



Análise de Sentimento e Mineração de Opinião: uma revisão bibliométrica da literatura

Sentiment Analysis and Opinion Mining: a bibliometric review of the literature

Flavio CECI [1](#); Guilherme Martins ALVAREZ [2](#); Alexandre Leopoldo GONÇALVES [3](#)

Recibido: 01/10/16 • Aprobado: 26/10/2016

Conteúdo

- [1. Introdução](#)
 - [2. Aspectos conceituais](#)
 - [3. Método utilizado](#)
 - [4. Execução e resultados](#)
 - [5. Conclusões](#)
- [Referências](#)

RESUMO:

Com a expansão da Web 2.0, os indivíduos passaram de apenas consumidores de informação para produtores de conteúdo, aumentando exponencialmente o número de publicações contendo suas experiências em relação a produtos e serviços. A análise de sentimento pode contribuir para a definição da imagem de um produto ou identificação do sentimento em uma sentença como positiva ou negativa. Dessa forma, a análise dessas informações pode propiciar benefícios para as organizações, fornecendo subsídios para a tomada de decisão. Sendo assim, este trabalho tem como objetivo realizar uma revisão bibliométrica da literatura acerca das relações entre os temas análise de sentimento e mineração de opinião, mapeando a produção intelectual desenvolvida até meados de 2016. A pesquisa foi realizada nas bases de resumos e citações de artigos Scopus® e Web of Knowledge®, fazendo uso das seguintes palavras-chave: "sentiment analysis" (análise de sentimento) e "opinion mining" (mineração de opinião). A partir da identificação destes estudos o presente trabalho contribui para a compreensão das possibilidades de aplicação da análise de sentimento e mineração de opinião nas organizações. Os resultados encontrados destacam um panorama geral do atual estágio dos estudos sobre as relações entre a análise de sentimento e a mineração de opinião.

ABSTRACT:

With the expansion of Web 2.0, individuals passed of only information consumers to content producers, exponentially increasing the number of publications containing their experiences in relation to products and services. The sentiment analysis can contribute to the image definition of a product or sentiment identification in a sentence as positive or negative. Thus, the analysis of this information can benefit these organizations, providing support for decision-making. Thereby, this study aims to carry out a bibliometric review of the literature about the relationship between the subjects sentiment analysis and opinion mining, mapping the intellectual production developed by the middle of 2016. The survey was conducted in bibliographic data bases Scopus® and Web of Knowledge®, using the following keywords: "sentiment analysis" (sentiment analysis) and "opinion mining" (opinion mining). From the identification of these studies this research contributes to the understanding of the application possibilities of sentiment analysis and opinion mining in organizations. The results highlight an overview of the current state of the studies on the relationship between sentiment analysis and opinion mining.

Key-words: Sentiment Analysis; Opinion Mining; Sentiment Classification; Natural Language Processing.

1. Introdução

As opiniões de clientes são um importante ativo para o processo de tomada de decisão de uma organização. Pode-se observar essa preocupação mesmo antes da popularização da Web. Os consumidores que se beneficiam da web como canal de compra, na grande maioria das vezes, fazem pesquisas em fóruns e blogs especializados, para conhecer a opinião de outros consumidores antes de efetuarem uma compra ou assinarem algum novo serviço (FELDMAN, 2013). Com o surgimento das plataformas da Web 2.0, os usuários da internet passam de apenas consumidores de conteúdo para também produtores, o que permite um aumento exponencial das publicações contendo opiniões (O'RELLY, 2005, MARTINEZ; FERREIRA, 2007). O presente cenário apresenta a necessidade de recursos computacionais para a classificação e análise de textos que auxiliem no gerenciamento de grandes coleções de documentos (DURIC; SONG, 2012; LI; TSAI, 2013).

De acordo com Atkinson, Salas e Figueroa (2015), a expansão do uso das mídias sociais gerou uma oportunidade para as pessoas publicarem as suas opiniões. Então, sendo as mídias sociais um espaço em que opiniões podem ser detectadas, as organizações precisam desenvolver um ferramental capaz de extrair as visões de seus clientes, a fim de interpretá-las e auxiliar a tomada de decisão. Para isso, pode-se utilizar como base a análise de sentimentos, que é uma das possibilidades para a tarefa de classificação de textos (BELLINI et al., 2012).

Segundo Serrano-Gerrero et al. (2015), a análise de sentimento é uma das temáticas mais recentemente pesquisadas no domínio do processamento de informação, e tem como objetivo explicitar elementos referentes às opiniões, em forma de texto. Pode-se aplicar a análise de sentimento para a identificação da imagem organizacional utilizando como base dados publicados em mídias sociais. A análise de sentimento é entendida como uma área com potencial para avaliar informações relacionadas a uma organização por meio do monitoramento de mídias sociais (JEBASEELI; KIRUBAKARAN, 2012).

Percebendo a importância da área da análise de sentimento e objetivando entender a sua relação com a área chamada de mineração de opinião, este artigo tem como objetivo desenvolver uma revisão bibliométrica a fim de, analisar como a presente área vem sendo desenvolvida, quais são as suas principais subáreas, o número de trabalhos que vem sendo publicados e os seus principais autores.

Este artigo está estruturado em cinco seções, onde a próxima seção tem como objetivo apresentar os aspectos conceituais que estão por trás da temática eleita para este artigo. A seção 3 apresenta mais detalhes sobre o método de pesquisa utilizado e o protocolo de busca. Na seção 4 são apresentados mais detalhes sobre a execução das etapas do método e os resultados alcançados. Por fim, na seção de conclusões são apresentadas as percepções dos autores sobre o desenvolvimento do artigo, bem como, os seus resultados.

2. Aspectos conceituais

Antes de aprofundar o estudo sobre os conceitos de análise de sentimentos e mineração de opinião é importante entender como os termos sentimento e opinião estão ligados a esta temática.

O termo sentimento é estudado e aplicado por muitas áreas, como por exemplo: a antropologia, a filosofia, as ciências biológicas, a psicologia, a ciência da computação, entre outras. Para Franz (2003), a base da psicologia utilizada para conceitualizar o sentimento veio dos estudos dos filósofos Aristóteles, Platão, Spinoza, Descartes, Kant e Hume. Sentimentos como designação das emoções, suscetibilidade e simpatias surgiram em 1771, com a utilização de alguns termos para este fim, como, por exemplo, interessante, saudade, apatia, tédio, entediado, acanhamento etc.

Thums (1999), afirma que sentir significa estar envolvido com algo, que pode ser uma ideia, um produto, outra pessoa, um problema, uma situação. Franz (2003) relata que os sentimentos não são apenas de natureza pessoal, pois também retratam fenômenos históricos e universais, podendo ser coletivos e comuns. O autor afirma ainda que Jung [4] trabalhou o conceito de sentimento como objeto de estudo, descobrindo o papel dos sentimentos a partir da observação de seus experimentos

de associação, nas quais se deparou com relações afetivas puras ("sim", "ruim", "gosta", entre outras).

Segundo Thums (1999), sentimento não representa apenas uma informação, pois pode ser entendido como motivação, sendo diretamente relacionado à tomada de decisão. Em relação à avaliação cognitiva do sentimento, Fialho (2011) relata que inicialmente ocorre uma avaliação primária, do tipo negativa ou positiva, e em um segundo momento, chega-se ao sentimento mais. "Os sentimentos como conteúdo da psique podem ser qualificados por um sinal positivo (+) ou negativo (-)" (FRANZ, 2003, p.150). Neste ponto pode-se observar a relação do uso do termo "sentimento" para o nome da área de análise de sentimento.

Na visão de Mejova (2011), o processo de definição do termo sentimento não é uma tarefa fácil; é sabido que este conceito está diretamente ligado à subjetividade e opinião. As opiniões possuem, de forma explícita ou implícita, um sentimento agregado, isto é, um sentido negativo ou positivo a respeito de um produto, evento ou situação (KIM; HOVY, 2004). De acordo com Pang e Lee (2008), na área de análise de dados, as opiniões são insumos essenciais para o processo de tomada de decisão organizacional, já que as mesmas representam as percepções acerca de algum produto e/ou serviço ofertado.

De acordo com Padmaja e Fatima (2013), as opiniões têm um papel importante para quase todas as ações humanas, desde a forma de pensar, o que fazer e como atuar. Elas retratam como a realidade é percebida por outras pessoas, isto é, como são as suas percepções a respeito de determinada coisa (ESULI, 2008).

Opiniões são utilizadas para explicitar pontos de vista, de maneira que os pensamentos de outras pessoas podem ser úteis para o processo de tomada de decisão. As organizações fazem uso desse tipo de instrumento por meio das opiniões relacionadas a seus produtos e serviços, direcionando suas ações estratégicas. Empresas aplicam pesquisas de mercado coletando as opiniões e sentimentos dos consumidores a respeito de seus produtos e os de seus concorrentes, a fim de compará-los e então planejar suas estratégias de negócio (ESULI, 2008).

Para que as opiniões possam ser utilizadas pelas organizações, faz-se necessário que estejam disponíveis de forma codificada em uma mídia, pois fontes de dados não estruturados podem conter uma grande quantidade de informações e de conhecimentos implícitos (AGARWAL et al., 2015). A *Web* passa, nesse contexto, a ser um importante canal para coletar e expressar pensamentos e sentimentos. Segundo Pang e Lee (2008), 81% dos usuários da *Web* realizam buscas sobre um produto antes de adquiri-lo. E desses 81%, 20% realizam esse tipo de pesquisa diariamente.

O desafio de classificar e analisar expressões com o objetivo de identificar sentimentos e opiniões não é recente na área da computação. Podem ser mencionados trabalhos como os apresentados por Kim e Hovy (2004), que retrata a preocupação de Hatzivassiloglou e McKeown (1997) com a extração e análise de opiniões, ou estudos como os de Riloff et al. (2003) e Wiebe et al. (2002), com o objetivo de identificar a subjetividade inclusa nas opiniões.

Percebe-se que esses desafios perduram até os dias atuais, em que as fontes e o consumo de informações evoluíram em paralelo aos progressos da *Web*. Inicialmente, textos publicados por sites, eram utilizados como fonte para essas análises. Mais tarde, os jornais foram sendo migrados para a *Web* e publicando notícias por esse meio de comunicação. Então, passou-se a se utilizar as notícias publicadas em jornais e revistas online como fonte para coleta de dados. Após do surgimento das mídias sociais, através das quais os usuários podem compartilhar, de maneira simples os seus conteúdos, estabeleceu-se um importante meio para ser utilizado como fonte de dados.

A mineração de opinião e análise de sentimento surgiram com o intuito de atuar na identificação de recursos computacionais para as tarefas de identificação, classificação e análise de opiniões e sentimentos.

A área de análise de sentimento tem como principal objetivo a polarização de uma sentença como positiva ou negativa (PANG; LEE, 2008). Na perspectiva de Narayanam, Liu e Choudhary (2009), a análise de sentimento é sinônimo da área de mineração de opinião, técnica na qual se tem como foco a classificação de um documento ou sentença como positiva ou negativa. Pode-se ainda considerar que a análise de sentimento ou mineração de opinião é a área da computação que estuda os sentimentos, as opiniões e as emoções expressas em um texto (LIU, 2010).

Verifica-se que os primeiros trabalhos relacionados com a área, tiveram seu surgimento entre as décadas de 80 e 90, com o objetivo de identificar a subjetividade latente em textos (WESTERKI, 2012), tendo como motivo para o desenvolvimento da área, a evolução dos métodos e técnicas de processamento de linguagem natural dos repositórios de dados da web social (PANG; LEE, 2008).

A análise de sentimento faz menção ao seu objetivo de estudo, como sendo, identificar o sentimento (se positivo ou negativo) de um texto, utilizando-se de algumas variações, por exemplo: mineração de opinião (em inglês, *opinion mining*), análise de sentimento (em inglês, *sentiment analysis*) ou análise de subjetividade (em inglês, *subjectivity analysis*). Também são utilizados os termos extração de avaliação (em inglês, *appraisal extraction*) e mineração de revisão (em inglês, *review mining*), mas esses termos estão mais associados com a área da computação afetiva (PANG; LEE, 2008).

Pode-se definir a análise de subjetividade como uma forma de separar opiniões individuais de fatos (WIEBE, 2000). De acordo com Banea, Mihalcea e Wiebe (2008), subjetividade, são as opiniões, emoções e os sentimentos expressos nos textos, podendo ser identificados e classificados. A análise ou mineração de subjetividade tem como objetivo identificar os termos subjetivos, podendo ser uma etapa de pré-processamento para a mineração de opinião ou para a análise de sentimento (ZHOU; CHAOVALIT, 2008).

A mineração de opinião tem seu foco na identificação e extração da subjetividade de documentos textuais, enquanto a análise de sentimento busca apresentar o sentimento da sentença (PADMJA; FATIMA, 2013). Na visão de Abbasi, Chen e Salem (2008), a análise de sentimento está dedicada em analisar documentos textuais contendo opiniões e emoções, enquanto a área de classificação de sentimento tem o propósito de determinar se um texto é objetivo ou subjetivo, e, quando subjetivo, se é positivo ou negativo.

3. Método utilizado

Para o desenvolvimento desse trabalho, utilizou-se o método de pesquisa bibliométrica como procedimento metodológico, o qual emprega ferramentas e técnicas baseadas em métodos estatísticos e matemáticos para mapear informações a partir de registros bibliográficos de documentos armazenados em bases de conhecimento (SANTOS; KOBASCHI, 2009).

A pesquisa caracteriza-se como aplicada, uma vez que está organizada em procedimento estruturado. Em relação ao enquadramento metodológico, classifica-se como pesquisa exploratória, bibliográfica e descritiva, pois apresenta diversas abordagens sobre os temas, análise de sentimento e mineração de opinião (RICHARDSON, 1999).

A revisão foi dividida em cinco etapas, como forma de estratégia, detalhadas a seguir:

- Primeira etapa: envolve a definição da temática e formulação da pergunta de pesquisa;
 - Segunda etapa: definição de termos ou palavras-chave, estratégias de busca, bases de dados e de outras fontes de informação a serem pesquisadas;
 - Terceira etapa: revisão e seleção dos artigos através da avaliação dos títulos e resumos identificados na busca inicial, estabelecendo critérios de inclusão e exclusão de artigos;
 - Quarta etapa: apreciação dos registros selecionados a partir da base de documentos;
 - Quinta etapa: apresentação dos resultados obtidos após a análise dos estudos.
-

4. Execução e resultados

A temática definida para a revisão bibliométrica está focada em análise de sentimento e mineração de opinião. Tem-se como objetivo a obtenção de artigos científicos relacionados ao tema e disponíveis em bases de documentos.

Após definir o tema e o objetivo da revisão, elaborou-se o seguinte questionamento como pergunta da revisão: Como a mineração de opinião e a análise de sentimento tem sido utilizadas pelas organizações nas tarefas de identificação e classificação de sentimentos e opiniões de documentos textuais?

O próximo passo foi a definição da fonte de dados, então, optou-se pelas bases *Scopus* e *Web of Knowledge* tendo também em vista a grande quantidade de artigos disponíveis e a abrangência de áreas suportadas.

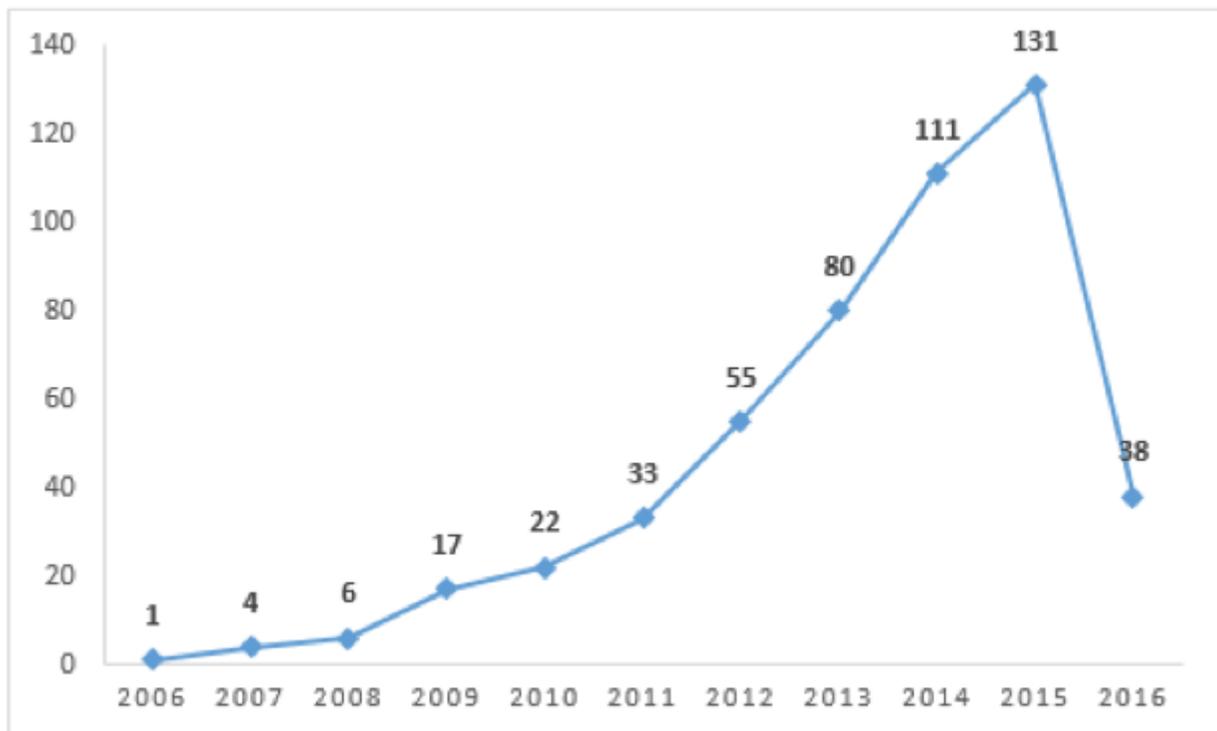
As consultas foram realizadas no período de julho de 2016 a setembro de 2016, nas bases *Web of Science* e *Scopus*, bases multidisciplinares, internacionais e de importância reconhecida cientificamente (ALMEIDA, 2006).

Visto que, a proposta se direciona a identificar as pesquisas já realizadas nesse tema, definiu-se como termos para a busca a combinação das palavras *'sentiment analysis'* e *'opinion mining'*. O conector entre os termos utilizados para a busca foi "E". Com isso, pretendeu-se recuperar os artigos relacionados aos dois termos de busca selecionados. É importante salientar que a base de conhecimento escolhida tem, na sua grande maioria, artigos em inglês.

A consulta montada ficou da seguinte maneira: "semantic analysis" AND "opinion mining". Para a consulta na *Web of Knowledge*, os campos de busca selecionados foram "Nome da publicação", "Tópico" e "Título". Na *Scopus* a busca foi realizada utilizando o "Título".

A busca na *Web of Knowledge* retornou 478 trabalhos. A consulta na *Scopus* apresentou 28 artigos. Em seguida todos os registros, recuperados a partir das buscas, foram exportados para o software *EndNote®*, onde foram realizados os procedimentos bibliométricos. Posteriormente, realizou-se a identificação dos estudos duplicados e exclusão dos estudos indexados sem autoria. A retirada destes estudos resultou na seleção de 498 trabalhos, distribuídos conforme demonstra a Figura 1.

Figura 1 – Artigos retornados a partir da busca distribuídos por ano

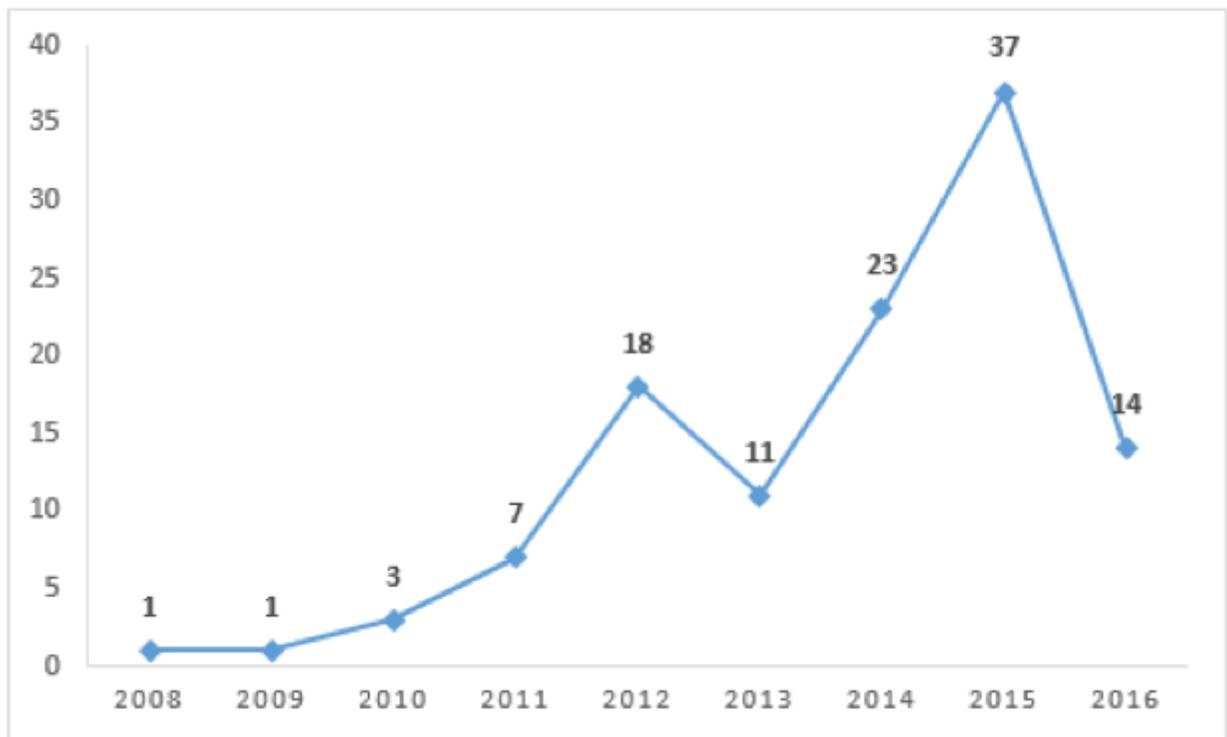


Fonte: Elaborado pelos autores

Dos 498 artigos resultantes das pesquisas efetuadas na *Web of Knowledge* e *Scopus*, apenas 114 estavam disponíveis para download, utilizando como alicerce o convênio da UFSC-CAPES com as bases de conhecimento em questão. Estes 114 trabalhos foram adotados, então, como sendo o conjunto de publicações utilizado para as análises bibliométricas posteriores.

A primeira análise, efetuada a partir da leitura dos artigos selecionados, segue a ordem cronológica de publicação. O artigo mais antigo selecionado foi publicado em 2008 e o mais recente, em 2016, ano corrente desta revisão. A Figura 2 demonstra como está distribuída a quantidade de artigos selecionados para leitura ao longo dos anos.

Figura 2 – Artigos coletados por ano



Fonte: Elaborado pelos autores

Ao analisar a Figura 1, percebe-se que, com exceção do ano de 2013, a quantidade de publicações relacionadas à análise de sentimento e mineração de opinião vem aumentando a cada ano. Essa informação também pode ser constada ao analisar a Figura 2, que apresenta todos os artigos disponíveis sobre o tema.

É interessante ressaltar que há uma queda na publicação de artigos sobre a temática em 2016, mas o motivo para isso é que esta revisão bibliométrica foi executada em meados de 2016.

Através da análise dos artigos recuperados nas bases de conhecimento, foi possível identificar 325 autores, provenientes de 38 países. A Tabela 1 apresenta os 10 países que obtiveram destaque nesse quesito.

Tabela 1 – Quantidade de autores por país

País	Total
China	29
Espanha	17
EUA	16
Índia	14
Itália	14
Inglaterra	9
Singapura	8
Escócia	5
Coreia do Sul	5
França	4

Fonte: Elaborado pelos autores

Com relação aos autores que mais publicaram sobre os temas da busca, verificou-se que 10 deles possuem três ou mais publicações no conjunto de artigos recuperados, conforme pode ser visto

através da Tabela 2.

Tabela 2 – Total de publicações por autor

Autor	Total
CAMBRIA, E.	7
BALAHUR, A.	6
HUSSAIN, A.	5
MARTINEZ-CÁMARA, E.	4
MARTÍN-VALDIVIA, M.T	4
MONTOYO, A.	4
TAN, S.	4
HAVASI, C.	3
MOLINA-GONZÁLEZ, M.D	3
PEREA-ORTEGA, J.M.	3

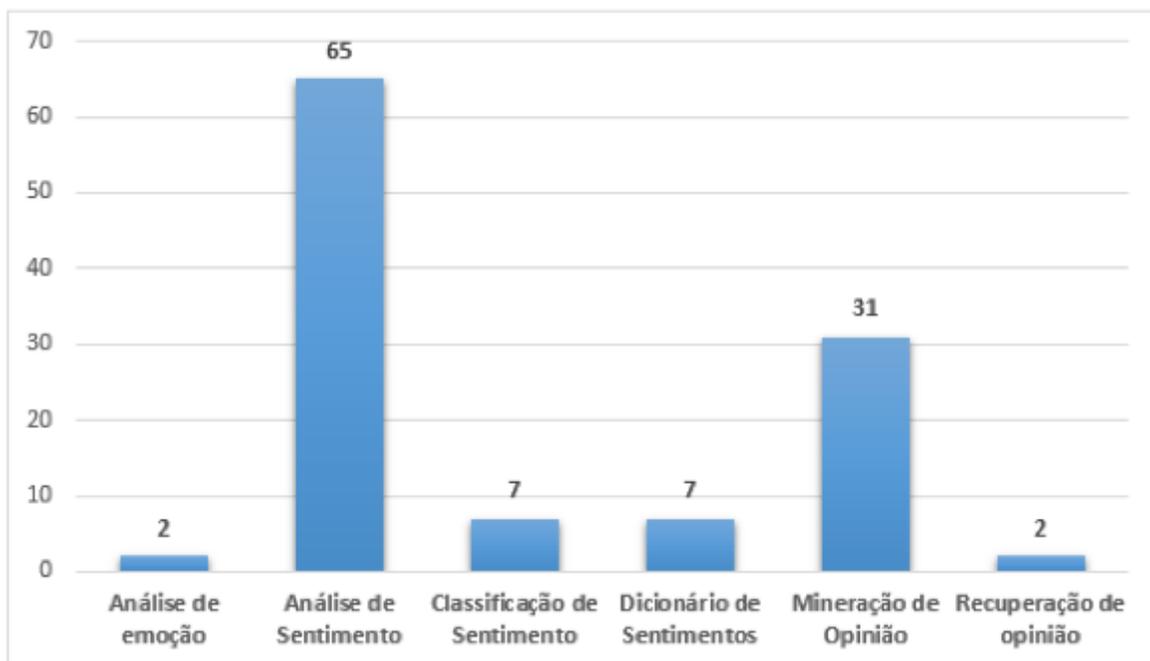
Fonte: Elaborado pelos autores

Ao separar os trabalhos dos autores apresentados na Tabela 2, tem-se 19 artigos distintos, portanto os 10 autores que mais publicaram, são responsáveis por 16,66% do conjunto de trabalhos analisados. Verificou-se também que desses 10 autores, muitos publicam juntos, como: Cambria, Havassi e Hussain (2012), Balahur e Perea-Ortega (2015), Molina-González et al. (2013), entre outros.

Para a revisão e seleção dos estudos apresentados na busca, utilizou-se como base para a avaliação, o título do artigo, seu resumo e sua introdução. Os 114 artigos disponíveis para download foram submetidos à leitura completa para extrair informações sobre a sua natureza, seu foco, técnicas utilizadas e outras informações que pudessem contribuir para este trabalho.

A partir da revisão dos artigos coletados, analisou-se o foco dos artigos frente às subáreas ou às áreas relacionadas com a análise de sentimento. A Figura 3 apresenta mais detalhes.

Figura 3 – Distribuição dos artigos por área temática



Fonte: Elaborado pelos autores

Ao analisar o gráfico apresentado na Figura 3, pode-se perceber que o foco na área da análise de sentimento continua em alta, seguido pela área de mineração de opinião. A utilização ou adequações de léxicos por um domínio de aplicação teve um aumento nos últimos três anos, até 2012 existiam apenas 5 artigos com esse foco e a partir de 2013 foram encontrados 24 artigos. Atribui-se esse aumento do número de publicações à percepção da necessidade de se levar em consideração elementos do domínio de aplicação para auxiliar na polarização final.

Dentre os trabalhos voltados à análise de sentimento, a sua grande maioria foca na tarefa de classificação de sentimento, utilizando técnicas de aprendizagem de máquina, Processamento de Linguagem Natural (NLP) e redes neurais, como exemplo, pode-se citar Boiy e Moens (2009), Huang e Qiu (2010), He, Alani e Zhou (2010), Cambria et al. (2012), Eiraki, Pisal e Singh (2012), Liu et al. (2013), Kansal e Toshniwal (2014), Lau, Li e Liao (2014), Peñalver-Martinez (2014), Agarwal et al. (2015), Atkinson, Salas e Figueroa (2015), Liu e Chen (2015), Zhao et al. (2015), Kumar e Jaiswal (2016) e Clavel e Callejas (2016).

Dentre os artigos com foco na mineração de opinião, muitos objetivam minerar opinião dos usuários em repositórios na web, como, redes sociais, fóruns e blogs, como pode-se observar nos trabalhos de Gokulakrishnan et al. (2012), Kontopoulos et al. (2013), Khan, Bashir e Qamar (2014), Eltayeb (2014), Molla, Biadgie e Sohn (2014), Korenek e Šimko (2014), Burnap et al. (2014), Preethi et al. (2015), Balahur e Perea-Ortega (2015), Bouazizi e Ohtsuki (2015), Feng et al. (2015), Alfaro et al. (2016) e Kuo et al. (2016).

Dentre os trabalhos com foco nos léxicos, alguns têm como objetivo traduzir léxicos concebidos em inglês para outro idioma, como é possível verificar nos trabalhos de Molina-Gonzalez et al. (2013), Hogenboom et al. (2014) e Martinez-Camara et al. (2014). Também encontram-se trabalhos voltados à construção de léxicos a partir de textos já polarizados, com o objetivo de facilitar a adequação do dicionário com o domínio de aplicação (CRUZ et al., 2014; RAO et al., 2014).

Dentre os artigos baseados em dicionários léxicos, a prática mais frequente é a da adequação ou expansão de um léxico já formado para características do domínio em questão. Pode-se citar alguns trabalhos dessa área, entre eles, Lee, Kim e Yun (2013), Robaldo e Di Caro (2013), Cho et al. (2014), Wu e Tsai (2014), Dragut et al. (2015), Cotelo et al. (2015) e Park, Lee e Moon (2015).

Através da leitura dos artigos, foi possível identificar as técnicas mais utilizadas para classificação, focadas na análise de sentimento. A Tabela 3 apresenta as técnicas ordenadas pelo seu uso nos artigos.

Tabela 3 – Total de técnicas utilizadas

Técnica	Total
SVM	28
NaïveBayes	14
POS Tagging	12
NLP	10
LDA	9
Clusterização	7
PMI	5
NER	4
Fuzzy	4
Random Forest	2

Fonte: Elaborado pelos autores

Verifica-se que, para a etapa de classificação, são utilizadas técnicas de diversas naturezas. O SVM (*Support Vector Machine*) que, segundo a revisão é a técnica mais utilizada, é conhecido como um método de aprendizagem supervisionada. *Naïve Bayes* é uma técnica de aprendizagem

supervisionada, da mesma forma que o SVM, e está em segundo lugar. *POS Tagging*, que é a terceira técnica mais utilizada, baseia-se em uma abordagem linguística.

As técnicas utilizadas no Processamento de Linguagem Natural (NLP) ficaram em quarto lugar. O quinto lugar foi o LDA (*Latent Dirichlet Allocation*), que é um modelo estatístico generativo. A técnica de clusterização é classificada com uma estratégia para aprendizagem não supervisionada e é a sexta técnica mais utilizada. A sétima técnica mais utilizada é a PMI (*Point Wise Mutual Information*), uma abordagem de natureza estatística. A oitava, chamada reconhecimento de entidades nomeadas (NER – *Named Entity Recognition*), fundamenta-se em uma abordagem linguística. A nona colocada é a Lógica Difusa (*Fuzzy Logic*), que é uma forma de lógica multivalorada. Em décimo lugar ficou o *Random Forest*, que é um método de aprendizagem para a classificação, regressão e outras tarefas, que opera através da construção de um grande número de árvores de decisão.

Como se pode observar pelas técnicas utilizadas nos trabalhos e apresentadas na Tabela 3, muitas são classificadas como abordagens de aprendizagem supervisionada. Por esse motivo, diversos trabalhos utilizam uma base de sentenças já polarizadas como uma 'partida fria' para tais técnicas.

Observa-se a existência de uma lacuna na área de classificação semântica e de análise de sentimento em relação ao reaproveitamento de um raciocínio já concretizado, de modo a aproximar-se do que existe na aprendizagem humana.

Verifica-se que as técnicas de classificação de texto podem ser utilizadas de maneira isolada ou combinada. Dessa maneira, dependendo da natureza da classificação, pode-se obter um resultado melhor.

Além das técnicas, observou-se que muitos trabalhos utilizam como base para sua análise, taxonomias, dicionários, corpus anotados ou ontologias. Dos 114 trabalhos coletados, 55 utilizam alguma base de conhecimento para auxiliar na classificação. Destacam-se os seguintes léxicos focados para a análise de sentimento:

- SentiWordNet;
- WordNetAffect; e
- MicroWordNetOpinion.

Dos léxicos relacionados, o mais utilizado é o SentiWordNet. Também foi identificada uma ontologia focada em análise de sentimento chamada EmotiNET.

A classificação semântica, mais precisamente, a análise de sentimento, pode ser aplicada a diferentes contextos. A partir da leitura dos artigos selecionados, é possível observar alguns casos de aplicação:

- Análise de pacientes com câncer;
- Análise da imagem de políticos (em período de eleição ou já eleitos);
- Opiniões a respeito de produtos ou serviços;
- *Marketing* e campanhas publicitárias;
- Inteligência competitiva;
- Análise de citações de trabalhos científicos;
- Análise de opiniões publicadas através de elementos de multimídia (áudio e vídeo);
- Detecção de crimes e de terrorismo;
- Análise do mercado financeiro;
- Identificação de situações críticas; entre outros.

A partir do conjunto de trabalhos recuperados, foi possível identificar 277 palavras-chave. A palavra-chave "sentiment analysis" (análise de sentimento) destaca-se com 84 citações nos estudos analisados e a palavra-chave "opinion mining" (mineração de opinião) vem logo em seguida com 76 ocorrências. A análise da nuvem de *tags* demonstrada na Figura 4, gerada a partir das palavras-chave dos trabalhos, evidencia os termos com maior número de citações.

Figura 4 – Nuvem de palavras-chave

da imagem de produtos e serviços ou na identificação do sentimento de uma sentença como positiva ou negativa.

Verifica-se que os autores têm utilizado a web para extrair os textos a serem analisados, através de mídias sociais, micro blogs, sites de e-commerce, entre outros. Além disso, grande parte dos esforços está direcionado para a criação e evolução de dicionários léxicos de domínio e tradução de léxicos já construídos em inglês para outros idiomas.

5. Conclusões

Este trabalho mapeou e analisou o campo de pesquisa que aborda os conceitos de análise de sentimento e mineração de opinião. Delineou-se um panorama sistemático com indicadores bibliométricos com foco na identificação dos trabalhos publicados e apontando as tendências dessa literatura.

Ao verificar e listar os principais estudos e autores nessa área de pesquisa, este trabalho também fornece subsídios para a continuidade dos estudos em outros campos, formação de redes de pesquisas, expansão de linhas de pesquisas que estabeleçam e investiguem em profundidade essas áreas.

Por meio desta revisão da literatura, pôde-se perceber que a área está sendo bastante pesquisada, ou seja, existem vários trabalhos sendo produzidos. A partir da leitura dos textos, pôde-se observar que os autores apontam uma série de oportunidades de pesquisa e trabalhos futuros.

As mídias sociais, blogs, tweets, serviços e sites de avaliação de produtos, são exemplos de meios através dos quais usuários têm manifestado suas ideias e opiniões. A mineração de opinião e a análise de sentimento são disciplinas que buscam identificar conteúdo expresso em opiniões, e determinar o sentimento, emoção, atitude ou percepção do público em relação ao objeto alvo desta opinião.

Através deste trabalho, percebe-se o aumento dos estudos na área da análise de sentimento e mineração de opinião, além da constante utilização de dados e informações presentes em recursos da Web 2.0 para a análise e posterior tomada de decisão por parte das organizações.

Com a análise realizada, o presente trabalho avança na compreensão do relacionamento entre as áreas de análise de sentimento e mineração de opinião, demonstrando uma carência de estudos que tratem claramente dos conceitos relativos aos modelos léxicos que abordem a linguagem não formal, bem como a utilização dessas áreas para apoiar tarefas intensivas em conhecimento nas organizações.

Referências

- ABBASI, A.; CHEN, H.; SALEM, A. Sentiment analysis in multiple languages- Feature selection for opinion classification in Web forums. **ACM Transactions on Information Systems**, v. 26, n. 3, p. 12-46, 2008.
- AGARWAL, Basant et al. Sentiment Analysis Using Common-Sense and Context Information. **Computational intelligence and neuroscience**, v. 2015, 2015.
- ALFARO, César et al. A multi-stage method for content classification and opinion mining on weblog comments. **Annals of Operations Research**, v. 236, n. 1, p. 197-213, 2016.
- ALMEIDA, E. C. E. **O portal de periódicos da Capes: estudo sobre a sua evolução e utilização**. 2006. 175 f. Dissertação (Mestrado) – Universidade de Brasília, Brasília. 2006.
- ATKINSON, John; SALAS, Gonzalo; FIGUEROA, Alejandro. Improving opinion retrieval in social media by combining featuresbased coreferencing and memory-based learning. **Information Sciences**, v. 299, p. 20-31, 2015.
- BALAHUR, Alexandra; PEREA-ORTEGA, José M. Sentiment analysis system adaptation for multilingual processing: The case of tweets. **Information Processing & Management**, v. 51, n. 4, p. 547-556, 2015.
- BANEA, C.; MIHALCEA, R.; WIEBE, Janyce. **A Bootstrapping Method for Building Subjectivity Lexicons for Languages with Scarce Resources**. **International Conference on Language Resources and Evaluation** (LREC 2008), Marrakech, 2008.

- BELLINI, F. et al. **The Role of the Prosumer in Social Networks and the Sentiment Analysis for the Customer Experience Management**. cersi.it, 2012.
- BOUAZIZI, Mondher; OHTSUKI, Tomoaki. Opinion Mining in Twitter How to Make Use of Sarcasm to Enhance Sentiment Analysis. In: **Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015**. ACM, 2015. p. 1594-1597.
- BOIY, Erik; MOENS, Marie-Francine. A machine learning approach to sentiment analysis in multilingual Web texts. **Information retrieval**, v. 12, n. 5, p. 526-558, 2009.
- BURNAP, Pete et al. Detecting tension in online communities with computational Twitter analysis. **Technological Forecasting and Social Change**, v. 95, p. 96-108, 2015.
- CAMBRIA, Erik; HAVASI, Catherine; HUSSAIN, Amir. SenticNet 2: A Semantic and Affective Resource for Opinion Mining and Sentiment Analysis. In: **FLAIRS conference**. p. 202-207, 2012.
- CAMBRIA, E. et al. **Sentic Computing for social media marketing**. *Multimed Tools Appl.* v.59, p.557-577, 2012.
- CLAVEL, Chloé; CALLEJAS, Zoraida. Sentiment analysis: from opinion mining to human-agent interaction. **IEEE Transactions on Affective Computing**, v. 7, n. 1, p. 74-93, 2016.
- CHO, Heeryon et al. Data-driven integration of multiple sentiment dictionaries for lexicon-based sentiment classification of product reviews. **Knowledge-Based Systems**, v. 71, p. 61-71, 2014.
- COTELO, J. M. et al. A modular approach for lexical normalization applied to Spanish tweets. **Expert Systems with Applications**, v. 42, n. 10, p. 4743-4754, 2015.
- CRUZ, F. L. et al. 'Long autonomy or long delay?'The importance of domain in opinion mining. **Expert Systems with Applications**, v. 40, n. 8, p. 3174-3184, 2013.
- DRAGUT, Eduard C. et al. Polarity Consistency Checking for Domain Independent Sentiment Dictionaries. **Knowledge and Data Engineering**, IEEE Transactions on, v. 27, n. 3, p. 838-851, 2015.
- DURIC, Adnan; SONG, Fei. Feature selection for sentiment analysis based on content and syntax models. In: **Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis**. Association for Computational Linguistics, 2012. p. 96-103.
- EIRINAKI, Magdalini; PISAL, Shamita; SINGH, Japinder. Feature-based opinion mining and ranking. **Journal of Computer and System Sciences**, v. 78, n. 4, p. 1175-1184, 2012.
- ELTAYEBY, Omar; MOLNAR, Peter; GEORGE, Roy. Measuring the Influence of Mass Media on Opinion Segregation through Twitter. **Procedia Computer Science**, v. 36, p. 152-159, 2014.
- FELDMAN, Ronen. Techniques and applications for sentiment analysis. **Communications of the ACM**, v. 56, n. 4, p. 82-89, 2013.
- FENG, Shi et al. A word-emoticon mutual reinforcement ranking model for building sentiment lexicon from massive collection of microblogs. **World Wide Web**, v. 18, n. 4, p. 949-967, 2015.
- FIALHO, Francisco, A. P. **Psicologia das Atividades Mentais: Introdução às Ciências da Cognição**. Florianópolis: Editora Insular, 2011.
- FRANZ, Marie-Louise von. **Tipologia de Jung**. São Paulo: Cultrix, 2003.
- GOKULAKRISHNAN, Balakrishnan et al. Opinion mining and sentiment analysis on a twitter data stream. In: **Advances in ICT for Emerging Regions (ICTer), 2012 International Conference on**. IEEE, 2012. p. 182-188.
- HATZIVASSILOGLU, Vasileios; MCKEOWN, Kathleen R. **Predicting the semantic orientation of adjectives**. Proceedings of the 35th annual meeting on Association for Computational Linguistics, 1997.
- HE, Yulan; ALANI, Harith; ZHOU, Deyu. Exploring English Lexicon Knowledge for Chinese Sentiment Analysis. In: **CIPS-SIGHAN Joint Conference on Chinese Language Processing**, Beijing, China, 2010
- HOGENBOOM, Alexander et al. Multi-lingual support for lexicon based sentiment analysis guided by semantics. **Decision support systems**, v. 62, p. 43-53, 2014.

HUANG, Z.; QIU, Yuhui. A multiple-perspective approach to constructing and aggregating Citation Semantic Link Network. **Future Generation Computer Systems**, n.26, p.400–407, 2010.

JEBASEELI, A.; KIRUBAKARAN, E. Opinion Mining of M Learning Reviews using Soft Computing Techniques. **International Journal of Computer Applications**, v. 54, n. 15, p. 44-48, 2012.

KANSAL, Hitesh; TOSHNIWAL, Durga. Aspect based Summarization of Context Dependent Opinion Words. **Procedia Computer Science**, v. 35, p. 166-175, 2014.

KHAN, Farhan Hassan; BASHIR, Saba; QAMAR, Usman. TOM: Twitter opinion mining framework using hybrid classification scheme. **Decision Support Systems**, v. 57, p. 245-257, 2014.

KIM, Soo-Min; HOVY, Eduard. **Determining the Sentiment of Opinions**. The 20th International Conference on Computational Linguistics. In Proceedings of COLING. Suíça, 2004.

KONTOPOULOS, E. et al. **Ontology-based sentiment analysis of twitter posts**. **Expert Systems with Applications**. v. 40, 4065–4074, 2013

KORENEK, Peter; ŠIMKO, Marián. Sentiment analysis on microblog utilizing appraisal theory. **World Wide Web**, v. 17, n. 4, p. 847-867, 2014.

KUMAR, Praveen; JAISWAL, Umesh Chandra. A Comparative Study on Sentiment Analysis and Opinion Mining. **International Journal of Engineering & Technology IJET**, v. 8, n. 2, p. 938 -943, 2016.

KUO, Yau-Hwang et al. Integrated microblog sentiment analysis from users' social interaction patterns and textual opinions. **Applied Intelligence**, v. 44, n. 2, p. 399-413, 2016.

LAU, Raymond YK; LI, Chunping; LIAO, Stephen SY. Social analytics: Learning fuzzy product ontologies for aspect-oriented sentiment analysis. **Decision Support Systems**, v. 65, p. 80-94, 2014.

LEE, Kong-Joo; KIM, Jee-Eun; YUN, Bo-Hyun. Extracting Multiword Sentiment Expressions by Using a Domain-Specific Corpus and a Seed Lexicon. **ETRI Journal**, v. 35, n. 5, p. 838-848, 2013.

LI, Weiyuan; XU, Hua. Text-based emotion classification using emotion cause extraction. **Expert Systems with Applications**, v. 41, n. 4, p. 1742-1749, 2014.

LIU, Yang et al. Riding the tide of sentiment change: sentiment analysis with evolving online reviews. **World Wide Web**, v. 16, n. 4, p. 477-496, 2013.

LIU, Shuhua Monica; CHEN, Jiun-Hung. A multi-label classification based approach for sentiment classification. **Expert Systems with Applications**, v. 42, n. 3, p. 1083-1093, 2015.

O'REILLY, Tim. **What is web 2.0. Design patterns and business models for the next generation of software**. 2005. Disponível em: <http://facweb.cti.depaul.edu/jnowotarski/se425/What%20Is%20Web%202%20point%200.pdf> . Acessado em: 12 dez. 2015.

MARTINEZ, Maria Laura; FERREIRA, Sergio Leal. **Da Web 2.0 ao Learning 2.0: Novas oportunidades e desafios para o design de interfaces de aprendizagem** (2007). Disponível em: http://www.degraf.ufpr.br/artigos_graphica/DAWEB.pdf. Acesso em: 20 jun. 2014.

MARTÍNEZ-CÁMARA, Eugenio et al. Integrating Spanish lexical resources by meta-classifiers for polarity classification. **Journal of Information Science**, p. 0165551514535710, 2014.

MEJOVA, Yelena. **Sentiment Analysis: An Overview**. Comprehensive exam paper, disponível em: <http://www.cs.uiowa.edu/~ymejova/publications/CompsYelenaMejova.pdf>. 2011.

MOLLA, Alemu; BIADGIE, Yenewondim; SOHN, Kyung-Ah. Network-Based Visualization of Opinion Mining and Sentiment Analysis on Twitter. In: **IT Convergence and Security (ICITCS), 2014 International Conference on**. IEEE, 2014. p. 1-4.

MOLINA-GONZÁLEZ, M. D.; MARTINEZ-CÁMARA, E.; MARTÍN-VALDIVIA, M. T.; PEREA-ORTEGA, J. M. Semantic orientation for polarity classification in Spanish reviews. **Expert Systems with Applications**, v. 40, n. 18, p. 7250-7257, 2013.

PADMAJA, S.; FATIMA, S. S. Opinion Mining and Sentiment Analysis – An Assessment of Peoples' Belief: A Survey. **International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC)** Vol.4, No.1, 2013.

PANG, Bo; LEE, Lillian. **Opinion Mining and Sentiment Analysis**. Foundations and Trends in Information Retrieval. Vol 2, p. 1-135, 2008.

PARK, Sungrae; LEE, Wonsung; MOON, Il-Chul. Efficient extraction of domain specific sentiment lexicon with active learning. **Pattern Recognition Letters**, v. 56, p. 38-44, 2015.

PEÑALVER-MARTINEZ, Isidro et al. Feature-based opinion mining through ontologies. **Expert Systems with Applications**, v. 41, n. 13, p. 5995-6008, 2014.

PREETHI, P. G. et al. Temporal sentiment analysis and causal rules extraction from tweets for event prediction. **Procedia Computer Science**, v. 48, p. 84-89, 2015.

RAO, Yanghui et al. Building emotional dictionary for sentiment analysis of online news. **World Wide Web**, v. 17, n. 4, p. 723-742, 2014.

RICHARDSON, R. J. **Pesquisa social: métodos e técnicas**. São Paulo: Atlas. 1999.

RILOFF, E. et al. **Learning Subjective Nouns Using Extraction Pattern Bootstrapping**. Proceedings of the CoNLL-03 conference. 2003.

ROBALDO, Livio; DI CARO, Luigi. Opinion Mining-ML. **Computer Standards & Interfaces**, v. 35, n. 5, p. 454-469, 2013.

SANTOS, R. N. M.; KOBASHI, N. Y. **Bibliometria, cientometria, infometria: conceitos e aplicações**. Pesq. Bras. Ci. Inf., Brasília, v.2, n.1, p. 155-172. 2009.

SERRANO-GUERRERO, Jesus et al. Sentiment analysis: A review and comparative analysis of web services. **Information Sciences**, v. 311, p. 18-38, 2015.

THUMS, Jorge. **Educação dos Sentimentos**. Porto Alegre: Editora da Ulbra e Editora Sulina, 1999.

WESTERSKI, Adam. **Semantic Technologies in Idea Management Systems: A Model for Interoperability, Linking and Filtering**. (Tese) Engenharia de Informática. Universidad Politécnica de Madrid, Madrid, 2012.

WIEBE, J. **Learning subjective adjectives from corpora**. Proceedings of the National Conference on Artificial Intelligence. p. 735-741, 2000.

WIEBE, J. et al. **NRRC summer study Jan Wiebe and group** (University of Pittsburgh) on 'subjective' statements. 2002.

WU, Chi-En; TSAI, Richard Tzong-Han. Using relation selection to improve value propagation in a ConceptNet-based sentiment dictionary. **Knowledge-Based Systems**, v. 69, p. 100-107, 2014.

ZHAO, Yanyan et al. Aspect-Object Alignment with Integer Linear Programming in Opinion Mining. **Plos ONE**. 10. p.1-18, 2015.

ZHOU, Lina; CHAOVALIT, Pamwadee. **Ontology-Supported Polarity Mining**. Journal of the American Society for Information Science and Technology. n. 59, pp. 98-110, 2008.

-
1. Doutor em Engenharia do Conhecimento pela Universidade Federal de Santa Catarina. Atua como professor titular dos cursos de graduação em Ciência da Computação e Sistemas de Informação e dos programas de pós-graduação em Engenharia de Projeto de Software, Banco de dados e Gerência de Projetos de TI da Universidade do Sul de Santa Catarina. É pesquisador e membro do grupo de pesquisa GPSC-Unisul. É coordenador do curso de Especialização em Banco de Dados da Unisul. <http://lattes.cnpq.br/0403935450457957>. E-mail: flavioceci@gmail.com
 2. Mestrando em Engenharia do Conhecimento pela Universidade Federal de Santa Catarina. Atua como pesquisador nas áreas de Análise de Sentimentos, Mineração Textual, Inteligência Artificial, Análise de Dados, Big Data e Inovação. <http://lattes.cnpq.br/3399404108804935>. E-mail: guilherme.alvarez@gmail.com
 3. Doutor em Engenharia de Produção pela Universidade Federal de Santa Catarina. Atualmente é Professor Adjunto lotado no Campus Araranguá/UFSC, e Professor Permanente do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento/UFSC e do Programa de Pós-Graduação em Tecnologias da Informação e Comunicação/UFSC. <http://lattes.cnpq.br/5138758521691630>. E-mail: a.l.goncalves@ufsc.br
 4. Carl Gustav Jung (1875 — 1961), psiquiatra e psicoterapeuta suíço, construiu a psicologia analítica. Ele definiu e implementou as definições de arquétipo, personalidade extrovertida e introvertida e inconsciente coletivo (LACHMAN, 2010)

