

Revisão sistemática sobre uso de ontologia para análise de sentimento em conteúdo da Web

Paulo de Oliveira Lima Júnior

Doutorado em Administração pela Universidade Federal de Lavras (UFLA) - Brasil. Professor do Centro

Federal de Educação Tecnológica de Minas Gerais (CEFET) - Nepomuceno, MG - Brasil.

<http://lattes.cnpq.br/4826999848494424>

E-mail: plimajunior@gmail.com

Submetido em: 12/09/2013. Aprovado em: 03/01/2017. Publicado em: 26/06/2017..

RESUMO

É crescente o número de trabalhos que procuram extrair informação e conhecimento a partir de dados não estruturados como textos publicados na web em blogs, microblogs, redes sociais e fontes de notícias. Em muitos casos, a avaliação deste tipo de dado não estruturado depende do contexto ou domínio específico de uma aplicação. Assim o uso de ontologia surge como uma ferramenta para dar suporte à classificação de dados e inferência de informação a partir de dados não estruturados. Os avanços trazem sistemas baseados em ontologias (ontology driven). Com o objetivo de verificar os avanços nesta área, o presente trabalho apresenta uma revisão sistemática sobre os temas análise de sentimento e ontologia. O objetivo específico é responder a questão: Qual é a contribuição do uso de Ontologia para Análise de Sentimento em dados não estruturados da web?

Palavras-chave: Análise de Sentimento. Ontologia. Web.

Systematic review on the use of ontology for sentiment analysis in Web content

ABSTRACT

A growing number of papers seek to extract information and knowledge from unstructured data such as published texts on blogs, microblogs, social networks and news sources. In many cases, the evaluation of this type of unstructured data depends on the specific context or domain of an application. Thus the use of ontology appears as a tool to support the classification of data and inference of information from unstructured data. The advances bring systems based on ontologies (ontology driven). In order to verify the advances in this field, this paper presents a systematic review on the subjects sentiment analysis and ontology. The specific objective is to answer the question: What is the contribution of the use of Ontology on Sentiment Analysis on unstructured Web data?

Keywords: *Sentiment Analysis. Ontology. Web.*

Revisión sistemática sobre la ontología de la uso para el análisis de sentimiento en el contenido Web

RESUMEN

Número creciente de estudios tratan de extraer información y conocimiento a partir de datos no estructurados, tales como textos publicados en la Web en blogs, microblogs, redes sociales y fuentes de noticias. En muchos casos, la evaluación de este tipo de datos no estructurados depende del contexto o área específica de una aplicación. Así, el uso de la ontología se presenta como una herramienta para la clasificación y deducción de información de datos a partir de datos no estructurados. Los avances traen sistemas basados en la ontología (ontology driven). Para comprobar el progreso en esta área, este trabajo presenta una revisión sistemática de los temas de análisis de sentimiento y la ontología. El objetivo específico es responder a la pregunta: ¿Cuál es la contribución del uso de la ontología para el Análisis de Sentimiento de datos no estructurados en la Web?

Palabras clave: *Análisis de sentimiento. Ontología. Web.*

INTRODUÇÃO

Os sistemas de informação tradicionais são projetados para subsidiar a tomada de decisões em diferentes setores. Um de seus componentes principais, o banco de dados, mantém os dados organizados para responder precisamente e em tempo, questões predefinidas de acordo com o modelo de negócio ao qual está inserido. Uma vez definidos os requisitos do negócio, um conjunto de dados quantitativos e qualitativos é coletado e estruturado para que análises estatísticas ou lógicas sejam aplicadas em busca de informação para promover inteligência competitiva e tomada de decisões.

Entretanto, alguns sistemas dinâmicos, como sistemas de previsão de tempo, análise de comportamento, análise de risco e diagnósticos médicos, demandam características além dos sistemas tradicionais, como lidar com dados em tempo real, qualitativos com alto grau de subjetividade em formatos diversos (áudio, vídeo e texto) e não estruturados.

Principalmente com o aumento de usuários da web e dispositivos móveis conectados, a produção de dados não estruturados cresce exponencialmente, exigindo dos sistemas um esforço maior para coleta, recuperação e análise. Com dados organizados em um modelo conceitual, a informação pode ser recuperada por consultas em linguagem padrão; entretanto, dado não estruturado, como um texto, por exemplo, contém informação carregada de desafios para uma máquina, entre eles, conteúdo com ambiguidade, ruídos e sentimentos expressos.

É crescente o número de trabalhos que procuram extrair informação e conhecimento a partir de textos publicados na Web em *blogs*, *microblogs*, redes sociais e fontes de notícias, o que impulsiona a pesquisa em Processamento de linguagem natural e mineração textual, onde surge também a análise de sentimentos, definida por Liu (2010) como o estudo computacional das opiniões, avaliações e emoções das pessoas, expressas em texto, relativas a entidades, eventos e seus atributos.

A análise de sentimentos é usada em sistemas onde é necessário saber a opinião de outros para a tomada de decisão, e vem sendo aplicada em diferentes domínios, entre eles: finanças, para identificar fatores de risco (HUANG; LI, 2011), prever ganhos futuros (LI, 2010), prever preços de ações (KRAVET; MUSLU, 2013), impacto na liquidez de mercado (MITRA et al., 2011) e detectar fraudes (HUMPHERYS et al., 2011), em *marketing* para identificar satisfação de clientes (AGUWA; MONPLAISIR; TURGUT, 2012) e recomendação produtos (GARCÍA-CUMBRERAS; MONTEJO-RÁEZ; DÍAZ-GALIANO, 2013) e na política (PANG e LEE, 2008) para mapear intenções de voto.

Porém, em muitos casos, a análise depende do contexto ou domínio específico da aplicação. Um texto sobre o clima de uma região indicando chuva pode ser positivo para o contexto do agronegócio e negativo para o domínio de entretenimento, por exemplo. Assim, o uso de ontologia surge como uma abordagem para dar suporte à organização de conhecimento e inferência de informação a partir de dados não estruturados, como textos da Web.

Nesse cenário, o presente trabalho apresenta uma revisão sistemática cujo objetivo específico é responder à questão: Qual é a contribuição do uso de ontologia para análise de sentimento em dados não estruturados da Web? Com o intuito de delimitar o escopo do trabalho, a revisão pretende buscar artigos que avançam a teoria e a prática com aplicações em diferentes domínios de conhecimento, e procura delinear como ontologias são utilizadas para análise de sentimento e sua contribuição.

O trabalho é dividido em 5 seções. Análise de sentimento e ontologia são contextualizadas na seção 2. A seção 3 apresenta a metodologia utilizada para a revisão sistemática, a seção 4 descreve as etapas da pesquisa com seus resultados, e a seção 5 traz as discussões e conclusões sobre os resultados.

ANÁLISE DE SENTIMENTO E ONTOLOGIA

Pesquisas em análise de sentimento têm diferentes abordagens: identificar a polaridade positiva ou negativa, (DAS; CHEN, 2007; TURNEY, 2002; DAVE; LAWRENCE; PENNOCK, 2003; PANG; LEE, 2008), reconhecer a opinião em classes mais específicas, como raiva e aversão (WILSON et al., 2005; YU; HATZIVASSILOGLU, 2003; WILSON; WIEBE; HOFFMANN, 2005) e identificar também a fonte da opinião em questões da forma: “O que X pensa sobre Y?” (CHOI et al., 2005).

Montoyo, Martinez-Barco e Balahur (2012) dividem as pesquisas e aplicações em quatro categorias: classificação de textos, criação de recursos para análise léxica, extração de opiniões e aplicações em análise de sentimento, cada uma com desafios específicos.

A classificação de textos, sentença ou palavra, de acordo com a polaridade ou orientação de seu conteúdo, em positivo, negativo ou neutro, é estudada, conforme Martín-Valdivia et al. (2013), por duas abordagens: aprendizado de máquina e orientação semântica.

Na abordagem por aprendizado de máquina, o problema é modelado como categorização de texto, com as classes positiva e negativa (em alguns casos neutra), e algoritmos treinados em uma coleção de dados previamente rotulada para classificar novos dados. Esta é a abordagem supervisionada, descrita em Pang e Lee (2008), Liu et al. (2013), Tsytsarau e Palpanas (2012). A abordagem por orientação semântica utiliza recursos léxicos como lista de palavras, características linguísticas como adjetivos e advérbios e dicionários com polaridade positiva ou negativa, de palavras (TURNEY, 2002; LIU; HU; CHENG, 2005; DING; LIU, 2007; HU; LIU, 2004; KAMPS et al., 2004). Existem diferentes recursos léxicos disponíveis com informação sobre a implicação emocional de palavras e polaridade; dentre os mais estudados, estão: General Inquirer (STONE; DUNPHY; SMITH, 1966), MPQA Subjective Lexicon (WILSON; WIEBE;

HOFFMANN, 2005), WordNet (FELLBAUM, 1998), SentiWordNet 3.0 (BACCIANELLA; ESULI; SEBASTIAN, 2010), Bing Liu’s Opinion Lexicon (HU; LIU, 2004) (LIU; HU; CHENG, 2005), sendo os dois últimos os mais citados (CRUZ et al., 2014).

O aprendizado de máquina depende da disponibilidade de dados rotulados para treinamento, o que pode ser uma tarefa inviável em alguns casos. Orientação semântica, no entanto, demanda vasta quantidade de recursos léxicos que geralmente dependem da língua, e o resultado é comprometido em caso de ausência de palavras do recurso léxico no texto (MARTÍN-VALDIVIA et al., 2013). Além dessas limitações, um problema comum às abordagens são os casos em que a análise de sentimento depende do contexto. Os termos ou significados podem variar sua polaridade de acordo com o domínio específico da aplicação.

O problema de nível semântico é comum na interoperabilidade entre sistemas e intercâmbio de dados, tarefas reforçadas com o avanço da Internet, que por sua vez impulsionou o surgimento da Web Semântica (BERNERS-LEE, 2001), onde ontologia é apresentada como uma taxonomia que define classes de objetos e seus relacionamento e um conjunto de regras de inferência através das quais computadores podem manipular termos mais efetivamente de forma útil e significativa para o ser humano.

Conforme Gruninger, (2002) e Dobson e Sawyer, (2006), uma ontologia representa o conhecimento em um modelo formal e estruturado, bem como promove a comunicação, o reuso e a organização do conhecimento e inferência computacional. Assim, como aponta Blanco, et. al., (2008), o principal objetivo de ontologias é estabelecer acordo ontológico para diminuir ambiguidade da linguagem. De acordo com Uschold, (1996), uma ontologia necessariamente expressa algum conjunto de visões do mundo de acordo com um dado domínio, e isto é concebido como um conjunto de conceitos como entidades, atributos e processos, suas definições e seus relacionamentos.

Golbeck (2003) destaca que algumas ontologias são abrangentes, cobrindo vasto conjunto de tópicos, enquanto outras são limitadas com especificações muito precisas sobre determinada área, mas todas apresentam elementos gerais que as definem:

- classes: categorias gerais de coisas em um domínio de interesse;
- propriedades: atributos de instâncias das classes;
- relacionamentos: entre classes e propriedades.

Com a abordagem teórica, surgem sistemas baseados em ontologias (*ontology driven*) e pesquisas que procuram incluir o nível semântico para resolver ambiguidades e configurar contexto em diferentes domínios de aplicações, entre eles a análise de sentimento.

METODOLOGIA E REVISÃO SISTEMÁTICA

Botelho, De Almeida Cunha e Macedo, (2011), ao apresentar um método de revisão integrativa em estudos organizacionais, subdividem os dois tipos de revisão de literatura com características e objetivos distintos: a revisão narrativa e a revisão bibliográfica sistemática. Enquanto a primeira é utilizada para descrever o estado da arte de um assunto específico, sob o ponto de vista teórico e contextual sem explicitar metodologia para busca de referências, fontes e critérios usados na avaliação e seleção dos trabalhos, a segunda utiliza métodos explícitos e sistemáticos para identificar, coletar, selecionar e avaliar os estudos com objetivo de responder a uma questão específica, formulada de maneira objetiva e delimitada. A revisão bibliográfica por sua vez é subdividida quanto à definição, propósito, escopo, amostra e análise em revisão integrativa, meta-análise, metassumário ou metassíntese e revisão sistemática.

Diante do objetivo deste trabalho, foi adotada a revisão sistemática definida em Botelho, De Almeida Cunha e Macedo (2011), uma vez que é citada pelos autores como um sumário de pesquisas passadas, com objetivo e abordagem rigorosa de estudos com hipóteses idênticas ou relativas,

limitada quanto ao escopo, tendo como propósito sumarizar evidência sobre um problema específico, envolvendo uma pesquisa quantitativa de análise narrativa ou estatística.

No que diz respeito a um roteiro específico para revisão sistemática, Conforto, Amaral e Silva (2011) trazem uma aplicação no desenvolvimento de produtos e gerenciamento de projetos. Esse tipo de revisão é pouco difundido nesta área, sendo mais utilizado em medicina, psicologia e ciências sociais. Com esse argumento, sua proposta tem 15 etapas distribuídas em 3 fases: entrada, processamento, saída. Seu roteiro é amplo e inclui etapas para definição do problema, condução das buscas e análise dos resultados. Apesar de considerar a definição do problema em profundidade e acrescentar características para passos posteriores à pesquisa, como a criação de alerta, que não é o objetivo deste trabalho, é adotado como ponto de referência para criação de uma metodologia específica. Assim, inspirado pelo modelo proposto em Conforto, Amaral e Silva (2011), este trabalho propõe uma metodologia dividida em 4 etapas: definição de busca, pesquisa em abrangência, amostra por refinamento, análise estatística da amostra.

A primeira etapa consiste em identificar e definir os termos e as palavras chaves a serem utilizadas para a pesquisa que, neste caso, devem ser relevantes para o tema pesquisado; identificar e definir as bases e ferramentas para consulta de trabalhos que nessa situação considera a disponibilidade de artigos *on-line*, opções para refinar a pesquisa e analisar os resultados, como gráficos e ordenamento por atributos específicos. A segunda etapa, pesquisa em abrangência, tem como objetivo analisar o resultado da consulta apenas com as palavras-chave, e a partir deste, delinear um panorama geral do tema principal, descobrir a linha do tempo de citações, trabalhos mais citados, meios de publicação, autores e outras características que serão subsídio para o refinamento realizado na próxima etapa.

A terceira etapa consiste em refinar o resultado da etapa anterior em busca e uma mostra específica com delimitações em parâmetros definidos segundo características observadas na análise anterior. A última etapa consiste em análise dos resultados da etapa anterior.

DEFINIÇÕES DE BUSCA

Os sistemas para análise de sentimento de dados qualitativos não estruturados disponíveis na Web possuem arquitetura que normalmente incluem as fases de seleção, coleta, extração de conteúdo, classificação e análise e se beneficiam dos avanços em duas grandes áreas: a ciência da computação e ciência da informação. Nessas áreas alguns temas despontam, como recuperação de dados da Web, mineração de dados, mineração textual, mineração de opiniões, análise de sentimentos, processamento de linguagem natural e técnicas específicas de aprendizado e classificação, como algoritmos SVM (*Support Vector Machine*) e Naive Bayes, entre outros. Com a finalidade de buscar um conjunto de trabalhos não restritos a uma técnica específica mas sim aplicações na área de análise de sentimento em dados da Web, as seguintes palavras-chave foram selecionadas para pesquisa: *Sentiment Analysis | Web*. Os termos são comumente utilizados em inglês, portanto as palavras-chave serão usadas neste idioma.

O trabalho procura verificar se os avanços em análise de sentimento incluem o uso de ontologias como uma ferramenta para dar suporte à classificação de dados e inferência de informação a partir de dados não estruturados, portanto, também é pesquisada a palavra-chave *Ontology*.

Existem várias ferramentas de busca específicas e bases de dados, cada uma com suas peculiaridades e facilidades para refinamento e análise dos resultados recuperados em uma consulta. Este trabalho se limita à utilização do portal *Web of Science*, por ser abrangente quanto ao número de bases de consulta e oferecer opções de refinamento quanto à área, data, tipo de documento e fonte de publicação, além de opções para análise estatística dos resultados.

PESQUISA EM ABRANGÊNCIA

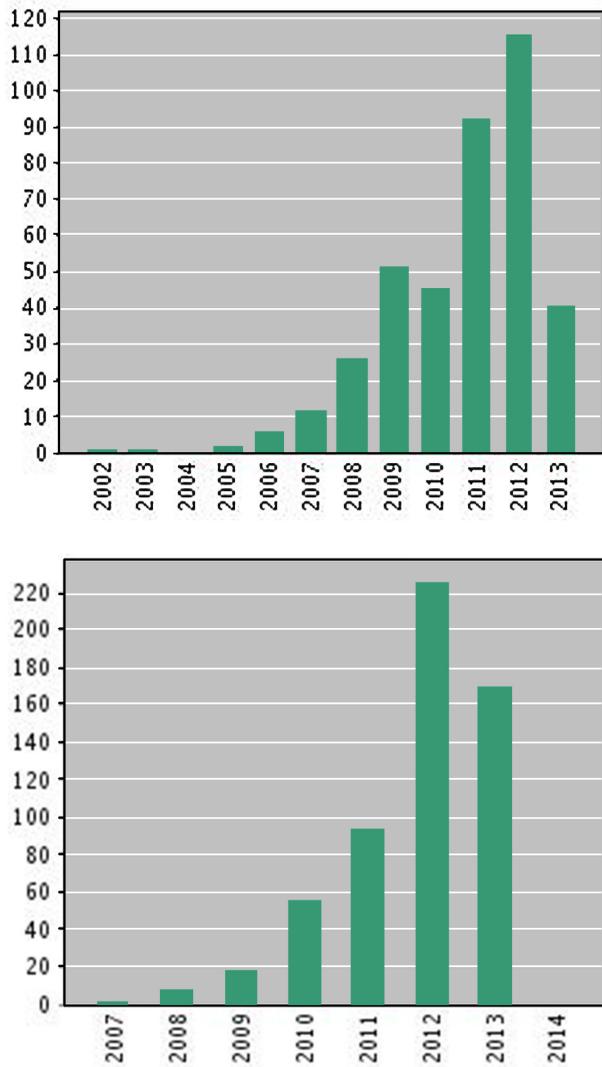
A pesquisa em abrangência foi a busca pelo termo *sentiment analysis* na ferramenta selecionada *Web of Science*, em todas as bases, sem restrições. Os parâmetros são ilustrados na tabela 1.

A consulta retornou 396 trabalhos. O gráfico da esquerda na figura 1 mostra a distribuição de itens publicados por ano, e o da direita mostra citações por ano.

Tabela 1 – Parâmetros para pesquisa em abrangência.

Palavras-chave:	Sentiment Analysis
Itens:	Título, Resumo e Palavras-chave
Bases de Dados:	Science Citation Index Expanded (SCI-EXPANDED) --1945-presente Social Sciences Citation Index (SSCI) --1956-presente Arts & Humanities Citation Index (A&HCI) --1975-presente Conference Proceedings Citation Index - Science (CPCI-S) --1990-presente Conference Proceedings Citation Index - Social Science & Humanities (CPCI-SSH) --1990-present
Sentença:	Tópico=(“sentiment analysis”) Tempo estipulado=Todos os anos. Bases de dados=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH.

Figura 1 – Gráficos com itens publicados por ano à esquerda, e citações por ano à direita.



Nota-se que as publicações tiveram aumento significativo a partir de 2007, e que as citações neste tema começam a crescer em 2009 com aumento significativo em 2012. Tem-se, portanto, o primeiro item para refinamento da pesquisa: data de publicação de 2008 até o presente (agosto de 2013).

A tabela 2 mostra a distribuição quanto ao tipo de documento. Apesar do número de *Proceedings Paper* ser significativamente maior, este trabalho filtrou os resultados por artigos (ARTICLE) pelas características de relevância apontadas por González-Albo e Bordons, (2011) e por apresentar número significativo de trabalhos: 141.

Tabela 2 – Distribuição quanto ao tipo de documento

Tipos de documento	Contagem do registro	% de 396
PROCEEDINGS PAPER	244	61,62%
ARTICLE	141	35,61%
REVIEW	13	3,28%
EDITORIAL MATERIAL	3	0,76%

Do resultado, 394 trabalhos foram publicados no idioma inglês.

Figura 2 – Publicações por fonte

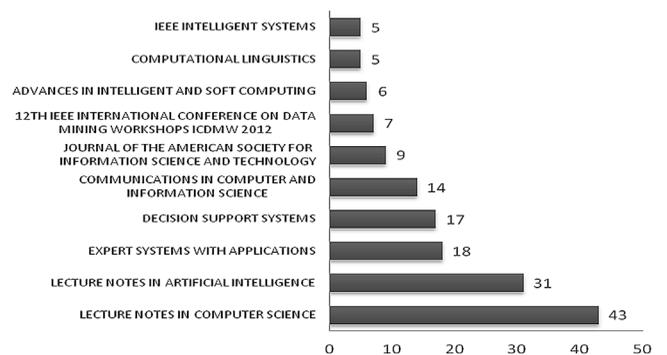
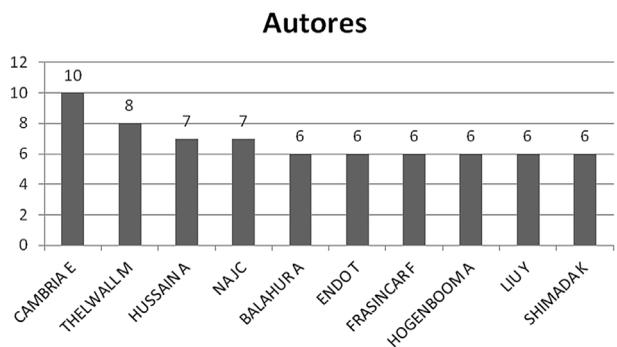


Figura 3 – Publicações por Autores



O maior número de publicações se deu no periódico *Lecture Notes in Computer Science*, com 10,86% das publicações. Destaque para *Expert System With Applications* (4,55%) e *Decision Support Systems* (4,29%), com fator de impacto JCR respectivamente 1.854 e 2.201, somando 8,84% das publicações.

Ambas classificadas pela Capes como Qualis A1 para a área de administração, turismo e ciências contábeis na classificação de periódicos 2012 e A1 para a área de avaliação interdisciplinar, o que mostra a abrangência do tema. Os principais autores são ilustrados no gráfico da figura 3.

Destaque para Cambria E. e Thelwall, M., ambos com publicações recentes, conforme tabela 3, que mostra as 10 publicações mais citadas do resultado. Destaque para o periódico *Computational Linguistics*.

Tabela 3 – Artigos mais citados

Citação/Título	Autores	Título da fonte	Ano da publicação	Total de citações	Média por ano
(Prabowo e Thelwall, 2009) Sentiment analysis: A combined approach	Prabowo, R.; Thelwall, M.	JOURNAL OF INFORMETRICS	2009	28	5.60
(Wilson, Wiebe e Hoffmann, 2009) Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis	Wilson, T.; Wiebe, J.; Hoffmann, P.	COMPUTATIONAL LINGUISTICS	2009	20	4.00
(Taboada <i>et al.</i> , 2011) Lexicon-Based Methods for Sentiment Analysis	Taboada, M; Brooke, J.; Tofiloski, M.; Voll, K.; Stede, M.	COMPUTATIONAL LINGUISTICS	2011	19	6.33
(Thelwall, Wilkinson e Uppal, 2010) Data Mining Emotion in Social Network Communication: Gender Differences in MySpace	Thelwall, Mike; Wilkinson, David; Uppal, Sukhvinder	JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY	2010	19	4.75
(Tan e Zhang, 2008) An empirical study of sentiment analysis for chinese documents	Tan, S.; Zhang, J.	EXPERT SYSTEMS WITH APPLICATIONS	2008	19	3.17
(Li e Wu, 2010) Using text mining and sentiment analysis for online forums hotspot detection and forecast	Li, N.; Wu, D. D.	DECISION SUPPORT SYSTEMS	2010	17	4.25
(Chmiel <i>et al.</i> , 2011) Negative emotions boost user activity at BBC forum	Chmiel, A; Sobkowicz, P.; Sienkiewicz, J.; Paltoglou, G.; Buckley, K.; Thelwall, M.; Holyst, Janusz A.	PHYSICA A-STATISTICAL MECHANICS AND ITS APPLICATIONS	2011	16	5.33
(Abbasi, Chen e Salem, 2008) Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums	Abbasi, A.; Chen, H.; Salem, A.	ACM TRANSACTIONS ON INFORMATION SYSTEMS	2008	16	2.67
(Archak, Ghose e Ipeirotis, 2011) Show me the Money! Deriving the Pricing Power of Product Features by Mining Consumer Reviews	Archak, N.; Ghose, An.; Ipeirotis, Panagiotis G.	KDD-2007 PROCEEDINGS OF THE THIRTEENTH ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING	2007	16	2.29
Learning to laugh (automatically): Computational models for humor recognition	Mihalcea, R; Strapparava, C	COMPUTATIONAL INTELLIGENCE	2006	15	1.88

A pesquisa em abrangência retrata um panorama geral das publicações sobre o tema principal deste trabalho: *Sentiment Analysis*. O resultado é abrangente, mas sua análise subsidia o refinamento para a próxima etapa em busca da resposta à questão colocada como objetivo do trabalho. Os parâmetros para refinamento são descritos na próxima etapa.

AMOSTRA POR REFINAMENTO

De acordo com Smith e Egger (1998), a revisão sistemática é buscada em fontes ou bancos de dados predeterminados, o que não significa que serão investigadas todas as fontes existentes no campo da pesquisa. Logo, esta etapa reduz a amostra pesquisada aplicando um filtro com parâmetros definidos após leitura dos resultados da etapa anterior e com as palavras-chave que convergem para o objetivo do trabalho. A tabela 4 descreve os parâmetros.

Tabela 4 – Parâmetros para refinamento da pesquisa

Atributo	Valor
Palavras-chave	Sentiment Analysis e Ontology e Web
Itens	Título, Resumo e Palavras-chave
Bases de Dados	Science Citation Index Expanded (SCI-EXPANDED) --1945-presente Social Sciences Citation Index (SSCI) --1956-presente Arts & Humanities Citation Index (A&HCI) --1975-presente Conference Proceedings Citation Index - Science (CPCI-S) --1990-presente Conference Proceedings Citation Index - Social Science & Humanities (CPCI-SSH) --1990-present
Tipo de Documento	Artigo
Anos	2009 a 2013 (agosto)
Idioma	Inglês
Sentença	Tópico=(“sentiment analysis”) OR Tópico=(“opinion mining”) AND Tópico=(ontology) AND Tópico=(web) Refinado por: Tipos de documento=(ARTICLE) AND Anos da publicação=(2012 OR 2013 OR 2009 OR 2011 OR 2010) AND Idiomas=(ENGLISH) Tempo estipulado=Todos os anos. Bases de dados=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH.
Ordenação	Por número de citações do mais para menos

O operador ‘e’ filtra os artigos, retornando somente aqueles que possuam, além do termo *sentiment analysis*, os outros dois: *ontology* e *web*. O uso das aspas nos termos com duas palavras garante que serão pesquisados somente trabalhos com as duas palavras na formação do termo. A pesquisa foi com o marcador Tópico que significa consultar a presença dos termos no título, resumo ou palavras-chave.

A pesquisa procurou trabalhos em um intervalo de tempo no qual as publicações e citações começaram a aumentar significativamente: de 2009 a 2013. O resultado da etapa anterior mostra uma diversidade de periódicos relevantes em que o tema é publicado, em diferentes áreas, portanto optou-se por não restringir tais atributos.

A consulta retornou uma amostra com apenas 6 artigos. O resultado é exibido na tabela 5.

Tabela 5 – Artigos selecionados para avaliação

Título	Autores	Título da fonte	Ano da publicação	Total de citações	Média por ano
Sentic Computing for social media <i>marketing</i>	Cambria, Erik; Grassi, Marco; Hussain, Amir; Havasi, Catherine	MULTIMEDIA TOOLS AND APPLICATIONS	2012	4	2.00
Feature-based Product Review Summarization Utilizing User Score	Yang, Jung-Yeon; Kim, Han-Joon; Lee, Sang-Goo	JOURNAL OF INFORMATION SCIENCE AND ENGINEERING	2010	3	0.75
Business Ontology for Evaluating Corporate Social Responsibility	Smeureanu, Ion; Diosteanu, Andreea; Delcea, Camelia; Cofas, Liviu	AMFITEATRU ECONOMIC	2011	1	0.33
Ontology-based sentiment analysis of twitter posts	Kontopoulos, Efstratios; Berberidis, Christos; Dergiades, Theologos; Bassiliades, Nick	EXPERT SYSTEMS WITH APPLICATIONS	2013	0	0.00
A Multi-agent Based System for Sentiment Analysis of User-Generated Content	Kechaou, Zied; Ben Ammar, Mohamed; Alimi, Adel M.	INTERNATIONAL JOURNAL ON ARTIFICIAL INTELLIGENCE TOOLS	2013	0	0.00
A framework for building web mining applications in the world of <i>blogs</i> : A case study in product sentiment analysis	Costa, Evandro; Ferreira, Rafael; Brito, Patrick; Bittencourt, Ig Ibert; Holanda, Olavo; Machado, Aydano; Marinho, Tarsis	EXPERT SYSTEMS WITH APPLICATIONS	2012	0	0.00

ANÁLISE DA AMOSTRA

As palavras *ontology* e *web* restringiram significativamente o resultado. O artigo mais antigo foi publicado em 2010, sendo o único da amostra no mesmo ano. Três artigos não tiveram citações. Uma justificativa é a publicação recente. Dois foram publicados no ano de realização desta pesquisa (2013) e um em 2012. Destaque para o artigo mais citado do conjunto, com quatro citações, publicado em 2012. Percebe-se que o conjunto de temas é recente.

A ocorrência de dois artigos da amostra no periódico *Expert Systems With Applications* reforça-o este como fonte relevante para pesquisas sobre o tema.

Com o objetivo de responder à questão proposta, a tabela 6 apresenta características observadas para análise e conclusões.

Tabela 6 – Síntese dos artigos selecionados

Título	Lacunas	Objetivo	Ontologia
(Cambria <i>et al.</i> , 2012) Sentic Computing for social media marketing	Apenas o valor de polaridade associado a uma opinião pode ser restritivo em mídias sociais, é necessária análise automática com aspectos afetivos. E as abordagens existentes são ineficientes, pois dependem de partes de texto onde as opiniões são explicitamente expressas como termos positivos e negativos e não consideram o contexto e domínio específico da opinião.	Desenvolver ferramenta para gerenciar informação de mídia social em nível semântico, capaz de capturar polaridade de opiniões e informação afetiva associada ao conteúdo gerado pelo usuário.	Desenvolveram a <i>human emotion ontology</i> (HEO) conjunto de conceitos e propriedades para representar emoções humanas em diferentes contextos
(Yang, Kim e Lee, 2010) Feature-based Product Review Summarization Utilizing User Score	Limitação nos métodos de análise de sentimento existentes para determinar polaridade de sentimento em palavras sensíveis ao contexto, e limitação nos métodos de pontuação quando apenas a pontuação geral do usuário é utilizada para avaliar as características de cada produto.	Utilizar informação sensível ao contexto para determinar a polaridade do sentimento e frequência de opinião é utilizada para pontuar as características de produtos em comentários sobre os mesmos na Web.	Não utilizam explicitamente o termo ontologia, mas um dicionário de palavras negativas e positivas para auxiliar a análise de polaridade dos comentários. O que se aproxima de ontologia enquanto formalização de um vocabulário específico.
(Smeureanu <i>et al.</i> , 2011) Business ontology for evaluating corporate social responsibility	Necessidade de determinar o desempenho de uma organização não só por fatores econômicos, mas também por responsabilidade social Dificuldade de considerar todos os fatores necessários para classificar o nível de responsabilidade social de uma empresa.	Determinar o nível de responsabilidade social de uma empresa analisando automaticamente com um software baseado em ontologia o seu site e outros que ilustram o papel da empresa no cenário econômico e social como sites com ofertas de emprego e informações financeiras.	Desenvolveram uma ontologia baseada na norma ISO 2600 para responsabilidade social, formalizando seus conceitos para que o software extraia termos ligados a responsabilidade social da corporação e possa classificá-la usando redes neurais.
(Kontopoulos <i>et al.</i> , 2013) Ontology-based sentiment analysis of twitter posts	Classificadores de sentimento em texto de <i>microblogs</i> frequentemente se mostram ineficientes, mensagens do Twitter, pela limitação de caracteres, tipicamente não tem palavras representativas e sintaticamente consistentes. Eles tratam cada mensagem como uma afirmação uniforme atribuindo um valor de sentimento para o todo.	Desenvolvimento de técnicas baseadas em ontologias para uma análise de sentimento mais eficiente de mensagens no Twitter, determinando o assunto discutido nas mensagens analisando cada uma em um conjunto de aspectos relevantes ao assunto.	Utilizam as técnicas de FCA e Ontology Learning para criar uma ontologia de domínio que será a base com conceitos e propriedades para proceder a análise de sentimentos do conjunto de mensagens do Twitter.
(Kechaou, Ben Ammar e Alimi, 2013) A multi-agent based system for sentiment analysis of user-generated content	Descobrir como as pessoas se sentem e reagem lidando com determinados tópicos em <i>blogs</i> , fóruns e outros meios na Web	Desenvolver um sistema multi-agente baseado em análise linguística principalmente destacando as principais contribuições dadas por estudos como este em combinação com análise sintática, semântica e subjetiva.	Apesar de conter a palavra ontologia nas palavras-chave, não foi observada utilização.
(Costa <i>et al.</i> , 2012) A framework for building web mining applications in the world of <i>blogs</i> : A case study in product sentiment analysis	Aplicativos com técnicas de mineração de dados para explorar bases de dados no contexto de mídias sociais são desenvolvidas de forma <i>ad hoc</i> principalmente pela falta de ferramentas adequadas gerando dificuldades para personalizar aplicações e demandando tempo para construir e manter tais aplicações.	Desenvolvimento de uma arquitetura de software de propósito geral para construir aplicações de mineração de dados na Web especificamente em <i>blogs</i> .	Utilizam ontologia para descrever <i>web services</i> , parte da arquitetura do software, com a finalidade de promover a capacidade de identificar serviços compatíveis não apenas os de uma interface específica A arquitetura tem uma camada de ontologia com 4 classes que podem ser estendidas para representação de conhecimento para mineração de dados

O mecanismo utilizado para definir uma ontologia na maioria dos trabalhos é a linguagem de marcação OWL (Ontology Web Language) construída sobre a metalinguagem XML (eXtensible Markup Language), possui recursos específicos para formalização de uma ontologia em meio digital. Cinco dos seis trabalhos da amostra coletada utilizam OWL.

No caso de Yang, Kim e Lee (2010) não há evidência de criação de uma ontologia, conceitualmente ou com uma linguagem para tal. Entretanto, o uso de um dicionário para formalizar os termos que indicam polaridade negativa ou positiva se aproxima do conceito de ontologia.

Os problemas de pesquisa na maioria dos artigos têm suas peculiaridades, mas compartilham uma lacuna comum em relação a metodologias que capturem o contexto para análise de sentimento, ou uma carga semântica adicional além da sintática. Os resultados em busca da resolução deste problema se mostram satisfatórios com a definição e o uso de ontologias.

Em Costa et al. (2012), além do uso de ontologias para identificar *web services* para que uma arquitetura computacional seja extensível, a própria arquitetura permite a construção de ontologias específicas para atuar como uma aplicação no campo de conhecimento do usuário.

Os *blogs* e *microblogs* como o *Twitter* aparecem como fonte significativa de dados para análise de sentimentos, ao passo que contêm mensagens textuais sobre diversos assuntos, constituindo uma massa de dados não estruturados rica em informações que refletem opiniões e sentimento.

CONCLUSÕES

Pelo conjunto de artigos, observa-se que o uso de ontologia é relevante para a análise de sentimentos em diferentes contextos com dados não estruturados extraídos da Web, principalmente como conceito para formalizar um vocabulário específico de um domínio e definições de entidades e relacionamentos. Os dados carregam componente semântico além do sintático, o uso de ontologias promove inferências refinadas por parte de algoritmos, que não ficam restritos a consultas sintáticas ou determinação de polaridade (positivo, negativo e neutro) por classificação, mas também podem identificar emoções humanas mais específicas, como em Cambira, et al. (2012), que definem a *Human Emotion Ontology*.

Os trabalhos utilizam ontologia como recurso para representar conhecimento de diferentes áreas, o que se aproxima mais da abordagem por orientação semântica com dicionários léxicos para análise de sentimento. Entretanto, apesar do aumento significativo de publicações sobre análise de sentimento a partir de 2009, a consulta por trabalhos que incluem também ontologia retornou uma amostra reduzida, o que sugere a investigação com esta abordagem em fase inicial.

Verifica-se que existem aplicações de análise de sentimento em várias áreas, mas a maioria dos meios de divulgação com trabalhos que combinam ontologia e análise de sentimento é relacionada à ciência da computação – inteligência artificial e sistemas especialistas, respectivamente, *International Journal on Artificial Intelligence Tools e Expert Systems with Applications*.

O tema é incipiente em ciência da informação, pois aparece uma publicação em meio relacionado na pesquisa em abrangência – *Journal of the American Society for Information Science na Technology*, entretanto está entre os trabalhos mais citados, o que aponta a relevância da área e sugere que o tema pode ser campo fértil para contribuições em ciência da informação.

Mesmo verificando ser tema incipiente, pela análise dos trabalhos selecionados, é possível responder à pergunta objetivo deste trabalho com a seguinte afirmação: o uso de ontologia para análise de sentimentos em dados não estruturados da Web contribui para capturar o contexto da informação e permite uma análise com características semânticas. Neste sentido, pode-se concluir que a sua utilização é uma opção viável para aplicações de análise de sentimento onde o contexto deve ser considerado.

É importante destacar que o sentido amplo da palavra ontologia pode levar a divergências e questionamentos quanto ao uso conceitual adequado do ponto de vista filosófico, mas mesmo que seja uma redução do conceito, é útil para a representação de dados e conhecimento em computação e ciência da informação.

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, 2008.
- ABECKER, Andreas et al. Towards a well-founded technology for organizational memories. In: *AAAI SPRING SYMPOSIUM ON ARTIFICIAL INTELLIGENCE IN KNOWLEDGE MANAGEMENT*, 1997. Proceedings... [S.l.: s.n.], 1997. P. 24-26.
- AGUWA, C. C.; MONPLAISIR, L.; TURGUT, O. Voice of the customer: customer satisfaction ratio based analysis. *Expert Systems with Applications*, v. 39, n. 11, p. 10112-10119, Sept. 2012.
- ARCHAK, N.; GHOSE, A.; IPEIROTIS, P. G. Deriving the pricing power of product features by mining consumer reviews. *Management Science*, v. 57, n. 8, p. 1485-1509, 2011.
- BACCIANELLA, S.; ESULI, A.; SEBASTIANI, F. SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: *INTERNATIONAL LANGUAGE RESOURCES AND EVALUATION*, 7., 2010, Malta. Proceedings... [Luxembourg]: ELRA, 2010. P. 2200-2204.
- BERNERS-LEE, Tim et al. The semantic web. *Scientific American*, v. 284, n. 5, p. 28-37, 2001.
- BLANCO, Carlos et al. A systematic review and comparison of security ontologies. In: *INTERNATIONAL CONFERENCE ON IEEE: availability, reliability and security, ARES 08*, 2008. [S.l.: s.n.], 2008. P. 813-820.
- BOTELHO, L. L. R.; CUNHA, C. C. de Almeida; MACEDO, M. O método da revisão integrativa nos estudos organizacionais. *Gestão e Sociedade*, v. 5, n. 11, p. 121-136, 2011.
- CAMBRIA, E. et al. Sentic computing for social media marketing. *Multimedia Tools and Applications*, v. 59, n. 2, p. 557-577, July 2012. Disponível em: <Go to ISI>://WOS:000304134000007 >. Acesso em: 15 dez. 2016.
- CHMIEL, A. et al. Negative emotions boost user activity at BBC forum. *Physica A: Statistical Mechanics and its Applications*, v. 390, n. 16, p. 2936-2944, 2011. Disponível em: <http://www.scopus.com/inward/record.url?eid=2-s2.0-79956350539&partnerID=40&md5=232b91c69f433b74932972d766bd04f8 >. Acesso em: dez. 2016.
- CHOI, Y. et al. Identifying sources of opinions with conditional random fields and extraction patterns. In: *CONFERENCE ON HUMAN LANGUAGE TECHNOLOGY AND EMPIRICAL METHODS IN NATURAL LANGUAGE*, 1., 2005, Stroudsburg. Proceedings... Stroudsburg: Association for Computational Linguistics, 2005. P. 355-362.
- CONFORTO, E. C.; AMARAL, D. C.; SILVA, S. Roteiro para revisão bibliográfica sistemática: aplicação no desenvolvimento de produtos e gerenciamento de projetos. In: *CONGRESSO BRASILEIRO DE GESTÃO DE DESENVOLVIMENTO DE PRODUTOS*, 8., 2011, Porto Alegre. Anais... Porto Alegre: [s.n.], 2011.
- COSTA, E. et al. A framework for building web mining applications in the world of blogs: A case study in product sentiment analysis. *Expert Systems with Applications*, v. 39, n. 5, p. 4813-4834, Apr. 2012. Disponível em: <Go to ISI>://WOS:000301155300017>. Acesso em: 15 dez. 2016.
- CRUZ, F. L. et al. Building layered, multilingual sentiment lexicons at synset and lemma levels. *Expert Systems with Applications*, v. 41, n. 13, p. 5984-5994, Oct. 2014.
- DAS, S. R.; CHEN, M. Y. Yahoo! for Amazon: sentiment extraction from small talk on the web. *Management Science*, v. 53, n. 9, p. 1375-1388, Sept. 2007.
- DAVE, K.; LAWRENCE, S.; PENNOCK, D. M. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In: *INTERNATIONAL CONFERENCE ON WORLD WIDE WEB*, 12., 2003, Budapest. Proceedings... New York: ACM, 2003. P. 519-528. Organization and structure of information using semantic web technologies. *Handbook of human factors in web design*, 2003.
- DING, X.; LIU, B. The utility of linguistic rules in opinion mining. In: *ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE*, 30., 2007, Amsterdam. Proceedings... New York: ACM, 2007.
- DOBSON, Glen; SAWYER, Peter. Revisiting ontology-based requirements engineering in the age of the semantic web. In: *INTERNATIONAL SEMINAR ON DEPENDABLE REQUIREMENTS ENGINEERING OF COMPUTERISED SYSTEMS AT NPPS*, 2006. Proceedings... [S.l.: s.n.], 2006. P. 27-29.

- EGGER, M.; SCHNEIDER, M.; DAVEY SMITH, G. Spurious precision?: meta-analysis of observational studies. *Bmj*, v. 316, n. 7125, p. 140-144, 1998.
- FELLBAUM, C. *WordNet*. New York: Wiley Online Library, 1998.
- GARCÍA-CUMBRERAS, M. Á.; MONTEJO-RÁEZ, A.; DÍAZ-GALIANO, M. C. Pessimists and optimists: improving collaborative filtering through sentiment analysis. *Expert Systems With Applications*, v. 40, n. 17, p. 6758-6765, Dec. 2013.
- GOLBECK, J. et al. Organization and structure of information using semantic web technologies. *Handbook of human factors in web design*, 2003.
- GONZÁLEZ-ALBO, B.; BORDONS, M. Articles vs. proceedings papers: do they differ in research relevance and impact?: a case study in the Library and Information Science field. *Journal of Informetrics*, v. 5, n. 3, p. 369-381, 2011.
- GRUBER, T. R. Toward principles for the design of ontologies used for knowledge sharing?. *International Journal of Human-Computer Studies*, v. 43, n. 5, p. 907-928, 1995.
- GRUNINGER, M.; LEE, J. Ontology-applications and design. *Communications of the ACM*, v. 45, n. 2, p. 39-41, 2002.
- HU, M.; LIU, B. Mining and summarizing customer reviews. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 10., 2004, Seattle. *Proceedings...* New York: ACM, 2004. P. 168-177.
- HUANG, K.-W.; LI, Z. A multilabel text classification algorithm for labeling risk factors in SEC form 10-K. *ACM Transactions on Management Information Systems*, v. 2, n. 3, p. 18-39, Aug. 2011.
- HUMPHERYS, S. L. et al. Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, v. 50, n. 3, p. 585-594, Feb. 2011.
- KAMPS, J. et al. Using wordnet to measure semantic orientations of adjectives. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 4., 2004, Lisboa. *Proceedings...* [Luxembourg]: ELRA, 2004.
- KECHAOU, Z.; BEN AMMAR, M.; ALIMI, A. M. A multi-agent based system for sentiment analysis of user-generated content. *International Journal on Artificial Intelligence Tools*, v. 22, n. 2, Apr 2013. Disponível em: <Go to ISI>://WOS:000317848400004 >. Acesso em: 15 dez. 2016.
- KONTOPOULOS, E. et al. Ontology-based sentiment analysis of twitter posts. *Expert Systems with Applications*, v. 40, n. 10, p. 4065-4074, Aug 2013. Disponível em: <Go to ISI>://WOS:000317162900023 >. Acesso em: 15 dez. 2015.
- KRAVET, T.; MUSLU, V. Textual risk disclosures and investors' risk perceptions. *Review of Accounting Studies*, v. 18, n. 4, p. 1088-1122, Dec. 2013.
- LI, N.; WU, D. D. Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems*, v. 48, n. 2, p. 354-368, 2010.
- LIU, B. Sentiment analysis: a multi-faceted problem. *IEEE Intelligent Systems*, v. 25, n. 3, p. 76-80, 2010.
- _____; HU, M.; CHENG, J. Opinion observer: analyzing and comparing opinions on the web. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 14., 2005, Chiba. *Proceedings...* New York: ACM, 2005. P. 342-351.
- LIU, Y. et al. Identifying helpful online reviews: a product designer's perspective. *Computer-Aided Design*, v. 45, n. 2, p. 180-194, Feb. 2013.
- MARTÍN-VALDIVIA, M.-T. et al. Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches. *Expert Systems with Applications*, v. 40, n. 10, p. 3934-3942, Aug. 2013.
- MITRA, G. et al. Automated analysis of news to compute market sentiment: its impact on liquidity and trading. 2011. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/d;jsessionid=9884AA0F751CF8ECA91D2BDA898FF0CA?doi=10.1.1.372.9679&rep=rep1&type=pdf>>. Acesso em: 20 abr. 2016.
- MONTOYO, A.; MARTINEZ-BARCO, P.; BALAHUR, A. Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. *Decision Support Systems*, v. 53, n. 4, p. 675-679, Nov. 2012.
- PANG, B.; LEE, L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, v. 2, n. 1-2, p. 1-135, 2008.
- PRABOWO, R.; THELWALL, M. Sentiment analysis: a combined approach. *Journal of Informetrics*, v. 3, n. 2, p. 143-157, 2009. Disponível em: <<http://www.scopus.com/inward/record.url?eid=2-s2.0-62349089279&partnerID=40&md5=3b307784d6ea8f60b0a5f6019b75a993>>. Acesso em: 15 dez. 2015.
- SMEUREANU, I. et al. Business ontology for evaluating corporate social responsibility. *Amfiteatru Economy*, v. 13, n. 29, p. 28-42, 2011.
- STONE, P. J.; DUNPHY, D. C.; SMITH, M. S. *The general inquirer: a computer approach to content analysis*. Cambridge: MIT, 1966
- TABOADA, M. et al. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, v. 37, n. 2, p. 267-307, 2011. Disponível em: <<http://www.scopus.com/inward/record.url?eid=2-s2.0-79958257877&partnerID=40&md5=5efd9716088009db6c9adb366be7860>>.
- TAN, S.; ZHANG, J. An empirical study of sentiment analysis for chinese documents. *Expert Systems with Applications*, v. 34, n. 4, p. 2622-2629, 2008.

THELWALL, M.; WILKINSON, D.; UPPAL, S. Data mining emotion in social network communication: gender differences in MySpace. *Journal of the American Society for Information Science and Technology*, v. 61, n. 1, p. 190-199, 2010. Disponível em: <<http://www.scopus.com/inward/record.url?eid=2-s2.0-72849107485&partnerID=40&md5=930c86a1404e3c09a1966c5c0d46c313>>.

TSYTSARAU, M.; PALPANAS, T. Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, v. 24, n. 3, p. 478-514, 2012.

TURNEY, P. D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: ANNUAL MEETING ON ASSOCIATION FOR COMPUTATIONAL LINGUISTIC, 40., 2002, Philadelphia. Proceedings... Stroudsburg: Association for Computational Linguistics, 2002. P. 417-424.

USCHOLD, M.; GRUNINGER, M. Ontologies: Principles, methods and applications. *The Knowledge Engineering Review*, v. 11, n. 2, p. 93-136, 1996.

WILSON, T. et al. OpinionFinder: a system for subjectivity analysis. In: HLT/EMNLP ON INTERACTIVE DEMONSTRATIONS, 2005, Vancouver. Proceedings... Stroudsburg: Association for Computational Linguistics, 2005. P. 34-35.

_____; WIEBE, J.; HOFFMANN, P. Recognizing contextual polarity in phrase-level sentiment analysis. In: CONFERENCE ON HUMAN LANGUAGE TECHNOLOGY AND EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2005, Sapporo. Proceedings... Stroudsburg: Association for Computational Linguistics, 2005. P. 347-354.

YANG, J.-Y.; KIM, H.-J.; LEE, S.-G. Feature-based product review summarization utilizing user score. *Journal of Information Science and Engineering*, v. 26, n. 6, p. 1973-1990, Nov. 2010. Disponível em: <Go to ISI://WOS:000284740500003>.

YU, H.; HATZIVASSILOGLOU, V. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2003, Sapporo. Proceedings... Stroudsburg: Association for Computational Linguistics, 2003. P. 129-136.