
Sumarização automática de opiniões baseada em
aspectos

Roque Enrique López Condori

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Roque Enrique López Condori

Sumarização automática de opiniões baseada em aspectos

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências - Ciências de Computação e Matemática Computacional. VERSÃO REVISADA

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Thiago Alexandre Salgueiro Pardo

**USP – São Carlos
Outubro de 2015**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

L864s López Condori, Roque Enrique
Sumarização automática de opiniões baseada em
aspectos / Roque Enrique López Condori; orientador
Thiago Alexandre Salgueiro Pardo. -- São Carlos,
2015.
144 p.

Dissertação (Mestrado - Programa de Pós-Graduação
em Ciências de Computação e Matemática
Computacional) -- Instituto de Ciências Matemáticas
e de Computação, Universidade de São Paulo, 2015.

1. Sumarização de Opiniões com base em Aspectos.
2. Processamento da Língua Natural. I. Salgueiro Pardo, Thiago Alexandre, orient. II. Título.

Roque Enrique López Condori

Automatic aspect-based opinion summarization

Master dissertation submitted to the Instituto de Ciências Matemáticas e de Computação - ICMC-USP, in partial fulfillment of the requirements for the degree of the Master Program in Computer Science and Computational Mathematics. FINAL VERSION

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Thiago Alexandre Salgueiro Pardo

USP – São Carlos
October 2015

*Prefiro me arriscar fazendo algo que gosto perdidamente
que se sentir perdido por não fazer o que amo
(Roque Balboa)*

Agradecimentos

Agradeço primeiramente a Deus pela força para realizar este trabalho.

À minha família, pelo amor e apoio. Em especial agradeço a meus pais Antonia e Roque, meus irmãos Noelia, Juan José e Andrea, e a Sonia, minha namorada.

Ao meu orientador, Prof. Thiago Pardo pela confiança depositada e por sua orientação ao longo deste projeto.

Aos meus colegas e amigos do NILC, por terem compartilhado comigo a sua amizade e conhecimento.

À SAMSUNG Eletrônica da Amazônia LTDA e CNPq pelo o apoio financeiro prestado ao meu mestrado.

Resumo

A sumarização de opiniões, também conhecida como sumarização de sentimentos, é a tarefa que consiste em gerar automaticamente sumários para um conjunto de opiniões sobre uma entidade específica. Uma das principais abordagens para gerar sumários de opiniões é a sumarização baseada em aspectos. A sumarização baseada em aspectos produz sumários das opiniões para os principais aspectos de uma entidade. As entidades normalmente referem-se a produtos, serviços, organizações, entre outros, e os aspectos são atributos ou componentes das entidades. Nos últimos anos, essa tarefa tem ganhado muita relevância diante da grande quantidade de informação *online* disponível na web e do interesse cada vez maior em conhecer a avaliação dos usuários sobre produtos, empresas, pessoas e outros. Infelizmente, para o Português do Brasil, pouco se tem pesquisado nessa área. Nesse cenário, neste projeto de mestrado, investigou-se o desenvolvimento de alguns métodos de sumarização de opiniões com base em aspectos. Em particular, foram implementados quatro métodos clássicos da literatura, extrativos e abstrativos. Esses métodos foram analisados em cada uma de suas fases e, como consequência dessa análise, produziram-se duas propostas para gerar sumários de opiniões. Essas duas propostas tentam utilizar as principais vantagens dos métodos clássicos para gerar melhores sumários. A fim de analisar o desempenho dos métodos implementados, foram realizados experimentos em função de três medidas de avaliação tradicionais da área: informatividade, qualidade linguística e utilidade do sumário. Os resultados obtidos mostram que os métodos propostos neste trabalho são competitivos com os métodos da literatura e, em vários casos, os superam.

Palavras-chave: Sumarização de Opiniões com base em Aspectos, Processamento da Língua Natural.

Abstract

Opinion summarization, also known as sentiment summarization, is the task of automatically generating summaries for a set of opinions about a specific entity. One of the main approaches to generate opinion summaries is aspect-based opinion summarization. Aspect-based opinion summarization generates summaries of opinions for the main aspects of an entity. Entities could be products, services, organizations or others, and aspects are attributes or components of them. In the last years, this task has gained much importance because of the large amount of online information available on the web and the increasing interest in learning the user evaluation about products, companies, people and others. Unfortunately, for Brazilian Portuguese language, there are few researches in that area. In this scenario, this master's project investigated the development of some aspect-based opinion summarization methods. In particular, it was implemented four classical methods of the literature, extractive and abstractive ones. These methods were analyzed in each of its phases and, as a result of this analysis, it was produced two proposals to generate summaries of opinions. Both proposals attempt to use the main advantages of the classical methods to generate better summaries. In order to analyze the performance of the implemented methods, experiments were carried out according to three traditional evaluation measures: informativeness, linguistic quality and usefulness of the summary. The results show that the proposed methods in this work are competitive with the classical methods and, in many cases, they got the best performance.

Keywords: Aspect-based Opinion Summarization, Natural Language Processing.

Publicações

Como resultado da pesquisa feita neste trabalho de mestrado, até o momento foram publicados os seguintes artigos:

- **Roque López**, Lucas Avanço, Pedro Balage, Alessandro Bokan, Paula Cardoso, Márcio Dias, Fernando Nóbrega, Marco Sobrevilla, Jackson Souza, Andressa Zacarias, Ariani Di Felippo, Eloize Seno e Thiago Pardo. **A Qualitative Analysis of a Corpus of Opinion Summaries based on Aspects**. In the Proceedings of the IX Linguistic Annotation Workshop (LAW-2015). Colorado, USA.
- **Roque López** e Thiago Pardo. **Experiments on Sentence Boundary Detection in User-Generated Web Content**. In the Proceedings of the XVI International Conference on Intelligent Text Processing and Computational Linguistics (CICLING-2015). Cairo, Egypt.
- Verônica Agostini, **Roque López** e Thiago Pardo. **Automatic Alignment of News Texts and their Multi-document Summaries: Comparison among Methods**. In the Proceedings of the XI International Conference on Computational Processing of Portuguese (PROPOR-2014). São Carlos, Brazil.
- Márcio Dias, Alessandro Bokan, Carla Chuman, Cláudia Barros, Erick Maziero, Fernando Nobrega, Jackson Souza, Marco Sobrevilla, Marina Delege, Lucía Castro, Naira Silva, Paula Cardoso, Pedro Balage, **Roque López**, Vanessa Marcasso, Ariani Felippo, Maria Graças e Thiago Pardo. **Enriquecendo o Corpus CSTNews - a Criação de Novos Sumários Multidocumento**. In the Proceedings of the I Workshop on Tools and Resources for Automatically Processing Portuguese and Spanish (ToRPorEsp-2014). São Carlos, Brazil.

Sumário

Lista de Figuras	xi
Lista de Tabelas	xiii
Lista de Abreviações	xv
1 Introdução	1
1.1 Contextualização e Motivação	1
1.2 Lacuna	7
1.3 Objetivos	9
1.4 Hipóteses	9
1.5 Metodologia de Trabalho	10
1.6 Estruturação do Documento	10
2 Fundamentação Teórica	13
2.1 Sumarização Automática de Textos	13
2.1.1 Sumarização Extrativa	14
2.1.2 Sumarização Abstrativa	16
2.2 Mineração de Opinião/Análise de Sentimentos	17
2.3 Sumarização de Opiniões	19
2.3.1 Tipos de Opiniões	20
2.3.2 Abordagens	21

2.4	Sumarização baseada em Aspectos	22
2.4.1	Etapas na Sumarização baseada em Aspectos	23
2.4.2	Formatos dos Sumários baseados em Aspectos	25
2.4.3	Sumarização Extrativa vs. Abstrativa em Opiniões	27
2.4.4	Sistemas de Sumarização de Opiniões Baseados em Aspectos	28
3	Trabalhos Relacionados	33
3.1	Abordagens Extrativas	33
3.1.1	Trabalho de Beineke et al. (2003)	34
3.1.2	Trabalho de Hu e Liu (2004)	35
3.1.3	Trabalho de Sharifi et al. (2010)	36
3.1.4	Trabalho de Tadano et al. (2010)	38
3.1.5	Trabalho de Shimada et al. (2011)	39
3.1.6	Trabalho de Xu et al. (2011)	41
3.1.7	Trabalho de Wang et al. (2013)	42
3.2	Abordagens Abstrativas	44
3.2.1	Trabalho de Carenini et al. (2006)	44
3.2.2	Trabalho de de Ganesan et al. (2010)	45
3.2.3	Trabalho de Labbé e Portet (2012)	47
3.2.4	Trabalho de Gerani et al. (2014)	49
3.2.5	Trabalho de Fabrizio et al. (2014)	51
3.3	Síntese dos Trabalhos Discutidos	52
3.4	Linha do Tempo	53
4	Criação e Análise do Córpus	57
4.1	A Importância de um Córpus de Sumários de Opiniões	57
4.2	Córpus Existentes de Sumários de Opiniões	58
4.3	Recursos	60
4.3.1	Opiniões de Livros	60
4.3.2	Opiniões de Produtos Eletrônicos	62
4.4	Anotação do Córpus	62
4.4.1	Sumários Extrativos	64

4.4.2	Sumários Abstrativos	65
4.5	Experimentos	66
4.5.1	Concordância dos Anotadores	66
4.5.2	Cobertura dos Aspectos	69
4.5.3	Orientação do Sentimento	71
4.6	Considerações Finais	72
5	Métodos de Sumarização de Opiniões	75
5.1	Considerações Iniciais	75
5.2	Métodos Extrativos	77
5.2.1	Método de Hu e Liu	77
5.2.2	Método de Tadano et al.	79
5.3	Métodos Abstrativos	82
5.3.1	Método de Ganesan et al.	82
5.3.2	Método de Gerani et al.	85
5.4	Propostas	91
5.4.1	Proposta Extrativa: Opizer-E	92
5.4.2	Proposta Abstrativa: Opizer-A	98
5.5	Considerações Finais	104
6	Experimentos e Resultados	107
6.1	Métodos de Avaliação	107
6.1.1	Avaliação da Informatividade	108
6.1.2	Avaliação da Qualidade Linguística	109
6.1.3	Avaliação da Utilidade	110
6.2	Resultados	110
6.2.1	Resultados da Avaliação da Informatividade	111
6.2.2	Resultados da Avaliação da Qualidade Linguística	114
6.2.3	Resultados da Avaliação da Utilidade	118
7	Considerações Finais	121
7.1	Contribuições	122

7.2	Limitações	124
7.3	Trabalhos Futuros	124
	Referências Bibliográficas	125
A	Critérios de Avaliação da DUC 2007	143

Lista de Figuras

1	Arquitetura da sumarização de opiniões baseada em aspectos	5
2	Exemplo de sumário extrativo de opiniões	6
3	Exemplo de sumário abstrativo de opiniões	7
4	Busca do <i>Iphone 5</i> no Google	7
5	Busca do <i>Iphone 5</i> no Buscapé	8
6	Exemplo de falta de coesão textual	16
7	<i>Hourglass</i> das Emoções	18
8	Opiniões sobre o celular Samsung	18
9	Sumário Estruturado (Hu e Liu, 2004)	26
10	Sumário Textual (Carenini et al., 2006)	26
11	Sumário Visual (Liu, 2012)	27
12	Opiniões sobre o Galaxy S3	28
13	Tela inicial de SumView	29
14	Resumo gerado pelo SumView	30
15	Tela de SEA	31
16	Grafo de palavras (Sharifi et al., 2010)	37
17	Sumário gerado (Sharifi et al., 2010)	37
18	Grafo do Opinois (Ganesan et al., 2010)	46
19	Arquitetura do sistema proposto pelo Labbé e Portet (2012)	48

20	Linha do Tempo	55
21	Exemplo de sumário extrativo	65
22	Exemplo de sumário abstrativo	66
23	Sumário produzido pelo método de Hu e Liu (2004)	79
24	Sumário produzido pelo método de Tadano et al. (2010)	81
25	Sumário produzido pelo método de Ganesan et al. (2010)	84
26	Grafo das relações RST entre os aspectos	86
27	Subgrafo das relações RST entre os aspectos	87
28	Grafo não dirigido das relações RST	88
29	Hierarquia dos aspectos do Galaxy S-III	89
30	Sumário produzido pelo método de Gerani et al. (2014)	91
31	Etapas das propostas	92
32	Exemplo de opinião	95
33	Sumário produzido pelo método Opizer-E	97
34	Sumário produzido pelo método Opizer-A	103

Lista de Tabelas

1	Média de resultados da TAC-2008 (Mithun e Kosseim, 2009)	3
2	Porcentagem de acertos nos métodos (Beineke et al., 2003)	35
3	ROUGE-N para os sumários automáticos	39
4	Acordo Semântico entre os resumos automáticos e os manuais (%)	40
5	Comparações do desempenho de diferentes abordagens em opiniões	42
6	Comparação da satisfação dos usuários (Wang et al., 2013)	43
7	Comparação dos métodos (Carenini et al., 2006)	45
8	Preferência dos usuários	50
9	Resultados da avaliação qualitativa	52
10	Trabalhos estudados	54
11	Composição do córpus ReLi	61
12	Resultados da concordância entre os anotadores	61
13	Conteúdo de OpiSums-PT	63
14	Entidades e aspectos considerados na anotação	64
15	Resultado da concordância entre anotadores	68
16	Cobertura dos aspectos nos sumários	70
17	Orientação do sentimento dos sumários	72
18	Aspectos únicos considerados nos métodos investigados	77
19	Templates traduzidos de Gerani et al. (2014)	90

20	Templates utilizados por Opizer-A	101
21	Resultados da ROUGE dos métodos extrativos	112
22	Resultados da ROUGE dos métodos abstrativos	113
23	Resultados da ROUGE dos métodos implementados	114
24	Resultados da avaliação da QL por produtos e livros	116
25	Resultados da avaliação da QL dos métodos implementados	117
26	Resultados da avaliação da utilidade dos métodos implementados	118

Lista de Abreviações

DUC *Document Understanding Conference*

EDU *Elementary Discourse Units*

MST *Maximum Spanning Tree*

NILC Núcleo Interinstitucional de Linguística Computacional

NLG *Natural Language Generation*

PLN Processamento de Língua Natural

POS *Part-Of-Speech*

QL Qualidade Linguística

ROUGE *Recall-Oriented Understudy for Gisting Evaluation*

RST *Rhetorical Structure Theory*

TAC *Text Analysis Conference*

TF-IDF *Term Frequency-Inverse Document Frequency*

Introdução

1.1 Contextualização e Motivação

A Internet mudou a vida das pessoas. Hoje em dia, se alguém quer cumprimentar um amigo pelo seu aniversário, o faz pelas redes sociais; se uma pessoa pretende ler uma notícia, ela busca os *websites* dos jornais; se alguém quer comprar um celular, ele pesquisa na web as opiniões sobre os diferentes tipos de celulares, para escolher o melhor.

Conhecer “*o que as outras pessoas pensam*” tem sido, e é, uma peça importante durante o processo de tomada de decisão (Pang e Lee, 2008). Atualmente, antes de adquirir um produto ou serviço, tornou-se habitual procurar, ler e analisar as opiniões *online* que outras pessoas estão dizendo sobre um produto ou um serviço de interesse.

Especificamente, no setor do *e-commerce*, as mídias sociais, tais como blogs, redes sociais e *sites* de opiniões dos clientes, estão tendo um papel importante no processo da compra, pois estas contêm muitas e diferentes opiniões. Além disso, estas opiniões geralmente são de livre acesso.

Nas últimas décadas, páginas de vendas (como Amazon¹ e Google Shopping²), sistemas

¹<http://www.amazon.com/>

²<http://www.google.com/shopping/>

de comparações (por exemplo, Buscapé³ e Bondfaro⁴) e páginas de avaliações (como Epinions⁵ e Reclame Aqui⁶) se tornaram repositórios gigantescos de opiniões expressas pelos navegadores. Segundo o estudo do Forrester-Research (2011), no ano 2011 foram geradas mais de 500 bilhões de opiniões sobre diversos produtos e serviços na Internet.

Este grande volume de informação tem sido o foco de interesse de várias tarefas do Processamento da Língua Natural (PLN), tais como: Sumarização de Opiniões, Agrupamento de Documentos, Sistema de Perguntas e Respostas, entre outras.

A sumarização de opiniões, também conhecida como sumarização de sentimentos, é a tarefa que consiste em gerar automaticamente um sumário a partir de um conjunto de opiniões sobre um alvo específico (Conrad et al., 2009), sendo as opiniões expressões subjetivas que descrevem sentimentos, emoções ou avaliações.

É importante ressaltar que ao longo deste texto, optou-se por utilizar o termo *opinião* para denotar uma opinião, sentimento, avaliação e emoção, como é feito também em Liu (2012). No entanto, estes conceitos não são equivalentes, como é mostrado em Munezero et al. (2014). Para simplificar a leitura deste texto, estes termos são utilizados indistintamente, porém são distinguidos quando é necessário.

A tarefa de sumarização de opiniões surgiu pelo crescimento contínuo de textos subjetivos, tais como: postagens, *tweets*, comentários, entre outros, nos quais os navegadores expressam suas opiniões sobre produtos, pessoas, eventos, empresas, etc. Esta nova tarefa visa analisar estas opiniões, para extrair e apresentar as mais relevantes para um usuário final.

Segundo o estudo de Mithun e Kosseim (2009), em comparação à sumarização tradicional (sumarização de documentos objetivos, por exemplo, notícias), a sumarização de opiniões apresenta um pior desempenho. Na Tabela 1, podem ser vistas as pontuações médias dos sistemas participantes na TAC 2008⁷, tanto para a sumarização tradicional (gênero notícias) quanto para a sumarização de opiniões (gênero *blog*). A TAC é a principal conferência e competição científica da tarefa de sumarização automática, a qual

³<http://www.buscapede.com.br/>

⁴<http://www.bondfaro.com.br/>

⁵<http://www.epinions.com/>

⁶[http://www.reclameaqui.com/](http://www.reclameaqui.com.br/)

⁷ *Text Analysis Conference* <http://www.nist.gov/tac/2008/index.html>

fornece um grande conjunto de dados de teste, procedimentos de avaliação e um fórum para compartilhar resultados.

Da Tabela 1, a informatividade avalia a relevância do conteúdo do resumo⁸; a qualidade linguística calcula a coerência e legibilidade do sumário; a capacidade de resposta avalia a utilidade do resumo. Os resultados mostram que as pontuações na sumarização tradicional são maiores do que na sumarização de opiniões.

Gênero	Informatividade (escala de 0 até 1)	Qualidade Linguística (escala de 1 até 5)	Capacidade de Resposta (escala de 1 até 5)
<i>Blogs</i>	0.21	2.13	1.61
Notícias	0.27	2.33	2.32

Tabela 1: Média de resultados da TAC-2008 (Mithun e Kosseim, 2009)

Esta diferença do desempenho entre a sumarização de opiniões e a tradicional pode ser causada por vários fatores. Por exemplo, ao contrário da sumarização tradicional, o sentimento (subjetividade) tem um papel muito importante na sumarização de opiniões. Adicionalmente, a maioria das opiniões está escrita em linguagem informal e, muitas vezes, não tem relação com o tópico principal. Além disso, a sumarização tradicional tem sido estudada mais do que a sumarização de opiniões, pois esta é uma tarefa mais recente.

As pesquisas em sumarização de opiniões são recentes. O primeiro trabalho foi publicado no final do ano 2003, realizado por Beineke et al. (2003). Outro trabalho pioneiro é o de Hu e Liu (2004), o qual foi a base de outros trabalhos de sumarização de opiniões. Ambas as pesquisas serão explicadas com mais detalhes mais à frente.

A partir das primeiras pesquisas, a quantidade de trabalhos sobre sumarização de opiniões vem crescendo rapidamente. Tais trabalhos são motivados pela importância desta tarefa na comunidade de marketing e academia. No marketing, a sumarização de opiniões é importante, porque com os sumários de opiniões as empresas poderiam saber os pontos fortes dos seus produtos e, mais importante ainda, as suas fraquezas. É importante na academia, porque resolve um problema atual crescente, que é a análise de muitos dados.

Com os sumários automáticos das opiniões, compradores têm a melhor informação para decidir realizar uma determinada compra ou não; pode-se comparar produtos e serviços

⁸Na tarefa de sumarização automática, os termos sumário e resumo são utilizados indistintamente.

de interesse; empresas podem melhorar seus produtos e embasar melhor suas tomadas de decisões; autores têm condições de aperfeiçoar suas obras; políticas de governos podem ser alteradas; órgãos e entidades governamentais podem monitorar e lidar com crises e demandas da população, etc.

Embora os trabalhos de sumarização de opiniões proponham diferentes métodos para gerar um sumário, esses métodos podem ser agrupados em três principais abordagens: sumarização tradicional, sumarização contrastiva e sumarização baseada em aspectos (Liu, 2012). A sumarização tradicional gera resumos sem levar em conta os sentimentos das opiniões. Devido a isso, o resumo gerado pode conter sentenças que não tenham relação com os principais tópicos. A sumarização contrastiva procura opiniões com pontos de vista contrários, mas não mostra o sentimento predominante nas opiniões. Assim, o sumário gerado mostra uma comparação das diferentes opiniões sem indicar qual é a preferência dos usuários. A sumarização baseada em aspectos gera resumos das opiniões em torno de um conjunto de aspectos ou características indicando a preferência dos usuários. Destas três abordagens, a mais estudada é a baseada em aspectos (Hu e Liu, 2004), pois gera um resumo com mais informação.

Um sistema de sumarização de opiniões baseado em aspectos tem como entrada um conjunto de comentários sobre um objeto (por exemplo, um *restaurante*) e produz um resumo que expressa o sentimento para alguns aspectos relevantes (por exemplo, a *comida* ou o *serviço*) (Zhu et al., 2009). Os objetos ou entidades normalmente referem-se a produtos, serviços, organizações, pessoas, eventos, entre outros, e os aspectos são atributos ou componentes das entidades (Zhang e Liu, 2013).

Formalmente, no nível de aspectos, Liu (2012) define uma opinião como uma quintupla $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$, onde e_i é uma entidade, a_{ij} é um aspecto do e_i , s_{ijkl} é o sentimento do aspecto a_{ij} da entidade e_i , h_k é o titular da opinião e t_l é o momento no qual a opinião é expressada por h_k . O sentimento s_{ijkl} pode ser positivo, negativo ou neutro, e é expressado em diferentes níveis de intensidade (por exemplo, de 1 a 3).

Segundo Kim et al. (2011), a arquitetura de um sistema de sumarização de opiniões baseada em aspectos contém três grandes etapas: Identificação dos Aspectos, Predição do Sentimento e Geração do Sumário. Na Figura 1, é mostrada uma breve descrição destas etapas. A etapa de identificação dos aspectos visa encontrar os tópicos importantes

presentes nas opiniões. A predição do sentimento determina a orientação sentimental (positiva ou negativa) em relação aos aspectos encontrados na primeira etapa. Por fim, a geração do sumário é a etapa responsável por determinar quais sentenças serão incluídas no resumo para gerar o sumário final. O presente trabalho de mestrado estará focado, principalmente, na terceira etapa.

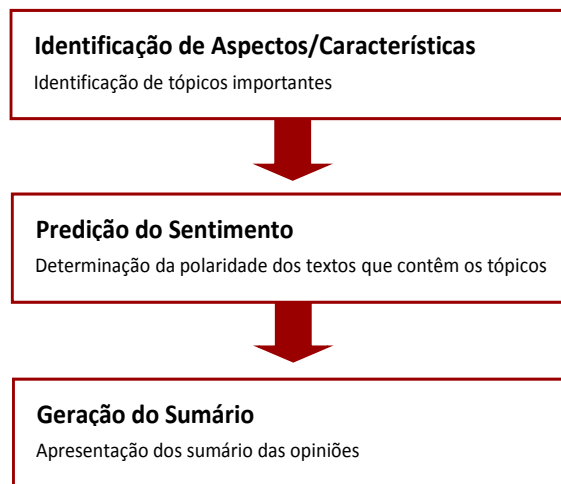


Figura 1: Arquitetura da sumarização de opiniões baseada em aspectos

Estas três etapas representam os maiores desafios da sumarização de opiniões baseada em aspectos. Na identificação de aspectos, entre outros problemas, tem-se que lidar com a presença de aspectos implícitos. Por exemplo, a sentença “*A bateria do celular é grande*” implicitamente menciona o aspecto “*tamanho*”, o qual não está presente no texto. Na predição do sentimento, a polaridade das expressões muda em contextos diferentes. Por exemplo, a palavra “*previsível*” tem uma conotação positiva ao se falar do funcionamento de aparelhos eletrônicos, mas negativa ao se falar de filmes. Na geração do sumário, deve-se identificar e analisar os qualificadores (expressões subjetivas sobre um aspecto), pois estes contêm as principais informações sobre um aspecto, e, portanto, devem estar presentes no sumário final.

Para ilustrar alguns desafios presentes na sumarização baseada em aspectos, na Figura 2 e na Figura 3, podem ser vistos dois sumários (escritos por um humano) sobre opiniões de smartphones. Nestas figuras, a entidade é o *smartphone* e os aspectos são a *duração da bateria* e a *câmera*. Pode-se facilmente notar a diversidade de fenômenos típicos da tarefa de sumarização que devem ser tratados, como o balanceamento entre opiniões positivas

e negativas, a fidelidade ao sentimento geral dos usuários (que podem ser mais positivos, negativos ou neutros) e o número de aspectos do produto que devem ser abordados e, neste caso, deve-se lidar com o *trade-off* entre quantidade de aspectos abordados, relevância dos aspectos e tamanho desejado do sumário.

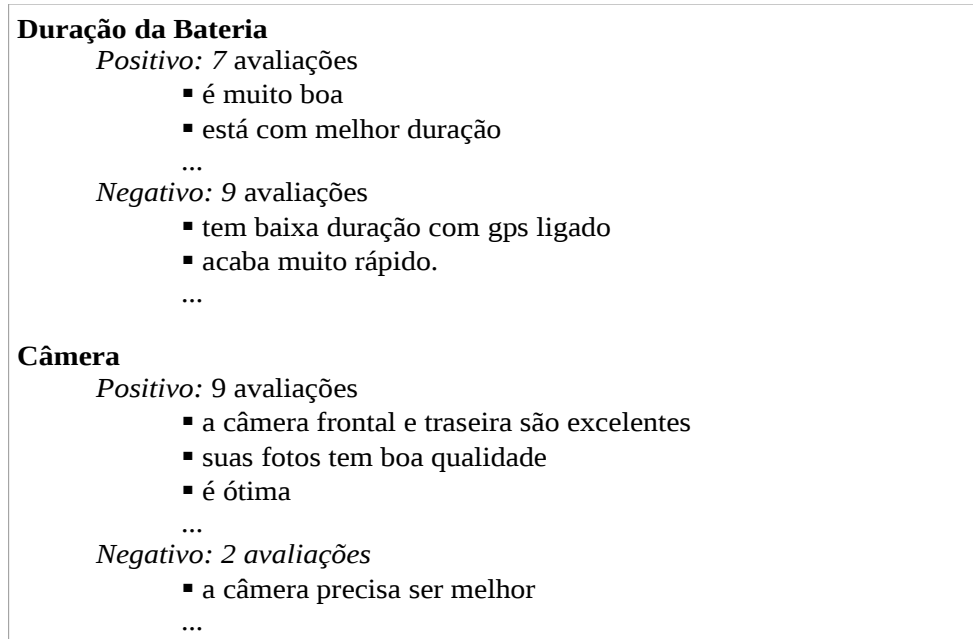


Figura 2: Exemplo de sumário extrativo de opiniões

A maioria dos trabalhos de sumarização de opiniões produzem resumos extrativos, os quais são construídos pela seleção e justaposição de sentenças ou partes de sentenças dos textos de origem. A Figura 2 mostra um sumário extrativo para os aspectos *duração da bateria* e *câmera*. Assim, para cada aspecto, exibe-se um grupo selecionado de avaliações indicando sua polaridade (positiva ou negativa).

Outro tipo de sumário de opiniões são os abstrativos, nos quais o texto é reescrito utilizando novas sentenças. O sumário da Figura 3, como se pode notar, é abstrativo, pois sintetiza os pontos importantes das avaliações feitas pelos usuários. Na Figura 3, as partes sublinhadas são as novas sentenças geradas. Segundo [Carenini et al. \(2006\)](#), com um sumário abstrativo é possível oferecer uma visão equilibrada entre os vários pontos de vista dos usuários sobre cada aspecto.

Conforme será mostrado nos próximos capítulos, há muitos trabalhos em sumarização de opiniões para a língua inglesa, mas nenhum trabalho para o Português. Neste cenário,

A opinião dos usuários está mais dividida em relação à duração da bateria. Ao mesmo tempo em que alguns acham que é muito boa, está com melhor duração, outros consideram que tem baixa duração com gps ligado, acaba muito rápido. A grande maioria dos leitores avaliou positivamente a câmera, afirmando que a câmera frontal e traseira são excelentes, suas fotos tem boa qualidade, é ótima. Alguns consideram que a câmera precisa ser melhor.

Figura 3: Exemplo de sumário abstrativo de opiniões

propõe-se, neste projeto de mestrado, investigar métodos de sumarização de opiniões, extrativos e abstrativos, baseados em aspectos, para a língua portuguesa do Brasil.



Figura 4: Busca do *Iphone 5* no Google

1.2 Lacuna

Como foi indicado na seção anterior, cada vez mais usuários deixam comentários expressando suas experiências com os produtos ou serviços. Por exemplo, se uma busca simples no Google for realizada procurando “*opiniões sobre o iphone 5*”, como é mostrado na Figura 4, são encontrados aproximadamente **1940000** resultados⁹.

⁹Busca realizada em 27 de Fevereiro de 2015.

Se a mesma busca for feita em um *site* especializado em comparações de produtos, o cenário é parecido. Se forem procuradas opiniões sobre o *Iphone 5* no Buscapé, são encontrados mais de **2269** comentários só para o primeiro produto (observar Figura 5). Esta quantidade é bem menor do que a anterior, mas continua sendo bastante. E, embora todas estas opiniões sejam muito úteis para os usuários, dificilmente uma pessoa poderá ler os **2269** comentários. Fazer uma análise manual deste grande volume de informação é uma tarefa quase impossível.

Filtros selecionados Celular e Smartphone > **iphone 5: encontramos 24 produtos**
Não era o que você procurava? [Veja aqui outros resultados com "iphone 5"](#) Como comprar Celular e Smartphone

Selecione produtos para **Comparar lado a lado** **Adicione à lista** Ordenado por: Popularidade

Produto	Preço	Comparar
Smartphone Apple iPhone 5 16GB Desbloqueado ★★★★★ 2.269 pessoas avaliaram Funções Extras: Agenda, Alarme, Calculadora, Calendário, Comando de Voz, Despertador, Gravador de Voz, Organizador, Reproduz Vídeo, Videoconferência, Viva Voz Resolução da Câmera: 8.0 MP Conexões: Bluetooth, EDGE, Fone de Ouvido, GPRS, USB 2.0, Wi-Fi	De: R\$ 1.599,00 até: R\$ 3.598,00 em 96 lojas	COMPARE 208 PREÇOS
Smartphone Apple iPhone 5 32GB Desbloqueado ★★★★★ 325 pessoas avaliaram Funções Extras: Agenda, Alarme, Calculadora, Calendário, Comando de Voz, Despertador, Gravador de Voz, Organizador, Reproduz Vídeo, Videoconferência, Viva Voz Resolução da Câmera: 8.0 MP Conexões: Bluetooth, EDGE, Fone de Ouvido, GPRS, Wi-Fi	De: R\$ 2.077,00 até: R\$ 3.998,00 em 48 lojas	COMPARE 145 PREÇOS
Smartphone Apple iPhone 5 64GB Desbloqueado ★★★★★ 232 pessoas avaliaram Funções Extras: Agenda, Alarme, Calculadora, Calendário, Comando de Voz, Despertador, Gravador de Voz, Reproduz Vídeo, Videoconferência, Viva Voz Resolução da Câmera: 8.0 MP Conexões: Bluetooth, EDGE, Fone de Ouvido, GPRS, Wi-Fi	De: R\$ 2.249,10 até: R\$ 4.149,00 em 46 lojas	COMPARE 134 PREÇOS

Figura 5: Busca do *Iphone 5* no Buscapé

Neste cenário, ler e analisar esta enorme quantidade de opiniões é um grande problema para os internautas e para as empresas. É um problema para os internautas, porque eles utilizam muito tempo lendo as opiniões, e nem sempre os comentários que leem são os mais importantes ou têm relação com as características do produto ou serviço que desejam. E é

um problema para as empresas, porque, além de precisarem de muito tempo para revisar as opiniões, elas podem não obter os *feedbacks* (opiniões) mais relevantes dos seus clientes.

A área de Sumarização de Opiniões pretende auxiliar neste problema. E, embora haja muitos trabalhos nesta área para a língua inglesa, a maioria deles tem resultados insatisfatórios. No caso do português do Brasil, como foi mencionado anteriormente, não há trabalhos em sumarização de opiniões, o que acentua, ainda mais, este problema.

1.3 Objetivos

Em linhas gerais, este trabalho de mestrado tem dois objetivos a serem atingidos:

- Investigar e desenvolver alguns métodos de sumarização de opiniões, extrativos e abstrativos, baseados em aspectos para a língua portuguesa do Brasil.
- Avaliar e comparar os sumários gerados pelos métodos extrativos e abstrativos em relação à informatividade, qualidade linguística e utilidade dos sumários, em um corpúsculo de opiniões em Português do Brasil.

1.4 Hipóteses

Para este trabalho de pesquisa, identificam-se as seguintes principais hipóteses:

- Os métodos automáticos de sumarização tradicional têm um pior desempenho, utilizando as medidas de informatividade, qualidade linguística e utilidade do sumário, na sumarização de opiniões.
- Acredita-se que, no corpúsculo de opiniões em Português do Brasil, os resumos gerados pelos métodos abstrativos de sumarização são mais concisos e menos redundantes que os sumários extrativos.
- Os usuários preferem os sumários abstrativos em vez dos extrativos, pois eles são mais úteis para a tomada de decisão de uma compra.

1.5 Metodologia de Trabalho

Para atingir os objetivos e verificar as hipóteses deste trabalho, como primeiro passo foi feito um estudo dos conceitos mais relevantes da tarefa de sumarização de opiniões com base em aspectos. Foi feita também uma revisão dos principais trabalhos dessa tarefa. Dos métodos clássicos revisados, foram analisados e selecionados alguns deles para ser implementados e adaptados para a língua portuguesa brasileira. A seleção foi baseada, principalmente, na diversidade dos métodos (tentou-se selecionar métodos com características diferentes) e na qualidade dos seus resultados.

Como consequência da análise feita nos métodos estudados, produziram-se duas propostas para gerar sumários de opiniões (um método extrativo e um método abstrativo). Estas duas propostas tentam utilizar as principais vantagens dos métodos clássicos para gerar melhores sumários.

Para avaliar o desempenho dos métodos implementados, foram utilizadas três medidas tradicionais na tarefa de sumarização automática: informatividade, qualidade linguística e utilidade do sumário. A fim de avaliar a informatividade dos sumários, foi construído um corpus de sumários humanos de opiniões, pois, até onde se sabe, não existia nenhum corpus desse tipo. No caso da qualidade linguística e utilidade dos sumários, foi requerida a participação de assessores humanos para avaliar os sumários com base nessas duas medidas.

1.6 Estruturação do Documento

O conteúdo da presente monografia está estruturado em mais seis capítulos, além deste, os quais são explicados a seguir.

No Capítulo 2, explicam-se os conceitos básicos que são utilizados na monografia, os quais incluem definições de sumarização de opiniões, análise de sentimentos, entre outros.

No Capítulo 3, é feita uma descrição dos trabalhos relacionados com o tema de pesquisa desta monografia, e são analisadas as características principais das distintas propostas.

No Capítulo 4, descrevem-se as características do corpus criado para as avaliações dos métodos automáticos. Além disso, mostram-se os resultados obtidos de algumas análises

sobre o *córpus*.

No Capítulo 5, são descritos, em profundidade, os métodos de sumarização de opiniões baseados em aspectos investigados neste trabalho. Neste capítulo, são descritas também duas propostas para gerar resumos de opiniões.

No Capítulo 6, explicam-se os métodos de avaliação utilizados nos experimentos. Os resultados obtidos nas distintas avaliações são descritos também neste capítulo.

Por último, no Capítulo 7, são apresentadas as considerações finais do presente trabalho de mestrado.

Fundamentação Teórica

Neste capítulo, é apresentada uma visão geral da tarefa de sumarização de opiniões baseada em aspectos e, para isso, são apresentados vários conceitos da área. Este capítulo está dividido em quatro seções. Na Seção 2.1 e na Seção 2.2, são descritas as duas maiores áreas da sumarização de opiniões: a sumarização de textos e a mineração de opiniões, respectivamente. Na Seção 2.3, são apresentadas as abordagens da sumarização de opiniões. Finalmente, na Seção 2.4, são descritas as etapas da sumarização de opiniões baseada em aspectos, assim como alguns sistemas deste tipo de sumarização.

2.1 Sumarização Automática de Textos

A Sumarização Automática de Textos é a tarefa que consiste na produção automática de uma versão mais curta (conhecida como sumário ou resumo) de um ou mais textos-fonte (Mani, 2001a). O sumário deve conter as informações mais relevantes dos textos-fonte. Muitas abordagens têm sido desenvolvidas para determinar de maneira automática quais informações devem ser incluídas no resumo.

Segundo Mani (1999) e Nenkova e McKeown (2011), os sumários podem ser classificados de várias formas. Quanto à informação que contêm, sumários podem ser classificados

em três tipos: indicativos, informativos e críticos/avaliativos. Os sumários indicativos contêm só os tópicos essenciais dos textos-fonte, não necessariamente contendo detalhes de resultados, argumentos e conclusões. Por exemplo, os *índices* são sumários indicativos. Os sumários informativos, que são os mais tradicionais, são considerados substitutos dos textos, devendo conter todas as informações principais. Os *abstracts* de artigos são exemplos deste tipo de sumário. Os sumários críticos, além de resumir o conteúdo dos textos-fonte, adicionam crítica em função do conteúdo. As *resenhas de livros* são exemplos de sumários críticos.

A sumarização automática também pode ser classificada como monodocumento ou multidocumento. Na primeira classe, o sumário é gerado automaticamente a partir de um único texto-fonte. Na segunda categoria, é produzido um sumário a partir de um conjunto de textos sobre um mesmo tópico.

Em relação à forma de composição dos sumários, têm-se a sumarização extrativa (quando se selecionam segmentos textuais inteiros) e a abstrativa (quando se realizam operações de reescrita). A seguir, estas abordagens são explicadas, assim como são citados brevemente alguns trabalhos desenvolvidos para essas abordagens.

2.1.1 Sumarização Extrativa

A sumarização extrativa gera um sumário selecionando os segmentos mais representativos (usualmente sentenças) dos textos-fonte, sem fazer nenhuma mudança nos segmentos. A ideia desta abordagem é extrair as sentenças que contenham muita informação e novidade.

Na seleção das sentenças mais relevantes, os métodos extrativos utilizam um mecanismo de ranqueamento para obter as sentenças com as melhores pontuações, ou seja, as mais importantes. Entre os métodos mais utilizados nesta abordagem, tem-se: métodos baseados em grafos (Erkan e Radev (2004), Mihalcea (2004), Wan e Xiao (2007)), baseados em aprendizado de máquina (Kupiec et al. (1995), Svore et al. (2007), Schilder e Kondadadi (2008)) e baseados em informação estatística (McCargar (2004), Gotti et al. (2007), Lloret e Palomar (2009)).

Em relação aos métodos baseados em grafos, a ideia consiste em representar o texto todo como um grafo, no qual os nós geralmente representam elementos do texto (por

exemplo, palavras) e as arestas descrevem as relações entre esses elementos (por exemplo, relações semânticas). Assim, aproveitando a topologia do grafo, podem-se identificar os elementos salientes do texto. [Erkan e Radev \(2004\)](#) propõem o LexRank, um sistema de sumarização multidocumento, no qual os nós representam as sentenças do documento e as arestas descrevem a similaridade entre sentenças. [Mihalcea \(2004\)](#) realiza uma análise de vários algoritmos baseados em grafos para a extração de sentenças relevantes na sumarização automática. [Wan e Xiao \(2007\)](#) apresentam um método de sumarização multidocumento diferenciando as arestas intra-documento e inter-documento.

Nos métodos baseados em aprendizado de máquina, [Kupiec et al. \(1995\)](#) propuseram uns dos primeiros trabalhos para esta abordagem. Nessa pesquisa, para cada sentença os autores calculam a probabilidade de ela pertencer ao sumário final. Tal probabilidade é calculada utilizando aprendizagem bayesiana ([Molina, 1931](#)). [Svore et al. \(2007\)](#) apresentam o NetSum, um sistema de sumarização monodocumento que utiliza o RankNet ([Burgess et al., 2005](#)) como algoritmo de aprendizado. [Schilder e Kondadadi \(2008\)](#) propõem o FastSum, um sumariador multidocumento que utiliza uma técnica de aprendizado chamada *Support Vector Regression* ([Drucker et al., 1997](#)).

Os métodos baseados em informação estatística partem da suposição que as palavras mais frequentes são as que indicam o tema principal de um documento. [McCargar \(2004\)](#) analisa as vantagens e desvantagens de várias técnicas estatísticas, tais como frequência da palavra e TF-IDF. Estas técnicas foram utilizadas também por [Gotti et al. \(2007\)](#) para fazer um ranqueamento das sentenças do texto. [Lloret e Palomar \(2009\)](#) usaram a frequência das palavras em combinação com o tamanho dos sintagmas nominais para calcular a relevância de uma sentença.

Para a língua portuguesa do Brasil, as pesquisas em sumarização extrativa são recentes, iniciando-se oficialmente em 2005 com o sistema GistSumm ([Pardo, 2005](#)). Outros sistemas de sumarização automática de textos são os propostos por [Castro Jorge e Pardo \(2010\)](#), [Cardoso et al. \(2011\)](#) e [Ribaldo et al. \(2012\)](#), os quais utilizam o cópulo CSTNews ([Cardoso et al., 2011](#)), um cópulo de textos jornalísticos em português do Brasil.

Independentemente da língua, como os métodos de sumarização extrativa não realizam mudanças nas sentenças dos sumários, estes resumos apresentam alguns problemas em sua qualidade, por exemplo, a falta de coesão. A falta de coesão refere-se à ausência da

relação de continuidade entre sentenças. Na Figura 6, observa-se a falta de coesão entre as sentenças 2 e 3 do sumário.

[A Receita Federal intensificou a fiscalização e o resultado foi um aumento do número de contribuintes que caíram na malha fina.]¹ [A expectativa da Receita é que até o final do ano mais de 300 mil contribuintes sejam autuados pela malha fina.]² [Elas podem recorrer administrativamente na própria administração tributária ou na Justiça.]³

Figura 6: Exemplo de falta de coesão textual

2.1.2 Sumarização Abstrativa

Diferente da abordagem extrativa, a sumarização abstrativa não apenas seleciona as sentenças dos textos-fonte, ela analisa os documentos e automaticamente gera novas sentenças. Esta abordagem tenta produzir novos textos a partir dos fragmentos originais identificados como importantes. Com esta característica, esta abordagem pode solucionar o problema de falta de coesão das abordagens extrativas.

Apesar de não ser uma questão nova, há relativamente poucos trabalhos sobre sumarização abstrativa. Esta ausência de trabalhos deve-se à grande dificuldade para gerar sumários abstrativos. Com efeito, esta abordagem requer muitos recursos, tais como gramáticas, léxicos e conhecimento do domínio na análise computacional.

Segundo [Ganesan \(2013\)](#), a sumarização abstrativa pode ser classificada em duas categorias: (1) abordagens com conhecimento a priori e (2) abordagens que utilizam técnicas de Geração de Linguagem Natural (GLN). Nas primeiras, definem-se manualmente *frames* ou *templates*, os quais são preenchidos com dados extraídos dos textos originais (([Harabagiu e Maiorano, 2002](#)), ([Jung e Jo, 2003](#))). Já no segundo caso, utilizam-se análises mais profundas de PLN para a geração de textos (([Radev e McKeown, 1998](#)), ([White et al., 2001](#))). Outros métodos, como compressão de sentenças ([Zajic et al., 2007](#)) e fusão de sentenças ([Barzilay e McKeown, 2005](#)), têm sido também aplicados para a geração de resumos abstrativos.

Em relação às duas principais abordagens, apesar da menor complexidade dos métodos com conhecimento a priori, nos trabalhos de [Van Deemter et al. \(2005\)](#) e [Embar et al. \(2013\)](#), mostra-se que os *templates* oferecem uma vantagem em cenários em que se precisa de informações específicas.

Para o português brasileiro, existem poucas iniciativas de trabalhos sobre sumarização abstrativa. Destaca-se o trabalho de [Seno e Nunes \(2010\)](#). As autoras propõem um método para a fusão de sentenças similares em português do Brasil utilizando uma abordagem independente de domínio. Esse método visa eliminar redundâncias em sistemas de sumarização multidocumento.

2.2 Mineração de Opinião/Análise de Sentimentos

A Mineração de Opinião, também conhecida como Análise de Sentimentos, é a tarefa de analisar e classificar as informações subjetivas e os sentimentos associados a um alvo específico ([Pang et al. \(2002\)](#), [Lloret et al. \(2012\)](#)). Geralmente estes sentimentos são classificados como *positivo* ou *negativo* segundo seu contexto.

Outro modelo também utilizado para classificar as emoções ou sentimentos é o *Hourglass* das Emoções (ver Figura 7). O *Hourglass* das Emoções é um modelo de categorização afetiva das emoções humanas proposto por [Plutchik \(2001\)](#). No modelo, os sentimentos são reorganizados em torno de quatro dimensões independentes, cujos níveis de ativação compõem todos os estados emocionais da mente ([Cambria et al., 2012](#)). Cada uma das quatro dimensões (Simpatia, Atenção, Sensibilidade e Atitude) são compostas por seis níveis de ativação, chamados “*Níveis Sentic*”. Estes níveis determinam a intensidade da emoção expressa e são rotulados em um conjunto de 24 sentimentos básicos.

Porém, devido à complexidade de identificar estas 24 classificações do sentimento, a grande maioria dos trabalhos em mineração de opinião foca-se em trabalhar com as categorias positivo e negativo. Por exemplo, o texto “*O Galaxy S3 é um excelente smartphone*” é considerado positivo, já que o autor mostra um comentário favorável a respeito do produto. Por outro lado, o texto “*O preço do Iphone 5 é absurdamente alto no Brasil*” é considerado negativo, porque apresenta uma ideia de rejeição a uma característica do produto.

Segundo [Pang e Lee \(2008\)](#), a classificação da orientação semântica pode ser feita em três níveis: nível de documento, nível de sentença e nível de entidade ou aspecto. Para mostrar o escopo dos três níveis, ilustra-se na Figura 8 uma opinião sobre um *celular Samsung* composto de quatro sentenças.



Figura 7: *Hourglass* das Emoções

No nível de documento, a tarefa consiste em classificar o documento completo como positivo ou negativo (Turney, 2002). Na Figura 8, no nível de documento, a opinião é classificada como positiva, pois em geral a avaliação é favorável. No nível de sentença, deve-se determinar se as sentenças expressam um sentimento positivo, negativo ou neutro. No exemplo, a primeira sentença é considerada neutra, e as três restantes são classificadas como positivas. No nível de entidade ou aspecto, o objetivo é analisar e descobrir os sentimentos (positivo ou negativo) dos aspectos ou entidades. Na Figura 8, identificam-se três aspectos: câmera, som e bateria. A câmera e o som têm uma avaliação positiva e a bateria tem duas avaliações, uma positiva e outra negativa.

[Eu comprei um celular Samsung há duas semanas.]¹ [Sua nova câmera realmente é eficiente.]² [O som é claro e a duração da bateria é longa, mas é um pouco grande.]³ [Fora este último, é um produto maravilhoso!]⁴

Figura 8: Opiniões sobre o celular Samsung

Existem duas principais abordagens na mineração de opinião: baseadas em léxicos e baseadas em aprendizado de máquina. Os métodos baseados em léxico (Ding et al. (2008), Thelwall et al. (2010), Taboada et al. (2011)) determinam a polaridade ou sentimento da opinião utilizando um dicionário de palavras sentimentais (por exemplo “ótimo”, “ruim” e “excelente”). Os métodos baseados em aprendizado de máquina (Turney (2002), Pang e Lee (2004), Lin e He (2009)) normalmente treinam classificadores de sentimento utilizando algumas características, como unigramas e bigramas. Entre as principais técnicas para o aprendizado, destacam-se: Naive Bayes (Kim et al., 2002), Máxima Entropia (Berger et al., 1996) e *Support Vector Machines* (Burges, 1998).

Para o Português do Brasil, há alguns trabalhos em construção de córpus com opiniões anotadas, criação de léxicos de sentimentos, e sistemas de classificação de sentimentos e visualização de informação. Carvalho et al. (2011) e Freitas et al. (2012) criaram corpora anotados em nível de aspecto dos domínios da política e de livros, respectivamente. Souza et al. (2011), Silva et al. (2012) e Balage Filho et al. (2013) apresentam os léxicos de sentimentos OpinionLexicon, SentiLex e LIWC, respectivamente, sendo que Balage Filho et al. (2013) avaliam comparativamente esses recursos na tarefa de classificação de sentimentos. Em relação a sistemas, Chaves et al. (2012) fazem a classificação de sentimentos no domínio da hotelaria. Siqueira e Barros (2010) investigam a extração de aspectos no domínio de lojas *online*.

2.3 Sumarização de Opiniões

A tarefa de Sumarização de Opiniões resulta da combinação das tarefas de Mineração de Opiniões e Sumarização Automática de Textos, e tem por objetivo gerar sumários ou resumos de opiniões. A Mineração de Opiniões encontra a orientação semântica das opiniões e a Sumarização Automática de Textos identifica as opiniões mais relevantes e gera um sumário delas.

Segundo Ku et al. (2006), a sumarização de opiniões é a tarefa que recupera todas as sentenças relevantes relacionadas com o tema principal do conjunto de documentos (opiniões), determina a polaridade de cada sentença relevante e, finalmente, resume as sentenças positivas e negativas.

A sumarização de opiniões pode ser vista como uma forma de sumarização tradicional multidocumento (sumarização de múltiplos documentos), pois trabalha com um conjunto de opiniões (documentos). Mas, além de trabalhar com vários documentos, a sumarização de opiniões tem peculiaridades diferentes da sumarização multidocumento tradicional. As principais diferenças da sumarização de opiniões em relação à sumarização tradicional são:

- Um resumo de opiniões centra-se no sentimento da opinião, nos aspectos (características) do produto ou serviço e também nas comparações de opiniões;
- O formato do sumário de opiniões, utilizado na maioria dos trabalhos, é mais estruturado;
- A sumarização de opiniões tem que lidar com as informalidades das opiniões/comentários (erros ortográficos, uso de gírias, etc.), provenientes principalmente da web.

Além disso, de acordo com [Pang e Lee \(2008\)](#), enquanto na sumarização tradicional a informação redundante é descartada, a sumarização de opiniões analisa e reporta o grau de redundância, pois a redundância indica a importância das opiniões.

2.3.1 Tipos de Opiniões

De acordo com o trabalho realizado por [Liu \(2012\)](#), as opiniões podem ser classificadas segundo a forma que elas estão expressas no texto, e podem ser de dois tipos: opiniões regulares e opiniões comparativas, explicadas abaixo.

A. Opiniões Regulares

Usualmente, as opiniões regulares são referidas simplesmente como opiniões. Uma opinião regular é uma atitude ou valoração positiva, negativa ou neutra de uma pessoa ou organização sobre uma entidade ou um aspecto dessa entidade ([Liu, 2011](#)). Por exemplo, a opinião “*A qualidade das ligações do Galaxy S3 é muito boa*” expressa um sentimento positivo sobre a qualidade das ligações.

B. Opiniões Comparativas

A opinião comparativa mostra uma relação de semelhança ou diferença entre duas ou mais entidades baseada em alguns dos aspectos comuns das entidades (Jindal e Liu (2006a), Jindal e Liu (2006b)). Para a opinião “*O Galaxy S3 tem melhor som que o Iphone 5*”, compara-se o *Galaxy S3* com o *Iphone 5*, baseando-se no som (aspecto) e expressa-se preferência pelo *Galaxy S3*.

2.3.2 Abordagens

Em geral, existem três grandes abordagens na sumarização de opiniões: Sumarização Tradicional, Sumarização Contrastiva e Sumarização baseada em Aspectos (Liu, 2012). Estas abordagens serão explicadas a seguir.

A. Sumarização Tradicional

A sumarização tradicional gera resumos sem levar em conta os aspectos nem os sentimentos das opiniões. Devido a isso, esta abordagem pode selecionar para o resumo sentenças que não tenham relação com os aspectos ou os sentimentos. Além disso, não considera o *lado quantitativo* de um sumário, o qual se refere à porcentagem ou número de opiniões positivas ou negativas dos aspectos ou entidades.

Existem vários trabalhos que utilizam esta abordagem para sumarizar opiniões. Beineke et al. (2003) propõem um método de aprendizado supervisionado para selecionar as sentenças mais importantes em um conjunto de comentários. Em (Seki et al., 2006), os autores apresentam um algoritmo de agrupamento de parágrafos para selecionar as sentenças mais relevantes. Wang e Liu (2011) utilizam dois métodos tradicionais de sumarização (baseado em grafos e ranqueamento) para gerar resumos de conversas sobre diferentes produtos.

B. Sumarização Contrastiva

Este tipo de sumarização foca-se em procurar opiniões com pontos de vista contrários. Por exemplo, uma pessoa pode dar a opinião “*A bateria do Iphone dura muito tempo*” e outra pessoa pode dizer o contrário: “*A bateria do meu Iphone tem pouco tempo de*

duração quando escuto música”. Estas duas opiniões podem dar ao leitor um ponto de vista comparativo das diferentes opiniões.

Das pesquisas sobre sumarização contrastiva, destaca-se o trabalho de [Kim e Zhai \(2009\)](#), no qual os autores propõem o sistema COS (do inglês, *Contrastive Opinion Summary*). Neste trabalho, dado um conjunto de sentenças positivas e outro de sentenças negativas, o COS extrai k pares de sentenças contrastivas que opinam sobre o mesmo tópico de um produto. Estes k pares de sentenças compõem o sumário final. [Lerman e McDonald \(2009\)](#) produzem resumos contrastivos de opiniões sobre dois produtos diferentes, destacando as diferenças das opiniões sobre eles. [Paul et al. \(2010\)](#) propõem um método similar, mas consideram a representatividade e contrastividade das sentenças.

C. Sumarização baseada em Aspectos

Um sistema de sumarização de opiniões baseado em aspectos tem como entrada um conjunto de comentários sobre um produto ou serviço e produz para cada aspecto um resumo das opiniões mais relevantes ([Blair-Goldensohn et al., 2008](#)). Por exemplo, para as opiniões do *Galaxy S3*, os possíveis aspectos poderiam ser: a bateria, a tela, o som e o preço.

A sumarização baseada em aspectos tem duas principais características. A primeira característica é que este tipo de sumarização centra-se nos aspectos ou entidades da opinião e na orientação do sentimento deles. A segunda característica é o *lado quantitativo* (a porcentagem ou número de opiniões positivas ou negativas). Na próxima seção, esta abordagem será explicada com mais detalhes, pois é o foco deste projeto.

2.4 Sumarização baseada em Aspectos

A sumarização baseada em aspectos, também conhecida como sumarização baseada em características ([Hu e Liu \(2004\)](#), [Liu et al. \(2005\)](#)) ou sumarização de sentimentos ([Titov e McDonald, 2008b](#)), divide o texto de entrada em aspectos (características ou sub-tópicos) e gera um sumário para cada aspecto. Segundo [Hu e Liu \(2004\)](#), um aspecto, em termos gerais, é uma característica ou função de um produto.

A sumarização de opiniões baseada em aspectos é muito útil quando as opiniões são

diferentes para cada aspecto, pois um sumário baseado em aspecto apresenta a distribuição das opiniões para cada aspecto separadamente, podendo ser mais informativo para o usuário. As abordagens baseadas em aspecto são muito populares e têm sido fortemente exploradas nos últimos anos (Hu e Liu (2004), Liu et al. (2005), Carenini et al. (2006), Lerman et al. (2009), Ganesan et al. (2010), Tadano et al. (2010), Shimada et al. (2011), Wang et al. (2013)).

Existem diferentes formatos para mostrar esta informação para diferentes níveis de especificidade. Porém, independentemente dos formatos de resumo, o objetivo da sumarização de opiniões baseada em aspectos é ajudar os usuários a apreenderem rapidamente a grande quantidade de opiniões de uma forma fácil.

2.4.1 Etapas na Sumarização baseada em Aspectos

Em geral, na sumarização de opiniões baseada em aspectos, existem três etapas: Identificação dos Aspectos, Predição do Sentimento e Geração do Sumário (Kim et al., 2011). Porém, alguns trabalhos integram estas três etapas em uma só (Mei et al. (2007), Titov e McDonald (2008a)).

A. Identificação de aspectos

Esta etapa visa identificar os aspectos ou tópicos em uma coleção de opiniões. Por exemplo, em um conjunto de comentários sobre celulares, possíveis aspectos poderiam ser: preço, peso, processador, etc.

Na maioria dos casos, os aspectos/características são substantivos ou sintagmas nominais. Devido a isso, alguns trabalhos (Hu e Liu (2004), Popescu e Etzioni (2005) e Lu et al. (2009)) utilizaram um etiquetador morfossintático (POS de *Part-Of-Speech* tagger, no inglês) para identificar os aspectos. Porém, em alguns casos, os aspectos não estão explícitos nas opiniões. Por exemplo, na opinião: “*O Galaxy S3 é muito leve*”, o aspecto é o *peso*, o qual não está presente na opinião. Neste cenário, a utilização de conhecimento do domínio ajuda na identificação dos aspectos (Zhuang et al., 2006).

B. Predição do Sentimento

A predição do sentimento determina a orientação semântica ou polaridade (positiva ou negativa) dos aspectos encontrados na primeira etapa. Esta etapa permite garantir que o sumário final contenha tanto opiniões positivas como negativas. Caso contrário, o sumário poderia conter opiniões que expressam só um dos sentimentos.

A predição do sentimento fornece uma visão geral do que as pessoas pensam sobre um produto ou serviço. Por exemplo, muitos usuários podem opinar que “*A velocidade do processador do Galaxy S3 é excelente*”, enquanto outros usuários poderiam não pensar o mesmo.

Como foi dito na Seção 2.2, aplicam-se abordagens baseadas em léxico e aprendizado de máquina para determinar a orientação semântica. No entanto, a questão principal, na predição do sentimento em nível de aspecto, consiste em determinar o escopo de cada expressão opinativa ou qualificador, ou seja, conhecer se a expressão opinativa qualifica o aspecto de interesse.

A maioria dos trabalhos utiliza um analisador gramatical para determinar o escopo dos aspectos. [Boiy e Moens \(2009\)](#) propõem utilizar um analisador gramatical considerando a posição da palavra em relação ao aspecto alvo na árvore gramatical. [Jiang et al. \(2011\)](#) apresentam uma proposta similar, mas não utilizam a posição da palavra analisada.

C. Geração do Sumário

Nesta última etapa, são gerados e apresentados os sumários dos dados processados nas etapas anteriores. Para cada uma das polaridades (positivo e negativo), é gerado um resumo, o qual pode estar em formato estruturado, textual ou visual.

A etapa de geração do sumário é a responsável por determinar quais sentenças serão incluídas no resumo, e gerar o resumo final. Os critérios mais utilizados para selecionar as sentenças do resumo são: sentenças com maior polaridade (as mais positivas e negativas), sentenças mais redundantes, e sentenças com opiniões mais opostas.

2.4.2 Formatos dos Sumários baseados em Aspectos

Na sumarização tradicional multidocumento, a maioria dos trabalhos tem focado principalmente em gerar sumários textuais. Porém, na sumarização baseada em aspectos, muitos pesquisadores têm proposto vários formatos de saída para os resumos automáticos. Os principais formatos de saída de um sumário baseado em aspectos são: estruturado em aspectos, textual e visual.

A. Sumário Estruturado em Aspectos

Estes sumários são agrupados em aspectos e pela polaridade deles. Logo, para alguns aspectos, são mostradas as opiniões ou sentenças mais relevantes de acordo com alguma heurística. Adicionalmente, estes sumários mostram o número de comentários positivos/negativos ou o número médio de estrelas. O número de estrelas refere-se à pontuação (geralmente de 1 até 5) que oferecem as páginas web para avaliar um produto. Muitos sistemas adotam o formato estruturado para a sumarização de opiniões baseada em aspectos.

A Figura 9 mostra um sumário estruturado em aspectos, sobre opiniões de câmeras. Como se pode ver, o sumário é organizado em função do aspecto “*imagem*”. Para este aspecto, são apresentadas as sentenças extraídas dos textos de origem agrupadas em sentenças positivas e negativas. Na figura, vê-se que há 12 sentenças que avaliam positivamente a imagem da câmera e somente 2 que a avaliam negativamente.

B. Sumário Textual

Os sumários textuais, em muitos casos, são gerados simplesmente listando-se algumas das opiniões ou sentenças do conjunto todo. As opiniões ou sentenças devem ser representativas dos aspectos. Na Figura 10, observa-se um sumário textual extrativo e um sumário textual abstrativo localizados na parte superior e inferior da imagem, respectivamente. Ambos os resumos falam sobre o produto *Canon G3*. No resumo extrativo, é mostrada uma lista simples de sentenças separadas pelo ponto. Entretanto, para o sumário abstrativo, é feita uma análise da preferência dos usuários.

Feature: picture

Positive: 12

- Overall this is a good camera with a really good picture clarity.
- The pictures are absolutely amazing - the camera captures the minutest of details.
- After nearly 800 pictures I have found that this camera takes incredible pictures.

...

Negative: 2

- The pictures come out hazy if your hands shake even for a moment during the entire process of taking a picture.
- Focusing on a display rack about 20 feet away in a brightly lit room during day time, pictures produced by this camera were blurry and in a shade of orange.

Figura 9: Sumário Estruturado (Hu e Liu, 2004)

Extract. Bottom line, well made camera, easy to use, very flexible and powerful features to include the ability to use external flash and lens/filters choices. It has a beautiful design, lots of features, very easy to use, very configurable and customizable, and the battery duration is amazing! Great colors, pictures and white balance. The camera is a dream to operate in automode, but also gives tremendous flexibility in aperture priority, shutter priority, and manual modes. I'd highly recommend this camera for anyone who is looking for excellent quality pictures and a combination of ease of use and the flexibility to get advanced with many options to adjust if you like.

Abstract. Almost all users loved the Canon G3 possibly because some users thought the physical appearance was very good. Furthermore, several users found the manual features and the special features to be very good. Also, some users liked the convenience because some users thought the battery was excellent. Finally, some users found the editing/viewing interface to be good despite the fact that several customers really disliked the viewfinder. However, there were some negative evaluations. Some customers thought the lens was poor even though some customers found the optical zoom capability to be excellent. Most customers thought the quality of the images was very good.

Figura 10: Sumário Textual (Carenini et al., 2006)

C. Sumário Visual

Os sumários visuais dão uma visão geral do sentimento das pessoas sobre um produto ou aspecto segundo sua polaridade (positiva ou negativa). Com estes sumários, as pessoas podem entender rapidamente o que gostam e o que não gostam do produto em discussão, apesar do grande número de aspectos em avaliação.

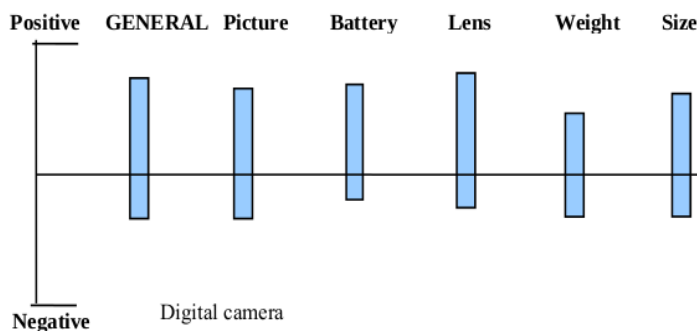


Figura 11: Sumário Visual (Liu, 2012)

A Figura 11 mostra um sumário visual de aspectos de uma câmera digital utilizando barras verticais. Na figura, cada barra acima do eixo das abscissas indica o número de opiniões positivas sobre um aspecto determinado. Por outro lado, a barra abaixo do eixo das abscissas mostra o número de opiniões negativas sobre o mesmo aspecto. Identificam-se os aspectos: geral (General), imagem (*picture*), bateria (*battery*), lentes (*lens*), peso (*weight*) e tamanho (*size*). O aspecto “geral” representa as opiniões sobre a câmera como um todo.

2.4.3 Sumarização Extrativa vs. Abstrativa em Opiniões

Na sumarização de documentos, as abordagens extrativas são fáceis de adaptar-se a domínios diferentes, pois elas limitam-se à extração de sentenças ou expressões, no entanto, os resumos gerados podem ser incoerentes. Por outro lado, as abordagens abstrativas proporcionam sumários mais sofisticados, que muitas vezes contêm material que enriquece o conteúdo original (Hahn e Mani, 2000).

Especificamente na sumarização de opiniões, tem havido poucos trabalhos comparando os métodos extrativos e abstrativos. Segundo Carenini et al. (2006), os métodos extrativos geralmente têm melhor desempenho para a sumarização tradicional, mas estes métodos

não são adequados para a captura de informações subjetivas, pois podem não expressar a distribuição das opiniões.

No trabalho de [Ganesan et al. \(2010\)](#), indica-se que, devido às pequenas variações de opiniões redundantes, os métodos extrativos são muitas vezes inadequados para resumir essas opiniões. Por exemplo, se forem consideradas as três sentenças da Figura 12 e a partir delas um resumo for gerado, com qualquer método extrativo, uma só sentença dessas três será escolhida como sumário. Em qualquer caso, o resumo gerado não vai refletir a variedade das opiniões.

1. A bateria do Galaxy S3 dura muito tempo, só teve que carregá-la uma vez cada três dias.
2. A bateria do Galaxy S3 é grande, mas é barata...
3. A bateria do Galaxy S3 é grande, mas dura muito tempo!

Figura 12: Opiniões sobre o Galaxy S3

Um sumário abstrativo como “*A bateria do Galaxy S3 é barata, dura muito tempo mas é volumosa*” seria muito melhor. Além disso, os sumários gerados pelos métodos extrativos são muito verbosos. Vê-los em dispositivos móveis (como *smartphones* ou *tablets*) é mais difícil. Um sumário abstrativo, conciso e informativo, poderia ser uma solução para este problema.

2.4.4 Sistemas de Sumarização de Opiniões Baseados em Aspectos

São poucos os sistemas de sumarização de opiniões baseados em aspectos. Entre os principais destacam-se: SumView e SEA. Estes dois sistemas trabalham com dados em inglês. Para o português brasileiro, não foi encontrado na literatura relacionada nenhum sistema de sumarização de opiniões baseado em aspectos.

A. SumView

O SumView é um sistema *web* de sumarização de comentários que extrai automaticamente as opiniões mais representativas de vários produtos¹. Desenvolvido por [Wang et al. \(2013\)](#), o SumView gera um sumário estruturado com base nos aspectos de produtos. Para

¹Disponível em <http://rev-sum.appspot.com/>

a produção do sumário, o sistema sugere cinco aspectos predefinidos para um produto. Além disso, oferece a opção de inserir manualmente um aspecto a mais.



Figura 13: Tela inicial de SumView

Na Figura 13, é apresentada a tela inicial do SumView. Nela, o sistema disponibiliza ao usuário a opção de escolher algumas marcas de panelas elétricas de arroz. Para gerar o sumário, o SumView oferece cinco aspectos: tempo (*time*), qualidade (*quality*), preço (*price*), tamanho (*size*) e limpeza (*clean*). Adicionalmente, o usuário pode inserir algum outro aspecto. Na Figura 14, visualiza-se o resumo que é mostrado para o usuário. Como pode-se notar, o sumário contém os cinco aspectos solicitados na tela inicial do SumView. Para cada aspecto, é mostrada só uma sentença.

Devido ao fato do SumView só mostrar uma sentença por aspecto, muitas vezes o sumário pode não conter os diferentes pontos de vista sobre o produto. Neste cenário, o resumo não refletiria o panorama real das opiniões sobre o produto. Essa é a principal desvantagem do SumView.

B. SEA

O sistema SEA (do inglês *Summarizer of Evaluative Arguments*)² é baseado no trabalho de Carenini e Moore (2006), o qual será explicado com mais detalhes no próximo

²Disponível em <http://www.cs.ubc.ca/~carenini/storage/SEA/demo.html>



Summary of Reviews

time: When programming however there is some confusion as to whether or not the time entered is the time it starts cooking RATHER than the time the rice will be finished.

quality: Quality: In general, the quality of the cooker is very good.

price: You may think the price is a bit steep for a rice cooker, but do not be deterred, it is worth the price many times over in convenience, reduced clean up and perfect rice every time!

size: I love the small size, about the size of a toaster, very easy to clean, just the right size for up to 4 people with reasonable appetites.

clean: Clean up is a breeze; a quick wipe down of the unit and warm soap and water for the bowl, and it's clean.

Figura 14: Resumo gerado pelo SumView

capítulo. Para a produção do sumário, o SEA utiliza uma hierarquia de aspectos definida previamente. Esta hierarquia agrupa os aspectos segundo seu nível de generalização. Por exemplo, os aspectos “flash” e “lentes” de um celular poderiam ser agrupados no aspecto “câmera”. O SEA produz tanto sumários textuais quanto visuais. No sumário textual, mostra-se uma descrição abstrativa de cada aspecto. No sumário visual, os aspectos são agrupados segundo a hierarquia definida.

A Figura 15 mostra um sumário gerado para um reprodutor de DVD (modelo Apex AD2600). Para este produto, identificam-se os aspectos “CD”, “DVD” e “DVD-R”, os quais são agrupados no aspecto “Formatos de Disco”. Os sumários textual e visual são mostrados no lado esquerdo e direito da imagem, respectivamente. O tamanho de cada quadro no resumo visual indica o número de menções a cada aspecto. Ao clicar no quadro de algum aspecto, mostra-se a quantidade de opiniões para esse aspecto, assim como a polaridade média das opiniões. A imagem mostra que o aspecto “CD” tem quatro avaliações e a sua polaridade média é -0.25.

Summary of customer reviews for: Apex AD2600 Progressive-scan DVD player

Most customers disliked the Apex AD2600¹. Although many customers found the user interface² to be good, many users thought the range of compatible disc formats³ was poor. Furthermore, many users disliked the video output⁴. However, many customers found the extra features⁵ to be good, even though many customers really disliked the supplied universal remote control⁶.

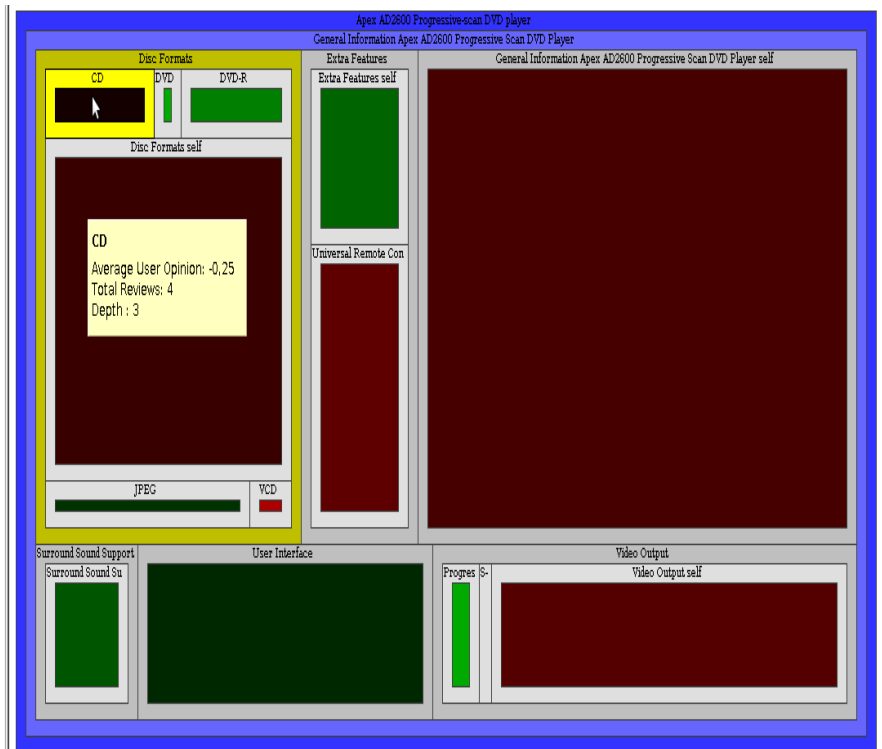


Figura 15: Tela de SEA

Trabalhos Relacionados

Neste capítulo, será apresentada uma revisão dos principais trabalhos relacionados à sumarização de opiniões. Como se mostrará nas próximas seções, há muitos trabalhos em sumarização para a língua inglesa, porém, nenhum trabalho para o português. Na Seção 3.1, são apresentados os trabalhos relacionados que utilizam uma abordagem extrativa. Na Seção 3.2, é feita uma descrição dos trabalhos que geram sumários abstrativos. Uma síntese dos trabalhos estudados é feita na Seção 3.3. Por fim, na Seção 3.4, mostra-se uma linha do tempo das pesquisas mais relacionadas a este trabalho de mestrado.

3.1 Abordagens Extrativas

A maioria dos trabalhos propostos para sumarizar opiniões utiliza uma abordagem extrativa. A seguir, explicam-se detalhadamente, em ordem cronológica as pesquisas mais relevantes na área de sumarização de opiniões, assim como as pesquisas fortemente relacionadas a este trabalho de mestrado.

3.1.1 Trabalho de Beineke et al. (2003)

O primeiro trabalho sobre sumarização de opiniões foi realizado por [Beineke et al. \(2003\)](#). Os autores propõem um método extrativo para selecionar uma única sentença do texto que reflete a opinião completa de seu autor. Como dados, eles utilizam 2500 comentários de filmes do *website Rotten Tomatoes*¹. Este *website* fornece uma sentença representativa para cada opinião, a qual eles consideram como o sumário da opinião.

O trabalho foca na sumarização de opiniões como um problema de classificação no nível de sentença, isto é, os autores determinam se a sentença selecionada automaticamente por seu método é ou não o sumário fornecido pelo *Rotten Tomatoes*. O resumo fornecido pelo *Rotten Tomatoes* é considerado o sumário de referência.

Para selecionar a sentença mais representativa de cada opinião, utilizam-se o algoritmo de Naive Bayes ([Kim et al., 2002](#)) e Regressão Logística ([Genkin et al., 2007](#)). Nestes dois algoritmos, utilizam-se como atributos a frequência das palavras e a posição da sentença. A frequência das palavras refere-se à quantidade de vezes que uma palavra aparece na sentença que é considerada o sumário de referência. A posição da sentença indica se a sentença é a primeira ou não da opinião.

Para avaliar o método proposto, calcula-se a porcentagem de acertos em predizer se uma sentença é ou não o sumário de referência. Na Tabela 2, os resultados dos experimentos são listados. [Beineke et al. \(2003\)](#) utilizaram um método *baseline* chamado “Aleatório”, o qual seleciona qualquer sentença como sumário. Utilizando as duas *features*, Naive Bayes e Regressão Logística obtiveram 23.1% e 25.8% em acertos, respectivamente. Por outro lado, usando só a posição da sentença, Regressão Logística obteve 14.5%. Os resultados obtidos foram baixos, devido à semelhança entre as sentenças candidatas ao sumário.

É importante indicar que este método não produz um resumo de um conjunto de opiniões como um todo, e sim gera um resumo para cada opinião. O critério de utilizar a posição da sentença como *feature* nem sempre é útil, pois, diferentemente dos textos objetivos (por exemplo, notícias), nas opiniões, a ideia principal muitas vezes se encontra na parte final da opinião.

¹<http://www.rottentomatoes.com/>

Método	Features	Porcentagem de acertos
Aleatório	Nenhuma	6.3%
Regressão Logística	Posição da sentença	14.5%
Naive Bayes	Posição da sentença e Frequência da palavra	23.1%
Regressão Logística	Posição da sentença e Frequência da palavra	25.8%

Tabela 2: Porcentagem de acertos nos métodos (Beineke et al., 2003)

3.1.2 Trabalho de Hu e Liu (2004)

O trabalho de [Hu e Liu \(2004\)](#) é a base de muitos outros trabalhos, por exemplo, [Liu et al. \(2005\)](#), [Shimada et al. \(2011\)](#) e [Wang et al. \(2013\)](#). Os autores propõem uma arquitetura de sumarização dividida em três passos principais: (1) identificação dos aspectos/características dos produtos presentes nas opiniões, (2) identificação das sentenças positivas e negativas para cada aspecto e, por fim, (3) produção do sumário utilizando a informação descoberta.

Como entrada, seu sistema recebe os nomes dos produtos e as páginas *webs* que contêm as opiniões, e a saída é um sumário estruturado, como foi mostrado na Figura 9. Para construir esse sumário, primeiramente identificam-se os aspectos (geralmente substantivos) utilizando um etiquetador morfossintático. Em seguida, procuram-se os qualificadores desses aspectos (usualmente adjetivos) e classifica-se a polaridade da sentença como positiva ou negativa. Para a classificação da polaridade, utiliza-se, inicialmente, uma pequena lista de adjetivos (etiquetados manualmente como positivos ou negativos). Depois, como geralmente os adjetivos compartilham a mesma polaridade com seus sinônimos, a base lexical *Wordnet* ([Miller, 1995](#)) é utilizada para procurar os sinônimos dos adjetivos iniciais e, assim, acrescentá-los a essa lista. Por fim, para formar o sumário, agrupam-se as sentenças em função do aspecto e da polaridade.

Na produção do sumário, para cada aspecto descoberto, opiniões relacionadas são agrupadas nas categorias positiva ou negativa segundo a polaridade delas. Uma contagem é realizada para mostrar quantas opiniões são positivas e negativas. No resumo, mostram-se todos os aspectos ranqueados segundo a frequência de aparição nos comentários.

Para os experimentos, utilizaram-se opiniões de cinco produtos eletrônicos: duas câmeras digitais, um reproduutor de DVD, um reproduutor MP3 e um celular. [Hu e Liu](#)

(2004) não avaliaram diretamente a qualidade dos sumários produzidos. Em vez disso, avaliaram as etapas anteriores, ou seja, a identificação de aspectos e a identificação da orientação das polaridades, afirmando que podem ser produzidos bons resultados na sumarização. Para a identificação de aspectos, a proposta obteve 0.68 e 0.56 em cobertura e precisão, respectivamente. Na identificação da orientação das polaridades, a proposta obteve 0.69 e 0.64 em cobertura e precisão, respectivamente.

A principal contribuição deste trabalho é a arquitetura que os autores propõem para a sumarização de opiniões baseada em aspectos. Além disso, é o primeiro trabalho a propor um formato estruturado nos sumários de opiniões. A principal desvantagem talvez seja que os autores não fizeram uma análise profunda na etapa de sumarização.

3.1.3 Trabalho de Sharifi et al. (2010)

Sharifi et al. (2010) desenvolveram dois algoritmos para a sumarização de postagens curtas da rede social Twitter². O primeiro, chamado algoritmo de Reforço da Frase (*Phrase Reinforcement*), utiliza um grafo não dirigido para representar as n-gramas sobrepostas em um conjunto de sentenças relacionadas (ver Figura 16). No grafo, o nó raiz representa o tópico (aspecto) e seus filhos do lado direito e esquerdo representam as palavras adjacentes à direita e à esquerda, respectivamente.

Por exemplo, se forem consideradas as seguintes publicações coletadas no dia da morte do comediante “*Soupy Sales*”:

- 1) *Aw, Comedian Soupy Sales died.*
- 2) *RIP Comedian Soupy Sales dies at age 83.*
- 3) *My favorite comedian Soupy Sales died.*
- 4) *RT @NY: RIP Comedian Soupy Sales dies at age 83.*
- 5) *RIP: Soupy Sales Died Today.*
- 6) *Soupy Sales meant silliness and laughs.*

O grafo gerado pelo algoritmo é mostrado na Figura 16. Nota-se que o nó raiz contém o aspecto, neste caso, “*Soupy Sales*”, e os nós filhos contêm as palavras adjacentes à direita e à esquerda, respeitando a posição delas no comentário. O grafo só considera os caracteres alfanuméricos e todas as palavras convertem-se em minúsculas. O número no

²<https://twitter.com/>

nó indica a quantidade de vezes que a palavra aparece no conjunto de comentários.

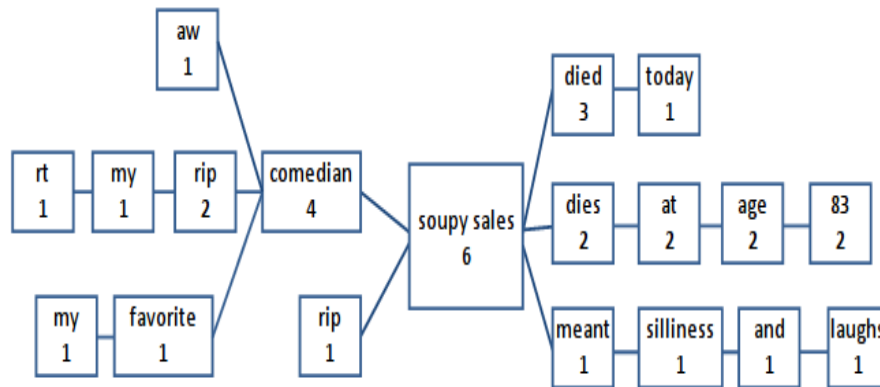


Figura 16: Grafo de palavras (Sharifi et al., 2010)

Para cada palavra presente no grafo, os autores calculam o peso dela utilizando a seguinte fórmula:

$$Peso(Nó) = Frequência(Nó) - Distância(Nó) * \log_2 Frequência(Nó) \quad (3.1)$$

Nesta formula, é considerada a frequência de aparição da palavra e também a distância dela em relação ao aspecto. Para gerar o sumário, o algoritmo procura os caminhos com os maiores pesos tanto à direita quanto à esquerda do aspecto. Para o exemplo, o sumário gerado é “rip comedian soupy sales dies at age 83” (ver Figura 17).

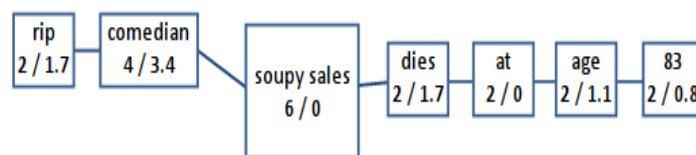


Figura 17: Sumário gerado (Sharifi et al., 2010)

No segundo algoritmo proposto, os autores utilizam a técnica TF-IDF (*Term Frequency-Inverse Document Frequency*) (Salton e McGill, 1986). O TF-IDF concede um peso maior às palavras que ocorrem frequentemente em um conjunto pequeno de documentos (postagens) e menor peso às palavras que ocorrem com pouca frequência ou ocorrem na maioria dos documentos. Para gerar o sumário, a ideia consiste em atribuir a cada sentença dos

comentários um peso que reflete a importância da sentença. Logo, as sentenças são ordenadas por seus pesos e as m sentenças principais são selecionadas como o resumo final. O peso de uma sentença é a soma dos pesos individuais das palavras. Para calcular o peso de uma palavra utiliza-se o TF-IDF. Neste caso, m é predefinido como o tamanho do sumário desejado.

Para os experimentos, os autores definiram 50 tópicos (aspectos) e, para cada um deles, compilaram 1500 comentários. Para avaliar os resultados, utilizaram a medida ROUGE-1 (Lin, 2004). O algoritmo TF-IDF produz melhores sumários do que o algoritmo de Reforço da Frase (0.30 e 0.33 de medida-F respectivamente).

Neste trabalho, os dois algoritmos propostos não consideram o sentimento das opiniões. Porém, os resultados obtidos são bons, pois na sumarização de opiniões um critério importante é a captura da redundância, a qual é feita pelos autores utilizando a frequência de palavras comuns.

3.1.4 Trabalho de Tadano et al. (2010)

Tadano et al. (2010) apresentam um método extrativo de sumarização de opiniões com base em três características ou *features*: (1) pontuações dos aspectos, (2) o valor TF-IDF e (3) o número de menções com tópico semelhante:

- Pontuações dos aspectos: Esta característica se refere à quantidade de estrelas assinaladas pelos usuários em uma escala de 0 até 5. Os autores assumem que os usuários escrevem seus comentários de acordo com as pontuações que dão. Ou seja, se um usuário outorga uma estrela a um aspecto, assume-se que o usuário escreveu um comentário negativo para esse aspecto.
- Valor TF-IDF: Esta característica é utilizada para calcular a importância de uma sentença. No primeiro passo, as sentenças dos comentários são divididas em lemas usando o analisador morfológico Mecab³. Depois, no segundo passo, o valor TF-IDF é calculado para cada uma das palavras das sentenças. Finalmente, a importância de uma sentença é calculada como a somatória dos TF-IDF das palavras da sentença.

³Disponível em <http://mecab.sourceforge.net/>

- Número de menções com tópico semelhante: O objetivo desta característica é agrupar as opiniões similares. Para isso, os autores utilizam o algoritmo K-Means (Steinhaus, 1956).

Com base nestas características, os autores propuseram três métodos de sumarização: O *Método1*, que considera só o valor do TF-IDF; o *Método2*, que utiliza o valor do TF-IDF e a importância do cluster; e o *Método3*, que considera o mesmo que o *Método2* mais a informação das pontuações dos aspectos. Nos três métodos, apresenta-se como sumário a sentença com melhor pontuação para cada aspecto.

Como conjunto de dados, os autores utilizaram comentários escritos em japonês sobre o jogo *New Super Mario Brothers*. Na Tabela 3, são mostrados os resultados obtidos para os três métodos propostos utilizando as medidas ROUGE-1 e ROUGE-2. A ROUGE (Lin, 2004) compara um sumário produzido automaticamente com um sumário de referência (geralmente humano), computando o número de N-gramas em comum. Nos experimentos, tanto para ROUGE-1 quanto para ROUGE-2, o *Método3* obteve os melhores resultados, mostrando, assim, que as três características são importantes na sumarização de opiniões.

		Vício	Conforto	Dificuldade	Gráficos	Música	Originalidade	Satisfação	Média
N=1	Método1	0.287	0.496	0.347	0.421	0.082	0.363	0.383	0.345
	Método2	0.435	0.539	0.333	0.186	0.137	0.372	0.414	0.362
	Método3	0.413	0.487	0.350	0.342	0.127	0.363	0.422	0.367
N=2	Método1	0.117	0.207	0.118	0.370	0.013	0.204	0.103	0.150
	Método2	0.297	0.203	0.127	0.039	0.040	0.176	0.112	0.144
	Método3	0.223	0.146	0.130	0.287	0.030	0.178	0.115	0.151

Tabela 3: ROUGE–N para os sumários automáticos

3.1.5 Trabalho de Shimada et al. (2011)

O trabalho de Shimada et al. (2011) está baseado na pesquisa anteriormente explicada (Seção 3.1.4). Shimada et al. (2011), além de utilizarem as três *features* de Tadano et al. (2010), integram informação objetiva ao sumário de opiniões. Essa informação objetiva é extraída da Wikipedia⁴.

⁴<https://www.wikipedia.org/>

Tendo o sumário de opiniões produzido pelo método de [Tadano et al. \(2010\)](#), os autores realizam o processo de integração da informação objetiva em duas etapas: (1) detecção e (2) alinhamento do KeySum (palavras-chave no sumário) e KeyWik (palavras-chave na Wikipedia).

O KeySum é o conjunto de substantivos que têm os valores TF-IDF mais altos no resumo. Por sua vez, o KeyWik é o conjunto de palavras-chave na Wikipedia e estas palavras são detectadas por meio de informações estruturais, como a itemização. Logo, se uma palavra-chave do KeySum está contida no KeyWik, o alinhamento é realizado. Assim, no sumário final, mostra-se, opcionalmente, para cada elemento do KeySum, seu elemento correspondente no KeyWik.

Para os experimentos, foram utilizadas 324 opiniões sobre três jogos de vídeo: *New Super Mario Brothers* (SM), *Mario Cart* (MC) e *Family Stadium* (FS). Para cada jogo, três anotadores selecionaram manualmente 50 sentenças, as quais formam o sumário. Devido ao grau de concordância entre os anotadores ter sido baixo (0.45), os autores utilizaram o *Acordo Semântico* em vez da ROUGE-N. Para calcular o acordo semântico, apresentam-se para um juiz duas sentenças (uma do sumário manual e outra do sumário automático). Dessa forma, o juiz julga se as sentenças são semelhantes ou não em termos de conteúdo.

O método *baseline* utilizado está baseado somente no valor TF-IDF e os aspectos considerados na avaliação são os mesmos que [Tadano et al. \(2010\)](#) utilizaram. Na Tabela 4, mostram-se os resultados obtidos. Como se pode ver, o método proposto apresenta melhor desempenho em comparação com o *baseline*.

		Vício	Conforto	Dificuldade	Gráficos	Música	Originalidade	Satisfação	Média
SM	Baseline	14.3	37.5	21.4	25.0	33.3	27.3	20.0	25.5
	Proposta	28.6	31.3	50.0	25.0	66.7	36.4	25.0	37.6
MC	Baseline	25.0	33.3	40.0	0.0	20.0	31.8	22.2	24.6
	Proposta	37.5	44.4	0.0	50.0	20.0	45.5	33.3	33.0
FS	Baseline	25.0	13.6	14.3	33.3	0.0	40.0	37.5	23.4
	Proposta	50.0	50.0	42.9	33.3	16.7	30.0	29.2	36.0

Tabela 4: Acordo Semântico entre os resumos automáticos e os manuais (%)

A pesquisa de [Shimada et al. \(2011\)](#) é a única que utiliza informação objetiva nos sumários, diferenciando-se dos outros trabalhos relacionados a este projeto de mestrado.

Além disso, esse trabalho propõe um método de avaliação que pode ser utilizado quando a concordância dos anotadores é baixa.

3.1.6 Trabalho de Xu et al. (2011)

Xu et al. (2011) indicam que, além da relevância do aspecto e a intensidade do sentimento, a representatividade e a diversidade devem ser considerados como requisitos importantes para gerar um bom sumário extrativo baseado em aspectos.

A representatividade refere-se ao fato que as sentenças selecionadas para o sumário devem ser opiniões de aspectos importantes (aspectos muito frequentes). Para um aspecto importante, uma sentença é representativa se ela compartilha opiniões comuns com várias outras sentenças.

Por sua vez, a diversidade refere-se ao resumo dever ter o máximo de cobertura e o mínimo de redundância em relação aos aspectos e as suas opiniões. O resumo deve abranger vários aspectos, em vez de focar-se em poucos e redundantes aspectos. Assim, é possível conhecer opiniões mais diversas.

Para obter as sentenças mais representativas para um determinado aspecto, Xu et al. (2011) criam um grafo, onde os nós representam as sentenças, e as arestas indicam a similaridade entre as sentenças. Logo, para fazer o ranqueamento das sentenças, os autores utilizaram o modelo *Markov Random Walk* (Lovász, 1996). Neste processo, uma sentença propaga a sua pontuação para as sentenças vizinhas em proporção com a similaridade correspondente. Assim, uma sentença semelhante com muitas outras pode receber mais pontuação.

Para ganhar diversidade, as sentenças são ordenadas de forma ascendente segundo sua pontuação. Logo, iterativamente, escolhe-se uma sentença e reduz-se a redundância penalizando as sentenças restantes similares à escolhida. Esse processo de remoção da redundância é repetido até alcançar o número de palavras definido como tamanho do resumo.

Para avaliar seu método proposto, os autores utilizam opiniões sobre restaurantes extraídas do site reviews.opentable.com. Três aspectos predefinidos pelo site são adotados para a avaliação: a comida, o serviço e o ambiente. Como método *baseline*, é considerado o método KL, o qual ranqueia as sentenças de acordo com a divergência de Kullback-Leibler

(Kullback e Leibler, 1951). Outro método *baseline* considerado é o ASMRW-NORR, o qual utiliza o *Markov Random Walk*, mas sem a remoção da redundância.

As comparações entre os diferentes métodos são mostrados na Tabela 5. A abordagem proposta tem o melhor desempenho entre todos os *baselines* em termos da métrica ROUGE. A Tabela 5 demonstra que a representatividade e a diversidade são requisitos importantes para gerar bons sumários.

	ROUGE-1 Cobertura	ROUGE-1 Precisão	ROUGE-1 Medida-F	ROUGE-L Cobertura	ROUGE-L Precisão	ROUGE-L Medida-F
KL	0.4606	0.4066	0.4236	0.4531	0.3998	0.4165
ASMRW-NORR	0.4376	0.4404	0.4308	0.4291	0.4317	0.4222
Proposta	0.5043	0.5021	0.4953	0.4971	0.4948	0.4881

Tabela 5: Comparações do desempenho de diferentes abordagens em opiniões

Com este trabalho, os autores mostram que, além de considerar a importância do aspecto e a intensidade do sentimento, é importante ter em conta a representatividade e a diversidade das opiniões. Outro ponto importante neste trabalho é o fato de ser um dos poucos que consideram o tamanho do sumário em função do número de palavras. Geralmente, na sumarização de opiniões, utiliza-se o número de sentenças como medida para o tamanho do sumário.

3.1.7 Trabalho de Wang et al. (2013)

Wang et al. (2013) propõem uma ferramenta, chamada SumView, para a sumarização de opiniões de produtos. Para a extração de aspectos, SumView baseia-se no trabalho feito por Hu e Liu (2004). Além disso, SumView oferece a opção de adicionar manualmente um aspecto.

Para a geração de sumários, os autores utilizaram uma modificação da técnica de agrupamento Fatoração de Matrizes não-Negativas (NMF, do inglês, *Non-negative Matrix Factorization*) (Li e Ding, 2006). Na sumarização de textos, a NMF é uma técnica que extrai as sentenças com maior probabilidade para um determinado tópico. Na proposta de Wang et al. (2013), a NMF agrupa as sentenças mais similares de um aspecto e seleciona a sentença com maior probabilidade como sumário. Ou seja, para cada aspecto o resumo é a sentença mais representativa.

Nos experimentos, foram utilizadas 173 comentários sobre painéis de arroz. Para avaliar os resultados, foi utilizado o critério de satisfação do usuário. Para isso, um grupo de 25 usuários avaliaram a qualidade dos sumários em uma escala de 1 até 5. Na avaliação, o SumView é comparado com os seguintes métodos de sumarização: (1) NMF; (2) Centroide, o qual extrai a sentença mais representativa utilizando a similaridade do cosseno; (3) Grafo, o qual usa o algoritmo PageRank (Page et al., 1998) para ranquear as sentenças; e (4) Corpernic, um software comercial para sumarização de documentos⁵. Como pode ser visto na Tabela 6, o SumView obteve uma pontuação de 3.72 como média, superando os outros métodos.

Systems	SumView	Corpernic	Centroide	Grafo	NMF
Usuário 1	4.2	3	2	2.4	2
Usuário 2	3.6	2.2	1.6	1.6	1.4
Usuário 3	3.4	1.6	1.6	1.8	1.8
Usuário 4	4	2.6	2	2.2	2
Usuário 5	4.6	2.2	2.2	2.6	2.6
Usuário 6	3.2	1.4	1.6	1.4	1.4
Usuário 7	3	1.4	1	1.6	1.6
Usuário 8	3.8	2.2	1.4	1.8	1.6
Usuário 9	3.4	1.6	1.6	2	2
Usuário 10	4.4	2.8	2.6	3	2.8
Usuário 11	3.4	1.4	1.8	1.4	1.6
Usuário 12	3.2	1.6	1.2	1.8	1.2
Usuário 13	3.8	2.2	1.8	2.4	2
Usuário 14	4.4	2.8	2	2.8	2.8
Usuário 15	3.8	2.2	1.6	2.4	2.2
Usuário 16	3.6	2.6	1.8	2.6	2.2
Usuário 17	3	1.4	1.6	1.6	1.8
Usuário 18	4.4	2.4	2	2.8	2.2 2
Usuário 19	3.8	2	2	2.2	2.2
Usuário 20	4	2	1.6	2.4	2.6
Usuário 21	3.2	2	1.4	2	2
Usuário 22	3.2	1.4	1.6	2.2	2.2
Usuário 23	4.4	3	2.2	3	2.8
Usuário 24	3.4	1.8	2	2	2
Usuário 25	3.8	2.8	2	3	3
Média	3.72	2.10	1.87	2.2	2.08

Tabela 6: Comparação da satisfação dos usuários (Wang et al., 2013)

⁵Disponível em <http://www.copernic.com/en/products/summarizer/>

3.2 Abordagens Abstrativas

São poucos os trabalhos que utilizam uma abordagem abstrativa para gerar sumários de opiniões. Esta ausência de trabalhos deve-se à grande dificuldade para gerar sumários abstrativos. A seguir, explicam-se detalhadamente, em ordem cronológica, as principais pesquisas de sumarização abstrativa.

3.2.1 Trabalho de Carenini et al. (2006)

Carenini et al. (2006) criaram sumários extrativos e abstrativos de textos opinativos. Para criar sumários extrativos, foi adaptado o sistema de sumarização multidocumento tradicional MEAD (Radev et al., 2003), incorporando atributos relacionados à presença de aspectos nas sentenças dos textos, enriquecendo, assim, o método original do sistema. Este novo método é chamado MEAD*.

Para a abordagem abstrativa, chamada SEA (do inglês *Summarizer of Evaluative Arguments*), os autores inicialmente extraíram os aspectos e seus modificadores, determinaram os aspectos mais importantes em função de suas frequências de ocorrência e calcularam a distribuição de opiniões para cada aspecto. Em seguida, utilizaram o sistema de geração de língua natural GEA (do inglês, *Generator of Evaluative Arguments*) (Carenini e Moore, 2006), o qual sintetiza o sumário a partir dos aspectos mais importantes que foram selecionados.

Os autores avaliaram seus resultados com um conjunto de opiniões sobre produtos (câmeras e reprodutores DVD), pedindo aos participantes do experimento que avaliassem os sumários produzidos do ponto de vista linguístico (com relação a critérios de legibilidade, gramaticalidade, coesão e coerência) e de conteúdo. A escala utilizada na avaliação ia de 1 a 5, onde 5 é a melhor pontuação.

Os autores avaliaram os sumários produzidos pelo MEAD, MEAD* e SEA. Também avaliaram sumários feitos por outras pessoas. Para avaliar o conteúdo, utilizaram três questões distintas, em uma escala de 1 a 5 também:

- (Cobertura) O resumo contém todas as informações que você teria incluído do texto original?

- (Precisão) O resumo não contém nenhuma informação que você não teria incluído a partir do texto original?
- (Exatidão) Toda a informação expressa no sumário reflete com exatidão as informações contidas no texto original?

Na Tabela 7, mostram-se os resultados obtidos. Os experimentos indicaram que, do ponto de vista linguístico, as duas abordagens propostas pelos autores foram equiparadas, mas, do ponto de vista de conteúdo, o sumário abstrativo superou o extrativo. Os autores também concluíram que ambos os sumários são inferiores aos produzidos por humanos e que são superiores aos produzidos pela versão original do MEAD (sem adaptação).

	SEA	MEAD*	MEAD	Sumário Humano
Gramaticalidade	3.43	2.71	3.14	4.29
Não redundância	3.14	3.86	3.57	4.43
Clareza Referencial	3.86	4.00	3.00	4.71
Foco	4.14	3.71	2.29	4.14
Estrutura	2.29	3.00	1.86	4.43
<i>Média</i>	<i>3.37</i>	<i>3.46</i>	<i>2.77</i>	<i>4.40</i>
Cobertura	2.33	2.57	1.57	3.57
Precisão	4.17	3.50	2.17	3.86
Exatidão	4.00	3.57	2.57	4.29
<i>Média</i>	<i>3.50</i>	<i>3.21</i>	<i>2.10</i>	<i>3.90</i>

Tabela 7: Comparação dos métodos (Carenini et al., 2006)

Este trabalho é, provavelmente, um dos trabalhos mais relacionados com esta pesquisa de Mestrado em termos de avaliação, pois realiza uma comparação das abordagens extrativas e abstrativas na sumarização de opiniões. Essa comparação dá uma ideia de como poderiam ser os resultados desta proposta de Mestrado.

3.2.2 Trabalho de de Ganesan et al. (2010)

Ganesan et al. (2010) propõem um método de sumarização abstrativa. Como foi explicado na Seção 2.1.2, geralmente estes métodos precisam de muitos recursos e eles são dependentes do domínio. Porém, o método de Ganesan et al. (2010) é independente do domínio e utiliza, de forma superficial, ferramentas de PLN.

Os autores criaram um grafo direcionado, chamado *Opinions graph* (ver Figura 18), no qual cada palavra única representa um nó e a aresta indica uma relação de precedência da palavra i com a palavra $i+1$. Uma aresta tem maior peso conforme a relação de precedência é repetida. Por exemplo, se forem consideradas as seguintes opiniões:

1. *My phone calls drop frequently with the iPhone.*
2. *Great device, but the calls drop too frequently.*

Opinions graph para estas opiniões é mostrado na Figura 18, na qual o primeiro número abaixo do nó indica o ID da opinião e o segundo número indica a posição da palavra na opinião. Como pode ser visto, os nós *calls*, *drop*, *frequently* e *the* estão presentes nas opiniões 1 e 2.

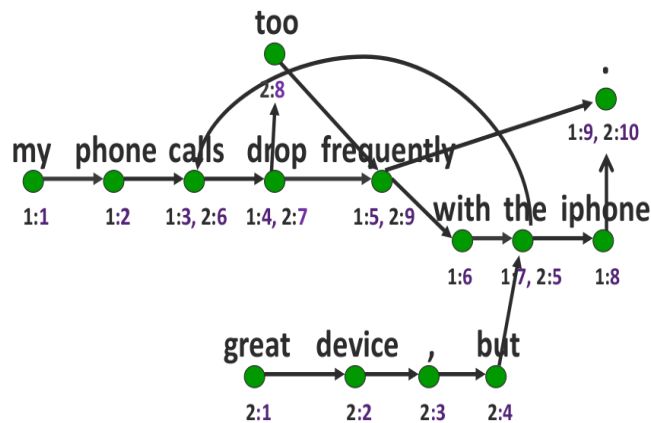


Figura 18: Grafo do Opinions (Ganesan et al., 2010)

O grafo tem três propriedades que são importantes para a geração de sumários abstrativos, as quais são apresentadas a seguir:

- Captura de redundância: Sequências altamente redundantes são capturadas pelos sub-grafos;
- Captura de subsequências com lacunas: Sequências com poucas variações são capturadas;
- Captura de estruturas *hubs*: Alguns nós podem ser considerados como *hubs*, os quais conectam outros muitos nós.

Para gerar sumários abstrativos, o primeiro passo é fazer um ranqueamento dos sub-grafos em função dos seus pesos. Como segundo passo, eliminam-se os sub-grafos duplicados ou muito similares (utilizando o índice de Jaccard ([Jaccard, 1901](#))). Os sub-grafos restantes são utilizados no sumário final, considerando o tamanho predefinido para o sumário. Este processo de sumarização é mais extrativo que abstrativo, porém, se comparado com abordagens extrativas, o *Opinosis* tem processos de fusão e compressão baseados nas três propriedades descritas anteriormente.

Para os experimentos, os autores propõem uma medida de *legibilidade*, a qual consiste em, dadas N sentenças do sumário gerado pelo sistema e M sentenças do sumário humano (feito manualmente), misturam-se todas essas sentenças e depois é solicitado que um assessor humano escolha as N sentenças que são menos legíveis do resumo. Se o assessor humano escolher muitas sentenças geradas pelo sistema, a legibilidade é baixa. Os testes de legibilidade mostram que mais de 60% das sentenças geradas pelo sistema não são diferentes das sentenças geradas pelos humanos.

Em comparação com o trabalho de [Carenini et al. \(2006\)](#), [Ganesan et al. \(2010\)](#) apresentam uma abordagem abstrativa mais simples, mas obtém bons resultados. Isso indica que é possível utilizar métodos não complexos para gerar sumários abstrativos.

3.2.3 Trabalho de Labbé e Portet (2012)

[Labbé e Portet \(2012\)](#) apresentam um estudo comparativo de três abordagens para sumarizar opiniões. A partir deste estudo, os autores propõem uma arquitetura para gerar sumários abstrativos personalizados segundo a preferência do usuário. Para os experimentos, foram utilizadas opiniões sobre hotéis do site [Tripadvisor](#)⁶.

A primeira abordagem estudada utiliza o sumarizador OTS⁷ (do inglês, *Open Text Summarizer*). O OTS é um sumarizador independente do domínio que utiliza o TF-IDF para selecionar o conteúdo do sumário. Nos experimentos com OTS, os autores indicam que informações não relevantes sobre os hotéis aparecem no sumário final. Devido a isso, eles concluem que a seleção de conteúdo baseada puramente em frequência, sem o envolvimento de algum conhecimento de domínio, não é promissora.

⁶ <http://www.tripadvisor.com>

⁷ Disponível em <http://libots.sourceforge.net>

Na segunda abordagem, a ideia principal consiste em extrair as informações mais relevantes que estejam relacionadas a um determinado tópico. Para isso, os autores utilizam o método proposto por [Labbé e Labbé \(2005\)](#) com a finalidade de obter as palavras que mais caracterizam o tópico. Essas palavras são utilizadas para ranquear as sentenças das opiniões: quanto mais palavras relacionadas ao tópico, maior a pontuação da sentença. Por fim, as sentenças com melhores pontuações são listadas como sumário. Segundo os autores, o resumo gerado mostra as sentenças mais relevantes para cada tópico. No entanto, existe muita redundância no sumário, e, além disso, opiniões constrativas não são obtidas com esta abordagem.

Na terceira abordagem, os autores utilizam o sistema RnR ([Rahayu et al., 2010](#)) para gerar os sumários. RnR captura e resume as opiniões expressas em um corpú de comentários destacando os tópicos negativos e positivos. Em termos de apresentação, o sistema produz um resumo para os dois tópicos mais frequentes. Além disso, RnR mostra as palavras sentimentais mais frequentes para cada tópico.

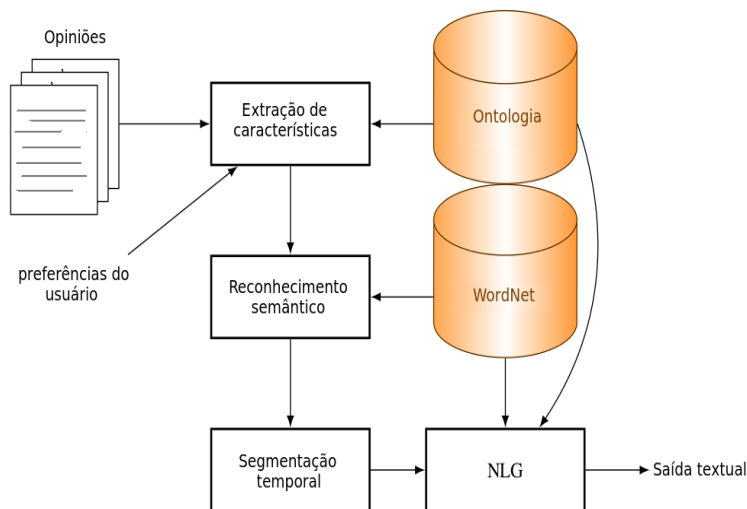


Figura 19: Arquitetura do sistema proposto pelo [Labbé e Portet \(2012\)](#)

Tendo como base o estudo feito, os autores propõem uma arquitetura (ver Figura 19) para gerar sumários abstrativos. Na Figura 19, como primeiro passo, o usuário solicita um sumário sobre um tópico do hotel ou o sistema extrai automaticamente os tópicos mais importantes utilizando uma ontologia sobre hotéis; em seguida, o sistema recupera todas as opiniões relacionadas a esse tópico. Uma vez extraídos os tópicos, o sistema

determina a polaridade das opiniões. Usando as avaliações, a segmentação temporal de [Charbonnier e Portet \(2012\)](#) é utilizada para identificar os principais períodos dos relatos sobre o hotel (por exemplo, “*no verão o hotel encontra-se mal ventilado*”). Por fim, um módulo de geração de língua natural sumariza as opiniões de cada tópico.

Neste trabalho, os autores não apresentam avaliações dos sumários nem muito detalhes da arquitetura, mas é importante ter em consideração a proposta, pois é uma das poucas abordagens abstrativas que podem ser utilizadas como alternativa à arquitetura de [Hu e Liu \(2004\)](#) (ver Seção 3.1.2).

3.2.4 Trabalho de Gerani et al. (2014)

O método proposto por [Gerani et al. \(2014\)](#) é um dos primeiros trabalhos que utilizam uma análise discursiva para gerar sumários abstrativos de opiniões. Para tal fim, este método utiliza a teoria discursiva RST (*Rhetorical Structure Theory*) ([Mann e Thompson, 1988](#)) para ranquear os aspectos presentes nas opiniões. Logo, o sumário final é gerado preenchendo alguns *templates* com os principais aspectos, considerando-se a polaridade deles.

A RST é uma teoria discursiva que divide o texto em unidades mínimas chamadas EDU (*Elementary Discourse Units*). Na RST, cada EDU é relacionada com as outras em uma estrutura de árvore por meio de relações discursivas (por exemplo, elaboração, contradição, etc.).

Como primeiro passo, este método utiliza as relações RST para agrupar e ranquear os aspectos que se encontram nas diferentes EDUs. Para fazer o ranking dos aspectos, é utilizado o valor PageRank ([Page et al., 1998](#)) de cada aspecto, assim como a força da polaridade deles utilizando a proposta de [Carenini et al. \(2013\)](#).

Em seguida, os principais aspectos são preenchidos em templates criados manualmente. A seleção dos templates é em função da polaridade e da frequência do aspecto. Por exemplo, se um aspecto foi mencionado muitas vezes nas opiniões com um sentimento negativo, um possível template nesse caso seria: “*Quase todos os usuários comentaram sobre o <aspecto> e eles não gostaram*”. Nesse trabalho, o número de aspectos a serem considerados no sumário é configurável. Assim, o sumário final é gerado selecionando um template para cada um dos aspectos principais.

Nos experimentos, os autores avaliaram 4 métodos de sumarização: (1) MEAD-LexRank (LR), um sumarizador que utiliza o algoritmo LexRank (Erkan e Radev, 2004) sob o framework de sumarização MEAD (Radev et al., 2003); (2) MEAD*, uma adaptação do MEAD que utiliza a informação dos aspectos; (3) uma adaptação do método proposto nesse trabalho que só considera a informação da força da polaridade (FP) e (4) o método completo.

Para avaliar esses quatro métodos, os autores realizaram dois estudos de caso baseados na preferência dos usuários: no primeiro estudo (“Estudo 1”), os usuários tinham que mostrar preferência por algum sumário (ou nenhum deles) com base no conteúdo do seus textos; para o segundo estudo (“Estudo 2”), os usuários, além de ler o conteúdo dos sumários, também leram as opiniões fonte e, com base nessas informações, fizeram seu julgamento.

Na Tabela 8, mostram-se os resultados obtidos pelos 4 métodos mencionados anteriormente. Como se pode observar, nos dois estudos, os usuários mostraram uma maior preferência pelo método proposto nesse trabalho, pois, nos sumários produzidos, considera-se tanto a importância dos aspectos quanto a polaridade deles.

Sistema I vs Sistema II	Sem preferência		Preferência Sistema I		Preferência Sistema II	
	Estudo 1	Estudo 2	Estudo 1	Estudo 2	Estudo 1	Estudo 2
LR vs MEAD*	7%	6%	35%	20%	58%	74%
LR vs FP	0%	0%	38%	21%	62%	79%
LR vs Proposta	0%	3%	26%	13%	74%	84%
MEAD* vs FP	0%	0%	38%	20%	62%	80%
MEAD* vs Proposta	0%	3%	25%	30%	75%	67%
SA vs Proposta	12%	11%	23%	32%	65%	57%

Tabela 8: Preferência dos usuários

A principal vantagem do método proposto por Gerani et al. (2014) é que permite encontrar e agrupar os principais aspectos das opiniões por meio das relações RST. Porém, a maior desvantagem desse método é a rigidez dos templates utilizados, pois não permite conhecer exatamente o que se diz nas opiniões fonte.

3.2.5 Trabalho de Fabrizio et al. (2014)

Fabrizio et al. (2014) desenvolveram Starlet-H, um método híbrido de sumarização abstrativa de opiniões, o qual combina as vantagens das abordagens extrativas e abstrativas. Similarmente ao trabalho de Gerani et al. (2014), este método também utiliza as informações das relações RST para gerar os sumários.

Por um lado, o módulo extrativo de Starlet-H tem como objetivo selecionar as frases (n-gramas) textuais mais representativas de cada aspecto para quatro tipos de polaridade: positivo, muito positivo, negativo e muito negativo. Para tal fim, foram treinados dois algoritmos de aprendizado de máquina (SVM e MaxEnt) em um corpus com anotações de etiqueta POS e polaridade das palavras.

Por outro lado, no módulo abstrativo, utilizaram-se seis relações RST (concessão, contraste, exemplo, resumo, lista, e justificativa) para construir quatro templates de sumários (um template para cada tipo de polaridade mencionada anteriormente). Cada um desses templates definiam as estruturas dos sumários e foram preenchidos com a informação extraída no módulo extrativo.

Para medir o desempenho de Starlet-H, os autores utilizaram quatro critérios para avaliar a qualidade dos sumários: (1) legibilidade, se o sumário é fácil de ler e entender; (2) integralidade, se o resumo captura todas as avaliações nas opiniões; (3) compactação, se o sumário é compacto e não tem informação repetida; e (4) exatidão, se o resumo é correto e expressa as avaliações nos comentários. Cada critério recebeu uma pontuação de 1 até 5 por um avaliador humano, sendo 5 a melhor nota.

Nos experimentos, além de testar o sumarizador Starlet-H, Fabrizio et al. (2014) fizeram testes com outros três métodos: MEAD, Starlet-A e Starlet-E. Starlet-A é uma variação de Starlet-H que não considera o módulo selecionador de frases textuais. Por sua parte, Starlet-E, é um método extrativo proposto por Di Fabrizio et al. (2011) que utiliza uma abordagem de aprendizado de máquina para a seleção de sentenças do sumário. Os resultados obtidos pelos 4 sumarizadores são mostrados na Tabela 9.

Como se pode observar na Tabela 9, o sumarizador Starlet-H obteve as melhores pontuações. Esses resultados demonstram que uma abordagem híbrida, entre métodos extrativos e abstrativos, pode ajudar a gerar melhores sumários do que as abordagens

	MEAD	Starlet-E	Starlet-A	Starlet-H
Legibilidade	2.95	3.17	3.64	3.74
Integralidade	2.88	3.29	3.29	3.58
Compactação	3.07	3.35	3.80	3.58
Exatidão	3.26	3.48	3.59	3.72

Tabela 9: Resultados da avaliação qualitativa

independentemente.

3.3 Síntese dos Trabalhos Discutidos

Nesse capítulo, foram apresentados trabalhos relacionados ao tema deste mestrado. Como foi mostrado, a maioria dos trabalhos sobre sumarização de opiniões utiliza uma abordagem extrativa, devido, principalmente, à complexidade dos métodos abstrativos. Logo, esta pesquisa pode contribuir, não só explorando métodos extrativos, mas também com um avanço no estado da arte para as abordagens abstrativas.

Na Tabela 10, é apresentado um resumo dos trabalhos de sumarização de opiniões citados neste capítulo. A ordenação dos trabalhos é dada pelo ano de publicação. A primeira coluna mostra o nome dos autores e o ano de publicação; a segunda indica a abordagem utilizada na sumarização; a terceira mostra a língua dos recursos utilizados nos experimentos; a quarta coluna descreve os métodos usados na identificação de aspectos; a quinta indica os métodos utilizados na predição do sentimento; a sexta coluna mostra o formato de saída dos resumos automáticos; finalmente, a última coluna descreve o método de avaliação dos sumários.

Dos métodos explicados neste capítulo, quatro métodos foram implementados e adaptados para a língua portuguesa brasileira. A seleção desses métodos foi baseada na qualidade dos seus resultados e na diversidade deles, pois, dado que os métodos utilizaram diferentes medidas de avaliação, não foi possível fazer uma comparação justa e selecionar os melhores. Dos métodos extrativos estudados foram selecionadas as propostas de [Hu e Liu \(2004\)](#) e [Tadano et al. \(2010\)](#), dos métodos abstrativos [Ganesan et al. \(2010\)](#) e [Gerani et al. \(2014\)](#) foram selecionados.

Como se explicará no Capítulo 5, de alguns dos métodos implementados, foram uti-

lizadas suas principais vantagens em duas novas propostas para gerar sumários automáticos de opiniões com base em aspectos.

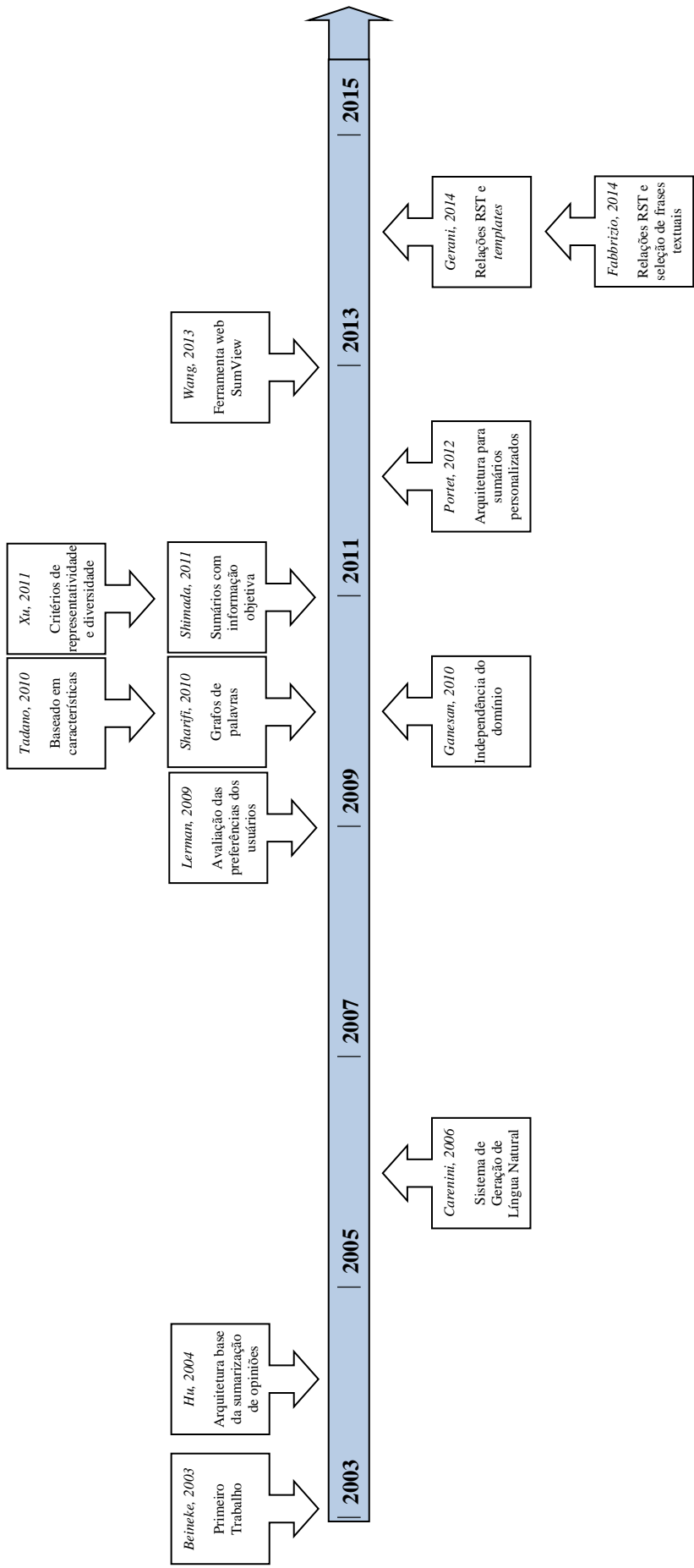
3.4 Linha do Tempo

Na Figura 20, é apresentada uma linha do tempo com alguns trabalhos citados nas seções anteriores, indicando a principal característica destes trabalhos. Na parte superior da imagem, encontram-se os trabalhos que utilizaram abordagens extrativas; na parte inferior da figura, mostram-se as pesquisas que usaram métodos abstrativos.

Trabalho	Abordagem	Língua	Identificação de Aspectos	Predição do Sentimento	Formato do Sumário	Métodos de Avaliação
Beineke et al. (2003)	Extrativa	Inglês	—	—	Textual	– Avaliam a identificação da sentença principal
Hu e Liu (2004)	Extrativa	Inglês	Etiquetador POS (substantivos)	Adjetivos manualmente polarizados	Estruturado	– Avaliam a identificação de aspectos e a identificação da polaridade
Caremini et al. (2006)	Abstrativa e Extrativa	Inglês	Ontologia manual	Hu e Liu (2004)	Textual	– Perguntas DUC-2005 – Perguntas sobre o conteúdo do sumário
Sharif et al. (2010)	Extrativa	Inglês	Manual	—	Estruturado	– ROUGE-1
Tadano et al. (2010)	Extrativa	Japonês	Manual	Número de estrelas da opinião	Textual	– ROUGE-1 – ROUGE-2
Ganesan et al. (2010)	Abstrativa	Inglês	Manual	—	Textual	– ROUGE-N – Teste de legibilidade
Shimada et al. (2011)	Extrativa	Japonês	Manual	Número de estrelas da opinião	Textual	– Acordo Semântico
Xu et al. (2011)	Extrativa	Inglês	MG-LDA	Léxico	Textual	– ROUGE-1 – ROUGE-L
Labbé e Portet (2012)	Abstrativa	—	—	—	—	—
Wang et al. (2013)	Extrativa	Inglês	Hu e Liu (2004) ou Manual	—	Estruturado	– ROUGE-N – Satisfação do usuário
Gerani et al. (2014)	Abstrativa	Inglês	Manual	Manual	Textual	– Preferência do usuário
Fabbrizio et al. (2014)	Abstrativa e Extrativa	Inglês	Manual	Manual	Textual	– Satisfação do usuário

Tabela 10: Trabalhos estudados

Abordagens Extrativas



Abordagens Abstrativas

Figura 20: Linha do Tempo

Criação e Análise do Córpus

Neste capítulo, será apresentado o processo de criação do córpus OpiSums-PT, assim como algumas análises sobre ele e os resultados dos mesmos. Para tanto, o capítulo está estruturado em seis seções. Na Seção 4.1, explica-se a importância que tem um córpus de sumários de opiniões e os objetivos que se perseguem com a criação de OpiSums-PT. Na Seção 4.2, descrevem-se alguns córpus existentes de sumários de opiniões. Os recursos linguísticos utilizados para a criação de OpiSums-PT são descritos na Seção 4.3. Na Seção 4.4, explica-se como foi criado o córpus de sumários de opiniões. Na Seção 4.5, são apresentados os experimentos e resultados da concordância entre os anotadores, a cobertura dos aspectos e a orientação do sentimento. Finalmente, na Seção 4.6, apresentam-se algumas considerações finais deste capítulo.

4.1 A Importância de um Córpus de Sumários de Opiniões

Na tarefa de sumarização de opiniões, para avaliar o desempenho dos métodos automáticos, geralmente é necessário ter um córpus de referência com resumos humanos. Com o dito córpus, poderia se comparar quão semelhantes são os resumos automáticos e os sumários humanos. Por meio dessa comparação, poderia ser possível identificar os

erros dos métodos automáticos e, conseqüentemente, melhorar o desempenho deles. Além disso, um cópüs de resumos de opiniões poderia ser usado em métodos de aprendizado de máquina como dados de treinamento para aprender padrões e, assim, extrair informações importantes das opiniões.

Infelizmente, são poucos os cópüs disponíveis na tarefa de sumarização de opiniões baseada em aspectos (Ganesan et al. (2010) Zhu et al. (2013) Kim e Zhai (2009)), o que dificulta o desenvolvimento das pesquisas nessa tarefa. A maioria dos cópüs existentes têm-se centrado no idioma Inglês. Para o Português Brasileiro, até onde sabemos, não há nenhum cópüs de resumos de opiniões disponível.

Devido a isso, neste trabalho de mestrado, criou-se um cópüs de sumários de opiniões com base em aspectos chamado OpiSums-PT (*Opinion Summaries in Portuguese*). OpiSums-PT contém múltiplos resumos humanos. A construção deste cópüs foi motivado por duas razões principais: (1) para lidar com a ausência de um cópüs de resumos de opiniões em Português do Brasil e (2) para avaliar como as pessoas geram resumos de opiniões. Particularmente, foi analisado quão semelhantes são os resumos humanos e quão importante são as informações da cobertura de aspectos e a orientação do sentimento no processo de criação dos sumários.

4.2 Cópüs Existentes de Sumários de Opiniões

Muitos trabalhos de sumarização de opiniões baseados em aspectos criaram seu próprio conjunto de dados coletando opiniões de páginas web de avaliações ou redes sociais. Destes recursos, poucos podem ser consideradas como conjuntos padrões. O conjunto de dados proposto em Hu e Liu (2004) é o recurso mais utilizado na tarefa de sumarização de opiniões baseada em aspectos. No entanto, esse cópüs não contém sumários manuais, porém, tem aspectos anotados e seu sentimento associado. Para avaliar os resumos automáticos nesse cópüs, os trabalhos que utilizaram esse recurso usaram questionários para selecionar os melhores resumos.

Em trabalhos anteriores, onde os resumos de opiniões foram criados manualmente, a anotação do cópüs não foi descrita com muito detalhe, pois não foi o foco principal desses estudos.

Em [Tadano et al. \(2010\)](#), três participantes anotaram 25 comentários (aproximadamente com 450 sentenças) de opiniões sobre um videogame. Dos 25 comentários, 50 sentenças foram selecionadas para o resumo. Nos experimentos, a medida ROUGE-1 entre os resumos dos anotadores foi 0.480, o que mostra que é difícil gerar o mesmo resumo de opiniões, mesmo entre humanos.

[Xu et al. \(2011\)](#) coletaram 32007 opiniões sobre três aspectos (comida, serviço e ambiente) de 173 restaurantes. Desse total de comentários, dez restaurantes foram selecionados para os testes e sete restaurantes para configurar alguns parâmetros do método automático proposto por Xu et al. Para cada aspecto de um restaurante, os autores criaram um resumo extrativo selecionando várias sentenças representativas da opinião. Cada resumo foi composto por 100 palavras em média.

Em [Carenini et al. \(2006\)](#), 28 anotadores criaram resumos abstrativos para um corpus de comentários sobre uma câmera digital e um leitor de DVD. Cada participante da anotação recebeu 20 comentários selecionados aleatoriamente e gerou um resumo de 100 palavras. Como instruções, os participantes assumiram que trabalhavam para um fabricante de produtos (câmera digital ou leitor de DVD). O objetivo dessas instruções era motivar os participantes a procurarem a informação mais importante para o sumário.

[Ganesan et al. \(2010\)](#) criaram um corpus de sumários abstrativos manuais usando opiniões sobre hotéis, carros e outros produtos eletrônicos. Para coletar os comentários, os autores utilizaram 51 tópicos (por exemplo, *Ipod:som* e *Toyota:conforto*). Cada tópico teve 100 sentenças redundantes relacionadas ao tópico. Ganesan et al. utilizaram um serviço de *crowdsourcing*, no qual cinco anotadores criaram cinco resumos diferentes para cada tópico. Após a criação dos sumários, os autores revisaram cada conjunto de resumos e deixaram de lado os resumos que tinham pouca ou nenhuma correlação com a maioria deles. Finalmente, cada tópico ficou com quatro sumários de referência, aproximadamente.

Diferentemente desses estudos, neste trabalho de mestrado foi realizada uma análise qualitativa dos sumários de opiniões. Além disso, foi realizada uma comparação entre os sumários extrativos e abstrativos em termos de concordância dos anotadores, cobertura de aspectos e orientação do sentimento. Até onde se sabe, não existem trabalhos semelhantes, provavelmente devido à dificuldade de gerar sumários humanos de opiniões.

4.3 Recursos

Para criar o OpiSums-PT, foram utilizados comentários de dois domínios: livros e produtos eletrônicos. Para o primeiro, foram utilizadas as opiniões do córpus ReLi (Freitas et al., 2013), uma coleção de opiniões sobre 13 livros. Para o segundo domínio, foram coletados comentários de quatro produtos eletrônicos do site Buscapé¹. O objetivo de utilizar estes dois domínios é ter um córpus com características diferentes nas opiniões. Nas seções seguintes, estes dois recursos são explicados com mais detalhes.

4.3.1 Opiniões de Livros

As opiniões de livros foram extraídas do córpus ReLi (Freitas et al., 2013). O córpus ReLi é composto por 1600 resenhas de 13 livros (de sete autores famosos da literatura clássica e contemporânea), totalizando cerca de 260000 palavras e 12000 sentenças. As resenhas foram escritas por usuários da web, de forma livre, em sites especializados de resenhas. A Tabela 11 apresenta a composição do ReLi, com a distribuição dos livros e autores.

As opiniões anotadas no ReLi referem-se exclusivamente ao livro resenhado ou aos aspectos dele (por exemplo, personagens, capítulos, linguagem, etc.). Assim, opiniões sobre outros livros ou filmes dos livros não foram consideradas. O córpus não considera também as opiniões que envolvem comparações, mesmo que um dos objetos comparados fosse o livro em questão.

Segundo Freitas et al. (2013), no ReLi, as resenhas foram anotadas no nível da sentença e do sintagma, e aconteceu em três etapas: (1) identificação e anotação da polaridade das sentenças que contêm opinião, (2) identificação do alvo da opinião, (3) identificação e anotação da polaridade do trecho (palavras ou expressões) que contêm opinião.

A anotação foi realizada, inicialmente, por três pesquisadores com formação em Linguística e devidamente treinados, produzindo resultados de concordância satisfatórios. No final da anotação, só um anotador terminou a tarefa. Na Tabela 12, são apresentados os resultados do estudo da concordância em uma amostra de 170 opiniões, entre os

¹<http://www.buscape.com.br/>

Autor	Título	Resenhas	Sentenças	Palavras
Stephenie Meyer	Crepúsculo	409	3266	62268
Thalita Rebouças	Fala sério, amiga!	161	910	16864
	Fala sério, amor!			
	Fala sério, mãe!			
	Fala sério, pai!			
	Fala sério, professor!			
Sidney Sheldon	O Outro lado da meia noite	230	1569	31712
	O Reverso da Medalha			
	Se houver Amanhã			
Jorge Amado	Capitães da Areia	187	1348	32117
George Orwell	1984	202	2228	5132
José Saramago	Ensaio sobre a Cegueira	271	1991	42152
J.D. Salinger	O Apanhador nos Campos de Centeio	140	1158	23725
Total		1600	12470	259978

Tabela 11: Composição do córpus ReLi

anotadores (A, B e C) em pares. Como se pode observar, os resultados da atribuição de polaridade são bem mais altos que os da identificação.

Anotadores	Identificação			Atribuição de Polaridade	
	Sentença	Objeto	Opinião	Sentença	Sintagma
A B	85.0%	71.1%	78.5%	98.1%	99.9%
A C	81.9%	75.3%	81.5%	98.4%	99.6%
B C	78.5%	71.3%	79.8%	98.4%	100%

Tabela 12: Resultados da concordância entre os anotadores

Para a anotação de nosso córpus, foram selecionados 10 comentários aleatoriamente para cada livro do ReLi, tentando ter opiniões positivas e negativas, assim como foi feito em outros trabalhos relacionados (Carenini et al. (2006), Tadano et al. (2010)) que usaram um número similar de opiniões como fonte de dados. Os comentários selecionados tinham que conter como máximo 300 palavras. Usamos este filtro porque, acredita-se, que as pessoas preferem ler opiniões mais concisas em vez de opiniões muito longas. Este critério também foi utilizado na seleção de opiniões de produtos eletrônicos.

4.3.2 Opiniões de Produtos Eletrônicos

As opiniões sobre produtos eletrônicos foram coletadas do Buscapé, um site onde os usuários comentam sobre diferentes produtos (por exemplo, smartphones, roupas, videogames, etc.). Estes comentários são escritos livremente em um template com três seções: pros, contras e opinião.

Para criar o córpus de sumários, foram coletadas um conjunto de opiniões sobre quatro produtos eletrônicos: dois smartphones (Samsung Galaxy S III e iPhone 5) e dois televisores (LG Smart TV e Samsung Smart TV). Para cada produto, foram selecionados aleatoriamente 10 comentários tentando ter opiniões positivas e negativas.

Este conjunto de comentários foi anotado por uma pessoa com forte conhecimento em Análise de Sentimento. A anotação consistiu na identificação dos aspectos dos produtos, por exemplo, a bateria e a foto para smartphones, e o som e o preço para os televisores. A identificação da polaridade dos segmentos que contêm o sentimento sobre os aspectos também foi anotada.

4.4 Anotação do Córpus

Segundo [Ulrich et al. \(2008\)](#), a sumarização abstrativa é o principal objetivo de muitos trabalhos de investigação, pois é como as pessoas fazem os sumários naturalmente, mas, a sumarização extrativa tem sido mais explorada e eficaz, dado que é mais fácil de desenvolver. Nesta anotação, foram gerados tanto sumários extrativos quanto abstrativos, para analisar como eles são criados e usados.

Foram criados múltiplos resumos de referência com o fim de reduzir a subjetividade em geral e qualquer possível viés e, também, para ter uma maior variedade. Para cada livro e produto eletrônico, foram gerados cinco sumários extrativos e cinco abstrativos. No total, 170 resumos foram criados manualmente. Na Tabela 13, mostra-se o conteúdo de OpiSums-PT em relação ao número de sentenças, *tokens*, *types* (palavras únicas) e a média deles por sumário.

A anotação foi realizada por 14 participantes com forte conhecimento em Linguística Computacional e Processamento de Língua Natural. Cada participante criou 12 sumários,

Característica	Sumários	Sumários
	Extrativos	Abstrativos
Sumários	85	85
Sentenças	534	430
<i>Tokens</i>	8435	8611
<i>Types</i>	1702	1833
Média de sentenças por sumário	6.3	5.1
Média de <i>tokens</i> por sumário	99.2	101.3
Média de <i>types</i> por sumário	71.1	72.4

Tabela 13: Conteúdo de OpiSums-PT

aproximadamente, durante todo o processo da anotação. Cada conjunto de cinco resumos (extrativos ou abstrativos) foi gerado por cinco anotadores diferentes.

Para gerar um resumo, extrativo ou abstrativo, cada anotador teve que ler 10 opiniões sobre os livros ou produtos eletrônicos. Este número de opiniões foi escolhido porque acredita-se que, quando as pessoas buscam opiniões, eles não leem grandes quantidades de opiniões, mas uma pequena amostra delas. Além disso, lidar com mais do que essa quantidade poderia ser inviável, tornando a tarefa muito complexa.

A tarefa de anotação foi realizada diariamente durante 13 dias, aproximadamente. Na primeira reunião, os anotadores receberam uma sessão de treinamento junto com o manual de anotação, para que se familiarizarem com a tarefa. Nesse documento, foram apresentadas todas as instruções, bem como os aspectos identificados nas opiniões de ReLi e Buscapé. Estes aspectos foram tomados das anotações dessas duas fontes de dados e foram mostrados aos participantes com a única intenção de que os anotadores se familiarizem a eles. A Tabela 14 mostra as entidades e os aspectos apresentados aos participantes na anotação de OpiSums-PT.

Nos outros dias da anotação, os anotadores criaram os resumos à distância e os enviaram por e-mail, como foi realizado em (Dias et al., 2014). A cada dia, um anotador gerou apenas um resumo (extrativo ou abstrativo). Este esquema foi adotado a fim de simplificar a tarefa dos anotadores e, conseqüentemente, para obter bons resumos, evitando-se cansaço e desatenção por excesso de trabalho.

Outra instrução brindada na anotação foi sobre o comprimento do resumo. Ambos os

Entidades	Aspectos
Livro	personagens, história, capítulos, diálogos, frases, o estilo de autor, vocabulário, texto
Smartphone	bateria, <i>design</i> , processador, tela, preço, câmera, peso, sistema operacional, internet, foto, vídeo, wi-fi, som, tamanho, fones de ouvido, velocidade, chip
TV	<i>design</i> , preço, câmera, qualidade da imagem, luminosidade, wi-fi, som, durabilidade, internet

Tabela 14: Entidades e aspectos considerados na anotação

resumos, extrativos e abstrativos, deveriam ser compostos por 100 palavras com uma tolerância de ± 10 palavras, aproximadamente. Foi escolhido o mesmo número de palavras para estes dois tipos de sumários para avaliar como eles são gerados sob restrições semelhantes. A taxa de compressão em porcentagem (por exemplo, 25 %) não foi usada porque a grande maioria dos trabalhos de sumarização de opiniões baseados em aspectos não utilizam este esquema (Carenini et al. (2006), Ganesan et al. (2010), Tadano et al. (2010)).

4.4.1 Sumários Extrativos

Para criar os resumos extrativos na anotação, foi solicitado aos anotadores que selecionassem as sentenças mais importantes das opiniões fonte. Não foi estabelecido um critério para determinar a importância de uma sentença, sendo decisão de cada anotador. Da mesma forma, não havia obrigação de se excluir sentenças com problemas de pronomes pendentes ou qualquer outro. Optou-se por esta autonomia, com o objetivo de que a criação dos resumos fosse o mais natural possível. O número de aspectos incluídos no resumo final foi escolhido por cada anotador.

O sumário final foi composto por sentenças completas. Os anotadores não tinham permissão para reescrever as sentenças das opiniões originais. Se uma sentença apresentasse erros de ortografia e/ou erros gramaticais, não deveria ser corrigida.

Cada sentença das opiniões tinha um identificador na parte final. Este identificador permitiu vincular a sentença do sumário com a opinião de origem. Assim, por exemplo, o identificador “<D20_S3>” referencia a terceira sentença da opinião (documento) 20. A Figura 21 mostra um exemplo de um sumário extrativo (em negrito, os identificadores

das sentenças).

```
Um Smartphone quase Perfeito! <D3_s1>  
O que gostei: Hoje é o melhor no mercado em relação  
ao seu processamento. <D2_s3>  
A bateria dura bastante e os aplicativos ja  
instalados sao otimos. <D7_s5>  
A camera é maravilhosa. <D7_s4>  
O que não gostei: Ele esquenta um Pouco na parte de  
baixo mas não chega a incomodar, na cor branca ele  
parece ser muito frágil e o S Voice ainda não  
funciona em português. <D3_s5>  
Esperava muito mais do Galaxy SIII pelo suspense que  
a Samsung promoveu. <D2_s1>  
Depois dessa, quem tem coragem de investir em média  
R$ 1.700,00 no Galaxy SIII ou tentar a sorte com o  
Galaxy S4? <D6_s9>
```

Figura 21: Exemplo de sumário extrativo

Como se pode observar na Figura 21, o resumo extrativo é composto por sete sentenças de diferentes opiniões (D2, D3, D6 e D7). Isso aconteceu muitas vezes nos resumos extrativos de OpiSums-PT, indicando que as sentenças relevantes para os anotadores foram escritas por diferentes usuários da web. Como consequência disso, a falta de coesão entre as sentenças dos sumários foi notória.

4.4.2 Sumários Abstrativos

A criação dos sumários abstrativos é mais difícil do que os sumários extrativos, pois implica gerar um novo texto. Nesta anotação, foi solicitado aos anotadores que gerassem resumos com o máximo de reescrita, a fim de obter resumos mais diferenciados em relação aos sumários extrativos.

Os resumos abstrativos deveriam indicar o cenário real das opiniões fonte (o sentimento geral predominante). Similarmente aos resumos extrativos, o número de aspectos a serem incluídos nos resumos abstrativos e a estrutura do texto foram decisões de cada anotador.

Na Figura 22, mostra-se um exemplo de sumário abstrativo sobre o livro Crepúsculo. Na primeira parte do texto, o autor do sumário dá o sentimento geral para este livro, e, em seguida, descreve o sentimento dos usuários para alguns aspectos do livro. Esta estrutura foi adotada pela maioria dos anotadores.

Em comparação com os resumos extrativos, estes sumários não apresentaram o problema da falta de coesão e mostraram explicitamente qual era o sentimento predominante nas opiniões fontes.

A grande maioria dos leitores avaliaram negativamente o livro Crepúsculo, pois em geral, eles argumentaram que o livro tem um romance exagerado. Entre as principais desvantagens do livro, os leitores mencionaram que os personagens são superficiais, a escrita é péssima e a história é chata. Muitos dos usuários não conseguiram terminar de ler o livro e não recomendariam ele para outras pessoas. Por outro lado, outra pequena parte dos leitores acharam que o livro Crepúsculo é bom, pois consideraram que ele é intenso, romântico, cheio de mistérios e brilhante. Estes leitores afirmaram que, embora Crepúsculo seja um livro fictício, ele mostra a importância de um verdadeiro amor.

Figura 22: Exemplo de sumário abstrativo

4.5 Experimentos

Depois da anotação, foram realizados alguns experimentos com o OpiSums-PT. Primeiro, foi calculada a concordância entre os anotadores para saber quão difícil é esta tarefa. Em segundo lugar, foi analisada a cobertura dos aspectos para estimar a proporção de aspectos que se conserva nos sumários finais. Finalmente, a orientação do sentimento nos resumos foi calculada para verificar se é proporcional ao sentimento geral das opiniões fonte.

Estes experimentos focam-se nessas três questões. Na tarefa de sumarização de opiniões baseada em aspectos, acredita-se que (1) as pessoas não geram sumários de opiniões muito semelhantes (como ocorre na sumarização tradicional), (2) que nem todos os aspectos são considerados no resumo final e (3) que as pessoas consideram a orientação do sentimento das opiniões para criar os sumários. No entanto, tanto quanto se sabe, não existem trabalhos anteriores que provaram essas questões. Neste estudo, exploram-se essas três questões.

4.5.1 Concordância dos Anotadores

Como parte dos experimentos, foi calculada a concordância entre os anotadores nos sumários extrativos e abstrativos. Para ambos, utilizou-se a medida ROUGE (Lin, 2004). Para os resumos extrativos, foi calculado também o coeficiente Kappa (Carletta, 1996) e a percentagem de sentenças comuns nos resumos. O coeficiente Kappa é uma medida clássica de concordância utilizada na área de PLN, a qual indica a concordância entre

anotadores ao mesmo tempo em que desconta a concordância ao acaso, por sorte.

Nos sumários extrativos, foi calculada a concordância Kappa para cada livro e produto eletrônico, tomando as sentenças das opiniões fonte e verificando quais delas foram incluídas nos resumos humanos. Em média, o valor Kappa obtido nos experimentos foi 0.185. De acordo com [Liu e Liu \(2008\)](#), os valores de Kappa para a sumarização de textos e a sumarização de reuniões foram em média 0.38 e 0.28, respectivamente. Em comparação com esses valores, a concordância Kappa obtida neste experimento para a sumarização de opiniões baseada em aspectos é bem menor. Isto é provavelmente devido ao fato de que na sumarização de opiniões existem muitas sentenças diferentes que expressam o mesmo significado. Assim, diferentes anotadores poderiam ter escolhido sentenças diferentes com conteúdo semelhante.

Para compensar esse problema da concordância Kappa, também foi utilizada a medida ROUGE-N. A medida ROUGE calcula a sobreposição de n-gramas entre os resumos e, portanto, poderia ajudar a identificar sentenças com conteúdo similar. Nestes experimentos, foi utilizada a medida ROUGE-1 (sobreposição de unigramas), pois é a mais utilizada na área.

Para cada anotador, foi computada a ROUGE-1 utilizando os resumos dos outros anotadores como sumários de referência, e, em seguida, calculou-se a média entre todas as medidas. A Tabela 15 mostra os valores da ROUGE-1 obtidos para cada livro e produto eletrônico nos sumários extrativos e abstrativos. Estes resultados são melhores do que os resultados da concordância Kappa e podem indicar que os anotadores selecionaram diferentes sentenças que tem conteúdo semelhante. Os resultados dos resumos extrativos são melhores do que os abstrativos, porque, nos sumários abstrativos os anotadores têm independência para utilizar diferentes palavras, sinônimos ou paráfrases.

Para os sumários extrativos, foi calculada também a porcentagem de sentenças comuns entre os resumos criados pelos anotadores, na Tabela 15, mostram-se os resultados (isso não foi possível para os sumários abstrativos, pois as sentenças não são as mesmas). A concordância total indica a proporção de sentenças comuns selecionadas por cinco anotadores; a concordância majoritária, por quatro ou três anotadores; e concordância minoritária, por dois anotadores. A concordância nula indica que os anotadores não concordaram na seleção de sentenças.

Por um lado, os resultados para estas métricas indicam que os anotadores escolhem as mesmas sentenças em poucos casos. Em média, apenas 1.1% (0.011) das sentenças foram selecionadas por todos os anotadores, e apenas 17.3% (0.173) delas, pela maioria dos anotadores. Acredita-se que isso acontece, principalmente, devido ao grande número de sentenças que os anotadores têm que ler para gerar o resumo (em média, 40 sentenças).

Livros/Produtos Eletrônicos	Sumários Extrativos					Sumários Abstrativos	
	Concordância Total	Concordância Majoritária	Concordância Minoritária	Concordância Nula	ROUGE-1	ROUGE-1	
	Capitães da Areia	0.000	0.267	0.200	0.533	0.405	0.218
Crepúsculo	0.000	0.286	0.357	0.357	0.414	0.239	
Ensaio sobre a Cegueira	0.000	0.043	0.217	0.739	0.250	0.251	
Fala sério. amiga!	0.077	0.154	0.154	0.615	0.606	0.299	
Fala sério. amor!	0.118	0.118	0.294	0.471	0.600	0.287	
Fala sério. mãe!	0.000	0.222	0.167	0.611	0.325	0.308	
Fala sério. pai!	0.000	0.143	0.143	0.714	0.418	0.352	
Fala sério. professor!	0.000	0.235	0.353	0.412	0.344	0.345	
O Apanhador nos Campos de Centeio	0.000	0.091	0.409	0.500	0.360	0.253	
O Outro lado da meia noite	0.000	0.136	0.182	0.682	0.392	0.232	
O Reverso da Medalha	0.000	0.100	0.250	0.650	0.339	0.305	
Se houver Amanhã	0.000	0.200	0.200	0.600	0.471	0.309	
1984	0.000	0.263	0.316	0.421	0.366	0.238	
Iphone 5	0.000	0.308	0.154	0.538	0.342	0.230	
Samsung Galaxy S III	0.000	0.100	0.200	0.700	0.235	0.276	
LG Smart TV	0.000	0.040	0.240	0.720	0.274	0.270	
Samsung Smart TV	0.000	0.238	0.333	0.429	0.451	0.270	
Média	0.011	0.173	0.245	0.570	0.388	0.275	

Tabela 15: Resultado da concordância entre anotadores

Por outro lado, em muitos casos, os anotadores escolheram diferentes sentenças (ver colunas de concordância minoritária e concordância nula) pois, como está relatado em (Rath et al., 1961), na tarefa de sumarização, não existe um único conjunto de sentenças representativas escolhidas pelas pessoas e há vários resumos possíveis, igualmente bons. Além disso, algumas características linguísticas especiais das opiniões, como a ironia ou o uso de gírias, fazem esta tarefa ainda mais difícil.

Em geral, todos os resultados apresentados na Tabela 15 mostram que é difícil gerar sumários de opiniões com base em aspectos semelhantes (extrativos ou abstrativos), mesmo entre os seres humanos, pois há muita subjetividade envolvida no processo de

sumarização. Estes resultados, apesar de serem baixos, podem ser utilizados como uma linha superior de desempenho para avaliar métodos automáticos de sumarização.

4.5.2 Cobertura dos Aspectos

Uma questão importante na sumarização de opiniões baseada em aspectos é a cobertura dos aspectos. A cobertura dos aspectos é um indicador de quantos aspectos das opiniões fonte são preservados no sumário final. A maioria dos trabalhos nesta tarefa estão focados na produção de um resumo para cada aspecto (Blair-Goldensohn et al. (2008), Tadano et al. (2010), Xu et al. (2011)). No entanto, se queremos um resumo geral, pode ser que essa abordagem não seja a ideal.

Neste trabalho de mestrado, foram produzidos sumários gerais com base em aspectos, ou seja, um resumo que contenha os aspectos mais importantes (segundo os anotadores) para um conjunto de opiniões fonte. Nos experimentos, para calcular a cobertura dos aspectos, foram também consideradas as entidades como aspectos, como fazem Gerani et al. (2014).

Para estimar a cobertura dos aspectos nos sumários extrativos, foram utilizados os aspectos anotados nas opiniões do ReLi e Buscapé. Em seguida, verificou-se quantos deles foram preservados nos resumos. Nos sumários abstrativos, foram utilizadas buscas semiautomáticas. Os aspectos foram buscados usando uma lista com seus nomes. Depois disso, foi feita uma revisão manual nos resumos, a fim de adicionar à lista possíveis sinônimos dos aspectos. Por exemplo, a palavra *narrativa* foi considerada como um sinônimo do aspecto *história*. Finalmente, determinou-se quantos aspectos estavam presentes nos resumos. Para cada livro e produto eletrônico, foi calculada a proporção de aspectos preservados nos cinco resumos, e, em seguida, foi calculada a média geral.

A Tabela 16 mostra a percentagem da cobertura dos aspectos para os sumários extrativos e abstrativos. Como se pode observar, em média os sumários abstrativos apresentam 72.6% dos aspectos e os sumários extrativos 58.3%, aproximadamente. Os resumos abstrativos possuem uma maior cobertura do que os sumários extrativos, pois os anotadores têm menos restrições para escrever um resumo abstrativo e, assim, eles podem incluir mais aspectos. Por outro lado, nos resumos extrativos, os anotadores estão limitados ao conteúdo das sentenças das opiniões fonte.

Nos experimentos, há poucos casos nos quais todos os aspectos foram incluídos nos resumos (livros “Fala Sério, amiga!” e “Fala Sério, professor!”). Nesses casos, as opiniões fonte apresentaram menos de três aspectos. Por outro lado, quando o número de aspectos nas opiniões fonte foi alto, poucos aspectos foram incluídos no resumo (por exemplo, produto Samsung Galaxy S III). Isso foi mais notório em produtos eletrônicos, pois eles têm opiniões mais técnicas que incluem muitos aspectos.

Livros/Produtos Eletrônicos	Sumários Extrativos	Sumários Abstrativos
Capitães da Areia	0.450	0.700
Crepúsculo	0.467	0.567
Ensaio sobre a Cegueira	0.300	0.600
Fala sério, amiga!	1.000	1.000
Fala sério, amor!	0.550	0.550
Fala sério, mãe!	0.400	0.767
Fala sério, pai!	0.800	0.900
Fala sério, professor!	0.700	1.000
O Apanhador nos Campos de Centeio	0.550	0.800
O Outro lado da meia noite	0.800	0.760
O Reverso da Medalha	0.650	0.800
Se houver Amanhã	0.640	0.680
1984	0.600	0.760
Iphone 5	0.444	0.578
Samsung Galaxy S III	0.333	0.400
LG Smart TV	0.514	0.714
Samsung Smart TV	0.720	0.760
Média	0.583	0.726

Tabela 16: Cobertura dos aspectos nos sumários

Os resultados da Tabela 16 indicam que, para um resumo geral baseado em aspectos, as pessoas consideram apenas alguns aspectos do texto, mas acima de 50% deles, em média. Não se encontrou outros trabalhos para comparar os resultados obtidos neste experimento sobre a cobertura dos aspectos, mas, estes resultados poderiam mostrar uma aproximação de quantos aspectos as pessoas consideram em um resumo de opiniões. Assim, os métodos automáticos de sumarização de opiniões poderiam usar estes resultados como um indicador de quantos aspectos devem ser incluídos nos resumos.

4.5.3 Orientação do Sentimento

No resumo, a comunicação aos leitores do sentimento nas opiniões sobre a entidade e seus aspectos não é uma simples questão de classificar o resumo como positivo ou negativo. Os leitores do sumário têm que saber se todas as opiniões que avaliaram a entidade foram feitas de forma semelhante ou variada. Assim, os resumos de opiniões devem preservar a distribuição da polaridade, tanto quanto possível, para refletir o sentimento geral sobre a entidade e seus aspectos.

Nos experimentos, foi avaliado o quanto os anotadores mantêm a orientação do sentimento nos resumos manuais. Para estimar o sentimento geral nas opiniões fonte, foram extraídos os segmentos que contêm sentimento das anotações do ReLi e do Buscapé. Foi calculada a porcentagem de segmentos positivos e negativos. Na Tabela 17, mostra-se a porcentagem de sentimentos positivo e negativo das opiniões fonte (coluna “Polaridade do Córpus”) para cada livro e produto eletrônico.

Para calcular o sentimento nos sumários extrativos, foi estimado o sentimento positivo e negativo usando as anotações de ReLi e Buscapé. Para os resumos abstrativos, foi calculado o sentimento com o método automático baseado em léxico proposto em [Taboada et al. \(2011\)](#), utilizando o léxico SentiLex ([Silva et al., 2012](#)), pois, segundo [Balage Filho et al. \(2013\)](#), ele tem os melhores resultados em comparação com outros léxicos do Português Brasileiro.

A Tabela 17 mostra os resultados da orientação do sentimento para cada livro e produto eletrônico. Em geral, os anotadores refletiram a distribuição do sentimento das opiniões fontes nos resumos. As proporções entre os sentimentos positivo e negativo não são exatamente as mesmas, mas são semelhantes. Isso mostra que as pessoas (anotadores) têm em conta o sentimento para criar os sumários e consideram ambas as classes, positiva e negativa, como elas aparecem nas opiniões fonte.

Há poucos casos em que a orientação do sentimento entre os resumos e as opiniões fonte foi oposta (marcados em negrito na Tabela 17). Isso indica que os anotadores se focaram apenas em uma parte das opiniões fonte, ignorando o sentimento geral.

Os sumários extrativos obtiveram melhores correlações do que os resumos abstrativos, porque as sentenças dos sumários extrativos são as mesmas das opiniões fonte, e também

Livros/Produtos Eletrônicos	Polaridade do Córpus		Sumários Extrativos		Sumários Abstrativos	
	Positivo	Negativo	Positivo	Negativo	Positivo	Negativo
Capitães da Areia	0.784	0.216	0.978	0.022	0.370	0.630
Crepúsculo	0.391	0.609	0.075	0.925	0.510	0.490
Ensaio sobre a Cegueira	0.812	0.188	0.880	0.120	0.471	0.529
Fala sério, amiga!	0.895	0.105	0.960	0.040	0.723	0.277
Fala sério, amor!	0.968	0.032	0.980	0.020	0.967	0.033
Fala sério, mãe!	0.510	0.490	0.680	0.320	0.569	0.431
Fala sério, pai!	0.842	0.158	0.877	0.123	0.950	0.050
Fala sério, professor!	0.621	0.379	0.791	0.209	0.686	0.314
O Apanhador nos Campos de Centeio	0.300	0.700	0.204	0.796	0.283	0.717
O Outro lado da meia noite	0.705	0.295	0.667	0.333	0.633	0.367
O Reverso da Medalha	0.667	0.333	0.521	0.479	0.558	0.442
Se houver Amanhã	0.867	0.133	0.952	0.048	0.716	0.284
1984	0.757	0.243	0.877	0.123	0.627	0.573
Iphone 5	0.975	0.025	0.971	0.029	0.810	0.190
Samsung Galaxy S III	0.584	0.416	0.272	0.728	0.460	0.540
LG Smart TV	0.622	0.378	0.674	0.326	0.753	0.247
Samsung Smart TV	0.556	0.444	0.502	0.498	0.536	0.464

Tabela 17: Orientação do sentimento dos sumários

porque o sentimento dos resumos abstrativos foi calculado automaticamente.

4.6 Considerações Finais

Neste capítulo foi apresentado o córpus OpiSums-PT, o primeiro córpus de sumários de opiniões com base em aspectos da língua portuguesa. Mostrou-se também neste capítulo, uma análise qualitativa sobre como as pessoas geram esses tipos de sumários. Como foi descrito, os resumos humanos são diversificados e as pessoas geram sumários somente para alguns aspectos mantendo a orientação do sentimento geral com poucas variações.

A criação deste córpus foi motivada, principalmente, pela importância que tem um córpus nesta tarefa e para auxiliar a avaliação dos métodos implementados neste trabalho ou outras futuras pesquisas na área opinião sumarização.

O córpus OpiSums-PT será utilizado neste trabalho para avaliar a informatividade dos sumários automáticos produzidos pelos métodos implementados, tanto os extrativos quanto os abstrativos.

A versão completa do OpiSums-PT está disponível na página do projeto Sucinto² sob uma licença *Creative Commons*. Junto com o cópuz, encontra-se também disponível o manual de anotação utilizado na construção de OpiSums-PT.

²<http://www.icmc.usp.br/pessoas/taspardo/sucinto/>

Métodos de Sumarização de Opiniões

Neste capítulo, serão descritos alguns métodos da literatura desenvolvidos neste trabalho de mestrado, bem como as duas propostas para gerar sumários de opiniões com base em aspectos. Para tal fim, o capítulo foi dividido em cinco seções. Na Seção 5.1, são apresentadas algumas considerações sobre os métodos desenvolvidos. Na Seção 5.2, é feita a descrição dos métodos extrativos implementados e suas adaptações para o Português do Brasil. Analogamente, na Seção 5.3, apresentam-se os métodos abstrativos estudados. Na Seção 5.4, são descritas e explicadas as duas propostas deste trabalho de mestrado. Por fim, na Seção 5.5, apresentam-se algumas considerações finais deste capítulo.

5.1 Considerações Iniciais

O presente trabalho de mestrado visa investigar métodos extrativos e abstrativos (ver Seção 2.1) baseados em aspectos para sumarizar opiniões. A motivação deste trabalho é comparar o desempenho de ambas as abordagens na sumarização de opiniões. Além disso, devido ao fato de que a sumarização abstrativa é pouco explorada devido à sua complexidade, uma motivação extra é dar um passo a mais no estado da arte para este tipo de sumarização e, assim, gerar sumários mais naturais. Para isso, foram selecionados

alguns métodos da literatura para esta tarefa.

A seleção foi baseada, principalmente, na diversidade dos métodos, ou seja, tentou-se selecionar métodos com características e modo de funcionamento diferentes com o fim de poder analisar diferentes abordagens na língua portuguesa brasileira.

Os métodos selecionados não podem ser considerados como trabalhos do estado da arte, pois a maioria dos trabalhos de sumarização de opiniões utilizaram diferentes medidas de avaliação e a comparação entre eles não é justa. Porém, os métodos desenvolvidos neste trabalho têm reportado bons resultados na literatura.

Foram desenvolvidos dois métodos extrativos e dois métodos abstrativos. Esses métodos foram adaptados para o Português do Brasil. Para alguns deles, a adaptação teve mudanças significantes em relação à versão original, devido à ausência de algumas anotações no *cópus*, ferramentas linguísticas ou, simplesmente, pelas características próprias da língua.

Para o desenvolvimento deste trabalho, foi utilizado o *cópus* OpiSums-PT, no qual os textos-fonte já têm identificados os aspectos e suas polaridades. A adoção deste *cópus* possibilita que a implementação para alguns métodos investigados foque-se exclusivamente na etapa de seleção de conteúdo para o sumário, podendo-se assumir como prontas as etapas de identificação de aspectos e predição do sentimento.

Dado que alguns textos-fonte do OpiSums-PT vêm do *cópus* ReLi e as anotações (entidades e aspectos) deles foram feitas com outros objetivos, decidiu-se realizar uma revisão manual a fim de agrupar alguns aspectos similares para a tarefa de sumarização. Por exemplo, no ReLi, os aspectos *enredo* e *história* são considerados como aspectos distintos. Neste trabalho, aspectos similares, como no exemplo anterior, foram considerados como um único aspecto. É importante indicar também que as entidades foram consideradas como aspectos, similar ao trabalho de [Gerani et al. \(2014\)](#). Na Tabela 18, são mostrados todos os aspectos únicos considerados nos métodos investigados. No caso da entidade “Livro”, em cada linha apresentam-se os aspectos similares identificados nas opiniões do *cópus* ReLi.

Entidades	Aspectos
Livro	história, enredo, trama, narrativa, tema personagens, protagonista capítulos, início do livro, metade de o livro, fim do livro falas, diálogos, frases escrita, linguagem, vocabulário, texto leitura estilo
Smartphone	bateria, <i>design</i> , processador, tela, preço, câmera, peso, sistema operacional, internet, foto, vídeo, wi-fi, som, tamanho, fones de ouvido, velocidade, chip
TV	<i>design</i> , preço, câmera, qualidade da imagem, luminosidade, wi-fi, som, durabilidade, internet

Tabela 18: Aspectos únicos considerados nos métodos investigados

5.2 Métodos Extrativos

5.2.1 Método de Hu e Liu

Como foi descrito na Seção 3.1.2, [Hu e Liu \(2004\)](#) propõem um método de sumarização de opiniões baseado em aspectos com três etapas: (1) identificação dos aspectos dos produtos, (2) identificação das sentenças positivas e negativas para cada aspecto e (3) produção do sumário.

Para a implementação do método de Hu e Liu, tendo em conta que o *corp*us OpiSums-PT possui, em seus textos-fonte, os aspectos identificados e devidamente polarizados, as etapas (1) e (2) foram omitidas. Na geração dos sumários, como primeiro passo, é feito um ranqueamento dos aspectos segundo sua frequência de aparição.

Para cada aspecto foi contabilizada a quantidade de sentenças nas quais o aspecto apresentava polaridade positiva ou negativa. A informação sobre o número total de sentenças positivas e negativas foi inserida no sumário. Logo, para cada aspecto foram selecionadas aleatoriamente algumas sentenças, tanto com polaridade positiva e negativa (ver Algoritmo 1).

Algoritmo 1: ALGORITMO DO HU E LIU

Entrada: Opiniões O , Aspectos A , Tamanho t

Saída: Sumário de Opiniões sum

início

 // Calcular a frequência dos aspectos

$d \leftarrow$ Dicionário

para $s_i \in O$ **faça**

para $a_i \in A$ **faça**

$d[(a_i, '+')] \leftarrow$ Obter frequência '+' em s_i

$d[(a_i, '-')] \leftarrow$ Obter frequência '-' em s_i

fim

fim

$A' \leftarrow$ Ordenar(d)

$sum \leftarrow \emptyset$

 // Gerar sumário

para $a_i \in A'$ **faça**

$sp \leftarrow$ Selecionar aleatoriamente $s_{a_i,+}$ // Sentença positiva

$sn \leftarrow$ Selecionar aleatoriamente $s_{a_i,-}$ // Sentença negativa

$sum \leftarrow sp + sn$

se $i = t$ **então**

 | Sair

fim

fim

fim

retorna sum

Ressalta-se que, nesta adaptação, o tamanho do sumário gerado é configurável em termos do número de aspectos e do número de sentenças por aspecto. Assim, dado um tamanho desejado (por exemplo, 150 palavras), o método implementado calcula o número de aspectos e sentenças a serem considerados no sumário final. Na proposta original do método de Hu e Liu, todos os aspectos são listados com todas suas sentenças.

Na Figura 23, é apresentado um resumo gerado pelo método de Hu e Liu. Como se pode observar, mostra-se uma sentença para cada aspecto. Neste caso, os aspectos com melhor ranqueamento são Samsung Smart TV, preço, durabilidade, câmera e qualidade da imagem. É importante ressaltar que o aspecto Samsung Smart TV, que também é a entidade, é o melhor aspecto ranqueado posto que foi o mais frequente nas opiniões. Para cada um desses aspectos, é mostrada a quantidade de sentenças positivas e negativas. Com este sumário, pode-se entender que as opiniões sobre o produto o Samsung Smart

TV estão mais relacionadas a esses aspectos. Em relação ao sentimento das opiniões, a quantidade de opiniões positivas e negativas está quase equilibrada.

<p>Aspecto: Samsung Smart TV Sentenças Positivas: 16 - Ótima televisão , ou melhor , central multimídia fiquei impressionado pela quantidade de recursos (Wi-Fi , Aplicativos e All Share) como tenho um celular Samsung (Omnia W) com Windows 7.5 tive uma interação completa. Sentenças Negativas: 11 - Custo-benefício não compensa.</p> <p>Aspecto: Preço Sentenças Positivas: 0 Sentenças Negativas: 2 - O que não gostei: Valor muito alto</p> <p>Aspecto: Durabilidade Sentenças Positivas: 0 Sentenças Negativas: 2 - A TV é maravilhosa, porém queimou com menos de dois anos de uso e o conserto na autorizada fica o preço de uma nova.</p> <p>Aspecto: Câmera Sentenças Positivas: 1 - A câmera embutida na tv com sensor de movimento e presença funciona bem e impressiona quem não conhece. Sentenças Negativas: 1 - O que não gostei: A câmera poderia ter movimentação horizontal de visão e não só apenas vertical.</p> <p>Aspecto: Qualidade da Imagem Sentenças Positivas: 1 - Excelente aparelho de TV com ótima qualidade de imagem e recursos. Sentenças Negativas: 0</p>

Figura 23: Sumário produzido pelo método de [Hu e Liu \(2004\)](#)

5.2.2 Método de Tadano et al.

O método de [Tadano et al. \(2010\)](#) gera um resumo para cada aspecto (separadamente) utilizando três características: (1) pontuações dos aspectos, (2) o valor TF-IDF ([Salton e McGill, 1986](#)) e (3) o número de menções com tópico semelhante (ver Seção 3.1.4 para mais detalhe).

Diferentemente de [Tadano et al. \(2010\)](#), a adaptação neste trabalho gera um único sumário geral agrupando a sentença mais representativa de cada um dos principais aspectos. A seleção dos principais aspectos foi com base na frequência de aparição nas opiniões, como foi feito também em [Hu e Liu \(2004\)](#).

Em relação à primeira característica que utiliza este método (pontuações dos aspectos), o método original considera o número de estrelas em um intervalo de 0 até 5. Na adaptação

implementada, foi considerada uma escala de 1 até 5, pois as opiniões de Buscapé no OpiSums-PT utilizam esse intervalo. No caso das opiniões do ReLi (que utilizam uma escala de 0 até 5), os comentários que tinham zero estrelas foram considerados com uma estrela, com a finalidade de uniformizar o intervalo.

Antes do cálculo do valor TF-IDF, para cada palavra foram identificadas as etiquetas morfossintáticas (por meio do tagger proposto por [Fonseca e Rosa \(2013\)](#)¹) e os lemas das palavras². O agrupamento das opiniões foi realizado utilizando o algoritmo K-Means ([Steinhaus, 1956](#)) e só foram consideradas as palavras dos textos que eram substantivos e adjetivos. A ideia de utilizar o algoritmo K-Means consiste em agrupar as sentenças mais similares em conteúdo e implicitamente, por meio dos adjetivos (por exemplo, bonito, ruim, útil), em polaridade.

Para a criação do sumário, de cada cluster é selecionada a sentença com maior TF-IDF, e, em seguida, essas sentenças são agrupadas segundo a quantidade de estrelas que tem a opinião que as continha. Finalmente, em cada grupo de estrelas (cinco grupos), as sentenças são ordenadas segundo a importância que obtêm considerando as três características explicadas anteriormente.

Da mesma forma que a adaptação de [Hu e Liu \(2004\)](#), o método implementado é configurável em função do número de aspectos e do número de sentenças por aspecto desejados no sumário. Para tal fim, os grupos (número de estrelas) que têm maior quantidade de sentenças têm mais prioridade. Assim, isso tenta garantir que o sumário vai mostrar a polaridade mais predominante nas opiniões (ver Algoritmo 2).

Na Figura 24, apresenta-se um sumário gerado pelo método de [Tadano et al. \(2010\)](#). O sumário contém sentenças para os cinco principais aspectos (Samsung Smart TV, preço, durabilidade, câmera e qualidade da imagem). Pode-se observar também que o resumo é composto por nove sentenças, entre positivas e negativas.

¹Tagger com acurácia de 96.48% em textos jornalísticos, aproximadamente.

²Foi utilizado o lematizador disponível em <http://www.nilc.icmc.usp.br/nilc/index.php/tools-and-resources>

```

Propaganda enganosa
Custo benefício não compensa.
O que não gostei: Valor muito alto
O que não gostei: Preço alto
Durabilidade 0
A TV é maravilhosa, porém queimou com menos de dois anos de uso e
o conserto na autorizada fica o preço de uma nova.
O que não gostei: A câmera poderia ter movimentação horizontal de
visão e não só apenas vertical.
A câmera embutida na tv com sensor de movimento e presença
funciona bem e impressiona quem não conhece.
Excelente aparelho de TV com ótima qualidade de imagem e
recursos.

```

Figura 24: Sumário produzido pelo método de [Tadano et al. \(2010\)](#)

Algoritmo 2: ALGORITMO DE TADANO ET AL.

Entrada: Opiniões O , Aspectos A , Tamanho t

Saída: Sumário de Opiniões sum

início

$d \leftarrow$ Dicionário

$O' \leftarrow \emptyset$

para $s_i \in O$ **faça**

para $a_i \in A$ **faça**

$d[a_i] \leftarrow$ Obter frequência em s_i

fim

$O' \leftarrow$ Tagger(s_i)

fim

$A' \leftarrow$ Ordenar(d)

$O'' \leftarrow \emptyset$

para $s_i \in O'$ **faça**

$t \leftarrow$ Extrair substantivos e adjetivos de s_i

$O'' \leftarrow$ TF-IDF(t)

fim

$E \leftarrow$ Dicionário

para $a_i \in A'$ **faça**

$C \leftarrow$ K-Means(O''_{a_i})

para $c_i \in C$ **faça**

$s_i \leftarrow$ Selecionar a sentença com máximo valor TF-IDF em c_i

$n \leftarrow$ Identificar a quantidade de estrelas de s_i

$E[(a_i, n)] \leftarrow s_i$

fim

fim

$sum \leftarrow \emptyset$

para $a_i \in A'$ **faça**

$sum \leftarrow$ Selecionar a sentença de E_{a_i} com mais elementos

se $i = t$ **então**

 Sair

fim

fim

retorna sum

5.3 Métodos Abstrativos

5.3.1 Método de Ganesan et al.

Para gerar os sumários de opiniões, o método de [Ganesan et al. \(2010\)](#) utiliza um grafo, no qual cada nó representa uma palavra única das opiniões e as arestas indicam uma relação de precedência da palavra i com a palavra $i+1$. Em seguida, dentro do grafo são procurados os sub-grafos (segmentos textuais das sentenças) mais redundantes que satisfazem um conjunto de sequências sintáticas. No método original, proposto para o Inglês, foram definidas quatro expressões regulares para encontrar essas sequências:

1. .* (Substantivo)+ .* (Verbo)+ .* (Adjetivo)+ .*
2. .* (Adjetivo)+ .* (/to)+ .* (Verbo).*
3. .* (Advérbio)* .* (Adjetivo)+ .* (Substantivo)+ .*
4. .* (Advérbio)+ .* (Preposição)+ .* (Substantivo)+ .*

Nessas sequências, os símbolos “+” indicam que as etiquetas morfossintáticas podem aparecer uma ou mais vezes; os símbolos “*” sinalam que as etiquetas morfossintáticas podem aparecer zero ou mais vezes; os símbolos “.” indicam que pode-se tratar de qualquer etiqueta morfossintáticas; e a expressão “(/to)” representa a palavra *to* do Inglês.

A finalidade dessas expressões regulares é capturar, dentro do grafo, as sequências de palavras mais opinativas das sentenças. Por exemplo, da sentença em Inglês “*My smartphone works excellent, in comparison with other ones*” a sequência “*smartphone works excellent*” seria extraída utilizando a primeira expressão regular, pois satisfaz a condição (“*smartphone_Substantivo works_Verbo excelente_Adjetivo*”). Assim, garantia-se que o sumário final gerado por este método estará composto, em sua maioria, por sequências de palavras opinativas.

Na adaptação desse método, foram utilizadas outras expressões regulares, pois as estruturas sintáticas para o Português são diferentes do Inglês (idioma utilizado no método original). Ademais, com essas expressões tentou-se encontrar as sequências de palavras meramente opinativas dentro do conjunto total de opiniões. A seguir, mostram-se as expressões regulares utilizadas no método adaptado.

1. .* (Substantivo)+ .* (Verbo)+ .* (Adjetivo)+ .*
2. .* (Adjetivo)+ .* (Substantivo)+ .*
3. .* (Substantivo)+ .* (Advérbio)+ .* (Adjetivo)+ .*
4. .* (Adjetivo)+ .* (Verbo)+ .* (Substantivo)+ .*
5. .* (Verbo)+ .* (Substantivo)+ .*

Essas sequências foram construídas depois de realizar algumas análises manuais das estruturas sintáticas mais frequentes das sentenças no *córpus OpiSums-PT*. Similar à versão original do método, estas expressões regulares tentam capturar os segmentos textuais opinativos presentes nas sentenças que conformam o grafo de opiniões. A seguir, são mostrados alguns exemplos de segmentos textuais opinativos que são identificados pelas expressões regulares definidas nesta adaptação:

1. Sua bateria_(Substantivo) é_(Verbo) legal_(Adjetivo)
2. Ótima_(Adjetivo) televisão_(Substantivo)
3. Tela_(Substantivo) muito_(Advérbio) bonita_(Adjetivo)
4. O mais bonito_(Adjetivo) é_(Verbo) o livro_(Substantivo)
5. Gostei_(Verbo) muito da tela_(Substantivo)

Na versão original, a implementação de [Ganesan et al. \(2010\)](#) recebe como entrada opiniões altamente redundantes sobre um determinado aspecto e gera um único sumário para esse aspecto. Dado que no *córpus OpiSums-PT* as opiniões descrevem vários aspectos, como primeiro passo foram separadas as sentenças por aspecto, ou seja, sentenças que mencionaram um mesmo aspecto foram agrupadas. Em seguida, foram identificados os principais aspectos segundo o critério de [Hu e Liu \(2004\)](#). Para cada aspecto principal, foram selecionados os segmentos textuais opinativos que satisfaziam as expressões regulares mencionadas anteriormente. A etiquetagem morfossintática foi realizada utilizando o tagger proposto por [Fonseca e Rosa \(2013\)](#). De todos os segmentos textuais opinativos identificados para um aspecto, foi selecionado o segmento textual mais frequente no grafo de opiniões. Finalmente, cada um dos segmentos textuais opinativos dos aspectos foram agrupados em um resumo geral (ver Algoritmo 3).

Algoritmo 3: ALGORITMO DE GANESAN ET AL.

Entrada: Opiniões O , Aspectos A , Tamanho t

Saída: Sumário de Opiniões sum

```
início
   $d \leftarrow$  Dicionário
  para  $s_i \in O$  faça
    para  $a_i \in A$  faça
       $d[a_i] \leftarrow$  Obter frequência em  $s_i$ 
    fim
  fim
   $A' \leftarrow$  Ordenar( $d$ )
   $G \leftarrow \emptyset$ 
  para  $a_i \in A'$  faça
     $G[a_i] \leftarrow \emptyset$ 
    para  $s_i \in O$  faça
       $G[a_i] \leftarrow$  Gerar grafo com palavras de  $s_i$ 
    fim
     $P[a_i] \leftarrow$  Identificar as expressões regulares definidas na Seção 5.3 em  $G[a_i]$ 
     $E[a_i] \leftarrow$  Selecionar os segmentos textuais mais frequentes em  $P[a_i]$ 
  fim
   $sum \leftarrow \emptyset$ 
  para  $a_i \in A'$  faça
     $sum \leftarrow$  Selecionar um segmentos textual de  $E[a_i]$ 
    se  $i = t$  então
      Sair
    fim
  fim
retorna  $sum$ 
```

Na Figura 25, mostra-se um sumário gerado pelo método de Ganesan et al. para o produto “Samsung Smart TV”. Como se observa, o sumário é bem menor comparado aos sumários anteriormente mostrados e parece mais um sumário extrativo do que abstrativo, pois as opiniões de entrada não foram muito redundantes, e isso é um requisito principal no método de [Ganesan et al. \(2010\)](#).

O que não gostei: e uma pena não tem hd interno.
O que não gostei: valor muito alto.
Dois anos de uso e o conserto na autorizada fica o preço de uma nova..
Não gostei: a câmera poderia ter movimentação horizontal de visão e não só apenas vertical..
Excelente aparelho de tv com ótima qualidade de imagem e recursos..

Figura 25: Sumário produzido pelo método de [Ganesan et al. \(2010\)](#)

5.3.2 Método de Gerani et al.

Como foi explicado anteriormente, o método de Gerani et al. (2014) utiliza as relações RST para ranquear os aspectos presentes nas opiniões fonte utilizando um grafo. Em seguida, o sumário final é gerado preenchendo alguns *templates* (criados manualmente) com os principais aspectos, considerando a polaridade deles.

Na adaptação realizada, foi utilizado o parser discursivo DiZer³ (Maziero et al., 2011) para encontrar as relações RST nas opiniões escritas em Português. Este parser permite identificar as relações RST (por exemplo, elaboração, contradição, etc.) que existem entre os segmentos de um texto, chamadas unidades mínimas de discurso (EDU do inglês *Elementary Discourse Units*). Cada EDU pode ser classificado como núcleo (informação principal) ou satélite (informação adicional).

Em cada EDU, verifica-se se há algum aspecto, pois a ideia deste método é que os aspectos herdaram as relações RST dos EDUs que os contêm. Em cada aspecto identificado, verifica-se também se ele é núcleo ou satélite. É importante ressaltar que o método de Gerani et al. (2014) está baseado no princípio que as relações RST entre aspectos podem ajudar a reconhecer quais são os aspectos mais relacionados no conjunto total das opiniões e essa informação permitirá criar melhores sumários.

Em seguida, um grafo dirigido é construído utilizando as relações RST entre os aspectos presentes em todos os EDUs. A direção das arestas entre os aspectos é determinada segundo a função discursiva (núcleo ou satélite) dos EDUs que contêm os aspectos. Os aspectos satélites apontam para os aspectos núcleos, pois estes últimos são considerados os mais importantes.

Na Figura 26, é mostrado um grafo para o produto “Samsung Galaxy S-III”. Este grafo nos permite identificar quais são os aspectos mais relacionados como os outros. Por exemplo, nessa figura, pode-se observar que o aspecto “Galaxy SIII” é o aspecto que tem mais relações RST com os outros aspectos. Como se pode observar também, cada aresta (relação RST) tem um peso associado.

³Parser discursivo com desempenho de 62.5% de acurácia na detecção de relações e 81% na determinação da nuclearidade em textos científicos.

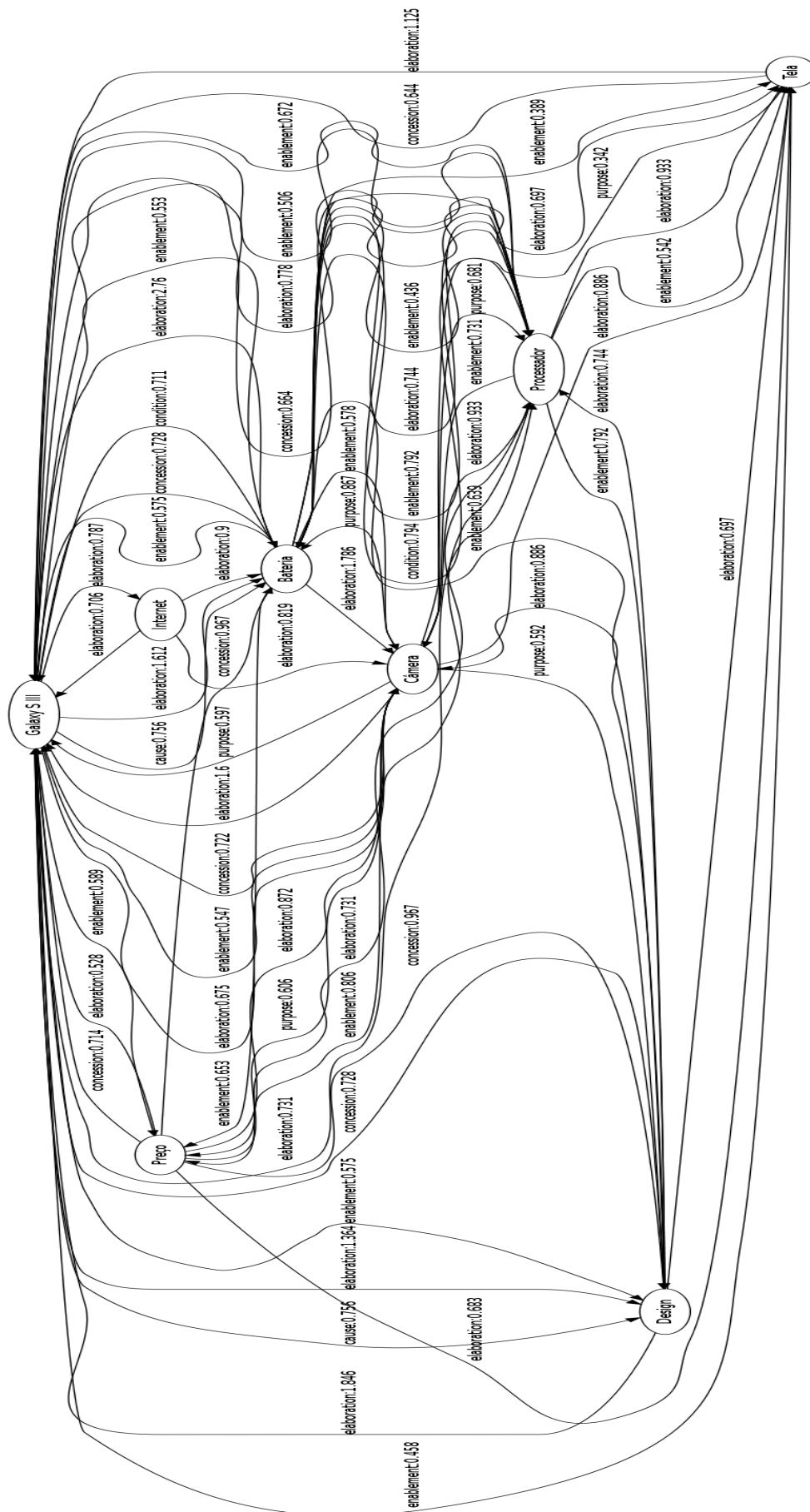


Figura 26: Grafo das relações RST entre os aspectos

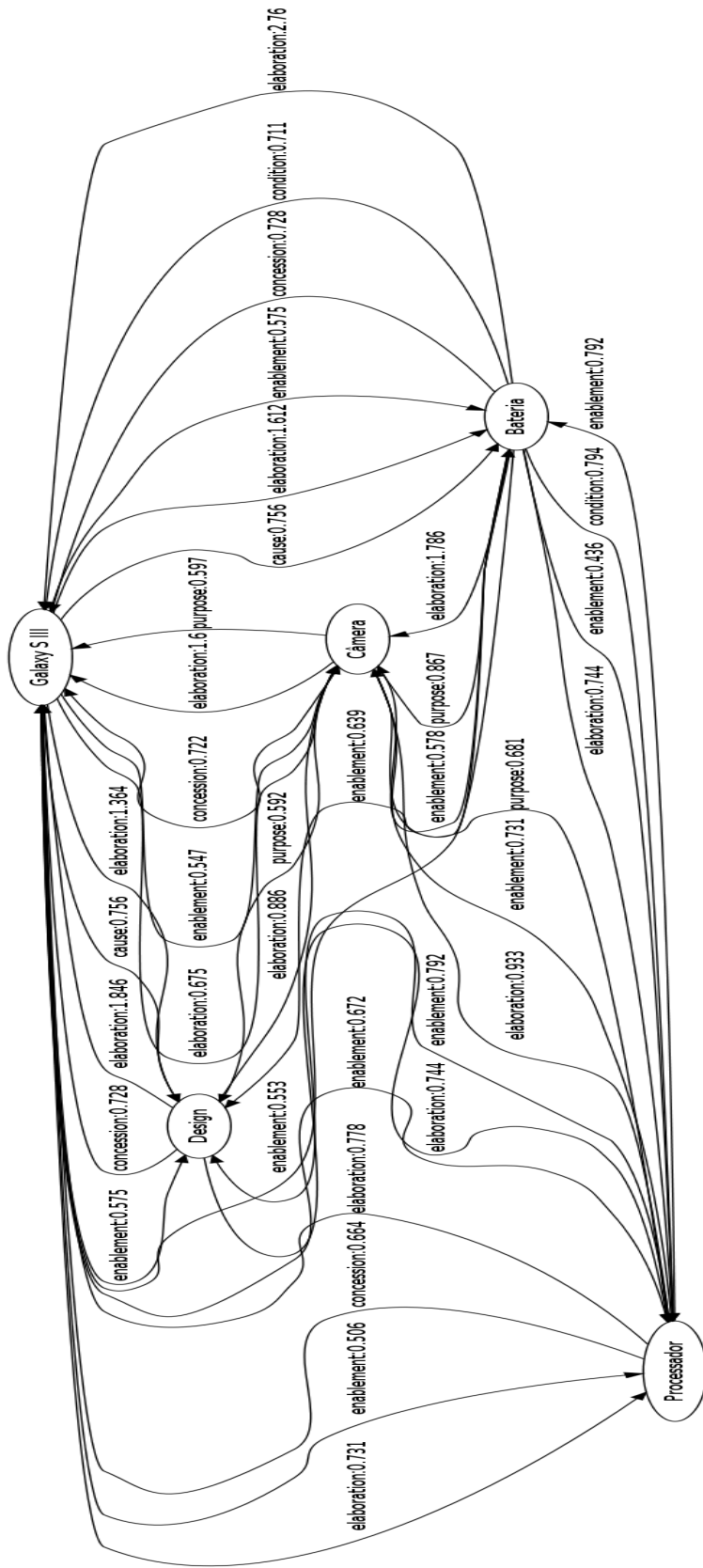


Figura 27: Subgrafo das relações RST entre os aspectos

Esse peso indica a força da relação RST entre os aspectos, e é calculado, principalmente, com base na distância (número de palavras) que existem entre os aspectos (em questão) na opinião. Em outras palavras, o peso atribuído a uma relação RST indica se essa relação é forte ou fraca, tendo em conta quão separados os aspectos estão no texto.

Para gerar um sumário de opiniões, no método de Gerani et al. (2014), é necessário indicar o número de aspectos a considerar no sumário desejado. Dado esse número N , do grafo geral extrai-se o subgrafo que contenha os N principais aspectos. A importância dos aspectos é calculada utilizando o algoritmo PageRank (Page et al., 1998) utilizando os pesos das relações RST. A Figura 27 mostra o subgrafo extraído (cinco principais aspectos) da Figura 26. A partir desse subgrafo, cria-se um grafo não dirigido, no qual o peso das arestas entre dois nós é a soma dos pesos de todas as relações RST entre eles (sem considerar a direção). Na Figura 28, mostra-se o grafo não dirigido para o exemplo anterior.

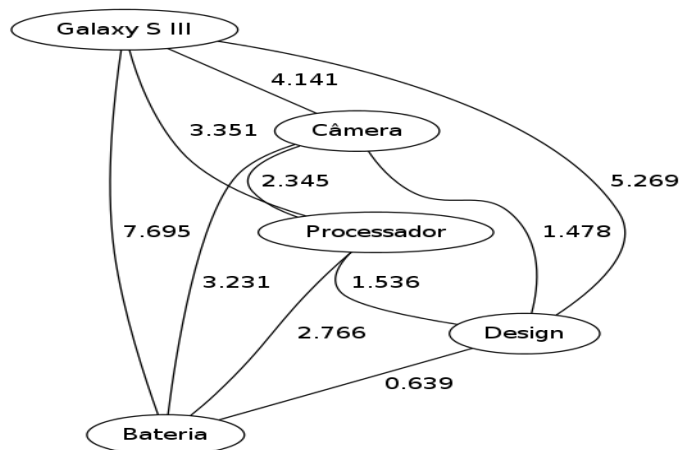


Figura 28: Grafo não dirigido das relações RST

Em seguida, calcula-se o *Maximum Spanning Tree* (MST) (Belzer et al., 1978) do grafo não dirigido com a finalidade de criar uma hierarquia dos aspectos. O MST de um grafo é uma árvore que liga todos os vértices do grafo usando as arestas com um custo total máximo. Por tanto, a árvore obtida permite identificar quais são as relações mais fortes do grafo e, é considerada por Gerani et al. (2014) como a hierarquia de aspectos. Na Figura 29, observa-se a hierarquia resultante para o grafo anterior não dirigido do produto “Samsung Galaxy S-III”. Essa hierarquia permite agrupar aos aspectos em função de como eles apareceram em todas as opiniões fonte. A ordem de descrição dos aspectos no sumário

segue a estrutura da hierarquia. Para o exemplo, a ordem dos aspectos no sumário seria: Galaxy S-III, câmera, processador, design e bateria.

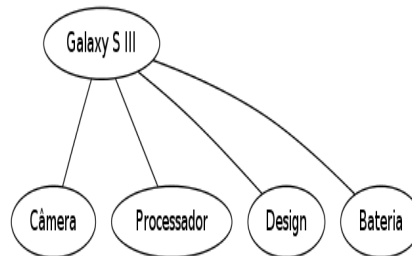


Figura 29: Hierarquia dos aspectos do Galaxy S-III

Finalmente, já com a ordem estabelecida para os principais aspectos, o passo seguinte consiste em preencher alguns templates criados manualmente com os principais aspectos. Para tal fim, foram traduzidos os templates originais (em inglês) para o português brasileiro. Na Tabela 19, mostra-se os templates utilizados na adaptação do método de Gerani et al. (2014).

A Tabela 19 está dividida em três partes: quantificadores, expressões com verbos e conectores. Os quantificadores permitem indicar a quantidade de usuários que mencionaram um aspecto; as expressões com verbos indicam o sentimento geral dos usuários; e os conectores servem para ligar duas sentenças. A seleção das opções em cada tipo de template é aleatória. Com relação às expressões com verbos, diferentemente do método original, nesta adaptação não foi utilizado o método de Carenini e Cheung (2008) para saber se as opiniões são controversas, pois as polaridades anotadas nas opiniões de OpiSums-PT só têm um sentimento associado (positivo ou negativo), e não uma força do sentimento atribuída (por exemplo, -1, +3, -2, entre outros) como no método original. Nesta adaptação, se a diferença entre as quantidades de segmentos positivos e negativos para um aspecto era menor ou igual a um, a opinião geral era considerada como controversa (ver Algoritmo 4).

Quantificadores
"1.0": ["Todos os usuários (<NÚMERO> de pessoas) que comentaram sobre o <ASPECTO>", "Todos os clientes (<NÚMERO> pessoas) que analisaram o <ASPECTO>"...]
"0.8": ["Quase todos os usuários comentaram sobre o <ASPECTO> e eles", "Quase todos os clientes mencionaram o <ASPECTO> e eles"...]
"0.6": ["A maioria dos usuários comentaram sobre o <ASPECTO> e eles principalmente", "A maioria dos compradores mencionaram o <ASPECTO> e eles"...]
"0.45": ["Quase a metade dos usuários comentaram sobre o <ASPECTO> e eles", "Quase 50.0% dos compradores mencionaram o <ASPECTO> e eles"...]
"0.2": ["Cerca de <PORCENTAGEM>% dos usuários comentaram sobre o <ASPECTO> e eles", "Em torno de <PORCENTAGEM>% dos compradores mencionaram o <ASPECTO> e eles"...]
"0.0": ["<NÚMERO> usuário(s) comentaram sobre o <ASPECTO> e em geral eles", "<NÚMERO> comprador(es) mencionaram o <ASPECTO> e eles"...]
Expressões com Verbos
"Controvérsia": ["tinham opiniões controversas sobre ele", "tinham opiniões variadas (positivas e negativas) sobre ele"...]
"-3": ["odiavam isso", "sentiram que era muito ruim", "acharam que era muito ruim"...]
"-2": ["não gostaram", "sentiram que era ruim", "acharam que era ruim"...]
"-1": ["não gostaram muito", "sentiram que era um pouco ruim", "acharam que era um pouco ruim"...]
"0": ["não expressaram opinião positiva nem negativa sobre ele"...]
"+1": ["gostaram dele", "sentiram que era bom", "acharam que era satisfatório"...]
"+2": ["gostaram muito", "realmente gostaram dessa característica", "acharam que era realmente bom"...]
"+3": ["adoraram", "sentiram que era excelente", "acharam que era ótimo"...]
Conectores
["Além disso, relacionado ao <ASPECTO>", "Ademais,", "Além disso, em relação ao <ASPECTO>", "Em relação ao <ASPECTO>", "Falando sobre o <ASPECT>", "Da mesma forma,", "Em contraste,"...]

Tabela 19: Templates traduzidos de Gerani et al. (2014)

Algoritmo 4: ALGORITMO DE GERANI ET AL.

Entrada: Opiniões O , Aspectos A , Tamanho t

Saída: Sumário de Opiniões sum

início

$T \leftarrow \emptyset$

para $o_i \in O$ **faça**

$R \leftarrow$ Identificar relações RST de o_i

para $r_i \in R$ **faça**

se r_i .contém(a_x, a_y) **então**

$T \leftarrow (a_x, peso_{xy}, a_y) // peso_{xy}$ representa o peso entre dois aspectos

fim

fim

fim

$G \leftarrow$ Criar grafo(T)

$P \leftarrow$ PageRank(G)

$G' \leftarrow$ Extrair subgrafo com os t principais aspectos de P

$G'' \leftarrow$ Calcular o Maximum Spanning Tree de G' não dirigido

$sum \leftarrow \emptyset$

para $a_i \in G''.nós()$ **faça**

$sum \leftarrow$ Selecionar templates de quantificadores para a_i

$sum \leftarrow$ Selecionar templates de expressões com verbos para a_i

$sum \leftarrow$ Selecionar templates de conectores para a_i

fim

fim

retorna sum

Na Figura 30, mostra-se o sumário produzido pelo método de Gerani et al. para o exemplo anterior (produto “Samsung Galaxy S-III”). Como pode-se observar, o conteúdo do sumário basicamente é constituído pelos templates da Tabela 19.

Todos os usuários (10 de pessoas) que comentaram sobre o Galaxy S III gostaram dele. Em relação à câmera, em torno de 20.0% dos compradores mencionaram a câmera e eles acharam que era satisfatória. Ademais, cerca de 40% dos usuários comentaram sobre o processador e eles expressaram opiniões controversas sobre essa característica. Além disso, em relação ao design, em torno de 30.0% dos compradores mencionaram o design e eles tinham opiniões variadas (positivas e negativas) sobre ele. Ademais, quase a metade dos usuários comentaram sobre a bateria e eles expressaram opiniões controversas sobre ela.

Figura 30: Sumário produzido pelo método de Gerani et al. (2014)

5.4 Propostas

Na sumarização de opiniões baseada em aspectos, os sumários extrativos oferecem a informação mais importante de cada aspecto de uma maneira específica e detalhada, o que pode ser muito útil para os leitores do sumário. Por outro lado, com os sumários abstrativos, é possível oferecer uma visão equilibrada entre os vários pontos de vista dos usuários sobre cada aspecto. Isso não é possível em sumários puramente extrativos.

Neste trabalho de pesquisa, foram implementados e adaptados quatro métodos clássicos de sumarização de opiniões com base em aspectos para a língua portuguesa do Brasil. Estes métodos foram analisados em cada uma de suas fases, e como consequência dessa análise, identificaram-se algumas lacunas que afetavam o desempenho deles. Entre os principais problemas encontrados nesses métodos, destacam-se o uso de poucas informações para selecionar as sentenças representativas e a utilização de templates puramente estáticos. Assim, essas lacunas motivaram também o desenvolvimento de duas propostas para gerar melhores sumários de opiniões.

Para a produção dos sumários, são propostos dois métodos novos: um método extrativo e um método abstrativo. Na Figura 31, são ilustradas as etapas das duas propostas. De maneira sucinta, o método extrativo proposto tem duas etapas: agrupamento de sentenças

e ranqueamento de sentenças. De forma análoga, o método abstrativo está dividido em duas etapas: agrupamento de segmentos textuais e geração com base em *templates*.



Figura 31: Etapas das propostas

Da Figura 31, na proposta extrativa, a primeira etapa visa agrupar as sentenças em nível de aspecto e de polaridade. Na segunda etapa, realiza-se um ranqueamento para selecionar as sentenças mais relevantes de cada aspecto. Com estas duas etapas, pretende-se gerar um sumário extrativo. Para o método abstrativo proposto, a primeira etapa visa agrupar e selecionar os segmentos textuais (n-gramas) mais relevantes para cada aspecto. Finalmente, na segunda etapa do método abstrativo, *templates* criados manualmente são preenchidos com os dados da etapa anterior segundo seu contexto.

5.4.1 Proposta Extrativa: Opizer-E

Neste método, chamado Opizer-E (*Opinion Summarizer - Extractive*), o objetivo consiste em gerar resumos extraindo uma pequena quantidade de sentenças das opiniões sobre os principais aspectos, de tal modo que a cobertura dos aspectos e a distribuição da polaridade sejam preservadas, tanto quanto seja possível. Para tal fim, identificaram-se duas etapas: agrupamento de sentenças e ranqueamento de sentenças.

Em geral, diferentemente dos métodos extrativos explicados anteriormente e de outras abordagens da literatura, Opizer-E combina dois tipos de informações (posição da

sentença e proximidade dos qualificadores) que não têm sido utilizadas em conjunto nos outros trabalhos. Estes tipos de informações serão explicados com mais detalhes mais à frente.

O formato do sumário utilizado nesta proposta é o formato estruturado em aspectos (ver Figura 9). A seleção deste tipo de formato se deveu, principalmente, a dois benefícios que ele oferece. Primeiramente, um sumário estruturado mostra, explicitamente, qual é o sentimento geral nas opiniões, indicando o número de opiniões positivas/negativas ou a quantidade de sentenças com esses sentimentos. Este é um grande benefício para o leitor do sumário, pois ele saberá, exatamente, quantas opiniões/sentenças são positivas ou negativas. O segundo benefício deste formato de sumário é que pode facilitar a leitura do sumário. Como é reportado na literatura, os sumários extrativos geralmente têm o problema da falta de coesão entre as sentenças do texto. Os sumários estruturados reduzem esse problema agrupando as sentenças por aspectos e mostrando-as como uma lista de itens. Isso, conseqüentemente, pode ajudar a facilitar a leitura do sumário. A seguir, as duas etapas do método extrativo proposto são explicadas com mais detalhe.

A. Agrupamento de Sentenças

Nesta proposta de pesquisa, trabalha-se com um conjunto de sentenças de opiniões que contêm múltiplos aspectos. Nesse cenário, algumas sentenças poderiam ser similares a outras. Esta etapa visa agrupar as opiniões que versam sobre o mesmo aspecto e tenham a mesma orientação da polaridade. Ou seja, o agrupamento é em nível de aspecto e polaridade.

Por exemplo, nas sentenças (1) “*A bateria do Iphone 5 dura pouco tempo*”, (2) “*A bateria do Iphone 5 é leve, eu gosto disso*” e (3) “*O tempo de duração da bateria do Iphone 5 é curto*”, a primeira e terceira sentença deveriam ser agrupadas no mesmo grupo, pois elas falam sobre o mesmo aspecto (bateria) e têm a mesma polaridade (negativa). Por outro lado, a segunda sentença ficaria em outro grupo, pois, embora ela descreva o aspecto bateria, a polaridade que possui é positiva.

Depois do agrupamento das sentenças por aspectos, utiliza-se a frequência de ocorrência dos aspectos para calcular a importância deles. Neste ranqueamento, também utilizado por [Hu e Liu \(2004\)](#) e [Wang et al. \(2013\)](#), quanto mais vezes ocorre um aspecto,

mais importante é ele.

B. Ranqueamento de Sentenças

Os sumários extrativos baseados em aspectos visam listar as sentenças mais relevantes para cada aspecto. Porém, estabelecer um critério de relevância não é uma tarefa fácil. Muitos trabalhos sugerem critérios distintos. Em [Beineke et al. \(2003\)](#), os autores fazem um ranqueamento das sentenças em função das posições delas nas opiniões. No trabalho de [Blair-Goldensohn et al. \(2008\)](#), as sentenças são ordenadas segundo a polaridade das sentenças, ou seja, os autores sugerem selecionar as sentenças mais positivas e mais negativas e utilizá-las no sumário final. [Sharifi et al. \(2010\)](#) utilizam o valor do TF-IDF para fazer o ranqueamento das opiniões.

O objetivo nesta etapa é fazer um ranqueamento das sentenças sobre um aspecto e mostrar as melhores no resumo final de acordo com o tamanho predefinido do sumário. Para tal fim, foram feitos experimentos com alguns dos critérios citados anteriormente e também com outros dois tipos de ranqueamentos propostos neste trabalho. O primeiro ranqueamento está baseado na informação da posição das sentenças na opinião que as contém, e o segundo ranqueamento utiliza a proximidade (distância) que existem entre os aspectos e os segmentos textuais que os avaliam.

O ranqueamento com base nas posições das sentenças está baseado na ideia de que a primeira sentença é mais importante do que as outras. Este critério tem sido utilizado em vários trabalhos de sumarização de notícias ([Seki \(2002\)](#), [Ouyang et al. \(2010\)](#), [Nóbrega et al. \(2014\)](#)) com resultados satisfatórios. [Beineke et al. \(2003\)](#) também utilizaram esse critério na sumarização de opiniões, outorgando um peso maior à primeira sentença e outro peso único para todas as demais sentenças. Diferentemente de [Beineke et al. \(2003\)](#), nesta proposta, cada sentença tem um peso diferente dependendo da sua posição e da quantidade de sentenças na opinião.

No Opizer-E, as sentenças têm maior importância se elas estão mais posicionadas na parte inicial da opinião, pois acredita-se que, quando os usuários dão uma opinião, eles preferem expressar o mais importante na parte inicial do texto. Foram feitos também experimentos outorgando maior peso às sentenças intermediárias e finais, mas os melhores resultados foram obtidos quando as sentenças iniciais tinham mais importância. Para

calcular o peso de cada sentença, foi utilizada a Equação 5.1.

$$PesoPosição (sentença) = \frac{TamanhoOpinião - Posição (sentença)}{TamanhoOpinião} \quad (5.1)$$

onde, $PesoPosição(sentença)$ é o peso que recebe uma sentença e seu valor oscila entre 0 e 1, sendo 1 o maior peso. $TamanhoOpinião$ refere-se ao número de sentenças de uma opinião e $Posição(sentença)$ indica a posição da sentença na opinião, sendo que a primeira sentença ocupa a posição 0.

Na Figura 32, mostra-se uma opinião com 4 sentenças. Para o aspecto “TV” (marcado em negrito), utilizando o ranqueamento por posição, a primeira sentença teria um peso igual a 1. De forma análoga, para o aspecto “câmera”, a terceira sentença teria peso igual a 0.5.

A melhor **tv** na atualidade!
União de tv, com interatividade e internet, a Samsung superou seus concorrentes inovando e trazendo recursos incríveis para uma tv.
A **câmera** embutida na tv com sensor de movimento e presença funciona bem e impressiona quem não conhece.
O contorno fino ressalta a qualidade da imagem da tela.

Figura 32: Exemplo de opinião

Em relação ao ranqueamento por proximidade, a ideia principal consiste em selecionar as sentenças que tenham os aspectos e seus qualificadores (segmentos textuais) mais próximos. Assim, com esse critério, pretende-se selecionar as sentenças que expressem opiniões mais pontuais e diretas sobre um aspecto determinado. Para tal fim, foi calculada a distância que existe entre os aspectos e seus qualificadores.

Neste trabalho, os qualificadores são segmentos textuais ou n-gramas que expressam um sentimento sobre um aspecto. A distância entre um aspecto e seus qualificadores é calculada em função das posições que ocupam dentro da sentença. Neste ranqueamento, quanto menor é a distância, maior é a importância da sentença. No caso em que os qualificadores estivessem em outra sentença, a distância considerada é a distância entre o qualificador e o início da sentença. Foram poucos os casos desse tipo reportados nos experimentos.

No exemplo anterior (Figura 32), os qualificadores aparecem sublinhados. No caso do aspecto “TV”, o qualificador “melhor” aparece bem próximo ao aspecto, o qual permite

que a primeira sentença tenha um peso forte (1.0). Agora, para o aspecto “câmera”, o qualificador “funciona bem e impressiona quem não conhece” encontra-se distante do aspecto, de forma que a terceira sentença obtém um peso menor (0.3) no ranqueamento por proximidade.

Por último, para calcular a importância final das sentenças para cada aspecto, foi utilizada a Equação 5.2, a qual considera os dois ranqueamentos explicados anteriormente. Foram feitos também experimentos utilizando os dois ranqueamentos separadamente, mas os melhores resultados foram obtidos quando se utilizaram os dois ranqueamentos conjuntamente.

$$\text{Importância (sentença)} = \alpha \times \text{PesoPosição} + (1 - \alpha) \times \text{PesoProximidade} \quad (5.2)$$

Na equação anterior, *Importância (sentença)* indica a importância de uma sentença para um determinado aspecto. O coeficiente α permite manter o *trade-off* entre os dois tipos de ranqueamentos. O melhor valor para α reportado nos experimentos foi 0.65. *PesoPosição* e *PesoProximidade* indicam o peso que tem uma sentença utilizando os ranqueamentos por posição e proximidade, respectivamente.

Na Figura 33, é apresentado um sumário produzido pelo método Opizer-E. Como se pode observar, mostra-se uma sentença positiva e negativa para cada aspecto. Na configuração padrão de Opizer-E, cada aspecto possui uma sentença positiva e negativa no sumário, com o fim de aumentar a cobertura dos aspectos no texto. Porém, esse número de sentenças por aspecto pode aumentar até cobrir a quantidade de palavras desejada para o sumário (ver Algoritmo 5).

Diferentemente do sumário gerado pelo método de Hu e Liu (2004) (ver Figura 23), este resumo contém sentenças que expressam opiniões pontuais e diretas sobre os diferentes aspectos presentes no sumário, com o qual é possível conhecer rapidamente o sentimento dos usuários sobre os aspectos do produto Samsung Smart TV.

```

Aspecto: Samsung Smart TV
  Sentenças Positivas: 16
  - A melhor tv na atualidade!
  Sentenças Negativas: 11
  - Produto queimou a tela 1 mês após a garantia.
Aspecto: Preço
  Sentenças Positivas: 0
  Sentenças Negativas: 2
  - O que não gostei: Valor muito alto
Aspecto: Durabilidade
  Sentenças Positivas: 0
  Sentenças Negativas: 2
  - Durabilidade 0
Aspecto: Câmera
  Sentenças Positivas: 1
  - A câmera embutida na tv com sensor de movimento e presença funciona bem e impressiona quem não conhece.
  Sentenças Negativas: 1
  - O que não gostei: A câmera poderia ter movimentação horizontal de visão e não só apenas vertical.
Aspecto: Qualidade da Imagem
  Sentenças Positivas: 1
  - Excelente aparelho de TV com ótima qualidade de imagem e recursos.
  Sentenças Negativas: 0

```

Figura 33: Sumário produzido pelo método Opizer-E

Algoritmo 5: ALGORITMO DE OPIZER-E

Entrada: Opiniões O , Aspectos A , Tamanho t

Saída: Sumário de Opiniões sum

início

$d \leftarrow$ Dicionário

para $s_i \in O$ **faça**

para $a_i \in A$ **faça**

$d[(a_i, '+')] \leftarrow$ Obter frequência '+' em s_i

$d[(a_i, '-')] \leftarrow$ Obter frequência '-' em s_i

fim

fim

$A' \leftarrow$ Ordenar(d)

$sum \leftarrow \emptyset$

para $a_i \in A'$ **faça**

$I \leftarrow$ Calcular a importância das sentenças $\in a_i$ com Equação 5.2

$sp \leftarrow$ Selecionar máxima $s_{a_i} \in I_+$ // Sentença positiva

$sn \leftarrow$ Selecionar máxima $s_{a_i} \in I_-$ // Sentença negativa

$sum \leftarrow sp + sn$

se $i = t$ **então**

 | Sair

fim

fim

fim

retorna sum

5.4.2 Proposta Abstrativa: Opizer-A

Os métodos abstrativos de sumarização de opiniões têm como objetivo sintetizar as informações mais relevantes presentes nas opiniões e refleti-las no texto do sumário. Neste trabalho, foram estudadas algumas abordagens para gerar sumários abstrativos, tais como técnicas de geração de linguagem natural (Radev e McKeown (1998), White et al. (2001)), compressão de sentenças (Zajic et al., 2007), fusão de sentenças (Barzilay e McKeown, 2005) e métodos baseados em *templates* (Harabagiu e Maiorano (2002), Jung e Jo (2003)). Depois do estudo, esta última abordagem foi selecionada para a implementação da proposta abstrativa deste trabalho de mestrado.

A escolha de um sistema abstrativo baseado em *templates* foi motivada pelas discussões no trabalho de Van Deemter et al. (2005) sobre a praticidade e a eficácia desta abordagem. Além disso, segundo Embar et al. (2013), o fato de os *templates* serem especificados manualmente provê uma vantagem em cenários em que se precisa de informações específicas.

Outra razão importante na escolha dessa abordagem foi o fato que, recentemente, outros trabalhos de sumarização de opiniões, como Gerani et al. (2014) e Fabbrizio et al. (2014), obtiveram resultados satisfatórios nas avaliações utilizando uma abordagem baseada em *templates*.

Neste trabalho, para a criação de sumários abstrativos de opiniões com base em aspectos, propõe-se o método Opizer-A (*Opinion Summarizer - Abstractive*), o qual é composto de duas etapas: agrupamento de segmentos textuais (qualificadores) e geração com base em *templates*.

Em geral, a principal diferença entre Opizer-A e os outros métodos abstrativos da literatura é que Opizer-A, além de utilizar as informações de polaridade dos aspectos mais importantes nos *templates*, utiliza também a informação dos qualificadores mais representativos dos aspectos. A seguir, as duas etapas de Opizer-A são explicadas com mais detalhes.

A. Agrupamento de Segmentos Textuais

Na sumarização automática de textos, um dos maiores desafios consiste em agrupar as informações mais similares dos textos-fonte, pois isso permite identificar os temas mais mencionados nos textos e, conseqüentemente, as informações mais representativas. Estas informações geralmente podem ser sentenças ou n-gramas dos textos. Dado que nesta abordagem devem ser gerados sumários abstrativos, optou-se por trabalhar com n-gramas ou segmentos textuais. Especificamente, foram selecionados os segmentos textuais que avaliam os aspectos, ou seja, os qualificadores, pois, nas opiniões, eles permitem conhecer, com mais precisão, o que os usuários falam sobre os aspectos. Por essa razão, esta etapa visa agrupar os segmentos textuais ou qualificadores mais similares e representativos de cada aspecto.

Como primeiro passo nesta etapa, todos os aspectos são ranqueados para determinar quais deles estarão presentes no sumário final. Para tal fim, foi utilizada a proposta de [Gerani et al. \(2014\)](#), a qual usa as relações RST entre os aspectos no ranqueamento. O principal motivo de utilizar essa proposta é que ela permite selecionar os aspectos mais relevantes considerando a sequência de ordem que eles deveriam ter no sumário para facilitar a leitura do texto.

Já com o ranqueamento dos aspectos calculado, os segmentos textuais (extraídos das anotações do corpus ReLi e Buscapé) são agrupados segundo os aspectos que contêm e a polaridade deles. Em outras palavras, para cada aspecto foram associados os segmentos textuais que os mencionavam em dois grupos de polaridade: positivo e negativo.

Em seguida, os segmentos textuais de cada aspecto e polaridade foram agrupados segundo seu conteúdo. Para tanto, foi utilizado o algoritmo K-Means ([Steinhaus, 1956](#)), com um valor k igual a 3 ou 4. Foi escolhido um valor pequeno, pois a ideia é garantir a seleção só dos principais tópicos/temas mencionados nos segmentos textuais. Ademais, acredita-se que não existem muitos tópicos/temas para um determinado aspecto nas opiniões. Por último, com esses valores para k , obtiveram-se os melhores resultados nos experimentos. No agrupamento com o algoritmo de K-Means, não foram consideradas as *stopwords*.

Finalmente, com os grupos ou *clusters* de cada aspecto e polaridade, foi selecionado o grupo que contivesse o maior número de elementos (segmentos textuais). Com isso,

pretendeu-se seleccionar a informação mais redundante e, portanto, mais importante. Do *cluster* escolhido, foi seleccionado o segmento textual com maior valor TF-IDF normalizado (ver Equação 5.3). Esse valor foi calculado utilizando os valores TF-IDF parciais das palavras que formavam o segmento textual, normalizado-se pelo seu tamanho (número de palavras). A ideia de normalizar era penalizar os segmentos textuais longos, pois isso alcança rapidamente o número de palavras permitido no sumário, reduzindo a possibilidade de colocar outras informações no texto.

$$SegmentoTextual = \max \left(\frac{\sum_{i=1}^n TF-IDF_i}{n} \right) \quad (5.3)$$

Da Equação 5.3, *SegmentoTextual* representa ao segmento textual seleccionado para conformar o sumário, n é o tamanho do segmento textual avaliado (número de palavras) e $TF-IDF_i$ representa o valor TF-IDF da i -ésima palavra.

B. Geração com base em *Templates*

A ideia geral desta etapa consiste em preencher alguns *templates* com os aspectos e segmentos textuais obtidos na etapa anterior e utilizá-los na produção dos sumários abstrativos. Os *templates* são estruturas linguísticas compostas de campos que permitem coletar as informações em função de um alvo específico (Jung e Jo, 2003).

No Opizer-A, os *templates* foram criados manualmente. Para determinar a estrutura e o conteúdo dos *templates*, foi feito um estudo dos trabalhos de Gerani et al. (2014) e Fabbri et al. (2014), os quais também utilizaram *templates* na geração de sumários de opiniões. Ademais, foram realizadas análises dos sumários humanos produzidos na anotação de OpiSums-PT, com o fim de extrair algumas formas de escrita comuns nos sumários. A partir desses estudos, foram criados *templates* para os principais tipos de avaliações, por exemplo, quando a maioria dos usuários avalia um aspecto da mesma forma ou quando há opiniões divididas.

Na Tabela 20, mostram-se os *templates* criados manualmente que são utilizados por Opizer-A. Como se pode observar, a Tabela 20 está dividida em cinco seções: primeira sentença, adjetivos com polaridade, expressões com verbos, conectores e explicadores. Os *templates* da seção “primeira sentença” definem a estrutura para o começo do sumário, onde os símbolos <> são os campos a serem preenchidos. Os *templates* das seções “adje-

tivos” e “expressões com verbos” permitem expressar um sentimento segundo a polaridade dos aspectos. Os “conectores” servem para ligar duas sentenças tendo em conta o sentimento delas. Finalmente, os “explicadores” dão suporte às sentenças por meio dos qualificadores. Mais adiante será explicada como é a seleção dos diferentes *templates*.

Primeira Sentença
"Controvérsia": ["As avaliações do <ASPECT> mostram um sentimento dividido", "O <ASPECT> tem opiniões positivas e negativas"...] "Polaridade": ["Em geral, as opiniões sobre o <ASPECT> foram", "A maioria das avaliações sobre o <ASPECT> são"...]
Adjetivos com Polaridade
"-2": ["muito negativas", "muito desfavoráveis"...] "-1": ["negativas", "desfavoráveis"...] "+1": ["positivas", "favoráveis"...] "+2": ["muito positivas", "muito favoráveis"...]
Expressões com Verbos
"Controvérsia": ["há opiniões controversas sobre ele", "expressaram-se opiniões controversas sobre essa característica"...] "-2": ["os usuários não gostaram disso", "foi avaliado como muito ruim", "as opiniões foram muito negativas"...] "-1": ["as pessoas não gostaram", "foi avaliado como ruim", "as opiniões foram negativas"...] "+1": ["os usuários gostaram disso", "as pessoas acharam que era satisfatório"...] "+2": ["as pessoas adoraram", "foi avaliado como excelente", "as opiniões foram excelentes"...]
Conectores
"Concordância": ["Da mesma forma", "Ao mesmo tempo", "De forma semelhante"...] "Não Concordância": ["Em contraste", "Contrariamente", "Ao contrário", "Porém"...] "Normal": ["Além disso", "Ademais"...] "Geral": ["Com relação ao <ASPECT>", "Em relação ao <ASPECT>", "falando sobre o <ASPECT>"...]
Explicadores
["pois <QUALIFICADOR>", "porque <QUALIFICADOR>"...]

Tabela 20: Templates utilizados por Opizer-A

Por exemplo, quando muitos usuários avaliam um aspecto da mesma forma, poderia se utilizar os *templates* “Em geral, as opiniões sobre o <ASPECT> foram” e “Pois <QUALIFICADOR>” da tabela anterior. Esses *templates* devem ser preenchidos com as informações relevantes obtidas na etapa anterior (agrupamento de segmentos textuais) para assim dar origem às sentenças que comporão o sumário final.

A seleção dos *templates* da Tabela 20 está baseada, principalmente, na polaridade dos aspectos. Para os *templates* das seções “primeira sentença”, “adjetivos com polaridade” e “expressões com verbos”, utilizou-se uma escala de -2 até 2 para determinar o sentimento dos aspectos, similar ao trabalho de Guzman e Maalej (2014) e à ferramenta Semantria⁴, sendo que os valores -2,-1,1, e 2 significam muito negativo, negativo, positivo e muito positivo, respectivamente.

O cálculo do score do sentimento de um aspecto é determinado pela frequência dos seus

⁴Disponível em <https://semantria.com/>

qualificadores positivos ou negativos normalizada pelo o número total de qualificadores. Assim, os scores no intervalo $[0, 0.25)$ são interpretados como muito negativo, os scores no intervalo $[0.25, 0.40]$ são interpretados como negativo, os scores no intervalo $[0.60, 0.75]$ são interpretados como positivo e os scores no intervalo $(0.75, 1]$ são interpretados como muito positivo. Quando o score da polaridade está no intervalo $(0.40, 0.60)$, considera-se como controvérsia, que indica que existe uma quantidade semelhante de qualificadores positivos e negativos sobre um mesmo aspecto (ver Equação 5.4).

$$Sentimento_i = \begin{cases} -2 & Score_i = [0, 0.25) \\ -1 & Score_i = [0.25, 0.40] \\ Controvérsia & Score_i = \langle 0.40, 0.60 \rangle \\ +1 & Score_i = [0.60, 0.75] \\ +2 & Score_i = \langle 0.75, 1 \rangle \end{cases} \quad (5.4)$$

No caso dos *templates* da seção “conectores”, a seleção se baseia na concordância entre a polaridade dos aspectos presentes em sentenças contínuas. Para os *templates* da seção “explicadores”, a seleção é aleatória.

Já com os *templates* definidos e a polaridade dos aspectos calculada, o passo seguinte consiste em organizar o texto do sumário. Isso é conhecido como *microplanning* na área de Geração de Linguagem Natural (Reiter, 1994).

No Opizer-A, a tarefa inicial no *microplanning* consiste em estruturar a primeira sentença do sumário. Essa sentença é muito importante, pois ela é a responsável por indicar qual é o sentimento predominante para o principal aspecto, que geralmente é a entidade. Para isso, o *template* inicial a ser preenchido corresponde à seção “primeira sentença”. Em seguida, segundo a polaridade do aspecto, é escolhido um *template* da seção “adjetivos com polaridade”. Com esses dois *templates*, a primeira sentença do sumário está pronta.

A partir da segunda sentença, o processo se repete para as demais sentenças. Assim, para criar uma nova sentença no sumário, é necessário verificar a polaridade do aspecto presente na sentença anterior. Com base nisso, escolhe-se um *template* da seção “conectores”, para então selecionar um *template* da seção “expressões com verbos”. Finalmente, é selecionado um *template* da seção “explicadores”. Esse *template* é preenchido com o

segmento textual (qualificador) mais representativo identificado na etapa anterior.

Nesse último passo, foi definido um conjunto pequeno de regras para adicionar alguns verbos na parte inicial dos qualificadores, com a intenção de melhorar a estrutura da sentença. Por exemplo, dado o *template* (da seção “explicadores”) “Pois <QUALIFICADOR>” e o qualificador “o melhor da atualidade”, seria conveniente adicionar o verbo “é” antes do qualificador, assim teríamos “pois é o melhor da atualidade”. As regras definidas basicamente analisavam a etiqueta morfossintática da primeira palavra do qualificador. Para tal fim, foi utilizado o tagger proposto por [Fonseca e Rosa \(2013\)](#).

Na Figura 34, mostra-se um sumário produzido pelo método Opizer-A. Na primeira parte do texto, é mencionado o sentimento geral para o principal aspecto (Samsung Smart TV), e, em seguida, descreve-se o sentimento dos usuários para alguns aspectos desse produto. No total, são cinco os aspectos descritos no sumário (Samsung Smart TV, preço, design, câmera e qualidade da imagem). Porém, essa quantidade é configurável segundo o número de palavras desejado para o sumário (ver Algoritmo 6).

A maioria das avaliações sobre a Samsung Smart TV são favoráveis. Ao contrário, em relação ao preço, os usuários não gostaram disso, pois é alto. Em contraste, em relação ao design, foi avaliado como excelente, pois é sofisticação e modernidade. Ademais, com relação à câmera expressaram-se opiniões controversas sobre essa característica porque funciona bem e impressiona quem não conhece, mas poderia ter movimentação horizontal de visão e não só apenas vertical. Além disso, em relação à qualidade da imagem, as opiniões foram excelentes porque é ótima.

Figura 34: Sumário produzido pelo método Opizer-A

Em comparação com o resumo gerado pelo método de [Gerani et al. \(2014\)](#) (ver Figura 30), o sumário produzido pelo Opizer-A, além de indicar o sentimento dos aspectos, mostra também qual é o principal motivo pelo qual cada aspecto é considerado como positivo, negativo ou neutro.

Algoritmo 6: ALGORITMO DE OPIZER-A

Entrada: Opiniões O , Aspectos A , Tamanho t

Saída: Sumário de Opiniões sum

```
início
   $T \leftarrow \emptyset$ 
  para  $o_i \in O$  faça
     $R \leftarrow$  Identificar relações RST de  $o_i$ 
    para  $r_i \in R$  faça
      se  $r_i.contém(a_x, a_y)$  então
         $T \leftarrow (a_x, peso_{xy}, a_y)$  //  $peso_{xy}$  representa o peso entre dois aspectos
      fim
    fim
  fim
   $G \leftarrow$  Criar grafo( $T$ )
   $P \leftarrow$  PageRank( $G$ )
   $G' \leftarrow$  Extrair subgrafo com os  $t$  principais aspectos de  $P$ 
   $G'' \leftarrow$  Calcular o Maximum Spanning Tree de  $G'$  não dirigido
   $H \leftarrow \emptyset$ 
  para  $a_i \in G''.nós()$  faça
     $S \leftarrow$  Selecionar segmentos textuais de  $a_i$ 
     $C \leftarrow$  K-Means( $S$ )
     $H[(a_i, '+' )] \leftarrow$  Selecionar seg. textual com Equação 5.3 do  $\max(|c_x|) \in C_+$ 
     $H[(a_i, '-' )] \leftarrow$  Selecionar seg. textual com Equação 5.3 do  $\max(|c_y|) \in C_-$ 
  fim
   $sum \leftarrow$  Selecionar templates da primeira sentença para  $a_0$ 
  para  $a_{i \geq 1} \in G''.nós()$  faça
     $p \leftarrow$  Obter polaridade de  $a_i$ 
     $sum \leftarrow$  Selecionar templates de adjetivos com polaridade para  $(a_i, p)$ 
     $sum \leftarrow$  Selecionar templates de expressões com verbos para  $(a_i, p)$ 
     $sum \leftarrow$  Selecionar templates de conectores para  $a_i$ 
     $sum \leftarrow$  Selecionar templates de explicadores para  $a_i$ 
     $sum \leftarrow$  Selecionar segmento textual de  $H[(a_i, p)]$ 
  fim
fim
retorna  $sum$ 
```

5.5 Considerações Finais

Neste capítulo foram descritos quatro métodos clássicos da literatura, dois extrativos e dois abstrativos. Foram explicadas também as duas propostas deste trabalho de mestrado para gerar sumários automáticos de opiniões, chamadas Opizer-E (método extrativo) e Opizer-A (método abstrativo).

Por uma parte, em relação aos métodos extrativos, Opizer-E tem a vantagem de gerar

um sumário em formato estruturado, o qual permite ao leitor saber rapidamente qual é o sentimento predominante nas opiniões. Além disso, Opizer-E combina dois tipos de informações diferentes para gerar os sumários. Como se mostrará no próximo capítulo, estas duas informações ajudaram a Opizer-E a obter os melhores resultados.

Por outra parte, dos métodos abstrativos, o benefício mais relevante de Opizer-A em comparação aos outros métodos é que esta proposta, além de mostrar explicitamente o sentimento para os principais aspectos, mostra também a informação mais redundante sobre os aspectos presente nas opiniões fonte.

Experimentos e Resultados

Neste capítulo, serão apresentados os experimentos realizados e os resultados obtidos nas avaliações. Para isso, o capítulo foi dividido em duas seções. Na Seção 6.1, são descritos os métodos de avaliação utilizados para medir o desempenho dos sumarizadores implementados. Na Seção 6.2, são descritos e explicados os resultados obtidos nos experimentos realizados. Foram feitos experimentos para avaliar a informatividade, qualidade linguística e utilidade dos sumários gerados pelos seis métodos implementados. Adicionalmente, foram realizados também testes com o sumariador RSumm (Ribaldo et al., 2012) (abordagem tradicional, não orientada a opiniões).

6.1 Métodos de Avaliação

Os sumários produzidos automaticamente foram avaliados com métricas tradicionais na área de sumarização de textos, tanto em termo de informatividade, quanto em termo de qualidade linguística. Segundo Mani (2001a), os métodos de avaliação de sistemas de sumarização têm que avaliar a informatividade e a qualidade linguística dos sumários. Nesta pesquisa, além dessas duas métricas, foi avaliada também a utilidade dos sumários. Essa medida tem sido utilizada em vários trabalhos recentes de sumarização de opiniões

(Wang et al. (2013), Fabbri et al. (2014), Gerani et al. (2014)). A seguir, são explicados os três métodos de avaliação utilizados neste trabalho de pesquisa.

6.1.1 Avaliação da Informatividade

A informatividade tem como objetivo avaliar a quantidade de informação do resumo, isto é, avaliar a quantidade de informação dos textos-fonte que é preservada no resumo (Mani, 2001b). Para medir a informatividade dos sumários automáticos, uma das medidas mais conhecidas e utilizadas na tarefa de sumarização é a ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*) (Lin, 2004), a qual foi adotada neste trabalho. A ROUGE compara a quantidade de n-gramas (conjunto de palavras em sequência) que tem em comum um sumário produzido automaticamente e um ou mais sumários humanos, chamados sumários de referência.

Os resultados da ROUGE são fornecidos em termos de precisão, cobertura e medida-f. A precisão, cobertura e medida-f definidas pela ROUGE são mostradas na Equação 6.1, Equação 6.2 e Equação 6.3, respectivamente.

$$Precisão = \frac{\# \text{ de } n\text{-gramas em comum com o sumário humano}}{\# \text{ de } n\text{-gramas do sumário automático}} \quad (6.1)$$

$$Cobertura = \frac{\# \text{ de } n\text{-gramas em comum com o sumário humano}}{\# \text{ de } n\text{-gramas do sumário humano}} \quad (6.2)$$

$$Medida-F = 2 \times \frac{Precisão \times Cobertura}{Precisão + Cobertura} \quad (6.3)$$

A precisão indica quanta informação do sumário de referência (humano) está no sumário automático, enquanto a cobertura indica quanta informação do sumário de referência foi coberto pelo sumário automático. Tais medidas são complementares e, por isso, costuma-se calcular a medida-f, a qual representa a média harmônica entre a precisão e a cobertura.

Como foi dito anteriormente, a ROUGE é muito utilizada na avaliação de sumários automáticos, pois é rápida e facilmente aplicável. No entanto, a principal desvantagem dessa medida é que, como apenas avalia a correspondência de n-gramas, ignora todos os aspectos relacionados à qualidade linguística dos sumários. Para abordar essas questões, existem outros métodos que avaliam a qualidade linguística dos sumários.

6.1.2 Avaliação da Qualidade Linguística

A avaliação da qualidade requer a participação de assessores humanos para julgar a qualidade do sumário, considerando alguns critérios linguísticos. Na DUC¹ e na TAC², grandes conferências e competições de sumarização, os sumários são avaliados considerando cinco critérios linguísticos: gramaticalidade, não redundância, clareza referencial, foco e estrutura, e coerência (Dang, 2005).

Este tipo de avaliação é realizada manualmente por humanos, que pontuam a qualidade do resumo, considerando estes critérios em uma escala de cinco níveis. No Apêndice A, é mostrado o questionário para avaliar os sumários descrevendo estes cinco critérios. A avaliação da qualidade não precisa fazer uma comparação com os resumos de referência (sumários feitos por humanos), diferentemente do que usualmente ocorre na avaliação de informatividade.

Neste trabalho, 26 pessoas, com conhecimento em PLN, participaram na avaliação da qualidade linguística dos diferentes sumários. Para isso, foi utilizado o questionário mostrado no Apêndice A. A avaliação foi online, sendo que os sumários e os questionários foram alocados no serviço Google Forms³.

Dessa forma, a avaliação dos critérios linguísticos ocorreu conforme uma escala de 1 a 5 pontos, em que “5” indica que o sumário é excelente com relação à propriedade em questão, “1” indica que o sumário é péssimo, e os pontos de “2” a “4” representam a gradação entre os dois extremos. A seguir, mostra-se a escala utilizada.

1. Péssimo
2. Ruim
3. Regular
4. Bom
5. Excelente

Embora esta medida não utilize sumários de referência, ela pode beneficiar aos sumários automáticos que sejam bastante diferentes dos sumários de referência. Porém, esta me-

¹ *Document Understanding Conference* <http://duc.nist.gov/pubs.html>

² *Text Analysis Conference* <http://www.nist.gov/tac/>

³ Disponível em <https://www.google.com/forms/>

didada não avalia quão útil é um sumário. A seguir, explica-se a medida de utilidade dos sumários adotada neste trabalho.

6.1.3 Avaliação da Utilidade

Este critério avalia a utilidade dos sumários de opiniões para os leitores. Em particular, o leitor deve avaliar o quão útil é o sumário em uma situação em que vai decidir se deve comprar um item. Dado que foi utilizado o *córpus OpiSums-PT*, um item pode ser um livro ou um produto eletrônico. Por exemplo, para avaliar a utilidade de um sumário sobre o produto Samsung Smart TV, foi feita a pergunta *“Se você deseja comprar o produto Samsung Smart TV, como você avalia a utilidade do sumário?”*

Para responder esta pergunta, foi utilizada também a escala descrita anteriormente, na qual uma pontuação de “5” indica que o sumário é muito útil (excelente), “1” indica que o sumário não é útil (péssimo), e os pontos de “2” a “4” representam a gradação entre os dois extremos.

A avaliação da utilidade dos sumários de opiniões ocorreu conjuntamente com a avaliação da qualidade linguística, por meio do serviço Google Forms, e também participaram 26 pessoas nesta avaliação.

6.2 Resultados

Como se mostrou no Capítulo 5, foram implementados e adaptados quatro métodos clássicos de sumarização de opiniões com base em aspectos para a língua portuguesa do Brasil. Estes métodos foram analisados em cada uma de suas fases. Como consequência dessa análise, produziram-se duas propostas para gerar sumários de opiniões. Estas duas propostas tentam utilizar as principais vantagens dos métodos clássicos para gerar melhores sumários. Além das avaliações aos seis métodos automáticos (os quatro métodos clássicos e os dois propostos), foram realizados também experimentos com o sumarizador RSumm.

O RSumm é um sumarizador multidocumento com base em grafos desenvolvido por Ribaldo et al. (2012). Para sumarizar, o RSumm cria um grafo multidocumento, onde cada sentença dos textos é representada por um nó no grafo e as arestas entre os nós indicam

quão semelhantes as sentenças são. Essa semelhança é calculada utilizando a medida do cosseno (Salton, 1989). Para selecionar o conteúdo do sumário, o RSumm oferece duas opções: Caminho denso e Caminho profundo. No caminho denso, a densidade de um nó é dada pela quantidade de conexões que tem com os demais nós. Dessa forma, o caminho é construído selecionando-se os vértices mais densos e considerando sua ordem no documento. O caminho profundo começa selecionando o vértice com maior densidade, para então selecionar os filhos que têm mais ligações. Para os experimentos, foi utilizado o caminho denso, pois é o que apresenta melhores resultados segundo as avaliações feitas por Ribaldo et al. (2012).

A fim de verificar a primeira hipótese deste trabalho, a qual indica que os métodos automáticos de sumarização tradicional têm um pior desempenho na sumarização de opiniões, também foram feitos experimentos utilizando as três medidas de avaliação: informatividade, qualidade linguística e utilidade do sumário.

Como será explicado nas seguintes seções, em geral, o RSumm, o qual é considerado o melhor método de sumarização tradicional para o Português Brasileiro, obteve os resultados mais baixos na sumarização de opiniões nas três medidas de avaliação utilizadas. Com base nesses resultados, a primeira hipótese desta pesquisa de mestrado foi comprovada. A seguir, são descritos os resultados obtidos nos três tipos de avaliações.

6.2.1 Resultados da Avaliação da Informatividade

Para a avaliação da informatividade, foi utilizada a medida ROUGE considerando os sumários do corpus OpiSums-PT como sumários de referência. Nos experimentos, tentou-se gerar sumários automáticos de um tamanho próximo a 100 palavras, já que os sumários de OpiSums-PT contêm aproximadamente essa quantidade de palavras e o objetivo era fazer uma comparação mais justa. Para esse fim, cada método automático implementado começava gerando um sumário com um número pequeno de palavras e ia-se incrementando essa quantidade até o texto conter uma quantidade próxima a 100 palavras. Cada sumário automático produzido foi comparado com os cinco sumários extrativos ou abstrativos de OpiSums-PT para o produto eletrônico ou livro em questão, para depois calcular a média de seus resultados.

Os sumários automáticos produzidos por métodos extrativos foram comparados com os

sumários extrativos humanos de OpiSums-PT e, do mesmo modo, os sumários abstrativos automáticos foram comparados com os sumários abstrativos humanos de OpiSums-PT. Optou-se por essa estratégia, pois, a geração destes dois tipos de sumários envolve diferentes processos.

Na Tabela 21, mostra-se os resultados dos métodos extrativos para a medida ROUGE-1, ROUGE-2 e ROUGE-L nos sumários de produtos eletrônicos e livros, considerando os sumários extrativos do corpus OpiSums-PT como sumários de referência. A ROUGE-1 e a ROUGE-2 são calculadas verificando a quantidade de unigramas e bigramas, respectivamente, comuns entre os sumários automáticos e os sumários de referência. De forma semelhante, a ROUGE-L, baseada no *Logest Common Subsequence* (Hirschberg, 1977), procura as maiores subcadeias comuns entre os dois tipos de sumários, executando portanto, uma avaliação similar à coocorrência de n-gramas.

Métodos		ROUGE-1			ROUGE-2			ROUGE-L		
		Precisão	Cobertura	Medida-F	Precisão	Cobertura	Medida-F	Precisão	Cobertura	Medida-F
Produtos	RSumm	0.273	0.284	0.278	0.154	0.156	0.154	0.254	0.263	0.257
	Hu e Liu (2004)	0.373	0.332	0.350	0.203	0.196	0.198	0.333	0.324	0.325
	Tadano et al. (2010)	0.341	0.335	0.337	0.217	0.218	0.217	0.330	0.325	0.327
	Opizer-E	0.374	0.351	0.362	0.255	0.238	0.246	0.366	0.344	0.354
Livros	RSumm	0.316	0.293	0.302	0.169	0.151	0.158	0.291	0.268	0.277
	Hu e Liu (2004)	0.382	0.384	0.380	0.217	0.216	0.216	0.345	0.342	0.342
	Tadano et al. (2010)	0.400	0.396	0.395	0.262	0.258	0.258	0.385	0.381	0.381
	Opizer-E	0.399	0.404	0.399	0.271	0.275	0.271	0.388	0.392	0.388

Tabela 21: Resultados da ROUGE dos métodos extrativos

Os resultados obtidos mostram que os valores mais altos da medida-f foram obtidos pelo método Opizer-E, tanto nas medidas ROUGE-1 (0.362 e 0.399), ROUGE-2 (0.246 e 0.271) quanto na ROUGE-L (0.354 e 0.388), os quais indicam que a seleção dos aspectos e das sentenças que compõem o sumário foram apropriadas. Acredita-se que as informações da posição da sentença e da proximidade dos qualificadores, utilizadas no Opizer-E, ajudaram a descartar as sentenças não relacionadas aos aspectos e também as sentenças sem polaridade. Por outro lado, como se observa também na Tabela 21, o método RSumm teve o desempenho mais baixo. O principal problema identificado nos sumários produzidos por esse método é que muitas sentenças não tem relação com os aspectos (contrariamente ao Opizer-E).

Os outros métodos de sumarização de opiniões (Hu e Liu (2004) e Tadano et al. (2010))

obtiveram resultados próximos a Opizer-E. No caso do método de [Tadano et al. \(2010\)](#), o fato de selecionar a polaridade do sumário em função da quantidade de estrelas atribuídas à opinião, afetou seu desempenho, pois nem sempre o número de estrelas da opinião reflete o que está escrito no texto.

Na Tabela 22, são apresentados os resultados dos métodos abstrativos implementados neste trabalho, considerando como sumários de referência os sumários abstrativos do corpus OpiSums-PT. Nessa tabela, evidencia-se que os valores da ROUGE-1, ROUGE-2 e ROUGE-L foram muito inferiores aos resultados obtidos pelos métodos extrativos, pois essas abordagens abstrativas não selecionam sentenças completas, mas criam um texto novo a partir das informações das sentenças.

Métodos		ROUGE-1			ROUGE-2			ROUGE-L		
		Precisão	Cobertura	Medida-F	Precisão	Cobertura	Medida-F	Precisão	Cobertura	Medida-F
Produtos	Ganesan et al. (2010)	0.144	0.232	0.178	0.031	0.051	0.038	0.121	0.196	0.150
	Gerani et al. (2014)	0.173	0.184	0.178	0.036	0.039	0.037	0.112	0.119	0.116
	Opizer-A	0.217	0.240	0.228	0.040	0.045	0.043	0.144	0.159	0.151
Livros	Ganesan et al. (2010)	0.120	0.309	0.168	0.030	0.076	0.042	0.106	0.276	0.149
	Gerani et al. (2014)	0.128	0.204	0.151	0.026	0.040	0.031	0.100	0.163	0.119
	Opizer-A	0.136	0.293	0.177	0.032	0.068	0.042	0.112	0.246	0.147

Tabela 22: Resultados da ROUGE dos métodos abstrativos

Como se observa na tabela anterior, o método de [Gerani et al. \(2014\)](#) teve o desempenho mais baixo, possivelmente devido à rigidez dos *templates* utilizados nesse método. Esses *templates* foram definidos manualmente e não compartilham muitas palavras comuns com as opiniões fonte, o que é penalizado fortemente pela medida ROUGE.

Ainda considerando a Tabela 22, observa-se que Opizer-A conseguiu resultados ligeiramente superiores aos demais métodos utilizando a medida ROUGE-1 e ROUGE-2. Acredita-se que isso aconteceu porque o Opizer-A coloca no sumário os qualificadores dos aspectos (segmentos textuais das sentenças). No entanto, o método de [Ganesan et al. \(2010\)](#) obteve o melhor desempenho para a ROUGE-L. Para gerar bons sumários abstrativos, o método de [Ganesan et al. \(2010\)](#) precisa de opiniões fonte altamente redundantes, caso contrário, os resumos gerados serão mais extrativos do que abstrativos. No OpiSums-PT, as opiniões fonte não são muito redundantes. Devido a isso, na maioria dos casos, os resumos gerados por [Ganesan et al. \(2010\)](#) foram formados por sentenças completas, fazendo com que a probabilidade de encontrar subcadeias maiores comuns entre os sumários automáticos e

os sumários de referência, fosse alta, permitindo assim obter uma boa medida ROUGE-L.

Por último, na Tabela 23, apresenta-se o resultado geral de todos os métodos implementados neste trabalho (a média dos resultados dos produtos eletrônicos e livros). É importante ressaltar o bom desempenho dos dois métodos propostos nesta pesquisa. O método Opizer-E obteve os melhores resultados entre todos os métodos. No caso de Opizer-A, embora não tenha superado em todas as medidas aos demais métodos abstrativos, alcançou resultados bem próximos do melhor desempenho do método proposto por Ganesan et al. (2010).

Métodos	ROUGE-1			ROUGE-2			ROUGE-L		
	Precisão	Cobertura	Medida-F	Precisão	Cobertura	Medida-F	Precisão	Cobertura	Medida-F
RSumm	0.306	0.291	0.296	0.165	0.152	0.157	0.282	0.267	0.273
Hu e Liu (2004)	0.380	0.372	0.373	0.214	0.211	0.212	0.342	0.337	0.338
Tadano et al. (2010)	0.386	0.382	0.382	0.251	0.249	0.249	0.372	0.368	0.368
Opizer-E	0.393	0.392	0.390	0.267	0.266	0.265	0.383	0.381	0.380
Ganesan et al. (2010)	0.125	0.291	0.170	0.030	0.070	0.041	0.110	0.257	0.149
Gerani et al. (2014)	0.138	0.199	0.158	0.029	0.040	0.032	0.103	0.153	0.119
Opizer-A	0.155	0.280	0.189	0.034	0.062	0.042	0.119	0.225	0.148

Tabela 23: Resultados da ROUGE dos métodos implementados

Como os resultados da Tabela 23 mostram o desempenho geral dos métodos implementados, aplicou-se o teste *t-Student* com 95% de confiança a esses dados a fim de verificar se havia alguma diferença estatisticamente significativa entre os desempenhos. Como esperado nos métodos extrativos, a diferença da Medida-F na ROUGE-1, ROUGE-2, ROUGE-L não é significativa entre os métodos de sumarização de opiniões com base em aspectos. Porém, a diferença dos resultados do RSumm (método tradicional) e o Opizer-E (método de sumarização de opiniões com base em aspectos) é estatisticamente significativa. O teste *t-Student* também foi aplicado para os métodos abstrativos. Por um lado, a diferença entre os métodos, na medida ROUGE-1 e ROUGE-2, é estatisticamente irrelevante. Por outro lado, para a medida ROUGE-L, existe diferença estatisticamente significativa entre Opizer-A e Gerani et al. (2014).

6.2.2 Resultados da Avaliação da Qualidade Linguística

Como foi dito anteriormente, 26 pessoas participaram na avaliação da qualidade linguística dos sumários. Para isso, elas tinham que ler os sumários produzidos e avaliá-los

utilizando os cinco critérios da DUC (ver Apêndice A). Devido ao fato de ser uma avaliação manual e custosa, em termos de tempo e complexidade, não foi possível utilizar todos os livros e produtos eletrônicos do corpus OpiSums-PT. Em vez disso, foi selecionada uma amostra deles.

A amostra continha sumários de três produtos eletrônicos (“Iphone 5”, “Samsung Galaxy S III”, “Samsung Smart TV”) e três livros (“Crepúsculo”, “Ensaio sobre a Cegueira”, “Fala sério, professor!”). A seleção desses itens foi baseada no desempenho que obtiveram os sumarizadores na avaliação da informatividade. Isto é, foram selecionados os produtos eletrônicos e livros nos quais os sumarizadores obtiveram o melhor, pior e desempenho médio na avaliação da informatividade. A ideia com isso era ter uma amostra com características diferentes.

Em relação aos métodos, foram avaliados os sumários produzidos por três métodos automáticos: RSumm, Opizer-E e Opizer-A. RSumm foi selecionado para esta avaliação porque é considerado o melhor método de sumarização tradicional para o Português Brasileiro. Opizer-E e Opizer-A foram escolhidos porque obtiveram os melhores resultados na avaliação da informatividade entre os métodos extrativos e abstrativos, respectivamente. Não foi possível avaliar todos os métodos implementados, pois se trata de uma avaliação manual, custosa em tempo e recursos. Além disso, foram avaliados também alguns sumários humanos abstrativos produzidos na anotação do OpiSums-PT. A seleção desses sumários foi aleatória e a intenção de avaliá-los era saber quão longe estão os sumários automáticos dos sumários humanos em relação à qualidade linguística.

Na Tabela 24, são mostrados os resultados obtidos pelos quatro métodos mencionados anteriormente para o conjunto de produtos eletrônicos e livros na avaliação da qualidade linguística. Dado que os sumários humanos avaliados obtiveram sempre o melhor julgamento, resalta-se em **negrito** os melhores resultados dos métodos automáticos para cada critério da DUC.

Como se pode observar nessa tabela, em relação à gramaticalidade, o método com melhores resultados foi o Opizer-E, tanto no domínio dos livros quanto dos produtos eletrônicos. Dado que o Opizer-E extrai as sentenças completas das opiniões fonte, a qualidade da gramaticalidade poderia ser afetada só quando as sentenças originais contivessem erros gramaticais (letras maiúsculas incorretas e conjugações dos verbos erradas,

Métodos		Gramaticalidade	Não Redundância	Clareza Referencial	Foco	Estrutura e Coerência
Produtos	RSumm	3.910	4.423	4.115	3.475	3.616
	Opizer-E	3.923	3.936	4.167	3.936	3.449
	Opizer-A	3.077	3.961	4.077	4.128	3.141
	Humano	4.731	4.744	4.808	4.782	4.731
Livros	RSumm	3.090	4.115	3.718	2.948	2.859
	Opizer-E	3.603	3.526	3.923	3.564	3.398
	Opizer-A	3.205	3.910	3.885	3.692	3.090
	Humano	4.564	4.641	4.718	4.628	4.590

Tabela 24: Resultados da avaliação da QL por produtos e livros

entre outras). Nesse sentido, os poucos erros gramaticais nas sentenças dos sumários favoreceram ao Opizer-E. Por outro lado, o método Opizer-A obteve os piores resultados. Um dos principais erros deste método reportados nos experimentos foi a presença de algumas frases (n-gramas) agramaticais na concatenação das partes fixas dos *templates* e as partes preenchidas. Por exemplo, a sentença “Ao mesmo tempo, em relação à bateria, as opiniões foram excelentes, pois é durabilidade impressionante” apresenta um erro em “pois é durabilidade impressionante”. Nesse caso, não foi escolhido o verbo correto para ligar as duas partes, o que tornou difícil a leitura do sumário. Outro erro identificado no Opizer-A foi a utilização do n-grama “devido a que”, que foi utilizado no contexto de explicar um motivo (similar a “porque” ou “pois”). Embora o leitor pudesse entender a ideia desse n-grama, a qualidade da gramaticalidade do sumário foi penalizada. Na avaliação da redundância, o sumarizador RSumm foi o melhor (4.423 e 4.115). Em relação aos métodos Opizer-E e Opizer-A, o método abstrativo gerou sumários menos redundantes. O Opizer-A foi penalizado, em alguns casos, pela utilização dos mesmos marcadores discursivos (“por exemplo”, “além disso”, “ademais” e “porém”, entre outros) em um sumário. O Opizer-E, por sua parte, gerou alguns sumários com sentenças repetidas. Esses casos aconteceram quando dois aspectos estavam presentes na mesma sentença e essa sentença era a única para esses dois aspectos. Assim, como não existiam outras sentenças para os dois aspectos, a sentença única se repetia em ambos os aspectos, afetando o desempenho de Opizer-E. Com respeito ao critério da clareza referencial, os resultados obtidos pelos métodos foram os mais próximos em comparação aos demais critérios. Para o domínio dos

produtos eletrônicos, os métodos automáticos obtiveram 4.115, 4.167 e 4.077 em média, e, para o domínio dos livros, os resultados foram 3.718, 3.923 e 3.885.

Da Tabela 24, o Opizer-A obteve os melhores resultados para o critério do foco do sumário. A principal desvantagem dos métodos extrativos, neste caso, Opizer-E e RSumm, é que eles selecionam sentenças completas e as colocam no sumário. Essas sentenças podem conter vários e diferentes temas, os quais desviam o texto do foco principal do sumário. Por outro lado, no Opizer-A, só são selecionadas algumas partes das sentenças (n-gramas), as quais são orientadas a um foco principal.

Com relação ao critério da estrutura e coerência, o método RSumm e Opizer-E obtiveram os melhores resultados no domínio de produtos e livros, respectivamente. No caso de Opizer-E, o fato de organizar o sumário em aspectos e por sentenças positivas e negativas ajudou a gerar um sumário melhor estruturado e coerente.

Na Tabela 25, mostra-se o desempenho geral dos métodos implementados. Esses resultados representam os valores médios dos resultados no domínio de livros e produtos eletrônicos (ver Tabela 24). Em geral, os sumários humanos obtiveram os melhores julgamentos em todos os critérios linguísticos avaliados, mostrando assim que os métodos automáticos de sumarização de opiniões implementados neste trabalho ainda podem ser melhorados. Acredita-se que, com a identificação dos erros explicados anteriormente, pode-se aprimorar esses métodos, o que constitui um dos trabalhos futuros desta pesquisa.

Métodos	Gramaticalidade	Não Redundância	Clareza Referencial	Foco	Estrutura e Coerência
RSumm	3.500	4.269	3.917	3.212	3.237
Opizer-E	3.763	3.731	4.045	3.750	3.423
Opizer-A	3.141	3.936	3.981	3.910	3.115
Humano	4.647	4.692	4.763	4.705	4.660

Tabela 25: Resultados da avaliação da QL dos métodos implementados

Por fim, a segunda hipótese deste trabalho indica que os resumos gerados pelos métodos abstrativos de sumarização de opiniões são mais concisos e menos redundantes que os sumários extrativos. Como se pode observar na Tabela 25, o método Opizer-A (método abstrativo) obteve melhor resultado do que o Opizer-E (método extrativo) no critério da não redundância. Assim, demonstrou-se que um sumário gerado por um método abstrativo

é menos redundante e mais conciso que um sumário extrativo, e com isso, foi validada a segunda hipótese deste trabalho.

6.2.3 Resultados da Avaliação da Utilidade

Com se explicou anteriormente, 26 pessoas participaram na avaliação da utilidade dos sumários. Para tal fim, os avaliadores tinham que ler os sumários e responder à pergunta, por exemplo, “*Se você deseja comprar o produto Samsung Smart TV, como você avalia a utilidade do sumário?*” (no caso de um sumário do produto Samsung Smart TV).

Na Tabela 26, são mostrados os resultados obtidos nos experimentos, tanto para os sumários dos produtos eletrônicos, quanto para os sumários dos livros. Mostra-se também nessa tabela, a média geral desses dois itens.

Métodos	Produtos	Livros	Média
RSumm	2.974	2.705	2.840
Opizer-E	3.244	3.103	3.173
Opizer-A	3.474	2.667	3.070
Humano	4.615	4.462	4.538

Tabela 26: Resultados da avaliação da utilidade dos métodos implementados

Como se pode observar, o desempenho de RSumm foi inferior aos dois sistemas, Opizer-A e Opizer-E. Isso ocorre, possivelmente, devido ao fato de que o RSumm não é um método orientado à sumarização de opiniões. Assim, o RSumm poderia ter selecionado para o resumo sentenças que não tivessem relação com os aspectos ou os sentimentos. Além disso, o RSumm não considera o “*lado quantitativo*” de um sumário, o qual se refere à porcentagem ou número de opiniões positivas ou negativas dos aspectos ou entidades, o que é muito útil no processo de tomada de decisão em uma compra.

O Opizer-A obteve o melhor resultado (3.474) nos sumários de produtos, mas, nos resumos de livros, o desempenho foi mais baixo (2.667). Isso aconteceu, principalmente, porque as opiniões de livros contêm poucos aspectos em relação às opiniões de produtos eletrônicos. Com poucos aspectos, o tamanho do sumário gerado pelo Opizer-A fica pequeno, pois existem poucas informações (aspectos) para produzir o texto. Contrariamente, com mais aspectos, é possível gerar sumários mais próximos ao tamanho desejado

do texto e, portanto, podem-se aproveitar melhor as informações presentes nas opiniões para gerar um sumário mais informativo e útil.

Em geral, o Opizer-E teve o melhor desempenho em comparação aos outros métodos automáticos. Este método agrupa as sentenças por aspecto, por polaridade e também indica a quantidade de sentenças positivas e negativas. Essa informação é muito útil antes de realizar uma compra, pois permite conhecer qual é o sentimento geral em relação aos aspectos e entidades.

Finalmente, a terceira hipótese deste trabalho de mestrado indica que os usuários (leitores) preferem os sumários abstrativos em vez dos resumos extrativos, pois eles são mais úteis antes do processo de uma compra. Conforme os resultados da Tabela 26, não é possível concluir como verdadeira essa hipótese. Como foi indicado anteriormente, com os sumários extrativos é possível também indicar a quantidade de opiniões/sentenças positivas e negativas (*“lado quantitativo”* de um sumário). Tal informação desempenha um papel muito importante no processo de uma compra e ressalta a utilidade do sumário.

Considerações Finais

No presente trabalho de mestrado, apresentou-se, até onde se sabe, a primeira investigação de sumarização de opiniões com base em aspectos para o idioma Português Brasileiro. Foram implementados e adaptados quatro métodos clássicos de sumarização de opiniões para esta língua. Estes métodos foram analisados em cada uma de suas fases e como consequência dessa análise, produziram-se duas propostas para gerar sumários de opiniões. Estas duas propostas, chamadas Opizer-E (método extrativo) e Opizer-A (método abstrativo), utilizam as principais vantagens dos métodos clássicos com o fim de gerar melhores sumários.

Além das avaliações dos seis métodos automáticos (os quatro métodos clássicos e os dois métodos propostos), foram realizados também experimentos com o sumarizador RSumm ([Ribaldo et al., 2012](#)), abordagem tradicional não orientada a opiniões. Os testes foram feitos em relação a três medidas de avaliação: informatividade, qualidade linguística e utilidade do sumário.

Em geral, o método proposto Opizer-E obteve os melhores resultados nos diferentes experimentos. Apesar de o Opizer-A não ter alcançado o melhor desempenho em todas as avaliações, os resultados obtidos foram muito próximos à melhor performance.

Os resultados dessas avaliações mostram também que, com base nas três medidas

de avaliação utilizadas, as abordagens tradicionais não são ideais para gerar sumários de opiniões, comprovando-se a primeira hipótese deste trabalho. Outros resultados dos experimentos mostram que os sumários dos métodos extrativos de sumarização de opiniões têm melhor informatividade do que os dos métodos abstrativos. Porém, em relação ao critério da não redundância, as abordagens abstrativas obtiveram melhores desempenhos que as abordagens extrativas, comprovando-se, dessa maneira, a segunda hipótese. Com relação à utilidade dos sumários, os métodos extrativos foram melhores do que os métodos abstrativos, com o que não foi possível comprovar a terceira hipótese desta pesquisa.

Por outro lado, do ponto de vista comercial, ainda não há exemplos, em português brasileiro, de aplicações que fazem uso de sumários textuais de opiniões. Acredita-se que este tipo de sumários pode auxiliar possíveis estratégias de marketing (conhecer os pontos fortes e fracos de algum produto) e pode ajudar também os compradores a ter uma melhor informação para realizar a compra correta. Igualmente, pode ajudar a indústria a melhorar os produtos produzidos, pois, por meio dos sumários, se conheceria quais são as necessidades e problemas que têm os clientes. Com os experimentos realizados nesta pesquisa, tenta-se dar o primeiro passo para futuras aplicações desse tipo.

Diante desses experimentos realizados, discutem-se também neste capítulo algumas contribuições e limitações deste trabalho. Ademais, propõem-se alguns trabalhos futuros.

7.1 Contribuições

São várias as contribuições obtidas nesta pesquisa. Além de ser a primeira investigação de sumarização de opiniões para a língua portuguesa, este trabalho permitiu o desenvolvimento teórico e prático dessa linha de pesquisa. Entre as principais contribuições feitas neste trabalho de mestrado, pode-se citar:

- Em termos de recursos, foi criado o *córpus* OpiSums-PT, o primeiro *córpus* de sumários de opiniões extrativos e abstrativos escritos em Português do Brasil. Este *córpus* pode auxiliar futuras pesquisas na área de sumarização de opiniões ou em outras linhas relacionadas, como análise de sentimentos ou geração textual.
- Análise qualitativa do processo de criação de sumários manuais de opiniões. Até onde se tem conhecimento, independentemente da língua, é a primeira análise feita

em sumários humanos. Mais informações sobre a análise realizada podem ser encontradas em [López et al. \(2015\)](#).

- Desenvolvimento e adaptação de quatro métodos tradicionais de sumarização de opiniões baseados em aspectos para a língua portuguesa brasileira. As versões originais desses métodos foram focadas para o idioma Inglês.
- Implementação de duas propostas (Opizer-E e Opizer-A), para gerar sumários (extrativos e abstrativos) de opiniões com base em aspectos. É importante ressaltar que os dois métodos obtiveram resultados promissores, sendo que o Opizer-E foi o melhor método e o Opizer-A obteve resultados muito próximos ao melhor método abstrativo.
- Avaliação dos sumários automáticos produzidos em relação a três medidas de avaliação: informatividade, qualidade linguística e utilidade do texto. A maioria dos trabalhos de sumarização de opiniões com base em aspecto tem utilizado uma ou duas medidas de avaliação. Neste trabalho, foram utilizadas três medidas, que permitiu uma análise mais completa dos sumários produzidos.
- Disponibilização de demonstrações online¹ dos sumarizadores implementados neste trabalho. As demonstrações encontram-se localizadas na página do projeto PROS@, pois este trabalho de mestrado está inserido nesse projeto maior, denominado PROS@ - “Processamento Semântico de Textos em Português Brasileiro”, do convênio entre o ICMC-USP e a empresa SAMSUNG Eletrônica da Amazônia.
- Por último, como resultado das pesquisas realizadas neste mestrado, em termos de publicações, têm-se, até o momento, dois artigos publicados em eventos internacionais da área de Linguística Computacional como primeiro autor ([López et al. \(2015\)](#) e [López e Pardo \(2015\)](#)). Adicionalmente, como parte do trabalho com outros pesquisadores relacionados com esta investigação, publicaram-se dois artigos ([Agostini et al. \(2014\)](#) e [Dias et al. \(2014\)](#)).

¹Disponível em <http://nilc.icmc.usp.br/semanticnlp/opizer/sumarizadores/>

7.2 Limitações

Mesmo com as contribuições feitas neste trabalho, foram identificadas também algumas limitações. Uma das principais limitações nesta pesquisa está relacionada às características do *cópus* OpiSums-PT, que é, até onde se tem conhecimento, o único *cópus* de sumários de opiniões disponível para a língua portuguesa do Brasil. O principal problema no *cópus* OpiSums-PT é que, em várias opiniões de produtos eletrônicos e livros, há poucas sentenças que contêm aspectos. Assim, por exemplo, existem aspectos que só têm uma ou duas sentenças relacionadas a eles em todo o conjunto de opiniões. Nesse caso, gerar sumários para um aspecto a partir de uma ou duas sentenças não é o cenário ideal para avaliar métodos automáticos de sumarização.

Outra limitação desta pesquisa foi a falta de um extrator automático de aspectos e polaridade em opiniões, o qual teria permitido saber qual é o verdadeiro desempenho de um sumarizador de opiniões completamente automático. Tem-se conhecimento de algumas pesquisas em andamento nessa área, mas, até o momento em etapas iniciais, sem métodos disponíveis para extrair aspectos nem suas polaridades. Finalmente, outra limitação observada ao longo do desenvolvimento deste trabalho está relacionada ao parser Dizer, o qual foi utilizado no método de [Gerani et al. \(2014\)](#) e no Opizer-A. Este parser teve vários erros na identificação das relações RST (devido às informalidades das opiniões) e é muito custoso em termos de tempo de execução.

7.3 Trabalhos Futuros

A partir deste trabalho de mestrado, foi possível identificar alguns outros cenários a serem investigados no futuro. Entre os principais trabalhos futuros que se desprendem desta pesquisa de mestrado, encontram-se:

- Estender os métodos Opizer-E e Opizer-A, integrando-os com um extrator automático de aspectos. Assim, esses métodos não precisariam de anotações prévias de aspectos e polaridades, podendo ser facilmente aplicáveis a qualquer conjunto de opiniões.

- Outra possibilidade futura para a extensão deste trabalho seria o aprimoramento de Opizer-A. Entre as principais alternativas, destaca-se a utilização de ferramentas de fusão de sentenças, tal como Zíper ([Seno e Nunes, 2010](#)).
- A fim de investigar com maior ênfase os métodos propostos neste trabalho (Opizer-E e Opizer-A), um dos passos seguintes é realizar experimentos e avaliações desses métodos em opiniões escritas na língua inglesa.
- Finalmente, outra questão interessante como trabalho futuro consiste em aumentar o corpus OpiSums-PT. A ideia com isso é tentar resolver o problema dos poucos aspectos presentes nas opiniões (explicado na seção de limitações) e também para poder utilizá-lo em investigações de métodos de Aprendizado de Máquina.

Referências Bibliográficas

- Agostini, V.; López, R.; Pardo, T. A. S. (2014). Automatic Alignment of News Texts and Their Multi-document Summaries: Comparison among Methods. In J. Baptista; N. Mamede; S. Candeias; I. Paraboni; T. Pardo; e M. Volpe Nunes (Eds.), *Computational Processing of the Portuguese Language*, Volume 8775 of *Lecture Notes in Computer Science*, pp. 220–231. Springer International Publishing.
- Balage Filho, P.; Pardo, T.; Aluísio, S. (2013). An Evaluation of the Brazilian Portuguese LIWC Dictionary for Sentiment Analysis. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology (STIL)*, pp. 215–219. Sociedade Brasileira de Computação.
- Balage Filho, P. P.; Pardo, T. A.; Aluisio, S. M. (2013). An Evaluation of the Brazilian Portuguese LIWC Dictionary for Sentiment Analysis. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*, pp. 241–272.
- Barzilay, R.; McKeown, K. R. (2005). Sentence Fusion for Multidocument News Summarization. *Computational Linguistics* 31(3), pp. 297–328.
- Beineke, P.; Hastie, T.; Manning, C.; Vaithyanathan, S. (2003). An Exploration of Sentiment Summarization. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, Stanford, US, pp. 12–15.
- Belzer, J.; Holzman, A. G.; Kent, A. (1978). *Encyclopedia of Computer Science and*

Technology: Volume 9-Generative Epistemology of Problem Solving to Laplace and Geometric Transforms, Volume 9. CRC Press.

Berger, A. L.; Pietra, V. J. D.; Pietra, S. A. D. (1996). A Maximum Entropy Approach to Natural Language Processing. Volume 22, Cambridge, MA, USA, pp. 39–71. MIT Press.

Blair-Goldensohn, S.; Neylon, T.; Hannan, K.; Reis, G. A.; McDonald, R.; Reynar, J. (2008). Building a Sentiment Summarizer for Local Service Reviews. In *Workshop NLP in the Information Explosion Era*, pp. 14–23.

Boiy, E.; Moens, M. F. (2009). A Machine Learning Approach to Sentiment Analysis in Multilingual Web Texts. Volume 12, pp. 526–558. Springer Netherlands.

Burges, C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. Volume 2, pp. 121–167. Kluwer Academic Publishers.

Burges, C.; Shaked, T.; Renshaw, E.; Lazier, A.; Deeds, M.; Hamilton, N.; Hullender, G. (2005). Learning to Rank Using Gradient Descent. In *Proceedings of the 22nd International Conference on Machine Learning, ICML '05*, New York, NY, USA, pp. 89–96. ACM.

Cambria, E.; Livingstone, A.; Hussain, A. (2012). The Hourglass of Emotions. In *Proceedings of the 2011 International Conference on Cognitive Behavioural Systems, COST'11*, Berlin, Heidelberg, pp. 144–157. Springer-Verlag.

Cardoso, P.; Maziero, E.; Castro Jorge, M.; Seno, E.; Di Felippo, A.; Rino, L.; Nunes, M.; Pardo, T. (2011). A Discourse Annotated Corpus for Single and Multi-Document Summarization of News Texts in Brazilian Portuguese. In *Proceedings of the 3rd RST Brazilian Meeting*, pp. 88–105.

Cardoso, P. C.; Pardo, T. A.; Nunes, M. (2011). Métodos para Sumarização Automática Multidocumento Usando Modelos Semântico-Discursivos. In *Proceedings of the 3rd RST Brazilian Meeting, Cuiabá/MT, Brazil, October*, Volume 26, pp. 59–74.

- Carenini, G.; Cheung, J. C. K. (2008). Extractive vs. NLG-based Abstractive Summarization of Evaluative Text: The Effect of Corpus Controversiality. In *Proceedings of the Fifth International Natural Language Generation Conference*, Morristown, NJ, USA, pp. 33–41. Association for Computational Linguistics.
- Carenini, G.; Cheung, J. C. K.; Pauls, A. (2013). Multi-Document Summarization of Evaluative Text. *Computational Intelligence* 29(4), 545–576.
- Carenini, G.; Moore, J. D. (2006). Generating and Evaluating Evaluative Arguments. *Artificial Intelligence* 170(11), 925–952.
- Carenini, G.; Ng, R.; Pauls, A. (2006). Multi-document Summarization of Evaluative Text. In *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL)*, pp. 305–312.
- Carletta, J. (1996). Assessing Agreement on Classification Tasks: The Kappa Statistic. *Computational Linguistics* 22(2), 249–254.
- Carvalho, P.; Sarmiento, L.; Teixeira, J.; Silva, M. J. (2011). Liars and Saviors in a Sentiment Annotated Corpus of Comments to Political Debates. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*, HLT '11, Stroudsburg, PA, USA, pp. 564–568. Association for Computational Linguistics.
- Castro Jorge, M. L. d. R.; Pardo, T. A. S. (2010). Experiments with CST-based multidocument summarization. In *Proceedings of the 2010 Workshop on Graph-based Methods for Natural Language Processing*, TextGraphs-5, Stroudsburg, PA, USA, pp. 74–82. Association for Computational Linguistics.
- Charbonnier, S.; Portet, F. (2012). A Self-Tuning Adaptive Trend Extraction Method for Process Monitoring and Diagnosis. *Journal of Process Control* 22(6), pp. 1127 – 1138.
- Chaves, M. S.; Freitas, L. A.; Souza, M.; Vieira, R. (2012). PIRPO: An Algorithm to Deal with Polarity in Portuguese Online Reviews from the Accommodation Sector.

- In *Proceedings of the 17th International Conference on Applications of Natural Language Processing and Information Systems*, NLDB'12, Berlin, Heidelberg, pp. 296–301. Springer-Verlag.
- Conrad, J. G.; Leidner, J. L.; Schilder, F.; Kondadadi, R. (2009). Query-based Opinion Summarization for Legal Blog Entries. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, pp. 167–176. ACM.
- Dang, H. T. (2005). Overview of DUC 2005. In *Proceedings of the Document Understanding Conference*.
- Di Fabbriozio, G.; Aker, A.; Gaizauskas, R. (2011). STARLET: Multi-document Summarization of Service and Product Reviews with Balanced Rating Distributions. In *Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops*, ICDMW '11, Washington, DC, USA, pp. 67–74. IEEE Computer Society.
- Dias, M.; Bokan, A.; Chuman, C.; Barros, C.; Maziero, E.; Nobrega, F.; Souza, J.; Sobrevilla, M.; Delege, M.; Castro, L.; Silva, N.; Figueira, P.; Balage, P.; López, R.; Felippo, A. D.; Graças Volpe, M.; Pardo, T. (2014). Enriquecendo o Córpus CSTNews - a Criação de Novos Sumários Multidocumento. In *Proceedings of the 1st Workshop on Tools and Resources for Automatically Processing Portuguese and Spanish - ToRPorEsp*, pp. 1–8.
- Ding, X.; Liu, B.; Yu, P. S. (2008). A Holistic Lexicon-based Approach to Opinion Mining. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, WSDM '08, New York, NY, USA, pp. 231–240. ACM.
- Drucker, H.; Burges, C. J. C.; Kaufman, L.; Smola, A.; Vapnik, V. (1997). Support Vector Regression Machines. In *Advances in Neural Information Processing Systems 9*, pp. 155–161. MIT Press.
- Embar, V. R.; Deshpande, S. R.; Vaishnavi, A. K.; Jain, V.; Kallimani, J. S. (2013). sArAmsha - A Kannada Abstractive Summarizer. In *Advances in Computing, Communications and Informatics (ICACCI)*, pp. 540–544.

- Erkan, G.; Radev, D. R. (2004). LexRank: Graph-based Lexical Centrality As Saliency in Text Summarization. *J. Artif. Int. Res.* 22(1), 457–479.
- Fabbrizio, G. D.; Stent, A.; Gaizauskas, R. (2014, June). A Hybrid Approach to Multi-document Summarization of Opinions in Reviews. In *Proceedings of the 8th International Natural Language Generation Conference (INLG)*, Philadelphia, Pennsylvania, U.S.A., pp. 54–63. Association for Computational Linguistics.
- Fonseca, E. R.; Rosa, J. L. G. (2013). Mac-Morpho Revisited: Towards Robust Part-of-Speech Tagging. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*, pp. 98–107.
- Forrester-Research (2011). Competitive Strategy In The Age Of The Customer. http://media2.bazaarvoice.com/documents/Bazaarvoice_Social_Commerce_Statistics_112612.pdf.
- Freitas, C.; Motta, E.; Milidiú, R.; Cesar, J. (2012). Vampiro que Brilha... rá! Desafios na Anotação de Opinião em um Corpus de Resenha de Livros. In *Proceedings do XI Encontro de Linguística de Corpus*, Instituto de Ciências Matemáticas e de Computação da USP, em São Carlos/SP.
- Freitas, C.; Motta, E.; Milidiú, R.; Cesar, J. (2013). Sparkle Vampire LoL! Annotating Opinions in a Book Review Corpus. In *11th Corpus Linguistics Conference*, pp. 128–146.
- Ganesan, K. (2013). Opinion Driven Decision Support System. PhD Thesis, University of Illinois at Urbana-Champaign, Champaign – USA.
- Ganesan, K.; Zhai, C.; Han, J. (2010). Opinosis: A Graph Based Approach to Abstractive Summarization of Highly Redundant Opinions. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, Beijing, China, pp. 340–348.
- Genkin, A.; Lewis, D. D.; Madigan, D. (2007). Large-Scale Bayesian Logistic Regression for Text Categorization. *Technometrics* 49, pp. 291–304.

- Gerani, S.; Mehdad, Y.; Carenini, G.; Ng, R. T.; Nejat, B. (2014). Abstractive Summarization of Product Reviews Using Discourse Structure. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1602–1613. Association for Computational Linguistics.
- Gotti, F.; Lapalme, G.; Québec, M.; Nerima, L.; Wehrli, E.; Genève, U. D. (2007). GOFAIsum: A Symbolic Summarizer for DUC. In *Proceedings of Document Understanding Workshop*.
- Guzman, E.; Maalej, W. (2014). How Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews. In *Requirements Engineering Conference (RE), 2014 IEEE 22nd International*, pp. 153–162.
- Hahn, U.; Mani, I. (2000). The Challenges of Automatic Summarization. *Computer* 33(11), 29–36.
- Harabagiu, A. M.; Maiorano, S. J. (2002). Multi-Document Summarization with GISTexter. In *Proceedings of International Conference on Language Resources and Evaluation*.
- Hirschberg, D. S. (1977). Algorithms for the Longest Common Subsequence Problem. *Journal of the ACM* 24(4), 664–675.
- Hu, M.; Liu, B. (2004). Mining and Summarizing Customer Reviews. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '04, New York, NY, USA*, pp. 168–177. ACM.
- Jaccard, P. (1901). Distribution de la Flore Alpine Dans le Bassin des Dranses et Dans Quelques Régions Voisines. In *Bulletin de la Société Vaudoise des Sciences Naturelles*, pp. 241–272.
- Jiang, L.; Yu, M.; Zhou, M.; Liu, X.; Zhao, T. (2011). Target-dependent Twitter Sentiment Classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, HLT '11, Stroudsburg, PA, USA*, pp. 151–160. Association for Computational Linguistics.

- Jindal, N.; Liu, B. (2006a). Identifying Comparative Sentences in Text Documents. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '06, New York, NY, USA, pp. 244–251. ACM.
- Jindal, N.; Liu, B. (2006b). Mining Comparative Sentences and Relations. In *Proceedings of the 21st National Conference on Artificial Intelligence*, pp. 1331–1336. AAAI Press.
- Jung, J.; Jo, G. S. (2003). Template-Based E-mail Summarization for Wireless Devices. In A. Yazıcı e C. Şener (Eds.), *Computer and Information Sciences - ISCIS 2003*, Volume 2869 of *Lecture Notes in Computer Science*, pp. 99–106. Springer Berlin Heidelberg.
- Kim, H. D.; Ganesan, K.; Sondhi, P.; Zhai, C. (2011). Comprehensive Review Of Opinion Summarization. Relatório técnico, University of Illinois at Urbana-Champaign.
- Kim, H. D.; Zhai, C. (2009). Generating Comparative Summaries of Contradictory Opinions in Text. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, New York, NY, USA, pp. 385–394. ACM.
- Kim, S.-B.; Rim, H.-C.; Yook, D.; Lim, H. (2002). Effective Methods for Improving Naive Bayes Text Classifiers. In *Proceedings of the 7th Pacific Rim International Conference on Artificial Intelligence: Trends in Artificial Intelligence*, PRICAI '02, London, UK, UK, pp. 414–423. Springer-Verlag.
- Ku, L. W.; Lee, L. Y.; Chen, H. H. (2006). Opinion Extraction, Summarization and Tracking in News and Blog Corpora. In *Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs*, pp. 100–107.
- Kullback, S.; Leibler, R. A. (1951). On Information and Sufficiency. *Annals of Mathematical Statistics* (1), pp. 79–86.
- Kupiec, J.; Pedersen, J.; Chen, F. (1995). A Trainable Document Summarizer. In *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '95, New York, NY, USA, pp. 68–73. ACM.

- Labbé, C.; Labbé, D. (2005). How to Measure the Meanings of Words? Amour in Corneille's Work. *Language Resources and Evaluation* 39(4), pp. 335–351.
- Labbé, C.; Portet, F. (2012). Towards an Abstractive Opinion Summarisation of Multiple Reviews in the Tourism Domain. In *SDAD 2012 The 1st International Workshop on Sentiment Discovery from Affective Data*, pp. 87–94.
- Lerman, K.; Blair-Goldensohn, S.; McDonald, R. (2009). Sentiment Summarization: Evaluating and Learning User Preferences. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '09, Stroudsburg, PA, USA, pp. 514–522. Association for Computational Linguistics.
- Lerman, K.; McDonald, R. (2009). Contrastive Summarization: An Experiment with Consumer Reviews. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion: Short Papers*, NAACL-Short '09, Stroudsburg, PA, USA, pp. 113–116. Association for Computational Linguistics.
- Li, T.; Ding, C. (2006). The Relationships Among Various Nonnegative Matrix Factorization Methods for Clustering. In *Proceedings of the Sixth International Conference on Data Mining*, ICDM '06, Washington, DC, USA, pp. 362–371. IEEE Computer Society.
- Lin, C.; He, Y. (2009). Joint Sentiment/Topic Model for Sentiment Analysis. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, New York, NY, USA, pp. 375–384. ACM.
- Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, Barcelona, Spain, pp. 74–81. Association for Computational Linguistics.
- Liu, B. (2011). *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. Data-Centric Systems and Applications. Springer.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Synthesis Digital Library of Engineering and Computer Science. Morgan & Claypool.

- Liu, B.; Hu, M.; Cheng, J. (2005). Opinion Observer: Analyzing and Comparing Opinions on the Web. In *Proceedings of the 14th International Conference on World Wide Web, WWW '05*, New York, NY, USA, pp. 342–351. ACM.
- Liu, F.; Liu, Y. (2008). What Are Meeting Summaries?: An Analysis of Human Extractive Summaries in Meeting Corpus. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, pp. 80–83. Association for Computational Linguistics.
- Lloret, E.; Balahur, A.; Gómez, J.; Montoyo, A.; Palomar, M. (2012). Towards a Unified Framework for Opinion Retrieval, Mining and Summarization. *Journal of Intelligent Information Systems* 39(3), 711–747.
- Lloret, E.; Palomar, M. (2009). A Gradual Combination of Features for Building Automatic Summarisation Systems. In *Text, Speech and Dialogue*, Volume 5729 of *Lecture Notes in Computer Science*, pp. 16–23. Springer Berlin Heidelberg.
- Lovász, L. (1996). Random Walks on Graphs: A Survey. In D. Miklós; V. T. Sós; e T. Szőnyi (Eds.), *Combinatorics, Paul Erdős is Eighty*, pp. 353–398. Budapest: János Bolyai Mathematical Society.
- Lu, Y.; Zhai, C.; Sundaresan, N. (2009). Rated Aspect Summarization of Short Comments. In *Proceedings of the 18th International Conference on World Wide Web, WWW '09*, New York, NY, USA, pp. 131–140. ACM.
- López, R.; Pardo, T.; Avançaço, L.; Balage Filho, P. P.; Bokan, A.; Cardoso, P.; Dias, M.; Nóbrega, F.; Cabezudo, M.; Souza, J.; Zacarias, A.; Seno, E.; Di Felippo, A. (2015). A Qualitative Analysis of a Corpus of Opinion Summaries based on Aspects. In *Proceedings of The 9th Linguistic Annotation Workshop*, Denver, Colorado, USA, pp. 62–71. Association for Computational Linguistics.
- López, R.; Pardo, T. A. (2015). Experiments on Sentence Boundary Detection in User-Generated Web Content. In *Proceedings of the 16th Computational Linguistics and Intelligent Text Processing*, Lecture Notes in Computer Science, pp. 227–237. Springer International Publishing.

- Mani, I. (1999). *Advances in Automatic Text Summarization*. Cambridge, MA, USA: MIT Press.
- Mani, I. (2001a). *Automatic Summarization*. Natural language processing. J. Benjamins Publishing Company.
- Mani, I. (2001b). Summarization Evaluation: An Overview. In *Proceedings of the Third Second Workshop Meeting on Evaluation of Chinese & Japanese Text Retrieval and Text Summarization, NTCIR-2*.
- Mann, W. C.; Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text* 8(3), 243–281.
- Maziero, E.; Pardo, T. A. S.; Cunha, I.; Torres-Moreno, J.-M.; SanJuan, E. (2011). DiZer 2.0-An Adaptable On-line Discourse Parser. In *Proceedings of the III RST Meeting (8th Brazilian Symposium in Information and Human Language Technology)*, pp. 50–57.
- McCargar, V. (2004). Statistical Approaches to Automatic Text Summarization. *Bulletin of the American Society for Information Science and Technology* 30(4), 21–25.
- Mei, Q.; Ling, X.; Wondra, M.; Su, H.; Zhai, C. (2007). Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs. In *Proceedings of the 16th international conference on World Wide Web, WWW '07*, New York, NY, USA, pp. 171–180. ACM.
- Mihalcea, R. (2004). Graph-based Ranking Algorithms for Sentence Extraction, Applied to Text Summarization. In *Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions, ACLdemo '04*, Stroudsburg, PA, USA, pp. 20–23. Association for Computational Linguistics.
- Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM* 38(11), 39–41.
- Mithun, S.; Kosseim, L. (2009). Summarizing Blog Entries Versus News Texts. In *Proceedings of the Workshop on Events in Emerging Text Types, eETTs '09*, Stroudsburg, PA, USA, pp. 1–8. Association for Computational Linguistics.

- Molina, E. C. (1931). Bayes' Theorem: An Expository Presentation. *The Annals of Mathematical Statistics* 2, pp. 23–37.
- Munezero, M.; Montero, C.; Sutinen, E.; Pajunen, J. (2014). Are They Different? Affect, Feeling, Emotion, Sentiment, and Opinion Detection in Text. *Affective Computing, IEEE Transactions* 5(2), 101–111.
- Nenkova, A.; McKeown, K. (2011). Automatic Summarization. *Foundations and Trends in Information Retrieval* 5(2-3), pp. 103–233.
- Nóbrega, F. A. A.; Agostini, V.; Camargo, R. T.; Felippo, A. D.; Pardo, T. A. S. (2014). Alignment-Based Sentence Position Policy in a News Corpus for Multi-document Summarization. In *Computational Processing of the Portuguese Language - 11th International Conference, PROPOR*, pp. 286–291.
- Ouyang, Y.; Li, W.; Lu, Q.; Zhang, R. (2010). A Study on Position Information in Document Summarization. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters, COLING '10*, Stroudsburg, PA, USA, pp. 919–927. Association for Computational Linguistics.
- Page, L.; Brin, S.; Motwani, R.; Winograd, T. (1998). The PageRank Citation Ranking: Bringing order to the Web. In *Proceedings of the 7th International World Wide Web Conference*, Brisbane, Australia, pp. 161–172.
- Pang, B.; Lee, L. (2004). A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, ACL '04*, Stroudsburg, PA, USA, pp. 271–278. Association for Computational Linguistics.
- Pang, B.; Lee, L. (2008). *Opinion Mining and Sentiment Analysis*. Volume 2, Hanover, MA, USA, pp. 1–135. Now Publishers Inc.
- Pang, B.; Lee, L.; Vaithyanathan, S. (2002). Thumbs Up?: Sentiment Classification Using Machine Learning Techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA, pp. 79–86. Association for Computational Linguistics.

- Pardo, T. A. S. (2005). GistSumm - GIST SUMMarizer: Extensões e Novas Funcionalidades. Série de Relatórios do NILC NILC-TR-05-05, Núcleo Interinstitucional de Lingüística Computacional (NILC), São Carlos-SP.
- Paul, M. J.; Zhai, C.; Girju, R. (2010). Summarizing Contrastive Viewpoints in Opinionated Text. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, EMNLP '10*, Stroudsburg, PA, USA, pp. 66–76. Association for Computational Linguistics.
- Plutchik, R. (2001). *The Nature of Emotions*, Volume 89.
- Popescu, A. M.; Etzioni, O. (2005). Extracting Product Features and Opinions from Reviews. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT '05*, Stroudsburg, PA, USA, pp. 339–346. Association for Computational Linguistics.
- Radev, D. R.; McKeown, K. R. (1998). Generating Natural Language Summaries from Multiple On-line Sources. *Computational Linguistics* 24(3), 470–500.
- Radev, D. R.; Teufel, S.; Saggion, H.; Lam, W.; Blitzer, J.; Qi, H.; Çelebi, A.; Liu, D.; Drabek, E. (2003). Evaluation Challenges in Large-Scale Document Summarization. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics, ACL '03*, Stroudsburg, PA, USA, pp. 375–382. Association for Computational Linguistics.
- Rahayu, D. A. P.; Krishnaswamy, S.; Labbé, C.; Alhakoon, O. (2010). Web Services for Analysing and Summarising Online Opinions and Reviews. In *ServiceWave*, Lecture Notes in Computer Science, pp. 136–149. Springer.
- Rath, G.; Resnick, A.; Savage, T. (1961). The Formation of Abstracts by the Selection of Sentences. *American Documentation* 12(2), 139–141.
- Reiter, E. (1994). Has a Consensus NL Generation Architecture Appeared, and is It Psycholinguistically Plausible? In *Proceedings of the Seventh International Workshop on Natural Language Generation, INLG '94*, Stroudsburg, PA, USA, pp. 163–170. Association for Computational Linguistics.

- Ribaldo, R.; Akabane, A. T.; Rino, L. H. M.; Pardo, T. A. S. (2012). Graph-Based Methods for Multi-document Summarization: Exploring Relationship Maps, Complex Networks and Discourse Information. In *Proceedings of the International Conference on Computational Processing of Portuguese*, Lecture Notes in Computer Science, pp. 260–271. Springer.
- Salton, G. (1989). *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.
- Salton, G.; McGill, M. J. (1986). *Introduction to Modern Information Retrieval*. New York, NY, USA: McGraw-Hill, Inc.
- Schilder, F.; Kondadadi, R. (2008). FastSum: Fast and Accurate Query-based Multi-document Summarization. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, HLT-Short '08, Stroudsburg, PA, USA, pp. 205–208. Association for Computational Linguistics.
- Seki, Y. (2002). Sentence Extraction by TF/IDF and Position Weighting from Newspaper Articles.
- Seki, Y.; Eguchi, K.; K, N.; Aono, M. (2006). Opinion-focused Summarization and its Analysis at DUC 2006. In *Proceedings of the Document Understanding Conference (HLT-NAACL 2006)*, pp. 122–130.
- Seno, E. R. M.; Nunes, M. (2010). *Um Método para a Fusão Automática de Sentenças Similares em Português*. Tese de Doutorado, Instituto de Ciências Matemáticas e de Computação – USP, São Carlos.
- Sharifi, B.; Hutton, M.-A.; Kalita, J. K. (2010). Experiments in Microblog Summarization. In *SocialCom/PASSAT*, pp. 49–56. IEEE Computer Society.
- Shimada, K.; Tadano, R.; Endo, T. (2011). Multi-Aspects Review Summarization with Objective Information. *Procedia - Social and Behavioral Sciences* 27, 140 – 149. Computational Linguistics and Related Fields.

- Silva, M. J.; Carvalho, P.; Sarmiento, L. (2012). Building a Sentiment Lexicon for Social Judgement Mining. In *Proceedings of the 10th International Conference on Computational Processing of the Portuguese Language, PROPOR'12*, Berlin, Heidelberg, pp. 218–228. Springer-Verlag.
- Siqueira, H.; Barros, F. (2010). A Feature Extraction Process for Sentiment Analysis of Opinions on Services. In *III International Workshop on Web and Text Intelligence*.
- Souza, M.; Vieira, R.; Buseti, D.; Chishman, R.; Alves, I. M. (2011). Construction of a Portuguese Opinion Lexicon from Multiple Resources. In *Proceedings of the 8th Brazilian Symposium in Information and Human Language Technology (STIL 2011)*.
- Steinhaus, H. (1956). Sur la Division des Corps Matériels en Parties. In *Bulletin of the Polish Academy of Sciences*, pp. 801–804.
- Svore, K.; Vanderwende, L.; Burges, C. (2007). Enhancing Single-Document Summarization by Combining RankNet and Third-Party Sources. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, Prague, Czech Republic, pp. 448–457. Association for Computational Linguistics.
- Taboada, M.; Brooke, J.; Tofiloski, M.; Voll, K.; Stede, M. (2011). Lexicon-based Methods for Sentiment Analysis. Volume 37, Cambridge, MA, USA, pp. 267–307. MIT Press.
- Tadano, R.; Shimada, K.; Endo, T. (2010). Multi-aspects Review Summarization Based on Identification of Important Opinions and their Similarity. In *PACLIC*, pp. 685–692. Institute for Digital Enhancement of Cognitive Development, Waseda University.
- Thelwall, M.; Buckley, K.; Paltoglou, G.; Cai, D.; Kappas, A. (2010). Sentiment in Short Strength Detection Informal Text. Volume 61, New York, NY, USA, pp. 2544–2558. John Wiley & Sons, Inc.
- Titov, I.; McDonald, R. (2008a). Modeling Online Reviews with Multi-grain Topic Models. In *Proceedings of the 17th international conference on World Wide Web, WWW '08*, New York, NY, USA, pp. 111–120. ACM.

- Titov, I.; McDonald, R. T. (2008b). A Joint Model of Text and Aspect Ratings for Sentiment Summarization. In *Association for Computer Linguistics*, pp. 308–316.
- Turney, P. D. (2002). Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, Stroudsburg, PA, USA, pp. 417–424. Association for Computational Linguistics.
- Ulrich, J.; Murray, G.; Carenini, G. (2008). A Publicly Available Annotated Corpus for Supervised Email Summarization. In *Proceedings of AAAI EMAIL Workshop*, pp. 77–87.
- Van Deemter, K.; Krahmer, E.; Theune, M. (2005). Real Versus Template-Based Natural Language Generation: A False Opposition? Volume 31, Cambridge, MA, USA, pp. 15–24. MIT Press.
- Wan, X.; Xiao, J. (2007). Towards a Unified Approach Based on Affinity Graph to Various Multi-document Summarizations. In *Research and Advanced Technology for Digital Libraries*, Volume 4675 of *Lecture Notes in Computer Science*, pp. 297–308. Springer Berlin Heidelberg.
- Wang, D.; Liu, Y. (2011). A Pilot Study of Opinion Summarization in Conversations. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, HLT '11, Stroudsburg, PA, USA, pp. 331–339. Association for Computational Linguistics.
- Wang, D.; Zhu, S.; Li, T. (2013). SumView: A Web-based Engine for Summarizing Product Reviews and Customer Opinions. Volume 40, Tarrytown, NY, USA, pp. 27–33. Pergamon Press, Inc.
- White, M.; Korelsky, T.; Cardie, C.; Ng, V.; Pierce, D.; Wagstaff, K. (2001). Multidocument Summarization via Information Extraction. In *Proceedings of the First International Conference on Human Language Technology Research*, HLT '01, Stroudsburg, PA, USA, pp. 1–7. Association for Computational Linguistics.

- Xu, X.; Meng, T.; Cheng, X. (2011). Aspect-based Extractive Summarization of Online Reviews. In *Proceedings of the 2011 ACM Symposium on Applied Computing, SAC '11*, New York, NY, USA, pp. 968–975. ACM.
- Zajic, D.; Dorr, B. J.; Lin, J.; Schwartz, R. (2007). Multi-candidate Reduction: Sentence Compression As a Tool for Document Summarization Tasks. Volume 43, Tarrytown, NY, USA, pp. 1549–1570. Pergamon Press, Inc.
- Zhang, L.; Liu, B. (2013). Aspect and Entity Extraction for Opinion Mining. In *Data Mining and Knowledge Discovery for Big Data*, Volume 1 of *Studies in Big Data*, pp. 1–40. Springer Berlin Heidelberg.
- Zhu, J.; Zhu, M.; Wang, H.; Tsou, B. K. (2009). Aspect-Based Sentence Segmentation for Sentiment Summarization. In *Proceedings of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion, TSA '09*, New York, NY, USA, pp. 65–72. ACM.
- Zhu, L.; Gao, S.; Pan, S. J.; Li, H.; Deng, D.; Shahabi, C. (2013). Graph-Based Informative-Sentence Selection for Opinion Summarization. In *Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 408–412. IEEE.
- Zhuang, L.; Jing, F.; Zhu, X.-Y. (2006). Movie Review Mining and Summarization. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management, CIKM '06*, New York, NY, USA, pp. 43–50. ACM.

Critérios de Avaliação da DUC 2007

Os critérios de avaliação da qualidade linguística são orientados para avaliar quão legíveis e fluentes são os resumos. Eles medem a qualidade do resumo, e não implicam comparação com um resumo modelo. Para este trabalho de mestrado, foram utilizados os critérios de avaliação da DUC usados ano 2007.

1. Gramaticalidade

O sumário não deve ter erros referentes a letras maiúsculas ou frases obviamente agramaticais, que tornam o texto difícil de ler.

1. Péssimo
2. Ruim
3. Regular
4. Bom
5. Excelente

2. Não Redundância

Não deve haver repetição desnecessária no sumário, como frases inteiras ou fatos que se repetem ou ainda o uso repetido de um substantivo ou um sintagma. Por exemplo, a repetição excessiva de “Luís Inácio Lula da Silva” seria ruim, mas o pronome “ele” seria bom para menções subsequentes.

1. Péssimo
2. Ruim
3. Regular
4. Bom
5. Excelente

3. Clareza Referencial

Deve ser fácil de identificar “o que” ou “a quem” os pronomes ou os sintagmas se referem no sumário. Se uma pessoa ou qualquer outra entidade é mencionada, deve estar claro qual é seu papel na história.

1. Péssimo
2. Ruim
3. Regular
4. Bom
5. Excelente

4. Foco

O sumário deve ter um foco principal. Assim, as frases devem conter somente informações que estão relacionadas ao restante do texto.

1. Péssimo
2. Ruim
3. Regular
4. Bom
5. Excelente

5. Estrutura e Coerência

O sumário deve estar bem estruturado e organizado. Ele não deve ser apenas um amontoado de informações relacionadas.

1. Péssimo
2. Ruim
3. Regular
4. Bom
5. Excelente