo, ao agilizar algumas etapas braçais, muitas vezes, fazendo aparecer problemas que não eram antes identificáveis com facilidade. Neste sentido, a Informática vem tentando, inclusive, contribuir também nos processos intelectuais da indexação. Contudo, a razão para que esse aspecto do tratamento documentário, esteja incluído na Informática Documentária é, segundo esses autores, consequência de seu valor para a recuperação efetiva de informações.

Quanto ao foco de cada autor, pode-se dizer que Saquel dá importância à organização e tratamento da informação com fins de recuperação garantida, dado o excesso informacional da área jurídica. Apresenta uma evidente preocupação com a qualidade no tratamento da informação jurídica para que o profissional e o cidadão comum possam fazer uso dela. Já para Deweze, importa a análise correta e adequada dos problemas da Documentação objetivando sua via de resolução informática. Parece claro que os problemas da Documentação de que fala Deweze estão relacionados com o acesso do usuário à informação. Ou seja, enquanto o primeiro focaliza o fim, o outro centra no processo, mas ambos desenvolvem as duas partes em suas obras, apontando preocupações semelhantes. Demonstram, apesar da distinção focal que, só com processos adequados atinge-se o fim desejado e, por sua vez, um processo só se justifica se visar um fim claramente definido.

Assim, Saquel não trata de informatização de bibliotecas, processamento de dados, tipologias de *softwares*, escolhas de sistemas de automação e tecnologias da informação em geral. No entanto, discorre sobre os serviços de informação (quase sempre, informatizados) de um centro de documentação em uma organização complexa.

Pode-se dizer que este autor não discorre tanto sobre a tecnologia mas sobre o processo documentário pois sua intenção é tratar da informação jurídica, o que remete à relação entre informação e Direito e não entre informação e tecnologia.

Como conseqüência da distinção dos autores, as origens dos termos *Informática Documental* e *Informatique Documentaire* seguem a mesma linha. Ou seja, na área jurídica, segundo Saquel, a origem do termo está ligada à complexidade subsequente ao crescimento das informações jurídicas e à dificuldade dos profissionais juristas em lidar com esse quadro a partir dos procedimentos tradicionais de tratamento e disseminação de informações. Já o termo *Informatique Documentaire* (obra de Deweze) é justificado pela crença de que a intervenção da Informática criou a necessidade de que os documentalistas estejam capacitados para tarefas de análise e aplicação da ferramenta informática e de concepção e utilização de tesouros. Fica evidente novamente que o primeiro tem o foco no usuário final da informação e o segundo tem o foco no intermediador da informação: com isso, Deweze é mais específico ao entrar detalhadamente nos aspectos e nas operações dos serviços e produtos de informação.

Quanto às definições apontadas por cada um: *Informática Jurídica Documental* é “o processamento da informação jurídica (legislação, doutrina e jurisprudência) em suportes computacionais para sua posterior recuperação”, e *Informatique Documentaire* é “o conjunto de aplicações da Informática à Documentação, técnica que se refere às intervenções da Informática nas diversas fases de produção e utilização de documentos: produção de textos, difusão pelo editor, gestão pela biblioteca, análise e indexação para a constituição de bases de dados bibliográficos e para a difusão seletiva, lógicas para a interrogação dessas bases de dados”. Como já foi dito, o primeiro diz o que é e para quê e o segundo diz o que é e do que trata. O trabalho sobre Informática Jurídica Documentária está mais voltado para a informação jurídica, enquanto o outro faz uma tentativa de definir, de fato, a Informática Documentária, como a intervenção da Informática à Documentação incluindo aí as diversas fases de produção e utilização de documentos: o primeiro é conceitual e trata do tratamento e disseminação da informação, o segundo é operacional e focaliza a produção e gestão da informação.

Saquel define análise documentária como: “operação que consiste em selecionar as idéias informativamente relevantes de um documento a fim de expressar seu conteúdo sem ambigüidades para recuperar a informação nele contida”. Aqui estão incluídas tanto a análise formal quanto a análise de conteúdo (ou, externa e interna, respectivamente). Esta definição é bastante significativa pois usualmente a análise de conteúdo é considerada a própria análise documentária, como se a representação descritiva (análise formal ou externa) também não fosse uma parte do processo de análise documentária. Deweze não conceitua explicitamente, mas intitula o capítulo sobre este item apropriadamente com o termo “análise de conteúdo”, dando a entender o mesmo que Saquel. A diferença é que Deweze discorre sobre Informática Documentária, procurando aquilo que o nome parece indicar e, portanto, desenvolvendo o tratamento temático da informação segundo suas diversas possibilidades automatizadas. Já Saquel desenvolve um conceito sobre Informática Documentária que não abarca a Informática, a não ser quando discorre sobre Recuperação da Informação ou indexação automática, mas mesmo assim, seu aprofundamento é mais conceitual e portanto, não se atém a tópicos da ferramenta informática.

Quanto à nomenclatura utilizada para indicar os profissionais em questão, constata-se que ambos referem-se ao profissional como documentalista ou profissional/especialista da informação. Já a área propriamente dita é nomeada por Saquel como

Ciência da Informação. Em momento algum cita a Biblioteconomia e Documentação e qualquer relação que estas áreas possam ter com a Ciência da Informação: apenas aponta que, com a explosão da informação e a especificidade do trabalho jurídico, faz-se necessária a prestação de serviços de informação que seja mais dinâmica, o que sugere uma referência às atividades tradicionais e burocráticas de longa data desenvolvidas pela Biblioteconomia. Deweze, por outro lado, trabalha com o termo Documentação e outros termos relativos a ele, como cadeia documentária, serviço ou centro de documentação, mas também utiliza o termo biblioteca. Não se aprofunda nos conceitos dos termos mas afirma que os processos documentários que descreve devem servir também ao estudante de Ciência da Informação. Entende-se que, como Saquel trabalha com a fundamentação teórica, refere-se sempre à Ciência da Informação, enquanto que, Deweze, mais centrado nas operações e nas características dos sistemas automatizados para tal, adota o termo mais técnico, Documentação.

Sobre o público que se pretende atingir, já foi mencionado que, muito provavelmente, o alvo para o curso de *Informática Jurídica Documental* seja o profissional da informação (bibliotecários e documentalistas) com interesse em especialização para a área jurídica. Já o público do livro sobre *Informatique Documentaire* está literalmente especificado pelo autor: profissionais de Informática, documentalistas, estudantes de Ciência da Informação e responsáveis por serviços de informação interessados em se aprofundar nos mecanismos de seu tratamento informático. Deweze parece almejar um público mais amplo por incluir os profissionais de Informática. Isto é verificável quando Saquel discorre sobre sistemas de bases de dados e seus usuários: inclui no grupo de usuários finais tanto bibliotecários e documentalistas quanto pesquisadores, excluindo os primeiros no grupo dos administradores das bases de dados. Deweze faz exatamente o contrário, considerando essencial que o documentalista possua alguns domínios de administração para bases de dados.

Importante citar ainda, em Saquel, o valor dado ao tratamento da informação com fins de resolução de problemas dos mais diversos cidadãos e, portanto, a relevância que ele confere à área documentária.

Ainda na mesma obra, é essencial para a caracterização da área Informática Documentária, a explicação que o autor fornece logo de início com relação à inclusão da Informática para a constituição da Ciência da Informação. Levanta a questão da prática que envolve esta ciência, sendo a Informática, juntamente com metodologias e aportes de outras áreas, o instrumento (ou ferramenta) que pode garantir que nenhum dado produzido se perca assim que seja dado a conhecer.

É relevante notar que as noções levantadas sobre Informática Documentária estão próximas das definições existentes para a Ciência da Informação. Ou seja, a Informática Documentária segue o caminho da superação das práticas da Biblioteconomia e das técnicas tradicionais da Documentação, contribuindo assim, para a elaboração da Ciência da Informação.

## **2.7 INFORMÁTICA DOCUMENTÁRIA**

Mario Saquel, autor do livro “Informática jurídica documental”, acredita que seria mais adequado chamá-lo de informação jurídica ou Ciência da Informação para a área jurídica, mas opta por Informática Jurídica Documentária devido a sua utilização já

majoritária na literatura. No Brasil, o termo Informática Jurídica foi encontrado na denominação de eventos e cursos universitários da área jurídica. O termo tem sido adotado para as questões da Informática e do Direito como o direito do autor e os delitos de Informática, os sistemas informatizados especialmente construídos para o conteúdo jurídico e, com menor incidência, os aspectos documentários da informação jurídica.

Já a Informática Médica, descrita no livro de Costa Carballo, não indica o aspecto documentário no sentido estrito, mas principalmente, os sistemas informáticos que possuem simulações dos conhecimentos e comportamentos de um especialista em determinado tema para ajuda ao diagnóstico e tratamento.

O estudo terminológico, realizado por Vicentini, em 1971, aponta para a evolução dos termos Biblioteconomia, Documentação e Informação e os conceitos e técnicas aí envolvidos. Descreve o impasse Biblioteconomia X Documentação que dominou por muito tempo o panorama internacional, sendo tema de publicações especializadas e de eventos e cita as várias tendências que fizeram parte desta discussão. Opta pelo ponto de vista de que esta separação é um retrocesso pois bibliotecários e documentalistas devem ser especializações de uma mesma profissão. Em seguida, por volta da Segunda Guerra, o termo Documentação passa a ser substituído por Informação. Sendo assim, o termo Informação surgiu, criou raízes e se ramificou, o que se percebe pelos vários campos e setores da Ciência da Informação. Numa destas ramificações está a origem da Informática que, por este motivo, confunde-se com a própria Ciência da Informação a julgar, por exemplo, por uma de suas primeiras aparições, a partir da fundamentação realizada pela pesquisa na ex-União Soviética, onde o termo *Informatika* contava com os conceitos do que se denominava *Information Science* nos Estados Unidos. Vários trabalhos foram publicados na Europa e Estados Unidos sobre Informática (*Informatics* no idioma inglês, *Informatique* em francês). O termo *Informatics* foi estabelecido com base no processo de formação do nome de outras ciências como *mathematics*, *semiotics*, *astronautics*, *electronics*. Saquel, no entanto, cita que o termo Informática tem derivação do francês *informatique*, que é composto de *information* e *automatique*, portanto, automação da informação. O estudo de Vicentini conclui pela definição do termo Informática como:

- a ciência que estuda o aspecto teórico e prático das atividades de coleta, processamento, armazenamento, recuperação e disseminação da informação por meio da mecanização e da automação;
- sistema de informação relacionado a um determinado campo do conhecimento, com o emprego de equipamento “cibernético”, como por exemplo, Informática Agrícola, Informática Aeronáutica etc.

Acreditava-se, a esta altura, que o termo Informática poderia unificar os conceitos Biblioteconomia, Documentação e Informação e que era mais adequado que a expressão Ciência da Informação para indicar os aspectos teóricos, sociais e aplicados desta área, a exemplo da pesquisa na União Soviética.

Como decorrência desta origem e dos desenvolvimentos posteriores, o termo constituído pela palavra “Informática” seguido de outra palavra designativa de uma área do conhecimento, passou a indicar o estudo e o uso de recursos tecnológicos eletrônicos para o tratamento, armazenamento e recuperação de informação desta área de conhecimento. Para enfatizar os princípios documentários no sentido estrito, este conceito é, muitas vezes, também adotado para compor a expressão, como no caso da Informática Jurídica Documentária.

Outras formas de denominação, no entanto, são utilizadas para indicar áreas da mesma natureza. Por exemplo, o geoprocessamento indica uma metodologia de tratamento de informações geográficas na forma de textos, números ou imagens (como o levantamento e cruzamento de dados socioeconômicos em espaços e períodos determinados) concretizada por tecnologias especificamente desenvolvidas para isto chamadas de sistemas de informações geográficas (SIG). Neste sentido, o geoprocessamento poderia ser adequadamente nomeado como Informática Geográfica.

## **CAPÍTULO 3 - TENTATIVAS DE PREENCHER LACUNAS HISTÓRICAS**

### **3.1 BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO: ABORDAGEM ESPECULATIVA**

A análise da produção sobre Informática Documentária em diversos idiomas apresentada no segundo capítulo reforça as características levantadas no histórico traçado no primeiro capítulo sobre as áreas conhecidas como Biblioteconomia, Documentação e Ciência da Informação. Era de se esperar que assim fosse uma vez que, a história que uma área de conhecimento e de atuação constrói, influencia cada uma de suas partes no decorrer do tempo e em função da forma e do local onde se manifestam.

Isto contribui para a existência de um quadro heterogêneo e, muitas vezes, confuso, que dificulta a efetivação de uma análise mais complexa destas realidades. Uma inconsistência observada no histórico traçado refere-se ao período entre o início da Documentação no final do século XIX e o surgimento da Ciência da Informação nos anos 50 do século seguinte. Não foram devidamente explicitados os fatos que caracterizaram este intervalo de tempo e que, supostamente, explicam a relação de existência entre uma área e outra. De fato, a Ciência da Informação proclamada então pelos Estados Unidos, apesar de carregar alguns pontos da Documentação, que tem origem européia, parece não assumir esta influência.

Para facilitar a compreensão deste quadro, recorreu-se a Buckland (1996), autor dos Estados Unidos mas versado na literatura em francês, que explorou o tema da distinção entre estas culturas na área da informação documentária e bibliotecas, inclusive com foco para a questão tecnológica. De fato, a pretensão do autor foi investigar o papel da Biblioteconomia, da Documentação e da Ciência da Informação nos Estados Unidos mas, para tanto, partiu para a investigação do papel destas áreas também na Europa. Buckland afirma, no entanto, que a discussão deve ser vista como uma análise especulativa.

Assim, buscando traçar a história cultural e intelectual da Biblioteconomia nos Estados Unidos, durante o século XX, são considerados três pontos:

- Experiências e inovações técnicas e tecnológicas realizadas pelos documentalistas europeus, desde a década de 30, parecem ter sido substancialmente ignoradas pela Biblioteconomia dos Estados Unidos até depois da Segunda Guerra Mundial;
- As inovações técnicas e tecnológicas foram uma força vital na Biblioteconomia nos Estados Unidos no final do século XIX e novamente depois da Segunda Guerra Mundial, mas não no intervalo entre estes períodos. A razão deste fenômeno parece não ser clara;

- Houve uma intensa mas geralmente insatisfatória controvérsia conhecida como Ciência da Informação X Biblioteconomia, depois da Segunda Guerra Mundial, que se dissipou no final dos anos 70. Falta investigar qual a origem desta controvérsia e porque isto aconteceu.

Buckland procura explicar estas questões a partir do papel da tecnologia na Biblioteconomia e do impacto da Graduate Library School, da University of Chicago, dos Estados Unidos.

O final do século XIX presenciou, tanto inovações técnicas e tecnológicas na Biblioteconomia dos Estados Unidos, como sua própria formação naquele país. Na primeira metade do século XX, houve uma expansão de serviços de bibliotecas por toda parte, especialmente nas áreas rurais, mas com base em uma estabilidade técnica e tecnológica, o que foi completamente diferente do período seguinte, ou seja, a partir da Segunda Guerra Mundial.

No entanto, ocorreram vários desenvolvimentos tecnológicos para Biblioteconomia e Documentação durante esta primeira metade do século XX até antes da Segunda Guerra Mundial, apontando que o interesse nas inovações técnicas e tecnológicas não era ausente neste período. Ao contrário, as características correntemente assumidas sobre a biblioteca eletrônica do século XXI, como o armazenamento compacto, a facilidade de reprodução, o acesso remoto ao texto integral, o hipertexto, os equipamentos de grande capacidade, a pesquisa sofisticada em sistemas de indexação complexos e outras noções contemporâneas, foram vistas e discutidas, ao menos como esboços, por idealistas práticos no Congresso Internacional sobre Documentação de 1935, antes da invenção dos computadores.

Contudo, apesar da existência destes experimentos, o interesse não era geral e eles não foram adotados significativamente. As duas principais inovações em bibliotecas nos Estados Unidos foram a fotoestática (projeção fotocopiadora em papel sensível) de 1912 e a microfilmagem. Por volta dos anos 30, a microfilmagem de jornais tornou-se comum e bibliotecas inovadoras estabeleceram laboratórios de fotoduplicação. Houve também a adoção de desenvolvimentos do século XIX, particularmente a duplicação a tinta e estêncil, telefones e máquinas de escrever e o uso de regras para catalogação, classificação e arquivo.

Vários pesquisadores estadunidenses escreveram sobre a tecnologia na literatura profissional de Biblioteconomia mas, em todos eles, as abordagens dos documentalistas europeus estavam ausentes. Além disso, a nova tecnologia de informação não foi ignorada fora dos círculos de bibliotecas.

Um dos elementos que ajuda a explicar por que estes desenvolvimentos foram grandemente ignorados nos círculos de bibliotecas antes da Segunda Guerra Mundial, é a influência da **Graduate Library School, da University of Chicago**, que foi o centro intelectual da Biblioteconomia dos Estados Unidos por volta dos anos 30 até os anos 60.

A obra de Pierce Butler, "Introduction to Library Science", de 1933, é considerada uma obra emblemática da assim conhecida Escola de Chicago e suas influências. O próprio Butler admitiu que a obra se tornaria obsoleta com o tempo, mas não houve produção que a substituísse. Desta obra, três pontos merecem destaque:

- Ausência de modelos, tecnologias, técnicas e habilidades de gerenciamento que forneçam um efetivo e eficiente serviço de biblioteca;

- Não há menção a grandes nomes que contribuíram com a Biblioteconomia como Panizzi, Cutter, Dewey, Jewett, Bliss e outros e a impressão de que suas contribuições não aconteceram (como se Adam Smith e Lord Keynes não fossem citados na área da Economia);

- Não trata dos aspectos científicos dos serviços de biblioteca (a despeito do título da obra) mas do *status* social dos bibliotecários e da função social das próprias bibliotecas e afirma que a atenção da Biblioteconomia deveria se voltar do processo para a função.

Sobre este último item, Buckland argumenta que a ciência examina fenômenos e processos mensuráveis. Função, como objetivo social, está na área dos gerenciadores, políticos, teólogos e outros, preocupados com prioridades éticas e sociais, mas não na área dos cientistas. Acredita, assim, que a obra de Butler demonstra uma abordagem limitada do escopo da Biblioteconomia em termos de conteúdo, superficial em termos científicos e, em particular, ausente de interesse em inovações técnicas e tecnológicas em bibliotecas.

A Escola de Chicago era famosa por ser interdisciplinar com muito interesse por Sociologia, Ciência Política e Educação, mas houve interesse limitado naquilo que as Humanidades poderiam contribuir para a Biblioteconomia. Houve alguma atenção para a administração de bibliotecas mas a retórica do “método científico” e a afirmação de ser interdisciplinar mascararam o fato de que seus interesses em Biblioteconomia eram limitados comparados com o tempo de Dewey e com os dias atuais. Pesquisadores da década de 70 criticaram a Escola de Chicago por não sustentar uma abordagem científica.

A direção da Graduate Library School apontava para um desejo geral de mudança do “técnico” (procedimentos) para algo mais “científico” (no sentido de erudito, preferencialmente, com ênfase quantitativa). Para outros, os bibliotecários deveriam reconhecer e ministrar as necessidades de grupos individuais no uso de livros, considerando as técnicas de biblioteca como menores. Em seu início, a Graduate Library School teve pouca familiaridade com Biblioteconomia.

A análise da produção dos anos 30 da revista *Library Quarterly*, publicada pela Graduate Library School, constata que os aspectos tecnológicos foram realmente evitados. Também a comparação dos volumes da *Library Quarterly* entre 1935 e 1939 com os anais dos congressos da FID para os mesmos anos, demonstra um contraste de interesses que evidencia escolas marcadamente diferentes.

No entanto, há indícios que demonstram que a Escola de Chicago tinha consciência da relevância das inovações técnicas e tecnológicas para serviços de biblioteca: Otlet e seus colegas do então Instituto Internacional de Documentação (IID) publicaram sobre o tema desde 1895; o American Documentation Institute foi fundado em 1937; a principal manifestação de inovação em tecnologia de biblioteca, neste período, foi a criação dos laboratórios de fotoduplicação em bibliotecas progressistas, principalmente preocupadas com microfilmagem de textos; pesquisadores da Escola de Chicago, especialistas nesta tecnologia de microfilmagem, tiveram um papel ativo no Congresso Mundial da Documentação Universal, em Paris, em 1937, no qual, outros notáveis europeus (Otlet, entre eles) também apresentaram seus trabalhos; a partir de 1938, teve início a publicação da revista dedicada à inovação tecnológica, o *Journal of Documentary Reproduction*, publicado em Chicago pela American Library Association; e um curso sobre microfilme foi introduzido na Graduate Library School em 1939. De fato, esta

Escola era consciente e interessada em tecnologia e em inovação tecnológica, a tal ponto que estimulou os documentalistas.

Em contraste com a Escola de Chicago houve, ao mesmo tempo, um interesse por aspectos tecnológicos dos serviços de biblioteca no Massachusetts Institute of Technology (MIT). Desde o início dos anos 30, Vannevar Bush, um professor do MIT e administrador acadêmico, tentou persuadir fundações e corporações a financiar sua máquina denominada seletor rápido para recuperação de informação em bibliotecas, e também, o uso de microfílm, células fotoelétricas e circuito digital para realizar recuperação rápida de textos. O famoso artigo de Bush, “As we may thing”, embora publicado em 1945, foi escrito no final dos anos 30.

Buckland imagina que uma Graduate Library School no MIT teria menor foco em pontos como leitura e papel das bibliotecas públicas. Mas a diferença significativa poderia ter sido uma maior ênfase para inovações técnicas e tecnológicas e para modelos de serviços de bibliotecas especializadas em informação em ciência, engenharia e indústria. Poderia se esperar um interesse em bibliotecas especializadas, e também, assim como os documentalistas, uma preocupação ativa com problemas de informação tanto dentro quanto fora de bibliotecas. Também considera que os cientistas da informação do pós-guerra, a propósito do desdenhamento das escolas de Biblioteconomia, devem ter tido vontade de se associar a uma Graduate Library School baseada na orientação técnica de um MIT. Além disso, as caras e ineficazes pesquisas e desenvolvimentos de engenheiros poderiam ter sido produtivas se não tivesse havido uma separação institucionalizada e atitudinal entre engenheiros e bibliotecários. Ao contrário, acredita que o que teria feito diferença, como agora se vê, seria uma Graduate Library School com orientação técnica, representada em um ambiente no qual aqueles com interesses técnicos, sociais e humanos poderiam coexistir, colaborar e se sentir em casa.

Nos anos 50, uma mudança tornou-se aparente na literatura dos Estados Unidos sobre Biblioteconomia em função da tentativa de reconstruir, analisar e interpretar o pós Segunda Guerra Mundial, ocasionando uma discussão baseada na dicotomia **Ciência da Informação X Biblioteconomia**. O principal elemento foi a presença de indivíduos voltados à tecnologia de fora da Biblioteconomia que estavam procurando conduzir a nova tecnologia para resolver velhos problemas. Isto poderia não ter gerado uma contenda se, naquele tempo, a Biblioteconomia tivesse sido bem povoada com indivíduos sofisticados tecnologicamente e com experiência e competência nos problemas e oportunidades de tecnologia em serviços de biblioteca, Documentação e gerenciamento de informação especializada. Contudo, a linha dominante, representada pela pesquisa de Chicago, estava longe de contemplar estes aspectos.

A disputa sobre Ciência da Informação e Biblioteconomia poderia ter ocasionado uma grande mudança, por meio da retomada das primeiras posições dos documentalistas europeus e, além deles, de Dewey, Cutter e outros e das inovações técnicas e tecnológicas do final do século XIX. A ênfase inicial em Ciências Sociais da Graduate Library School de Chicago, apesar de desejável, poderia ter sido diluída ou contrabalanceada com o tempo.

Na década de 50, Margaret Egan e Jesse Shera, membros da Graduate Library School, apontam a atenção dos bibliotecários, durante os anos anteriores, sobre a revolução da comunicação de massa e seu provável efeito sobre os serviços de biblioteca para o leitor em geral, enquanto poucos estavam se preocupando com a revolução da organização e serviços de biblioteca, que foi tratada por outro campo, assim nomeado “comunicação de informação especializada”. Shera criticou a falta de atenção

para os aspectos técnicos do trabalho em biblioteca e a rápida acumulação de técnicas especializadas desenvolvidas pelos documentalistas e especialistas de informação. E apontou a própria Graduate Library School como uma das principais responsáveis pela repulsa pelas habilidades técnicas.

Com o tempo, uma mudança lenta para Biblioteconomia e Ciência da Informação ocorreu nas escolas de Biblioteconomia dos Estados Unidos. Acreditava-se que havia a necessidade de incluir novos estudos de Ciência da Informação e Documentação no currículo básico, o que passou a ser um desafio contínuo desde os anos 60. O que estava sendo introduzido era, mais ou menos, uma versão atualizada da espécie de material que esteve presente nos congressos da FID dos anos 30.

A tecnologia de serviços de biblioteca esteve estável por mais de meio século e os bibliotecários sentiram mais crescimento que mudança. A transformação que ocorreu após a Segunda Guerra Mundial tornou-se uma disputa pois o que estava em questão era mais que o uso de tecnologia. Cientistas da informação eram percebidos por muitos bibliotecários como uma espécie de ameaça até o final dos anos 70 quando, o debate Ciência da Informação X Biblioteconomia dispersou em segmentos diferentes aquilo que era uma ênfase construtiva sobre teoria, modelos e serviços.

Além disso, especialmente, durante a Depressão, os fundos diminuíram e as inovações tecnológicas nem sempre eram consideradas uma prioridade.

Outra possibilidade deste quadro é que a aparente ausência de interesse em **Documentação nos Estados Unidos** era, ao menos em parte, uma questão de semântica. A documentalista e bibliotecária francesa Suzanne Briet, personalidade de referência em Documentação, escreveu sobre esta área nos anos 50 e realizou um *tour* pelos Estados Unidos observando bibliotecas. Ela comenta que, embora o termo Documentação seja pouco conhecido nos Estados Unidos, suas técnicas eram habilmente praticadas na forma de serviço de referência e serviço de biblioteca especializada, ambos, dentro e em separado das grandes bibliotecas gerais. Sua explicação era que, por causa destas práticas terem sido desenvolvidas na Biblioteconomia dos Estados Unidos antes que na Europa, não houve uma necessidade, como na Europa, para o termo “centro de documentação”. Esta explicação dá vazão a várias questões sobre os bibliotecários especializados estadunidenses e seu relacionamento com os documentalistas europeus.

Por fim, fica claro que as três questões inicialmente propostas estão intimamente relacionadas.

Uma razão para que os modelos e a tecnologia fossem de interesse limitado na Biblioteconomia nos Estados Unidos na primeira metade do século XX, é que a maior influência acadêmica foi de um grupo, conhecido como Escola de Chicago, que estava empenhado em outro sentido. Nos anos 30, esta Escola e a dos documentalistas europeus representavam linhas marcadamente distintas. Depois da Segunda Guerra Mundial, houve um período de tensão pois a abordagem dominante da Biblioteconomia dos Estados Unidos, orientada para o não tecnológico e para as Ciências Sociais e representada pela Escola de Chicago, foi desafiada, rivalizada e transformada pelo retorno do interesse em modelos e tecnologia. Os assuntos que tinham interessado aos documentalistas europeus emergiram como uma poderosa força na Biblioteconomia dos Estados Unidos 20 anos depois que na Europa. Havia, neste momento, novas e poderosas máquinas e surgiu, depois de alguns anos, um novo nome: Ciência da Informação. Os indivíduos que conduziram a mudança comumente vinham de fora da

Biblioteconomia e houve pouca associação com a Europa, então devastada pela Guerra. Os documentalistas europeus dos anos 30, que tinham escrito principalmente em francês e alemão, foram amplamente esquecidos.

A temporária falta de ênfase em modelos e tecnologia contribuiu para uma prolongada falta de identidade e direção nos departamentos acadêmicos de estudos de biblioteca e de informação. O propósito de uma escola profissional baseada na universidade deve estar centrado em modelos para os melhores serviços. A ausência desta preocupação central levou à carência de objetivos, além de sustentar a continuidade de treinamento em procedimentos, uma preocupação com o profissionalismo e pouca base convincente para uma agenda de pesquisa. Com isto, uma visão coerente para pesquisa e para a educação profissional na universidade, é também ausente.

### **Sobre a distinção cultural estadunidense e francesa em Informática Documentária**

Como em muitos outros momentos desta pesquisa, o sentido de localização histórica é gratificante por promover a contextualização de problemas e a construção de identidade científica. Sendo assim, a análise da obra de Pierce Butler, de 1933, que marcou muitos dos preceitos da Escola de Chicago, a qual, por sua vez, foi linha dominante em Biblioteconomia até os anos 60, aponta as origens de entendimentos e posturas da área, difíceis de compreender sem estas referências. De fato, algumas das características da Biblioteconomia presentes na obra de Butler são: ausência de modelos, técnicas, tecnologias e habilidades de gerenciamento; não referência a pesquisadores elementares da área; e discussão sobre o papel social das bibliotecas e do *status* profissional do bibliotecário. Estes itens parecem tão atuais para a realidade local (brasileira) quanto o seu reconhecimento pode contribuir para superar a falta de linguagem própria na Biblioteconomia.

Além disso, parece que o enfoque humanístico, construído a partir da discussão sobre o papel social das bibliotecas, era e é ausente de modelos para este fim, o que causou o efeito contrário, uma vez tratar-se de uma área que também é técnica e tecnológica. Ou seja, o resultado é uma técnica que não se desenvolve, é pouco contextualizada e não sofre interpretações nem é discutida, transformando os aspectos sociais em discussões sem sentido. Não há construção de vínculos entre o lado humanístico e o técnico, levando ao esvaziamento de cada um deles. Talvez isto seja decorrente, entre outros, do fato, não citado por Buckland, de que a ênfase para as Humanidades realizada pela Escola de Chicago como um todo era essencialmente estatística e centrada nas questões das massas, em oposição à abordagem filosófica que predominava na Europa e que, até hoje é marca de países como França e Alemanha. Tanto que, os documentalistas europeus, apesar de suas pesquisas e experimentações técnicas e tecnológicas, não relegaram as preocupações com a leitura e com a Biblioteconomia não especializada. Esta mesma característica foi descrita anteriormente com relação ao desenvolvimento, pelos russos, na década de 60, da área que denominaram de *Informatika*. A abordagem russa, assim como a dos documentalistas europeus, incluía tanto as áreas especializadas quanto as áreas gerais e a leitura pública pois, desenvolveu os elementos que possibilitaram o reconhecimento de interesses distintos a partir de uma base comum.

Shera e Egan discorreram sobre a primeira ruptura entre Biblioteconomia e Bibliografia, por volta de 1850, impulsionada pela proliferação dos artigos de periódicos. Os princípios da Bibliografia foram sistematizados pelo projeto de Otlet e sua efetivação enquanto área, a Documentação, realizou-se, de fato, durante a Segunda Guerra Mundial, que exigiu o tratamento analítico de documentos diversos. A segunda ruptura ocorreu reforçada pela sedimentação da Documentação, a partir dos anos 50, que conduziu à formação de uma espécie de Biblioteconomia “humana” de um lado e “especializada e tecnológica” de outro.

Parece que a Biblioteconomia e a Documentação eram áreas mais únicas na Europa que a Biblioteconomia e a Ciência da Informação nos Estados Unidos. Ou seja, o que a princípio sugere uma divisão na Europa é, de fato, a constatação de interesses distintos em uma mesma área, os mesmos interesses que os Estados Unidos fragmentaram em várias partes e que dificilmente formam um todo, a não ser talvez sob o ponto de vista essencialmente funcionalista. Assim, reforça-se que a relação entre Biblioteconomia e Ciência da Informação é decorrência de um processo de continuidade da discussão sobre a oposição entre Biblioteconomia e Documentação. Esta divisão causa incômodo ainda hoje para aqueles que buscam compreender teoricamente a área como um todo e desenvolver uma prática fundamentada e abrangente.

Interessa notar que, a consideração de Suzanne Briet de que a diferenciação destas culturas refere-se apenas a uma questão de semântica, deixa margem a dúvidas. Ela afirma que os aspectos documentários estavam presentes nos Estados Unidos nos serviços de referência e nos serviços das bibliotecas especializadas, mas estes serviços não foram explicitados para que se pudesse detectar as formas e a intensidade do uso das técnicas e princípios da Documentação. No caso dos serviços de referência das bibliotecas gerais, como estes serviços poderiam ser desenvolvidos a partir de buscas elaboradas, se o processo de tratamento do conteúdo dos documentos não era refinado?

A visão abrangente de Buckland ajuda a entender a surpreendente literatura sobre Informática Documentária na França. Explica, em parte, as razões de a produção francesa apontar para a essência, ou seja, para o conteúdo dos documentos (e as formas de representação e de acesso destes conteúdos), mas acima de tudo, de fazê-lo de forma pioneira desde os anos 30 e, provavelmente, não ter sido superada por outra cultura em suas peculiaridades. Assim, quando os Estados Unidos voltaram-se para o interesse na tecnologia aplicada à Biblioteconomia, mesmo que tivessem se preocupado com o que os europeus já haviam produzido, fariam e fizeram diferente a construção em Informática Documentária. Uma diferença está no fato de que abordagens holísticas sobre uma área correm menos riscos de apresentar lacunas ou partes que não se conversam. Assim, nos Estados Unidos, o tratamento de conteúdo e a manutenção de seus registros em máquinas permaneceram com uma área e o gerenciamento e preservação dos suportes físicos com outra. Estas áreas, assim desenvolvidas, são pouco dialogáveis e de difícil complementação.

No entanto, a partir da década de 60, os Estados Unidos atropelaram o mundo com o desenvolvimento da automação de bibliotecas. Assim, quando se trata de bibliotecas estritamente, as literaturas atuais da Espanha, França e Reino Unido apontam unanimemente os Estados Unidos como referência. Há que se admitir a eficiência e competência adquirida pelos Estados Unidos na realização daquilo para o que se propôs: a gestão de acervos e seu uso massivo pela população e a disseminação de bases de dados especializadas a segmentos específicos. Também se deve considerar que são os mesmos Estados Unidos que estão envolvidos no grande desafio de aproximar estas

duas áreas (apesar da suspeita sobre seus métodos), dado que a evolução tecnológica tornou transparente sua identidade comum. Assim, os sistemas para gerenciamento de bibliotecas são, cada vez mais, sistemas para gerenciamento de documentos (sejam os documentos da biblioteca ou não), além de serem crescentemente centrados nas necessidades de seus usuários. Resta saber como unir aquilo que se tornou tão distinto, apesar de semelhante em suas bases. Além das questões técnicas e tecnológicas, o aspecto cultural é de tal monta que, a elaboração de pelo menos um século e meio (considerando o período de 1850 como o da primeira ruptura), necessitará de muitas décadas de re-elaboração. No entanto, apesar da tendência de aproximação, são justamente as marcas culturais que estão em questão, uma vez ser inevitável e até desejável que características locais de visão de mundo permaneçam. Assim, diferenças continuarão existindo, no entanto, as mais dominantes influenciarão países como o nosso, de menor tradição científica e tecnológica.

### **3.2 PRODUÇÃO DE DOCUMENTOS/ INFORMAÇÕES DOCUMENTÁRIAS: APRESENTAÇÃO/ REPRESENTAÇÃO DO CONTEÚDO E FORMAS DE ACESSO**

Foi a partir dos avanços da área conhecida como Documentação, desenvolvida na Europa, que a informação passou a ser compreendida como o objeto de estudo e de trabalho dos bibliotecários e documentalistas. Além disso, também foram os documentalistas europeus os pioneiros no reconhecimento do documento como o suporte (impresso, sonoro, eletrônico etc.) que contém a informação. Apesar do caráter científico e inovador destes avanços, foi nos Estados Unidos que a evidência da informação como objeto se disseminou pelo mundo, por meio do surgimento de uma nova área, a Ciência da Informação.

No entanto, muito antes, os copistas das bibliotecas medievais, considerados por alguns como os primeiros bibliotecários, efetivavam a reprodução dos manuscritos. Os bibliotecários de então eram responsáveis pela (re)produção dos documentos, intervindo em sua forma e, possivelmente, até em seu conteúdo. Com o advento da tecnologia da impressão na metade do século XV, ocorreu lentamente uma alteração na relação destes “bibliotecários” com seu objeto de trabalho. A tarefa de reprodução de manuscritos foi retirada da atividade geral de organização e preservação de documentos que era realizada na biblioteca, passando a acontecer em oficinas especializadas. Com isto, os livros chegavam à biblioteca apenas para serem organizados e preservados, atividades que, suspeita-se, conduziram com o tempo ao distanciamento dos “bibliotecários” do conteúdo destes documentos, ou seja, da informação neles contida.

Com a disseminação das bibliotecas públicas como equipamento do projeto da modernidade, visando ampliar o acesso à educação e à cultura, iniciada no século XVII e acentuada no final do século XIX e início do século XX, em especial nos Estados Unidos e Reino Unido, o empenho dos bibliotecários centrou-se no acesso, minimizando a atenção aos processos documentários. Este fenômeno reforçou o distanciamento iniciado com o advento da imprensa.

Assim, as teorias e as técnicas documentárias procuraram resgatar, desde esse período, o reconhecimento do conteúdo dos documentos como princípio teórico para a área e como parâmetro técnico das atividades de armazenamento e acesso.

Atualmente, a Ciência da Informação presencia o retorno da produção ampla e efetiva de documentos/ informações documentárias como aspectos não apenas teóricos, mas incluídos nas atividades dos bibliotecários e profissionais de áreas correlatas. De fato, a Editoração e as Artes Gráficas atualizadas, especialmente aquelas voltadas para o ambiente Web, estão crescentemente incluídas como aplicações da Ciência da Informação e da prática do bibliotecário.

Sendo assim, principalmente no ambiente Web, várias transformações vêm sendo verificadas. Muitos documentos surgem apenas no meio eletrônico, outros são criados simultaneamente no meio impresso e no meio eletrônico, alguns estão sendo migrados do primeiro para o segundo e, muitos, provavelmente nunca cheguem a ser acessados pelo ambiente eletrônico. Isto promove um crescimento de documentos em mídias diferentes (muitas vezes, os mesmos documentos) que geram formas distintas de apresentação/ representação e acesso e um grande desafio para a atividade integrada de tratamento, armazenamento e disseminação de informações documentárias.

Depois da predominância do livro e apesar da diversidade de outros materiais impressos e dos materiais audiovisuais, volta a necessidade ampla do uso do termo “documento”, já proclamada pelos documentalistas há um século. Este termo passou a ser adotado para as informações em ambiente eletrônico para dar conta da diversidade de formas, sejam textuais, sonoras, auditivas, visuais ou mistas, que a tecnologia eletrônica atual propicia. Com esta nova configuração informacional, o conceito de documento se confunde com o de informação pois o suporte e seu conteúdo estão imbricados. Em função disto, foi fortalecida a discussão sobre o tratamento da unidade informacional, o que implicou em transformações na forma de realizar este tratamento.

Diferente dos copistas, centrados na reprodução dos documentos e não no seu tratamento para acesso, atualmente, a inserção do profissional bibliotecário na produção dos documentos eletrônicos é consequência da necessidade de tratá-los adequadamente para disponibilização em repositórios informacionais virtuais. Ou seja, tanto para os documentos vindos de outras mídias, quanto para os documentos nascidos no meio eletrônico, deve haver uma intervenção em sua forma de apresentação/ representação e de acesso para que possam ser adequadamente armazenados e disponibilizados neste novo ambiente. A necessidade medieval de reprodução de documentos com fins de preservação, transformou-se em um desejo da pós-modernidade por acesso e uso que levou à necessidade de produção, tratamento e organização.

Em decorrência do desenvolvimento tecnológico e da cultura de cada época e local, a apresentação/ representação e a forma de acesso a conteúdos de documentos foram realizadas de forma bastante distinta. Em geral, os primeiros documentos eram representados e acessados em função da disposição física dos mesmos com o uso de sistemas de classificação, desde os mais empíricos até chegar aos clássicos sistemas de classificação (como CDD e CDU), criados no final do século XIX, e caracterizados pela pretensão de contemplar todo o conhecimento humano e pela sua grande disseminação pelo mundo até os dias atuais. Também neste período, foram criadas e disseminadas formas de representação e de acesso concebidas para serem armazenadas separadas dos próprios documentos que representavam. Estas formas foram consagradas pelas fichas catalográficas, muitas vezes ainda utilizadas. Com o início da representação dos documentos pelo meio eletrônico, efetivada pela automação dos catálogos de bibliotecas e pelas bases de dados documentários, substancialmente na década de 60, estas representações, antes registradas no meio impresso, passaram a existir eletronicamente, facilitando e aprimorando diversas atividades que forneceram consistência às

informações tratadas e alto poder de recuperação. Contudo, muitas das técnicas anteriores, como os sistemas clássicos de classificação e as fichas catalográficas são adotados no ambiente eletrônico, com maiores ou menores adaptações.

Com o advento dos documentos eletrônicos, muitas bases de dados ou catálogos de bibliotecas passaram a anexar, junto ao registro ou representação de cada documento, o próprio documento. Assim, o usuário pesquisa o catálogo ou a base de dados, acessa o registro do documento desejado e aciona um comando que promove a visualização deste documento. No entanto, com a tecnologia eletrônica, o registro ou representação do documento e o próprio documento permanecem ancorados no mesmo ambiente tecnológico ou, em outras palavras, estão contidos no mesmo suporte. Historicamente, a representação do conteúdo de documentos objetivou a recuperação deste conteúdo e o acesso ao documento, os quais aconteciam em momentos distintos. Atualmente, é possível representar o conteúdo dos documentos neles mesmos e, por conseqüência, acessar o documento e a sua representação simultaneamente e no mesmo suporte. Isto é realizado com o uso das linguagens de marcação de textos.

Nestas linguagens, além da facilitação do acesso, a representação é realizada a partir da estruturação dos documentos, prescindindo da digitação de campos em planilhas, muitas vezes necessárias, mesmo em sistemas de cooperação catalográfica. É por este motivo que a produção deste documento é mais adequada de ser realizada pelo mesmo serviço, uma vez tratar-se de um processo único. Vale considerar que a referência à produção de documentos não está aqui indicando a produção intelectual de seu conteúdo, mas a sua apresentação/ representação e formas de acesso, ou seja, refere-se ao documento como informação documentária e não como o conhecimento nele embutido.

Estas mudanças apontam para as noções de pré e pós-representação segundo o momento da construção do documento, sintetizadas por Biojone (2001). Na pós-representação, o documento já existe ou já está acabado, para o que, se recorre aos métodos tradicionais de catalogação, classificação, indexação e elaboração de resumos. Na pré-representação, é realizada uma marcação dos pontos de acesso no documento no momento de sua construção, o que implica que esta construção é realizada em função da necessidade de apresentação/ representação e acesso. Nos dois casos, a partir da extração de informações, cada texto original gera um documento ou informação documentária (ficha catalográfica, registro de computador ou documento eletrônico). A diferença, no entanto, é que com a marcação, o próprio texto é transformado em documento.

A representação e as formas de acesso que, como citado acima, vem se realizando há tempos, recebeu denominação específica com o documento eletrônico. Sendo assim, o termo “metadado”, apesar de seu sentido mais amplo, passou a indicar o dado que representa o conteúdo de documentos eletrônicos, com fins de recuperação deste conteúdo.

No entanto, argumenta-se a favor da padronização de metodologias (e, antes disso, da utilização de alguma metodologia) para a construção destes documentos eletrônicos. Esta nova tecnologia envolve diversos profissionais como editores e bibliotecários e implica o uso de uma metodologia comum para possibilitar a produção compartilhada e o acesso facilitado, evitando documentos incompatíveis entre si. O trabalho cooperativo de produção de registros bibliográficos realizado pelos bibliotecários tem seu universo ampliado, uma vez que a tecnologia eletrônica propiciou (e exigiu) a

atuação de diversos profissionais, como programadores e analistas, *designers*, bibliotecários e outros.

Dentre as metodologias que fazem parte deste novo contexto, estão algumas linguagens internacionais de marcação como o SGML, criado em 1986, e o XML, de 1996. Dada a proliferação do uso da tecnologia HTML, é importante considerar sua habilidade na apresentação de documentos mas sua baixa adequação para a estruturação destes documentos. Sendo o SGML conhecido por sua complexidade e o HTML pela sua pobreza semântica, o XML é o meio termo entre os dois. Portanto, este último não é um sucessor do HTML, mas do SGML, com suas características de linguagem para tratamento de documentos e algumas simplificações.

Pode-se concluir que o universo vulgarizado da Internet pelo uso do HTML como ferramenta para disponibilização de qualquer informação, vem exigindo tecnologia e metodologia adequadas para armazenamento e acesso, e os procedimentos técnicos aí envolvidos são a ampliação das técnicas documentárias e de bibliotecas e a agregação de outras especialidades profissionais. O XML é bastante adequado para o intercâmbio de dados e vem sendo utilizado neste sentido. No entanto, foi criado com a pretensão de tornar-se um meio universal de publicação. Sendo assim, da mesma forma que os formatos de registro bibliográfico como o Formato MARC, na prática, são mais adotados para intercâmbio que como método de tratamento de informações bibliográficas, espere-se que o XML não seja desvirtuado de sua pretensão original.

Acredita-se que a adoção de tecnologias em Biblioteconomia e Documentação apresenta uma linha de continuidade que se contrapõe às aparentes rupturas amplamente proclamadas. Sendo assim, apesar de ter havido um período de latência em que o bibliotecário praticamente não interferiu na produção dos documentos, não é fato que apenas com o ambiente Web este quadro se alterou. Tradicionalmente, bibliotecas e centros de documentação elaboraram boletins de alerta bibliográfico, muitas vezes, atendendo a perfis para disseminação seletiva da informação (DSI), mesmo antes do advento dos computadores. Também é antiga a inserção de profissionais bibliotecários na produção de trabalhos acadêmicos, livros e periódicos, pois as formas de representação e de acesso ao conteúdo destes documentos exigem procedimentos documentários como critérios de elaboração de índices e princípios normalizados de organização bibliográfica. Ou seja, neste momento, já havia uma intervenção na forma da elaboração dos documentos que antecedeu o surgimento dos documentos eletrônicos. O que mudou com os documentos eletrônicos foi a efetivação simultânea das etapas de produção que antes eram separadas. Em um primeiro momento, era realizada a edição do documento para publicação impressa (área da Editoração e das Artes Gráficas), assim como, a criação de índices para a publicação de livros e revistas, além da normalização de trabalhos acadêmicos (área da Biblioteconomia e Documentação). Em um segundo momento, a partir dos documentos então criados, ocorriam os processos tradicionais de representação descritiva e temática como a catalogação, a classificação, a indexação e a elaboração de resumos (área da Biblioteconomia e Documentação). Atualmente, o desenvolvimento tecnológico permite que todos estes processos sejam unificados e racionalizados, ou seja, a transformação do texto em documento/ informação documentária conta com a inter-relação dos processos de formatação do conteúdo (apresentação/ representação) e de geração de índices (formas de acesso). Sendo assim, a continuidade tecnológica se manifesta, não por rupturas, mas pelo aperfeiçoamento e integração de procedimentos para a elaboração de produtos e prestação de serviços.

Em suma, entende-se que a tecnologia eletrônica reforçou a informação como objeto de estudo e trabalho de bibliotecários e os reaproximou da atividade de produção de documentos, do reconhecimento da diversidade documental e da necessidade da representação do seu conteúdo com fins de recuperação e acesso. Mesmo que todos estes pontos já tenham sido preconizados pela Documentação, a sua efetivação está ainda no quadro das tendências, configurando-se em um longo período de transição. No entanto, o que está em questão não é a transformação de toda e qualquer informação para o meio eletrônico, mas o aprimoramento de teorias, métodos e práticas dos profissionais da informação e de seus usuários para um universo mesclado, onde as mais diversas formas documentais devem coexistir. A convivência com os ambientes híbridos parece ser o grande desafio, porém, acredita-se que a compreensão histórica do papel da produção de documentos/ informações documentárias pelos bibliotecários, deve ajudar a cumpri-lo.

## CONSIDERAÇÕES FINAIS

### **Biblioteconomia, Documentação, Ciência da Informação e a Informática Documentária**

A motivação inicial desta pesquisa foi a de sistematizar o conhecimento sobre o escopo dos recursos de Informática e de telecomunicações envolvidos no tratamento, armazenamento, recuperação e disseminação da informação documentária. O quadro inicial pesquisado apresentou-se excessivamente funcionalista, a vinculação entre a Biblioteconomia e a Informática era superficial e estes pontos estavam em desacordo com a experiência pessoal acadêmica e empírica sobre os sistemas documentários. Foi então localizado o termo francês *Informatique Documentaire* que, tanto pelo nome quanto pelo conteúdo, atendeu à visão teórica que se tinha em mente e forneceu os elementos para a experiência prática vivenciada.

No entanto, de imediato ficou evidente que, apesar destas conveniências, a abordagem francesa não parecia encontrar familiaridade na produção teórica e nos serviços de informação da realidade brasileira. Percebeu-se que a linha desenvolvida pelos Estados Unidos e pela produção em língua inglesa de uma forma geral, possuía características marcadamente distintas da francesa, sendo a primeira aquela que caracteriza as práticas locais.

A história apresentada sobre as teorias e as práticas leva a concluir que o quadro atual deve-se ao fato de que a Documentação francesa foi suplantada pela Ciência da Informação, modelo teórico disseminado pelos Estados Unidos mas que entrou na agenda da produção científica internacional. Atualmente verifica-se relativa aproximação destes modelos teóricos e das características dos sistemas de automação que lhes são decorrentes.

Todas as áreas que promovem coleta, tratamento, armazenamento, disseminação, transferência e assimilação de informações têm sido afetadas pelas transformações sociais e tecnológicas. No entanto, algo relativo à Biblioteconomia e Documentação e a seu pretenso projeto de "cientificização", a Ciência da Informação, parece seguir mais lento e difícil e isso merece alguma atenção, especialmente, numa realidade como a brasileira.

### **Compreensão histórica e conceitual da Informática Documentária**

Uma parte das dificuldades encontradas para a construção deste trabalho esteve primeiramente centrada na necessidade de busca por conceitos básicos, não dos processos e das técnicas, mas antes, dos contextos de desenvolvimento e uso destes processos e técnicas. Esta busca pelos conceitos básicos levou à pesquisa histórica.

Sendo assim, mesmo considerando que todo trabalho acadêmico exige, para ser rigoroso e contemplar certo nível de racionalidade científica, a sistematização de conceitos básicos e de pesquisa história, suspeita-se da relativa ausência de literatura

efetiva sobre estas abordagens que permitisse que esta pesquisa contemplasse mais rapidamente estes tópicos para atingir e desenvolver o tema central.

Neste sentido, pode-se dizer que ao menos dois aspectos são importantes para o entendimento das ferramentas tecnológicas necessárias para a atuação em Biblioteconomia e Documentação: a história que conta e contextualiza a criação e o desenvolvimento destas ferramentas e a conceituação das mesmas por meio da abstração de suas características e funcionalidades. De fato, um aspecto leva ao outro e vice-versa pois estão intrinsecamente relacionados. Entende-se que estas abordagens são de extrema relevância, em especial, quando se trata de tecnologias pois elas tendem a se vulgarizar pelo uso: ao se tornarem dispositivos que dão continuidade às funções exercidas pelos órgãos e sistemas humanos, como as lunetas que permitem ao homem enxergar aquilo que seu olho não vê ou as memórias de computador que armazenam mais do que a memória humana é capaz, conduzem o homem ao esquecimento sobre sua origem e funcionamento.

Por desconhecer ou esquecer o que são e quais são as funções ideais de determinadas tecnologias, dificultando seu uso efetivo a partir da adaptação a suas necessidades, é que se considera essencial a compreensão histórica e conceitual destas tecnologias para os profissionais de informação. Sem esta abordagem, a aprendizagem sobre os aspectos básicos das tecnologias eletrônicas, como características de *hardware* e *software*, torna-se não apenas desnecessária como enfadonha, uma vez que estas tecnologias fornecem mais sentido na prática que em descrições teóricas.

A análise conceitual e histórica das ferramentas tecnológicas adotadas pela área da Biblioteconomia e Documentação pode auxiliar no reconhecimento de que a informação é o objeto desta área e que a comunicação implica em determinação de contexto. Em tempos de tecnologização global, a Ciência da Informação que vem sendo discutida tem por desafio distinguir informação e tecnologia e, ao mesmo tempo, construir as interfaces entre elas que levem de fato ao uso disseminado e inteligente de ambos.

## **Avaliação terminológica dos conceitos relacionados ao termo Informática Documentária**

Verificou-se que os termos Informática Documentária e Automação de Bibliotecas, no Brasil, pertencem ao mesmo campo associativo, sendo o segundo termo, a forma mais amplamente utilizada para nomear o conceito. Outra expressão adotada no país é automação de serviços/ unidades de informação, forma ambígua ao parecer indicar a unidade física biblioteca ou centro de documentação e/ou os serviços realizados por estas unidades, assemelhando-se à Automação de Bibliotecas, ou ainda, englobando também serviços e redes de informação, considerados apenas virtualmente.

A partir da literatura estrangeira, constata-se que o termo Informática Documentária é mais amplo que Automação de Bibliotecas ou unidades de informação pois não só inclui estes mas também contempla outros traços que são pouco utilizados no contexto brasileiro. Por exemplo, no caso da produção, tratamento e disseminação de documentos eletrônicos, não estão em questão as bibliotecas ou unidades de informação e a automação do acervo destes centros, mas sim a informação, a qual já nasce “informatizada”.

Também o termo tecnologias da informação é dado, muitas vezes, como equivalente à Informática Documentária. No entanto, trata-se de um dos elementos básicos desta. A tecnologia da informação pode ser definida como o conjunto de tecnologias que possibilita, de forma econômica, a manipulação da informação desde o momento em que foi gerada. A Informática Documentária utiliza-se de algumas das tecnologias de informação mais amplamente disseminadas, mas necessita de outras adaptadas a suas próprias necessidades, as quais se referem não só a questões econômicas e racionais mas, principalmente, às possibilidades de tratamento da informação segundo sua natureza lingüística/ conteudística.

Assim pode-se concluir que, se Informática Documentária é distinta de Automação de Bibliotecas e de tecnologias de informação, estas últimas também não são similares.

Acredita-se que esta confusão conceitual não pode estar indicando uma área sedimentada e consolidada. No Brasil, sequer há literatura significativa para análise substantiva: há sim, vários autores com produção relevante, no entanto, ainda não existe acúmulo de conhecimento sobre o tema.

Quanto à produção estrangeira, pode-se encontrar fundamentação teórica e metodologias construídas. Há principalmente duas correntes: uma que inclui a gestão de acervos (*Library Automation*) e as técnicas e metodologias para a recuperação da informação (*Information Retrieval*) e outra que engloba a produção, tratamento, armazenamento, recuperação e disseminação das informações (*Informatique Documentaire*). Outros idiomas utilizados em países de menor tradição científica encontram-se, como é comum em situações semelhantes, à mercê de uma ou outra abordagem. Sabe-se, no entanto, da primazia que a linha inglesa, divulgada pelos Estados Unidos, costuma ter nestes casos, promovendo uma grande disseminação de seus métodos, sem que ocorra necessariamente uma escolha consciente, em função da ausência de conhecimento e discernimento sobre outra abordagem.

Constatou-se que o termo francês *Informatique Documentaire*, escolhido como abordagem referencial para este trabalho, de fato, pauta-se em uma visão holística ao englobar todos os conceitos e metodologias encontrados relativos ao processo automatizado do fluxo informacional.

Já no idioma inglês, foi possível perceber ao menos dois termos para dar conta do conceito acima referido: *Library Automation* e *Information Retrieval* (e não apenas o primeiro considerado inicialmente). Ou seja, é necessário discorrer sobre estes dois termos para contemplar, em parte, o campo de significação da *Informatique Documentaire* nomeada pela literatura em francês.

Estas nomenclaturas, aparentemente meras arbitrariedades assumidas por cada língua, decorrem em desenvolvimentos teóricos e práticos distintos e marcados por espectros culturais peculiares. Entre outros, podem ser citados:

- A visão francesa considera como finalidade do serviço de documentação o fornecimento da informação útil ao usuário em tempo adequado, para o que se deve procurar integrar três funções: gestão da biblioteca, análise de conteúdo e pesquisa retrospectiva. Sendo assim, são consideradas as especificações que visam a informatização do fluxo do acervo e os itens necessários para o tratamento do conteúdo dos documentos que compõem este acervo. Esta abordagem é tão antiga quanto o próprio conceito de Informática Documentária e os programas que a implementaram. No entanto, apenas recentemente, o mercado de *softwares* dos Estados Unidos e de outros

países pautados na abordagem em idioma inglês têm iniciado a inclusão, em seus programas de automação de bibliotecas, de módulos para gerenciamento de tesouro e aplicativos de indexação, percebendo a necessidade destas funções, provavelmente por conta da maturidade adquirida nestes espaços de oferta de serviços de informação. Além disso, na Informática Documentária, a Automação de Bibliotecas é apenas um dos tópicos que, de forma integrada, compõe a área.

- Quanto ao termo *Information Retrieval*, alguns poucos autores o consideram como a tradução literal do termo *Informatique Documentaire*. No entanto, uma das diferenças entre eles está no fato de que o primeiro não inclui a Automação de Bibliotecas, uma vez que esta área desenvolve-se em domínio específico. Além disso, constatou-se que a Informática Documentária parte de princípios lógico-semânticos para o processamento da informação e concepção dos *softwares* documentários. A Recuperação da Informação (*Information Retrieval*), por outro lado, desenvolveu muitos estudos centrados em experiências que tratam a informação de forma estatística, tendo mais tarde se preocupado extensamente com a questão da representação do conhecimento segundo a abordagem cognitivista. De fato, como foi apresentado, na linha francesa, a Recuperação da Informação é uma parte da Informática Documentária.

- Também os processos de GED e de produção e edição de documentos são tratados por domínios próprios na visão em idioma inglês, mas são componentes do conceito mais abrangente representado pela Informática Documentária segundo a abordagem francesa. Provavelmente, também nestes processos, a segmentação esteja fornecendo uma visão parcial do tratamento documentário, como no caso da não inclusão da análise de conteúdo na explicitação dos processos informatizados (entendendo-se por análise de conteúdo, o tratamento com fins de representação dos aspectos lógico-semânticos do documento e não apenas a indicação superficial dos temas de que trata).

A relação de reciprocidade entre os diversos elementos constituintes do fluxo documentário no modelo francês fornece um movimento sinérgico entre estes elementos que os torna complementares e flexíveis ao mesmo tempo. Ou seja, os três itens principais do serviço de documentação na abordagem francesa, a gestão da biblioteca, a análise de conteúdo e a pesquisa retrospectiva, são considerados essenciais para o atendimento de demandas de usuários, desde que adequadamente integrados e harmonizados entre si. Paradoxalmente, a abordagem inglesa, por desenvolver estas funções em domínios específicos e independentes, aparenta caminhar em sentido oposto ao da integração, proporcionando uma sucessão de fragmentos desvinculados que potencializam o surgimento de lacunas no processo.

Sendo assim, parece razoável a suposição inicial de que a área de conhecimento definida para Informática Documentária pela abordagem francesa, engloba fundamentos básicos e pertinentes para suprir à Biblioteconomia e Documentação um escopo conceitual e metodológico abrangente, em contraposição ao fragmentado modelo da língua inglesa.

Além do aspecto global e integrador da visão francesa, cuja preferência neste trabalho está relacionada à proposição de um modelo para Informática Documentária, entende-se que os processos de tratamento e recuperação da informação com base em parâmetros teóricos lógico-semânticos trazem mais profundidade e rigor científico à área. De fato, estes parâmetros estão vinculados à essência do aspecto informacional. Além disso, considerando que os serviços de informação envolvem questões de ordem pragmática, a linha adotada, por mais que filosófica como é peculiar aos modelos

européus, não dispensa ou minimiza os aspectos tecnológicos propriamente ditos, ou seja, a implementação dos princípios citados visando sua aplicação e funcionalidade.

Faz-se necessário esclarecer que, o que está em questão são os princípios norteadores para a constituição de cada sistema de automação fornecidos por áreas relativamente distintas como a Biblioteconomia e a Documentação. Considera-se a Biblioteconomia tradicional como a área que inclui a gestão de acervo visando seu acesso pelo usuário e, a Documentação, acrescentando a esta atividade, o tratamento do conteúdo de documentos diversos para fins de recuperação. Portanto, os princípios documentários contribuíram para o enriquecimento da prestação de serviços de informação, por meio da realização da análise de conteúdos dos documentos, da implementação do tratamento de tipologias documentais distintas e do requinte nas formas de recuperação da informação. Acredita-se que o entendimento destas áreas exige uma distinção epistemológica necessária para a reflexão aqui proposta sobre as metodologias de aplicação de tecnologia. Para tanto, o histórico traçado no capítulo 1 e finalizado no capítulo 3.1, procurou caracterizar cada uma destas linhas por meio das condições culturais, técnicas e tecnológicas de origem no tempo e no espaço.

Sendo assim, o gerenciamento automatizado de acervos de bibliotecas, mesmo com baixo nível de tratamento de conteúdo, vem sendo desenvolvido continuamente e, em especial nos Estados Unidos, é considerada questão resolvida. Já os processos automatizados de tratamento e recuperação de informações documentárias apresentam sistematização metodológica pouco disseminada, inclusive, no Brasil. No entanto, entende-se que o domínio das metodologias documentárias seja essencial às novas demandas que vêm surgindo como decorrência da grande quantidade de informações e da dificuldade para seu acesso.

O quadro abaixo procura dispor os termos utilizados para indicar o conceito de Informática Documentária segundo as abordagens predominantes realizadas em idioma francês e inglês e organizados em tópicos que abarcam as etapas e as características do fluxo informacional automatizado.

Pelo termo *Informatique Documentaire* entende-se a área responsável pela aplicação da Informática e das telecomunicações às diversas fases de produção, tratamento, armazenamento, recuperação e disseminação de informações documentárias. O termo *Library Automation* inclui a especificação dos processos automatizados necessários à gestão dos serviços de bibliotecas, sejam locais ou virtuais. E o termo *Information Retrieval* pode ser definido como o conjunto de questões que envolve a recuperação efetiva de informações de um sistema, incluindo ou não, a representação dos documentos que a antecede, e a avaliação dos resultados em função das interrogações realizadas.

<b>IDIOMAS/ REGIÃO</b>	<b>FRANCÊS/ VERTENTE EUROPÉIA</b>	<b>INGLÊS/ VERTENTE NORTE- AMERICANA</b>	<b>INGLÊS/ VERTENTE NORTE- AMERICANA</b>
TERMOS UTILIZADOS	<i>Informatique Documentaire</i>	<i>Library Automation</i>	<i>Information Retrieval</i>
Abrangência temática da informação	relacionada ao conhecimento genérico e a áreas especializadas	relacionada ao conhecimento genérico ou ao tratamento simultâneo de várias áreas especializadas do conhecimento	relacionada a áreas especializadas do conhecimento
Tipos de documentos/ informações	documentos impressos, incluindo aqueles não convencionais e multimeios, representados por registros bibliográficos, catalográficos ou de texto integral (documentos eletrônicos), além de informações cadastrais e estatísticas	principalmente documentos impressos tradicionais como livros e periódicos, mas também inclui multimeios, representados por registros catalográficos (documentos eletrônicos são, em geral, acessados mas não tratados pela biblioteca)	documentos impressos, incluindo aqueles não convencionais e multimeios, representados por registros bibliográficos, catalográficos ou de texto integral (documentos eletrônicos), além de informações cadastrais e estatísticas
Caracterizaçã o do serviço de informação	objetiva fornecer dados sobre documentos em qualquer suporte, disponíveis em acervo local ou não	objetiva fornecer dados sobre documentos, especialmente aqueles disponíveis em acervo local mas, crescentemente, fornece acesso a bases de dados externas referenciais ou de textos integrais	centra-se na noção de conjunto de dados dispostos para possibilitar recuperação eficiente de informações, conceito que, muitas vezes, desconsidera a questão do contexto de um serviço de informação, seja local ou virtual

<b>IDIOMAS/ REGIÃO</b>	<b>FRANCÊS/ VERTENTE EUROPÉIA</b>	<b>INGLÊS/ VERTENTE NORTE- AMERICANA</b>	<b>INGLÊS/ VERTENTE NORTE- AMERICANA</b>
TERMOS UTILIZADOS	<i>Informatique Documentaire</i>	<i>Library Automation</i>	<i>Information Retrieval</i>
Produção de informações/ documentos	trabalha na formatação/ editoração de informações/ documentos no meio eletrônico para posterior impressão ou não	normalmente não trabalha na formatação/ editoração de informações/ documentos no meio eletrônico; esta atividade é mais comum em bibliotecas que produzem boletins informativos e de alerta ou que realizam normalização e formatação de trabalhos acadêmicos	trabalha na formatação/ editoração de informações/ documentos no meio eletrônico
Aquisição de documentos e acesso a informações	efetua aquisição de documentos e possibilita acesso a estes documentos e a outros disponíveis integralmente ou não	efetua aquisição de documentos e possibilita acesso a estes documentos e a outros disponíveis integralmente ou não	possibilita acesso a documentos no meio eletrônico disponíveis integralmente ou não

<b>IDIOMAS/ REGIÃO</b>	<b>FRANCÊS/ VERTENTE EUROPÉIA</b>	<b>INGLÊS/ VERTENTE NORTE- AMERICANA</b>	<b>INGLÊS/ VERTENTE NORTE- AMERICANA</b>
TERMOS UTILIZADOS	<i>Informatique Documentaire</i>	<i>Library Automation</i>	<i>Information Retrieval</i>
Tratamento de informações	catalogação, classificação, indexação e elaboração de resumos de documentos convencionais e eletrônicos; a indexação é realizada por meio de vocabulários controlados, como os tesauros	tombamento, catalogação, classificação e indexação de documentos convencionais e, em menor medida, de documentos eletrônicos, por catalogação cooperativa ou não (quase nunca são elaborados resumos); a indexação é realizada, principalmente, por atribuição de palavras-chave ou uso de cabeçalhos de assunto	trabalha essencialmente com a análise automática de textos (de início com títulos e resumos de documentos, atualmente com textos integrais), algumas vezes com base em dicionários ou tesauros
Estruturação dos dados	desenvolvimento ou adoção de formatos catalográficos ou bibliográficos, para textos, sons, imagens e multimídia	adoção de formatos catalográficos comuns com fins de intercâmbio de dados	desenvolvimento ou adoção de formatos catalográficos ou bibliográficos, para textos, sons, imagens e multimídia
Recuperação de informações	interrogação em bases de dados referenciais ou em texto integral	interrogação nos catálogos locais e em bases de dados externas, referenciais ou de texto integral	interrogação em bases de dados referenciais ou em texto integral
Circulação de informações	empréstimo de documentos e disseminação de informações eletrônicas	empréstimo de documentos e, com menor incidência, disseminação de informações eletrônicas	disseminação de informações eletrônicas

<b>IDIOMAS/ REGIÃO</b>	<b>FRANCÊS/ VERTENTE EUROPÉIA</b>	<b>INGLÊS/ VERTENTE NORTE- AMERICANA</b>	<b>INGLÊS/ VERTENTE NORTE- AMERICANA</b>
TERMOS UTILIZADOS	<i>Informatique Documentaire</i>	<i>Library Automation</i>	<i>Information Retrieval</i>
Utilização das tecnologias da informação para a representação das informações	indexação e gerenciamento de vocabulário controlado como tesauro, além de análise de informações por recursos como lematização e analista morfológico	catalogação e atribuição de assuntos aos documentos	em geral, análise e recuperação de informações segundo ocorrência e frequência e/ou com ajuda de dicionários ou tesauros
Função geral das tecnologias da informação	conjunto de aplicações para possibilitar o fluxo de informações	conjunto de aplicações para administrar as tarefas da biblioteca	conjunto de aplicações para possibilitar a recuperação de informações

Sendo assim, o gerenciamento de informações proposto pelo modelo estadunidense tem como foco os resultados, nos estudos sobre recuperação da informação, e a eficiência e eficácia de operações, no âmbito das bibliotecas, enquanto o modelo francês não assume estes elementos como pressupostos pois devem resultar do aperfeiçoamento dos processos de organização da informação.

A literatura dos Estados Unidos e Reino Unido indica que a biblioteca faz controle de seu acervo pela automação de seus catálogos a acessa dados de conteúdo de documentos (ocasionalmente dos seus próprios documentos) pelas bases de dados acessíveis remotamente. Acredita-se que este seja um modelo fragmentado pois o registro descritivo fica separado do registro de conteúdo para o mesmo documento. Na literatura francesa, o centro de documentação realiza de forma informatizada o controle de acervo e a análise de conteúdo, além de acessar bases de dados sobre documentos que não possui. Entende-se que este seja um modelo integrado pois, tanto as bases de dados do acervo local como as bases de dados externas, agregam tratamento de conteúdo. No modelo francês, o gerenciamento do sistema depende do conteúdo, que é sua essência, e menos da eficiência dos serviços. Ou seja, uma vez considerada a essência, as operações podem ser adequadamente definidas, desenvolvidas e implantadas, e níveis de eficiência e eficácia têm possibilidade de serem alcançados.

Para contemplar esta atividade holística e baseada no tratamento do conteúdo, devem ser sistematizados os aspectos da Informática que contemplam a aplicação da indexação e das linguagens documentárias. É necessário verificar, neste sentido, se a segmentação calcada pela busca de resultados e da eficiência das operações não gera falhas de configuração da área, ocasionando falta de compreensão sobre os aspectos envolvidos na inter-relação entre as partes. Por exemplo, se vocabulários controlados e

tesauros são tratados apenas teórica e metodologicamente, sem instrumentação tecnológica e, por outro lado, se as tecnologias de informação englobam apenas os conceitos básicos de Informática, suspeita-se que lacunas estejam presentes.

### **Informática Documentária no contexto dos diversos tipos de serviços de informação no Brasil**

No universo das bibliotecas públicas, a Informática Documentária não encontrará espaços para aplicação enquanto predominar o discurso de sua função social sem agregar instrumental para alcance desse objetivo. Ou seja, dispensa-se o técnico procurando dar um ar nobre à atividade biblioteconômica mas não é construído um modelo sobre o tipo de serviço para a própria.

Sobre a biblioteca escolar, é comum a afirmação de que o requinte no tratamento documentário não é adequado pois o público caracteriza-se por estar em processo de formação e, não sendo especializado, necessita de um sistema simplificado para a garantia de acesso fácil. No entanto, é justamente este estágio de formação de conceitos e pouco conhecimento de mundo, onde se encontram as maiores dificuldades para a implantação de um sistema simples, uma vez que, os conceitos que buscam explicar nossa sociedade e os recursos tecnológicos eletrônicos são, ambos, complexos. Um aspecto facilitador neste contexto é o fato de que o tratamento documentário deve ser idealmente baseado no projeto pedagógico da escola, o que implica a adoção dos conceitos desenvolvidos nas atividades de ensino-aprendizagem para a representação do conteúdo dos documentos. De fato, acredita-se que uma biblioteca escolar que não se relaciona com as atividades pedagógicas (mesmo que seja para subvertê-las) é apenas um espaço para o depósito organizado do acervo. Neste sentido, a realidade das bibliotecas públicas apresenta um quadro mais difícil: é bem mais heterogênea ao incluir públicos de todas as idades, com diferentes formações e necessidades, inclusive entre os membros de um mesmo bairro e, em especial, em uma sociedade caracterizada pelas discrepâncias sociais como a brasileira.

As bibliotecas escolares caracterizam-se idealmente por gerenciar o acervo e por desenvolver a formação para a leitura, as habilidades para a pesquisa, a assimilação das informações e a capacidade de produção de textos. Além destas atividades centradas nos alunos, o corpo docente também necessita realizar pesquisas e o corpo administrativo pode fazer uso do acervo para leitura, dependendo da política e condições institucionais. Para dar conta destas atividades, a Informática Documentária contribui no bom uso das ferramentas informáticas para o tratamento documentário e para a pesquisa. Em especial, no aprendizado da pesquisa de informações, a alfabetização tecnológica inclui a habilidade para a busca em bases de dados, cujo grau de dificuldade implícito nos exercícios propostos se dá pelo próprio nível de ensino em que se encontram os grupos de alunos. Também conta com a orientação no acesso às informações disponíveis na Internet e na discriminação sobre sua qualidade e uso. E, até na formação para a leitura, a indexação e a pesquisa podem ser definidas em função das categorias de gênero literário, faixa etária a que se destina, tema, nome de personagens, período histórico, local onde se passa a história e outros. Se estas são as categorias adotadas nas atividades de leitura, os livros estarão representados a partir dos mesmos princípios, promovendo e reforçando o entendimento destas categorias e o acesso aos livros.

Sendo assim, o bibliotecário escolar representa um bibliotecário-educador, no entanto, um educador não para o conhecimento mas para o uso da informação nele contido e, ao mesmo tempo, pode e deve ser o catalisador/ otimizador dos recursos informáticos na escola.

A biblioteca universitária, por sua vez, quase sempre é caracterizada por ter a riqueza e o ônus de um extenso acervo que exige muitos recursos para ser gerenciado. No entanto, os princípios e ferramentas da Informática Documentária colaboram para que, a gestão de acervo e das demais atividades administrativas e de tomada de decisão, sejam realizadas de forma eficiente e eficaz. Mas, principalmente, por meio de um tratamento documentário elaborado e da busca de fontes de informação por acesso remoto ou local, otimizados pelo uso de recursos informáticos, pode-se obter mais agilidade e qualidade no atendimento. Assim como na biblioteca escolar, deve ser garantido apoio e acompanhamento intelectual à pesquisa de alunos e professores em consonância com a proposta acadêmica. Este apoio intelectual à pesquisa exige certas capacidades cognitivas do mediador que, por sua vez, não prescindem da sustentação em um sistema documentário refinado. De fato, não é coerente exigir habilidades intelectuais de bibliotecários de referência (mediadores) e, até mesmo de usuários, se o sistema implantado não imprime parâmetros tecnológicos e lingüísticos do mesmo nível.

As bibliotecas especializadas, os centros de documentação e outros serviços de informação podem ser entendidos como os serviços em empresas e de organizações em geral, constituindo acervos locais ou não, atendendo a áreas especializadas ou necessidades de informações diversas, como as de ordem utilitária. Em função dos modelos predominantes de gestão de acervos e, crescentemente, dos serviços de pesquisa de informações na Internet, a organização destas bibliotecas, centros e demais serviços de informação tem apresentado lacunas e carência de modelos. Isto se verifica, incrivelmente, mesmo em áreas científicas e tecnológicas onde o investimento é sempre maior em função de seu papel estratégico. A compreensão integradora que a área da Informática Documentária fornece parece ser um elemento importante para auxiliar na reversão deste quadro.

## **Sistemas de informação no âmbito da administração de empresas**

Como decorrência da abordagem aqui dada ao tema, algumas aplicações não foram apontadas em função do baixo rigor com que termos e produtos são manipulados e disseminados a partir de importações pouco cuidadosas de conceitos e técnicas de outras áreas.

Na área dos Sistemas de Informações, quase sempre voltada para o ambiente empresarial, termos como "tecnologias da informação" e "sistemas de informações" carecem de conceituação precisa. Acredita-se que uma das causas esteja na sustentação da área pelo poder comercial, em detrimento dos seus conceitos, características e usos ideais. Muitas vezes, o foco destes produtos apresenta-se ao mercado a partir do apelo da atualidade tecnológica mais que em função de suas funcionalidades e adequações.

Muitos destes sistemas de informação estão baseados no acúmulo sistemático de dados quantitativos como forma de obter elementos suficientes para ajuda à tomada de decisão. Em geral, no entanto, são sistemas que gerenciam controle de estoque e fluxo

de trabalho, além dos programas contábeis e financeiros. São mais adequadamente considerados sistemas de informática com fins administrativos ou operacionais do que sistemas de informação. Quanto aos sistemas de apoio à decisão, vale lembrar que, os sistemas de automação de biblioteca incluem módulos de informações gerenciais, mas esta é apenas uma parte do sistema.

Os sistemas de informação, segundo a abordagem da Ciência da Informação e Documentação, são configurações informacionais (em geral, representadas no ambiente computacional) onde conteúdos são tratados e armazenados com fins de atender necessidades por informação, sejam de ordem utilitária, para tomada de decisão nos negócios, para a educação inicial e continuada, para a pesquisa ou para o lazer e a cultura.

Como os sistemas de informação voltados às empresas centram-se na lógica do mercado, não necessariamente são respaldados cientificamente. Isto parece apontar para uma base teórica localizada na Ciência da Informação (como é possível encontrar em algumas linhas de pesquisa).

### **Fontes de informação, informação e conhecimento**

Considera-se que as atividades da Biblioteconomia e Documentação abarcam a coleta e disseminação de um conjunto de informações relevantes para determinado fim em uma área de conhecimento, mas não devem se confundir com esta. Estas atividades visam a informação e não o conhecimento pois é a informação que possibilita lidar instrumentalmente com o conhecimento, ou seja, de forma estruturada e organizada.

Da mesma forma, deve-se melhor definir fontes ou recursos de informação. Entende-se que as fontes de informação de uma área são os tipos de publicação mais comuns, as bases de dados com significativo conteúdo sobre o tema, além de instituições e especialistas produtores naquela área de conhecimento. Já a natureza e as características da informação de uma área referem-se à forma de produção e uso do conhecimento desta área, a qual gera categorias lógico-semânticas peculiares que se evidenciam no processo da representação da informação. Sendo assim, as fontes de informação são apenas os suportes por meio dos quais o conhecimento se manifesta. As fontes ou recursos informacionais devem ser considerados pois influenciam a forma de realizar buscas e de armazenar informação, mas não têm relação direta com o conteúdo informacional que, de fato, determina o processo de tratamento e o de recuperação.

Ferramentas lingüísticas e tecnológicas são adotadas para auxiliar na manipulação de informações em diversas áreas para a prestação de serviços. A Biblioteconomia e a Documentação, atualmente representadas pela Ciência da Informação, possuem relativa linguagem própria mas caracterizam-se por sua função de mediação. A julgar por outras áreas da Comunicação, tendem a confundir os conteúdos envolvidos na teoria e prática da mediação com os conteúdos da área de conhecimento com a qual a mediação é realizada. A carência de discernimento entre a própria área e a área com que realiza mediação é tanto maior quanto menor for a compreensão sobre a própria área.

Por este motivo, é relevante que se realize a discriminação entre as fontes de informação (suportes de informação), a informação (unidade do conhecimento) e o conhecimento (conjunto de saberes de uma área).

### **Informática Documentária, o profissional bibliotecário e o ensino**

Neste sentido, a proposta inicial da pesquisa enfatizou a informação como o objeto da área da Biblioteconomia e Documentação. No entanto, atualmente, por mais que este seja o discurso disseminado massivamente entre os próprios bibliotecários, a informação ainda é vista como a fonte de informação (como foi descrito acima) e não como a unidade do conhecimento. Apenas se a informação for apropriada conceitual e metodologicamente pelo bibliotecário em sua prática, é que os serviços bibliotecários serão reconhecidos como decorrentes do uso de instrumentos de mediação e de significação, seja no contexto das áreas especializadas ou do conhecimento genérico.

Estes pontos são levantados frente à perplexidade causada pela extensa literatura atual da área que enfatiza a tecnologia em função de sua grande ação transformadora enquanto, a atuação em Biblioteconomia faz uso desta tecnologia de modo meramente formal ao transferir técnicas de representação e recuperação descritiva do conteúdo de documentos para o ambiente eletrônico, ao mesmo tempo em que, não explora criativamente a compreensão das categorias de informação a partir dos recursos tecnológicos mais diversos no atendimento às demandas da sociedade por informação.

No Brasil, ao que consta, os modernos e atualizados sistemas de gerenciamento de bibliotecas estão conquistando o mercado e algumas reflexões presentes na literatura e nos cursos de Biblioteconomia estão acompanhando e respaldando este processo. No entanto, apesar de estes sistemas apresentarem recursos que apontam para o avanço dos aspectos relacionados ao tratamento e recuperação da informação, a ênfase está centrada na padronização das informações com vista ao compartilhamento de grandes volumes de dados por meio das atividades cooperativas. Os aspectos de tratamento e recuperação da informação, mesmo no nível em que são contemplados nos sistemas de automação mais presentes no país são, em geral, sub-utilizados em suas potencialidades pela comunidade de operadores (bibliotecários e outros profissionais). Estes fatos são recorrentemente percebidos na manipulação de *softwares* como o programa gratuito CDS-ISIS que é altamente difundido no país, mas também no uso de sistemas de automação de bibliotecas comerciais brasileiros ou estrangeiros, mesmo aqueles considerados os mais avançados em termos mundiais.

É necessário constatar que, em geral, faltam conhecimentos e práticas profissionais para atuar frente a listas de controle de termos, aplicativos de indexação e gerenciadores de tesouro, com uso de recursos tecnológicos ou não. Como uma de suas conseqüências, há carência de massa crítica para o desenvolvimento, seleção e operação de sistemas de informação automatizados de processamento, armazenamento e disseminação da informação documentária, seja por parte de bibliotecários ou de analistas e programadores. Por outro lado, por uma contingência, desde o início da automação documentária, são os próprios sistemas que têm contribuído para o amadurecimento e formação em Informática Documentária dos bibliotecários e, em menor medida, dos analistas e programadores.

A argumentação de que os bibliotecários seriam todos bibliotecários de sistemas atualmente reforça a noção de fluxo documentário, segundo o qual, se os processos são automatizados, qualquer bibliotecário que atua em ao menos um destes processos, deve ter compreensão global da automação implementada para desempenhar com qualidade e eficiência seus serviços. No entanto, independente da existência de automação, como considerar a parte independente do seu todo, ou ao contrário, como considerar o todo sem a observação de suas partes específicas? A administração de um serviço de informação, por exemplo, deve contemplar todas as etapas do processo documentário, mas a capacidade de gestão está implícita a cada uma das decisões que um indexador ou um bibliotecário de referência deve tomar ao desempenhar suas funções. Já a análise documentária não faz parte apenas das atividades do indexador e daqueles que criam e gerenciam tesouros mas, se estamos realmente pautados na noção de fluxo documentário, devemos ressaltar que os princípios lógico-semânticos envolvidos na construção de vocabulários e na indexação, são os mesmos da seleção e pesquisa de informações, da organização de informações em um portal eletrônico e da produção de documentos eletrônicos ou impressos, entre outros.

Em cursos de Biblioteconomia, com o início da automação dos processos de biblioteca, a Informática foi inserida como disciplina a parte e/ou penetrou outras disciplinas promovendo sua divisão em modos de operação denominados manual e automatizado (como nas disciplinas de catalogação e fontes de informação). Esta introdução fragmentada de noções de Informática deflagrou a ausência de modelos teóricos consistentes. Por exemplo, como dividir as fontes de informação entre manuais e automatizadas se o atendimento ao usuário se faz pelo uso de qualquer fonte de conteúdo pertinente e de acesso possível, independente do suporte e da tecnologia utilizados para tratamento e disponibilização? Ou ainda, se há elementos conceituais e regras envolvidos no processo de catalogação, que tipo de alterações a informatização trouxe? Isto demonstra que, se não há compreensão teórica sobre o processo, uma técnica tradicionalmente adotada, pode perder a referência com a introdução de inovações tecnológicas. Sendo assim, acredita-se que a Informática deve fazer parte de todas as disciplinas, mesmo que haja uma em que seja tratada isoladamente, como fundamentos básicos ou história da Informática na Biblioteconomia e Documentação.

Por outro lado, quando se afirma que os profissionais bibliotecários devem estar atentos às novas demandas do mercado e que os cursos de formação precisam também colaborar para isto, deve-se considerar que uma área do conhecimento não se configura única e exclusivamente pela forma como se dá seu desenvolvimento profissional junto às demandas do mercado. No entanto, como é característico às tecnologias eletrônicas, sua evolução excessivamente rápida dificulta a produção e sedimentação de conhecimento e, dado o ritmo mais lento que é intrínseco à reflexão teórica, as demandas por serviços de informação do mercado podem ser bons indicadores. Contudo, são os métodos de pesquisa científicos e contextualizados que possibilitam as devidas re-significações de uma área. Além disso, muitas vezes, o mercado se apropria de novos recursos tecnológicos como decorrência de pesquisas em ciência aplicada de fins mercadológicos ou não.

Outro aspecto é o da imposição tecnológica de nossa sociedade a profissionais e usuários da informação. Acredita-se que, dadas as tendências globalizantes e neoliberais em que estamos envolvidos, desconsiderar a evolução tecnológica é dar ênfase ao processo desigual de apropriação de conhecimentos e habilidades, além de adotar uma postura passiva frente às transformações e às possibilidades criativas de inovação e aperfeiçoamento de serviços e produtos de informação.

## REFERÊNCIAS BIBLIOGRÁFICAS

ASSOCIATION FRANÇAISE DE NORMALISATION (AFNOR) (1987). **Vocabulaire de la documentation**. 2. ed. Paris.

BIBLIOGRAPHIE du cours de maîtrise: Informatique Documentaire (2001). Paris: Université de Paris X. Disponível em: <<http://panini.u-paris10.fr/infocom/biblio-434.html>>. Acesso em: 6 jul. 2001.

BIOJONE, Mariana Rocha (2001). **Forma e função dos periódicos científicos na comunicação da ciência**. São Paulo. Dissertação (Mestrado em Ciências da Informação e Documentação) – Escola de Comunicações e Artes, Universidade de São Paulo.

BORGMAN, Christine L. (1997). From acting locally to thinking globally: a brief history of library automation. **Library Quarterly**, Chicago, v. 67, n. 3, p. 215-249.

BRADFORD, S. C. (1961). **Documentação**. Rio de Janeiro: Fundo de Cultura. (Biblioteca Fundo Universal de Cultura: Estante de Documentação). 292 p. (Tradução de: "Documentation", London, Crosby Lockwood & Son Ltd., 1953).

BUCKLAND, Michael (1996). Documentation, Information Science and Library Science in the USA. **Information Processing & Management**, v. 32, n. 1, p. 63-76.

CHAUMIER, Jacques (1973). **As técnicas documentais**. Lisboa: Europa América. 108 p. p. 64. (Tradução de: "Les techniques documentaires", Paris, 1971).

CHAUMIER, Jacques (1982). **L'accès automatisé à l'information**: informatique documentaire et bases de données. Paris: Entreprise Moderne d'Éditions. 147 p.

CHAUMIER, Jacques (1994). La nouvelle offre logicielle pour la recherche documentaire. **Documentaliste: Sciences de l'Information**, v. 31, n. 1, p. 3-8.

CHERNYI, A. I.; GILYAREVSKII, R. S.; MIKHAILOV, A. I. (1973). **Fundamentos de la informática**. La Habana: IDICT/Academia de Ciencias de Cuba. 2 v. (Tradução de: "Osnovi Informatiki", Moscú, Nauka, 1968).

CHONEZ, André (1993). Pour une histoire de l'informatique documentaire en France. **Documentaliste: Sciences de l'Information**, v. 30, n. 4-5, p. 248-253.

COELHO NETO, José Teixeira (1995). **Moderno e pós-moderno**. 3. ed. São Paulo: Iluminuras. 227 p. p. 13-39.

COSTA CARBALLO, Carlos da (1995). **Introducción a la informática documental**: fundamentos teóricos, prácticos y jurídicos. 2. ed. rev. Madrid: Síntesis. (Ciencias de la Información). 367 p.

COURSE description and information (2001). Long Island: Palmer School. Disponível em: <<http://palmer.cwpost.liu.edu/>>. Acesso em: 23 out. 2001.

CUNHA, Miriam Vieira da (1999). A formação em Ciência da Informação na França, no Canadá e na Dinamarca: comparação com o sistema brasileiro. **Encontros Bibli: revista de Biblioteconomia e Ciência da Informação**, Santa Catarina, n. 8.

- DACHELET, Roland (1990). **Etat de l'art de la recherche en informatique documentaire**: la representation des documents et l'accès à l'information. Le Chesnay: INRIA. (Programme 8 : Communication Homme-Machine). 32 p.
- DEPARTAMENTO de Ciencias de la Computación e Inteligencia Artificial (2002). Granada: Universidad de Granada. Disponível em: <[http://decsai.ugr.es/orden\\_es.html](http://decsai.ugr.es/orden_es.html)>. Acesso em: 23 jan. 2002.
- DEWEZE, André (1994). **Informatique documentaire**. 4. ed. Paris: Masson. (Manuels informatiques Masson). 292 p.
- DOCENCIA de primer y segundo ciclo (2002). Madrid: Universidad Carlos III. Disponível em: <<http://rayuela.uc3m.es/pryscicl.htm>>. Acesso em: 23 jan. 2002.
- ENCICLOPÉDIA MIRADOR INTERNACIONAL (1994). São Paulo ; Rio de Janeiro: Encyclopaedia Britannica do Brasil Publicações Ltda.
- FAURE, J. C. (1982). Organisation et stockage de l'information scientifique et technique. In: INFORMATIQUE et information scientifique et technique. Le Chesnay: INRIA.
- GARCÍA GUTIÉRREZ, Antonio Luis; LUCAS FERNANDEZ, Ricardo (1987). **Documentación automatizada en los medios informativos**. Madrid: Paraninfo. 264 p.
- GIAPPICONI, Thierry (1998). Les ressources bibliographiques de la Bibliothèque Nationale de France: la politique bibliographique de la Bibliothèque de Fresnes. **Bulletin des Bibliothèques de France**, Paris, v. 43, n. 6.
- GREGORY, Gwen (2000). The importance of automation literacy. **Information Today**, v. 17, n. 1, p. 61.
- GUINCHAT, Claire; MENOUE, Michel; BLANQUET, Marie-France (1994). **Introdução geral às ciências e técnicas da informação e documentação**. 2. ed. corr. aum. Brasília: IBICT. 540 p. (Tradução de: "Introduction générale aux sciences et techniques de l'information et la documentation", UNESCO, 2. ed., 1990).
- HOPKINSON, Alan (2001). CDS-ISIS Information. **Information Development**, v. 17, n. 1, p. 9-11.
- INFORMÁTICA Documental y Lingüística Computacional (2002). Las Palmas: Universidad de Las Palmas. Disponível em : <<http://www.gedlc.ulpgc.es/doctorado/progdoc2.htm>>. Acesso em: 23 jan. 2002.
- JACOB, Elin K.; SHAW, Debora (1998). Sociocognitive perspectives on representation. **Annual Review of Information Science and Technology (ARIST)**, Medford/NJ, v. 33, p. 131-185.
- JAEGER, Herman K. de (1971). **L'informatique documentaire**: une solution photo-electronique: le systeme Sesam. Ghent, Belgium: Association Scientifique et Technique pour la Recherche en Informatique Documentaire. 12 p.
- KATO, Mary A. (1987). **No mundo da escrita**. 2. ed. São Paulo: Ática. (Fundamentos). 144 p. p. 13-14.
- KHOURI ST-PIERRE, Anastassia (1989). L'impact de la mutation de l'informatique documentaire au Quebec. **Documentation & Bibliothèques**, v. 35, p. 73-77.

KOBASHI, Nair (1996). Análise documentária e representação da informação. **Informare**: Cadernos de Pós-Graduação em Ciência da informação, Rio de Janeiro, v. 2, n. 2, p. 5-27.

LA DESS Informatique Documentaire (2002). Lyon: ENSSIB. Disponível em: <<http://www.enssib.fr/autres-sites/dessid/dessid98/competen.htm>>. Acesso em: 23 jan. 2002.

LAHARY, Dominique (1997). Bibliothéconomie. In: CACALY, Serge (Coord.). **Dictionnaire encyclopédique de l'information et de la documentation**. Paris: Nathan. 635 p. p. 78-81.

LEMOIS, Antonio Agenor Briquet de (1998). Bibliotecas. In: CAMPELLO, Bernadete Santos; CALDEIRA, Paulo da Terra; MACEDO, Vera Amália Amarante (Orgs.). **Formas e expressões do conhecimento**: introdução às fontes de informação. Belo Horizonte: Escola de Biblioteconomia da UFMG. 414 p. p. 347-366.

LÉNART, Michèle (1996). L'évolution récente des logiciels de recherche et de gestion documentaires. **Documentaliste**: Sciences de l'Information, v. 33, n. 6, p. 302-309.

LÉNART, Michèle (1998). Les logiciels de recherche documentaire et de gestion de bibliothèque: evolution récente et offre de serveurs web. **Documentaliste**: Sciences de l'Information, v. 35, n. 6, p. 292-303.

LÉNART, Michèle; BOURDIN, Joseph (1994). **Recherche documentaire et gestion de bibliothèque** : um logiciel unique? L'offre du marché. Paris: ADBS Éditions. 339 p. (Coll. Sciences de l'Information. Série Études et techniques).

LESK, Michael (1995). **The seven ages of Information Retrieval**. Presented on: Conference for the 50<sup>th</sup> anniversary of "As we may think", 12-14 October 1995, MIT, Cambridge, Massachusetts. 11 p. Disponível em: <<http://www.ifla.org/VI/5/op/udtop5/udtop5.htm>>. Acesso em: 3 out. 2001.

MANGUEL, Alberto (1997). **Uma história da leitura**. São Paulo: Companhia das Letras. 405 p.

MANZANOS, Norberto (1999). "Por qué ISIS?" In: MANUAL para instructores de Winisis. Buenos Aires: Centro Atómico Constituyentes (CAC)/ Comisión Nacional de Energía Atómica (CNEA). 425 p. p. 27-34. Disponível em: <<http://www.cnea.gov.ar/cac/ci/isis/isidams.htm>>. Acesso em: 10 out. 2001.

MARCONDES, Carlos Henrique (1994). Automação das funções de biblioteca e pacotes de software: características e vocações. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 23, n. 1, p. 65-77.

MARTIN, Philippe (1997). Format. In: CACALY, Serge (Coord.). **Dictionnaire encyclopédique de l'information et de la documentation**. Paris: Nathan. 635 p. p. 227-230.

MARTINS, Wilson (1996). **A palavra escrita**: história do livro, da imprensa e da biblioteca. São Paulo: Ática. 519 p.

MEADOW, Charles T. (1992). **Text information retrieval systems**. San Diego: Academic Press, Inc. p. 2-19.

MENDONÇA, Ercília Severina (2000). A lingüística e a ciência da informação: estudos de uma interseção. **Ciência da Informação**, Brasília, v. 29, n. 3, p. 50-70.

MORENO GONZÁLEZ, José A. (1985). Qué fue del concepto soviético “Informatika”? **Documentación de las Ciencias de la Información**, Madrid, n. 18.

NIEUWENHUYSEN, Paul (1991). Computerised storage and retrieval of structured text information: CDS-ISIS version 2.3. **Program Automated: Library & Information System Trends**, v. 25, n. 1, p. 1-18.

O QUE É UNICODE? (2002). Disponível em: <<http://www.unicode.org/unicode/standard/translations/portuguese.html>>. Acesso em: 1 fev. 2002.

OHIRA, Maria Lourdes Blatt (1994). Biblioinfo: base de dados sobre automação em bibliotecas (informática documentária): 1986-1994. **Ciência da Informação**, Brasília, v. 23, n. 3, p. 369-371.

PALLIER, Denis (1997). Bibliothèque spécialisée. In: CACALY, Serge (Coord.). **Dictionnaire encyclopédique de l'information et de la documentation**. Paris: Nathan. 635 p. p. 98-99.

PINHEIRO, Lena Vânia Ribeiro (1999). Campo interdisciplinar da Ciência da Informação: fronteiras remotas e recentes. In: \_\_\_\_\_. **Ciência da Informação, ciências sociais e interdisciplinaridade**. Brasília: IBICT.

PLANES de estudio (2001). Madrid: Escuela Universitaria de Biblioteconomía y Documentación/ Universidad Complutense de Madrid. Disponível em: <<http://berceo.eubd.ucm.es/>>. Acesso em: 14 out. 2001.

RAYWARD, W. Boyd (1997). The origins of Information Science and the International Institute of Bibliography/ International Federation for Information and Documentation (FID). **Journal of the American Society for Information Science**, v. 48, n. 4, p. 289-300.

REYNOLDS, Dennis (1985). **Library automation: issues and applications**. New York: R. R. Bowker Company.

RODRIGUES, Adriano Duarte (1994). **Comunicação e cultura: a experiência cultural na era da informação**. Lisboa: Presença. p. 49-83.

ROWLEY, Jennifer (1998). **The electronic library**. London: Library Association. 4. ed. of “Computers for libraries”. (Traduzido para: **Informática para bibliotecas**. Brasília: Briquet de Lemos/ Livros, 1994. 307 p.).

SAFFADY, William (1999). **Introduction to automation for librarians**. 4. ed. Chicago: American Library Association. 339 p.

SAGREDO, Félix et al (1994). **Tecnologias documentales: memorias ópticas**. Madrid: Tecnidoc. 317 p.

SAGREDO, Félix; NUÑO, María Victoria (1994). En los orígenes de la Biblioteconomía y Documentación: Ebla. **Documentación de las Ciencias de la Información**, Madrid, n. 17.

SANTOS, Boaventura S. (1989). **Introdução à ciência pós-moderna**. Rio de Janeiro: Graal. 176 p. p. 17-45.

SAQUEL, Mario (2001). **Informática jurídica documental**. 7. ed. Santiago, Chile: Infojuris. Disponível em: <<http://www.chile.cc/msaque!>>. Acesso em: 22 jan. 2002.

SARACEVIC, Tefko (1995). Interdisciplinary nature of information science. **Ciência da Informação**, Brasília, v. 24, n. 1, p. 36-41.

SHERA, Jesse H.; EGAN, Margaret E. (1961). Exame do estado atual da biblioteconomia e documentação. In: BRADFORD, S. C. **Documentação**. Rio de Janeiro: Fundo de Cultura. (Biblioteca Fundo Universal de Cultura: Estante de Documentação). 292 p. p. 15-64. (Tradução de: "Documentation", London, Crosby Lockwood & Son Ltd., 1953).

SHERA, Jesse. H. (1980). Sobre biblioteconomia, documentação e ciência da informação. In: GOMES, Hagar Espanha (Org.). **Ciência da informação ou informática?** Rio de Janeiro: Calunga. 105 p. p. 91-105.

SOUZA, Clarice Muhlethaler de et al (1994). BIBID: base de dados da literatura brasileira de informática documentária. In: CONGRESSO LATINO-AMERICANO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO, 2., CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO, 17., 1994, Belo Horizonte. **Anais...** Belo Horizonte: ABMG. 820 p. p. 15-24.

SUTTER, Éric (1997). Recherche documentaire. In: CACALY, Serge (Coord.). **Dictionnaire encyclopédique de l'information et de la documentation**. Paris: Nathan. 635 p. p. 486-488.

VERNOTTE, France (1997). Centre de documentation et d'information (CDI). In: CACALY, Serge (Coord.). **Dictionnaire encyclopédique de l'information et de la documentation**. Paris: Nathan. 635 p. p. 125-127.

VICENTINI, Abner Lellis Correa (1971). **De la bibliotecología a la informática**: evolución del concepto documentación. Bogota: Colciencias. 79 p. (Bibliotecología y Documentación, 5). (Traduzido de: VICENTINI, Abner Lellis Correa. Da biblioteconomia à informática. **Revista do Serviço Público**, Brasília, v. 3, n. 105, p. 252-295, set./dez. 1970).

WERSIG, Gernot (1997). Information science: the study of postmodern knowledge usage. **Information Processing & Management**, v. 29, n. 2, p. 155-166.

YVES MARCOUX (2002). Montréal: Université de Montréal. Disponível em: <<http://mapageweb.umontreal.ca/marcoux/enseign/>>. Acesso em: 26 jan. 2002.

## ANEXO 1 - LISTA DE SIGLAS

AACR2= Anglo-American Cataloguing Rules, 2. ed.

ADBS= Association des Professionnels de l'Information et de la Documentation (antiga Association Française des Documentalistes e des Bibliothécaires Spécialisées)

AGILE= Auto-Graphics Interactive Library Exchange

AIRS= ASTRA Information and Retrieval System

ASCII= American Standard Code for Information Interchange

ATMS= Advanced Text Management System (da IBM)

BASIC= Beginners' All-Purpose Symbolic Instruction Code

BCD= Binary Coded Decimal

BIT= binary digit

BLAISE= British Library Automated Information Service

CAIRS= Computer Assisted Information Retrieval System

CANMARC= MARC do Canadá

CAR= computer assisted retrieval

CCF= Common Communication Format

CDA= compact disc audio

CDD= Classificação Decimal de Dewey

CD-I Vídeo= compact disc-interactive vídeo

CDI= centre de documentation et information (França)

CD-ROM= compact disc-read only memory

CDS-ISIS= Computerized Documentation System - Integrated Set of Information Systems  
(o MicroISIS e o MINISIS são as formas comumente usadas para a versão do CDS-ISIS, respectivamente, para micro e minicomputadores)

CDU= Classificação Decimal Universal

CEPAL= Comisión Económica para América Latina y el Caribe

COBOL= Common Business Oriented Language

COM= computer output microfilm

CPU= central processing unit

CRT= cathode ray tube

DESS= Diplome d'Études Supérieures Spécialisées (França)

DPI= dots per inch (pontos por polegada)

DSI= disseminação seletiva da informação

DVD= digital video disk

DVI= digital video interactive

EAN= European Article Numbering

EBCDIC= Extended Binary Coded Decimal Interchange Code

EBSI= École de Bibliothéconomie et des Sciences de l'Information (Montreal, Canadá)

EDI= electronic data interchange

ENSSIB= École Nationale Supérieure des Sciences de l'Information et des Bibliothèques (Lyon, França)

FID= Federação Internacional de Informação e Documentação

FORTTRAN= formula translation

GED= gerenciamento eletrônico de documentos

GPSS= General Purpose System Simulator

GUI= graphical user interface

HTML= Hypertext Markup Language

HyTime= Hypermedia Time-Based Structuring Language

IBERMARC= MARC da Espanha

IDAMS= Internationally Developed Data Analysis and Management Software

IFLA= International Federation of Library Associations and Institutions

INTERMARC= MARC da França

ISBD= International Standard Bibliographic Description

ISBN= International Standard Book Number

ISO= International Standard Organization

ISSN= International Standard Serial Number

KWIC= keyword in context

LAN= local area network

LCD= liquid crystal display

LCMARC= base de dados de registros MARC da Library of Congress

LILACS= Literatura Latino-Americana em Ciências da Saúde

LISP= List Processor

Mac-OS= Macintosh Operating System

MARC – Machine Readable Cataloguing

MeSH= Medical Subject Headings

MIT= Massachussetts Institute of Technology (Estados Unidos)

MS-DOS= Microsoft Disk Operating System

NLM= National Library Medicine

OCLC= Online Computer Library Center (Estados Unidos)

OCR= optical character recognition

ODA= Office Document Architecture

OPAC= online public access catalog

ORBIT= Online Retrieval of Bibliographical Information

OS= Operating System

PC= personal computer

PDF= Portable Document Format (da Adobe)

PICA= Dutch Project for Integrated Catalogue Automation

PROLOG= Programming in Logic

QBE= Query By Example

RAM= random access memory

RLIN= Research Libraries Information Network

ROM = read only memory

RPG= Report Program Generator

RTF= Rich Text Format (da Microsoft)

SGBD= sistema de gerenciamento de bases de dados

SGD= sistemas de gestão documentária

SGML= Standard Generalized Markup Language

SIBIL= Système Intégré pour les Bibliothèques Universitaires de Lausanne

SIG= sistema de informações geográficas

SQL= Structured Query Language

SRI= sistema de recuperação da informação

STAIRS= Storage and Information Retrieval System

UKMARC= MARC do Reino Unido

UNIMARC= Universal MARC

UNISIST= United Nations International System for Information in Science and Technology

USMARC= MARC dos Estados Unidos, desenvolvido pela Library of Congress

UTLAS= University of Toronto Library Automation System

WAN= wide-area network

WLN= Western Library Network

WORM= write one read many

WWW= World Wide Web

XML= eXtensible Markup Language

## ANEXO 2 - GLOSSÁRIO

**Abstracts**= Resumos.

**Algoritmo**= Série de regras que dão origem a uma sucessão lógica de operações dirigidas à resolução de um problema específico (COSTA CARBALLO, 1995). Conjunto de operações executadas em uma ordem dada para obter um resultado (DEWEZE, 1994).

**Arquivo invertido**= Semelhante a um índice, depende de pelos menos de dois arquivos: arquivo texto que contém os registros atuais e arquivo índice que fornece acesso a esses registros. O arquivo de índice contém um registro para cada termo indexado de todos os registros da base de dados, ordenados alfabeticamente. Cada termo é acompanhado pela informação sobre sua frequência de ocorrência na base de dados, pelo arquivo no qual está localizado, pelo arquivo com o qual é completado e pela possibilidade de informação de localização tais como parágrafo (ou campo) com o qual é localizado (ROWLEY, 1998).

**Autopostage**= Indexação complementar de um documento, que consiste em atribuir automaticamente descritores pertencentes às mesmas cadeias hierárquicas que os descritores atribuídos diretamente pelo indexador. Esta operação pode ser efetuada na entrada de documentos no banco de dados subindo a cadeia hierárquica, ou no momento da interrogação do banco de dados descendo a cadeia hierárquica (AFNOR, 1987).

**Bancos de dados restritos**= Incluem séries econômicas, constantes físico-químicas, resistência de materiais, estruturas químicas etc., por meio da consulta aos dados brutos ou à associação destes dados aos *softwares* de exploração (modelos econômicos para os estudos de mercado, cálculo de estruturas etc.) (DEWEZE, 1994).

**Base de dados bibliográficos ou referoteca**= Composta por registros que fazem referência aos documentos primários como livros, revistas, etc. (DEWEZE, 1994).

**Base de dados factuais**= Encontra-se diretamente a informação pesquisada como os anuários de empresas com suas características e produtos, repertórios de textos jurídicos, anuários telefônicos etc. (DEWEZE, 1994).

**bit-map**= Representação ponto por ponto dos elementos constitutivos dos caracteres (DEWEZE, 1994).

**Browsing**= folhear ou percorrer seqüencialmente registros de uma base de dados ou mover-se entre *sites* e páginas com base em *hyperlinks*. Os *browsers* possibilitam percorrer registro a registro, mas a identificação de recursos específicos necessita de uma ferramenta de pesquisa (ROWLEY, 1998).

**B-trees** (árvores balanceadas)= Os índices de pesquisa (arquivos invertidos) sob a forma de *B-trees* possibilitam um acesso rápido aos termos mesmo em índices grandes e com muitas chaves (CHAUMIER, 1994).

**Byte**= Conjunto de *bits* que codifica determinado caractere nos computadores (SAFFADY, 1999).

**DSI** (difusão seletiva da informação)= Consiste em fazer chegar periodicamente ao usuário os registros correspondentes a sua área de interesse, com ajuda de um conjunto

de palavras-chave ou descritores que constituam seu perfil documentário. É efetuada no conjunto de registros integrados à base em um dado período de tempo (DEWEZE, 1994).

**Engenharia documentária**= Conjunto de procedimentos e metodologias para a concepção da produção e gestão de documentos e dos serviços de informação (CACALY, 1997).

**Engenharia lingüística**= Conjunto de procedimentos e metodologias para a concepção e a realização de *softwares* que permitem o tratamento automático da linguagem natural (CACALY, 1997).

**Extranet**= Designa um acesso do tipo Intranet, só que não limitado à empresa mas estendido a seus interlocutores privilegiados como fornecedores e clientes (LÉNART, 1996).

**Fotocomposição** (*photocomposition*)= Procedimento de edição eletrônica a partir de um texto sobre suporte magnético produzindo um filme para impressão (CHAUMIER, 1982).

**Grossistas** (*grossistes*)= Distribuidores de informação produzida por terceiros, em oposição aos distribuidores da informação que produziram (CHAUMIER, 1982).

**Hiperônimos**= Palavras que se referem a uma noção genérica, como animal doméstico, mamífero (DEWEZE, 1994).

**Hipertexto**= Busca os documentos que contenham um termo determinado e selecionado em um documento visualizado (COSTA CARBALLO, 1995).

**Hipônimos**= Palavras que se referem a uma noção específica, como gato, vaca (DEWEZE, 1994)

**Hipotexto**= Realiza o mesmo que o hipertexto mas somente nos documentos selecionados na última pesquisa (COSTA CARBALLO, 1995).

**Índice KWIC** (*keyword in context*)= Apresentação das palavras significativas do documento como formas de entrada (DEWEZE, 1994).

**Intranet**= Reflete o uso da tecnologia de Internet para o sistema de comunicação interna de uma organização (ROWLEY, 1998).

**Lematização** (*lemmatisation*)= Operação que consiste em conduzir termos para a sua raiz eliminando as marcas de conjugação, no caso dos verbos, ou de declinação, para pronomes e adjetivos (LÉNART, 1996).

**Metadados**= Dado sobre o dado; é adotado especialmente para indicar o dado que representa o conteúdo de documentos eletrônicos.

**Online** e **offline**= Indicam, respectivamente, a presença ou ausência de conexões elétricas ou outras entre dispositivos de computação (SAFFADY, 1999).

**Palavras**= Cadeias de caracteres incluídos entre dois espaços (DEWEZE, 1994).

**Precisão**= Conjunto de documentos recuperados que são relevantes (LESK, 1995).

**Pull**= Corresponde, para o usuário, à ação de obter (puxar) a informação pela consulta, impressão ou teletransferência (*download*) (LÉNART, 1998).

**Push**= Consiste em encaminhar (empurrar) as informações ao usuário e, se possível, em função de suas necessidades próprias (LÉNART, 1998).

**Real time** (tempo real) e **batch** (em lote)= Referem-se ao modo de processamento de dados (relação entre transações ou outros eventos e processamento de informações sobre estes eventos pelo computador) e não especificamente ao fato de haver conexão elétrica ou não. No primeiro caso, em oposição ao segundo, quase não há espaço de tempo entre as transações e seu registro no sistema (SAFFADY, 1999).

**Revocação** (*recall*)= Conjunto de documentos relevantes recuperados (LESK, 1995).

**SGBD** (sistema de gerenciamento de bases de dados)= Conjunto complexo de programas desenhado para facilitar a criação e utilização de bases de dados (SAFFADY, 1999).

**Telemática** (*télématique*)= Conjunto de serviços de natureza ou de origem informática podendo ser fornecido através de uma rede de telecomunicação (CHAUMIER, 1982).

**Truncamento** (*truncature*)= Possibilidade de pesquisa sobre o radical de um termo a partir de uma interrogação em um sistema documentário (CHAUMIER, 1982).

**Turnkey**= Usado em oposição a *software only*, é o pacote completo com máquinas, aplicações, instalações, treinamento e suporte.

**Unicode**= Sistema padrão de codificação de letras e outros caracteres dos computadores. Os computadores designam números para gravar as letras e os outros caracteres na memória, o que levou à criação de centenas de sistemas diferentes de codificação para as diversas línguas ou até para uma mesma língua. O Unicode foi inventado para fornecer um único número para cada caractere, independente da plataforma, do programa e da língua. O Consórcio Unicode é uma organização sem fins lucrativos fundada para desenvolver, ampliar e promover o uso do Padrão Unicode que especifica a representação do texto em *software* e padrões modernos (O QUE É UNICODE?, 2002).

**Videotexto** (*vidéotex*)= Sistema que permite visualizar sobre uma tela de televisão a informação alfanumérica e gráfica transmitida pela rede telefônica. O usuário dispõe de uma via de retorno que fornece um sistema interativo (CHAUMIER, 1982).

**Workflow**= sistema encarregado de guiar e controlar de forma automática todos os componentes de um processo de trabalho como pessoas, tarefas, documentos e normas, graças à execução de um *software* instalado em uma rede que controla a representação automatizada dos processos de trabalho.

**Z39.50**= protocolo que permite ao usuário acessar OPACs e bases de dados na Web usando uma interface comum, independente do *hardware* e do *software*. Este padrão habilita um sistema local a atuar como um cliente de outro sistema, de tal forma que o usuário local possa pesquisar o sistema remoto por meio da interface local. No entanto, ainda não possibilita a pesquisa em múltiplos sistemas ao mesmo tempo e a mescla de resultados em uma mesma apresentação dos dados (BORGMAN, 1997).

## ANEXO 3 - BIBLIOGRAFIA TEMÁTICA COMPLEMENTAR

### Ciência da Informação e Informática:

Dada a origem do conceito de Ciência da Informação estar diretamente relacionado às primeiras aplicações da Informática, os textos abaixo discorrem sobre a imbricação desta origem, os desenvolvimentos soviéticos nesta área sob a nomenclatura *Informatika* e diversos avanços epistemológicos em Ciência da Informação.

CHERNYI, A. I.; GILYAREVSKII, R. S.; MIKHAILOV, A. I. **Fundamentos de la informática**. La Habana: IDICT/Academia de Ciencias de Cuba, 1973. 2 v. (Tradução de: "Osnovi Informatiki", Moscú, Nauka, 1968).

GOMES, Hagar Espanha (Org.). **Ciência da Informação ou informática?** Rio de Janeiro: Calunga, 1980. 112 p. (Série Ciência da Informação).

MORENO GONZÁLEZ, José A. Qué fue del concepto soviético "Informatika"? **Documentación de las Ciencias de la Información**, Madrid, n. 18, 1985.

SARACEVIC, Tefko. Interdisciplinary nature of information science. **Ciência da Informação**, Brasília, v. 24, n. 1, p. 36-41, 1995.

VICENTINI, Abner Lellis Correa. **De la bibliotecología a la informática**: evolución del concepto documentación. Bogota: Colciencias, 1971. 79 p. (Bibliotecología y Documentación, 5). (Traduzido de: VICENTINI, Abner Lellis Correa. Da biblioteconomia à informática. **Revista do Serviço Público**, Brasília, v. 3, n. 105, p. 252-295, set./dez. 1970).

WELLISCH, Hans. From Information Science to Informatics: a terminological investigation. **Journal of Librarianship**, v. 4, n. 3, July 1972.

WERSIG, Gernot. Information science: the study of postmodern knowledge usage. **Information Processing & Management**, v. 29, n. 2, p. 155-166, 1997.

YUEXIAO, Z. Definitions and sciences of information. **Information Processing & Management**, v. 24, n. 4, p. 479-482, 1988.

### Biblioteconomia e Informática:

Referências da bibliografia brasileira sobre a aplicação da Informática à Biblioteconomia, com abordagem para a automação dos serviços de biblioteca e para a informatização de processos documentários. A maioria refere-se a textos que agregam a segunda abordagem à primeira mas alguns centram-se unicamente nesta.

BARSOTTI, Roberto. **A Informática na Biblioteconomia e Documentação**. São Paulo: Pólis ; APB, 1990. 125 p. (Coleção Palavra-Chave, 2).

KREMER, Ligia Leindorf Bartz; MARCHIORI, Patricia Zeni. Automação documentária: contribuições conceituais para a prática. **Revista de Biblioteconomia de Brasília**, v. 20, n. 1, p. 15-26, jan./jun. 1996.

MARCONDES, Carlos Henrique. Automação das funções de biblioteca e pacotes de software: características e vocações. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 23, n. 1, p. 65-77, jan./jun. 1994.

OHIRA, Maria Lourdes Blatt. Biblioinfo: base de dados sobre automação em bibliotecas (informática documentária): 1986-1994. **Ciência da Informação**, Brasília, v. 23, n. 3, p. 369-371, set./dez. 1994.

SEMINÁRIO SOBRE DOCUMENTAÇÃO E INFORMAÇÃO, 24-27 nov. 1971, Rio de Janeiro, Fundação Getúlio Vargas. **Da documentação à informática**. São Paulo: FGV, 1974.

SOUZA, Clarice Muhlethaler de et al. BIBID: base de dados da literatura brasileira de informática documentária. In: CONGRESSO LATINO-AMERICANO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO, 2., CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO, 17., 1994, Belo Horizonte. **Anais...** Belo Horizonte: ABMG, 1994. 820 p. p. 15-24.

### ***Informatique Documentaire:***

Bibliografia que apresenta a origem do termo Informática Documentária na França, sua história e significado, e a evolução das metodologias que o constituem face às tecnologias desenvolvidas desde então.

BUFFELAN, J. P. **Informatique juridique documentaire**. La Garenne-Colombes: Éditions de l'espace européen, 1991.

CACALY, Serge (Coord.). **Dictionnaire encyclopédique de l'information et de la documentation**. Paris: Nathan, 1997. 635 p.

CHAUMIER, Jacques. **L'accès automatisé à l'information**: informatique documentaire et bases de données. Paris: Entreprise Moderne d'Éditions, 1982. 147 p.

CHONEZ, André. Introduction à l'informatique documentaire (I): présentation, définition. **Documentaliste**: Sciences de l'Information, v. 10, n. 2, p. 83-87, juin 1973.

CHONEZ, André. Introduction à l'informatique documentaire (II): les bases théoriques. **Documentaliste**: Sciences de l'Information, v. 10, n. 4, p. 203-207, dec. 1973.

CHONEZ, André. Introduction à l'informatique documentaire (III): les instruments. **Documentaliste**: Sciences de l'Information, v. 11, n. 1, p. 19-25, mars 1974.

CHONEZ, André. Pour une histoire de l'informatique documentaire en France. **Documentaliste**: Sciences de l'Information, v. 30, n. 4-5, p. 248-253, juil.-oct. 1993.

DACHELET, Roland. **Etat de l'art de la recherche en informatique documentaire**: la représentation des documents et l'accès à l'information. Le Chesnay: INRIA, 1990. (Programme 8 : Communication Homme-Machine). 32 p.

DEWEZE, André. **Informatique documentaire**. 4. ed. Paris: Masson, 1994. (Manuels informatiques Masson). 292 p.

HEID, Madeleine. L'informatique documentaire et l'information économique. **Documentaliste**: Sciences de l'Information, v. 20, n. 2, p. 70-72, mars-avr. 1983.

INFORMATIQUE documentaire: langage documentaire ou langage naturel?

**Documentaliste**: Sciences de l'Information, v. 11, n. 1, p. 47-54, mars 1974.

INSTITUT DE RECHERCHE D'INFORMATIQUE ET D'AUTOMATIQUE. **L'informatique documentaire**. Les Chesnay, 1971. (Cahier de l'Inria). 278 p.

JACOB, R. L'informatique documentaire au Congrès IFIP 1968 (Edimbourg, 5-10 août 1968). **Documentaliste**: Sciences de l'Information, v. 6, n. 2, p. 21-23, juin 1969.

RICHARDOT, Bruno. L'informatisation des centres de documentation et d'information: le nouveau paysage documentaire des CDI et la merveilleuse fécondité pédagogique de l'informatique documentaire. **Inter-CDI**: revue des centres documentaires et bibliothèques, n. 145, jan./fev. 1997, p. 9-10.

### ***Library Automation:***

Para a concepção em língua inglesa sobre o tema Informática Documentária são utilizados, não de forma completamente equivalente, os termos dos títulos abaixo, cujo conteúdo permite traçar parte da linha adotada pelos Estados Unidos e Reino Unido, para o tema desta pesquisa.

BORGMAN, Christine L. From acting locally to thinking globally: a brief history of library automation. **Library Quarterly**, Chicago, v. 67, n. 3, p. 215-249, 1997.

REYNOLDS, Dennis. **Library automation**: issues and applications. New York: R. R. Bowker Company, 1985.

ROWLEY, Jennifer. **The electronic library**. London: Library Association, 1998. 4. ed. of "Computers for libraries". (Traduzido para: **Informática para bibliotecas**. Brasília: Briquet de Lemos/ Livros, 1994. 307 p.).

SAFFADY, William. **Introduction to automation for librarians**. 4. ed. Chicago: American Library Association, 1999. 339 p.

### ***Information Retrieval:***

Ainda na concepção em língua inglesa e, complementando o conceito de *Library Automation*, o termo *Information Retrieval* abrange boa parte do tema Informática Documentária.

BLAIR, D. C. **Language and representation in information retrieval**. Amsterdam: Elsevier Science Publishers, 1990. 335 p.

ELLIS, David. **New horizons in information retrieval**. London: Library Association, 1990. 133 p.

JACOB, Elin K.; SHAW, Debora. Sociocognitive perspectives on representation. **Annual Review of Information Science and Technology (ARIST)**, Medford/NJ, v. 33, p. 131-185, 1998.

LESK, Michael. **The seven ages of Information Retrieval**. Presented on: Conference for the 50<sup>th</sup> anniversary of "As we may thing", 12-14 October 1995, MIT, Cambridge, Massachussets. 11 p. Disponível em: <<http://www.ifla.org/VI/5/op/udtop5/udtop5.htm>>. Acesso em: 3 out. 2001.

MEADOW, Charles T. **Text information retrieval systems**. San Diego: Academic Press, Inc., 1992.

### ***Informática Documental:***

Na produção do idioma espanhol sobre o tema da pesquisa, encontra-se essencialmente a vertente em francês que inclui, entre outros, as necessidades informáticas para o tratamento semântico da informação. Seguem abaixo também as obras que foram traduzidas para este idioma, apontando para apropriações realizadas por esta cultura.

CARIDAD, Mercedes; MOSCOVO, Purificación. **Los sistemas de hipertexto e hipermedios: una nueva aplicación documental**. Madrid: Fundación Germán Sánchez Ruipérez ; Pirámide, 1991.

COSTA CARBALLO, Carlos Manuel da. **Introducción a la Informática Documental: fundamentos teóricos, prácticos y jurídicos**. Madrid: Síntesis, 1995.

DEWEZE, André. **Informática documental**. Barcelona: Masson, 1988. (Traduzido de: DEWEZE, André. **Informatique documentaire**. Paris: Masson, 1986?. (Méthode + programmes)).

LOSANO, M. G. **Curso de informática jurídica documental**. Madrid: Tecnos, 1987.

REYNOLDS, Dennis. **Automatización de bibliotecas: problemática y aplicaciones**. Madrid: Fundación Germán Ruiperez, 1989. 840 p. (Traduzido de: **Library automation: issues and applications**. New York: R. R. Bowker Company, 1985).

SAFFADY, William. **Informática documental para bibliotecas**. Madrid: Díaz de Santos, 1986. 319 p. (Traduzido de: **Introduction to automation for librarians**. Chicago: American Library Association, 1983.).

SAGREDO, Félix et al. Los hipertexto e hipermedia: una revolución creativa en la informática documental. In: \_\_\_\_\_. **Tecnologías documentales: memorias ópticas**. Madrid: Tecnidoc, 1994. 317 p. p. 219-240.

SAQUEL, Mario. **Informática jurídica documental**. 7. ed. Santiago, Chile: Infojuris, 2001. Disponível em: <<http://www.chile.cc/msaquel>>. Acesso em: 22 jan. 2002.