International Journal of Advances in Computer Science & Its Applications – IJCSIA
Volume 4: Issue 4 [ISSN 2250-3765]
Publication Date : 27 December,2014

A Sentiment Analysis Model Using Ontology-enriched Conceptual Graph and Operational Rules

[Tung Thanh Nguyen, Tho Thanh Quan]

***Abstract*—This paper presents a sentiment analysis model which combines conceptual graph enriched by domain ontology which expert-specified operational sentiment rules.Whereas conceptual graph and domain ontology allows automatic shalow parsing of natural sentences, operatioanal sentiment rules allows the sentiment system captures the linguistic knowledge provided by experts and integrates this knowledge seamlessly into the sentiment analysis process. We found that domain ontologies are extremely helpful when applied in specific domains. In the meantime, operational sentiment rules make our system achieve better performance as compared to data mining techniques.**

Keywords—Sentiment analysis, conceptual graph, domain ontology, operational sentiment rules.

I. Introduction

A. Sentiment Analysis

Sentiment analysis([1][1]) or *opinion mining* ([2]) is the task that aims to infer the *sentiment orientation* in a document ([3]). There are three levels of sentiment analysis as follows: (i) document-based level; (ii) sentence-based level; and (iii) aspect-based level.

Data mining/machine learning techniques are commonly used to infer sentiment opinion of a whole document. Those techniques always require raining sets. The popular techniques used in literature include *Naive Bayes* ([4],[5],[6],[7],[8]); *Support Vector Machine* (SVM)([5], [6], [7], [9],[10]); *Maximal Entropy* ([11],[12],[13]); n-gram model ([8]). Data mining/machine learning techniques achieved good performance for document-based sentiment analysis, but generally failed to handle sentence-based level, due to the lack of linguistic processing. To handle this problem, Yu and Hatzivassiloglou ([14]) suggested to select a set of adjectives as the *seeds* used to infer sentiment implication of a sentence. This set of *seeds* can be expanded using *distributional similarity* ([15]). McDonald et.al. ([16]) proposed to use *Conditional Random Field* method, which was later extended for a set of labeled sentences ([17]).

Naryanana et. al. ([18]) argued that each kind of level should be treated differently when performed sentiment analysis. Not only that, the *discourse information* would also need to be taken into account to analyze multiple linked sentences. The common approach to handle this problem is to use a lightweight NLP techniques such as shallow parsing to perform *pattern mining* for compound sentence ([19],[20],[21]). Zirn et. al. ([22]) also proposed to use *Markov logic network* to handle this problem.

In this paper, we focus on sentiment analysis at aspect-level posed the problems of recognizing aspects and inferring aspect-associated sentiment opinions. Thus it is deemed more complex than the two above levels. *Topic modeling* was commonly applied to handle both problems at the same time ([23]). Mei et. al. ([24]) proposed to use *Probabilistic Latent Semantic Analysis(PLSA*), while most of recent works were based on *Latent Dirichlet Analysis* (*LDA*) ([25],[26],[27],[28]).

B. Conceptual Graph for Natural Language Processing

Conceptual graph (CG) is a logical system can express meaningin a form that is humanly readable, and computationally tractable. Thus, it is used commonly in translating from natural language to computer oriented formalism. Many searching systems used CG to proccess natural language queries. [29] and [30] used CG to represent a query as a graph of objects then retrieve meaning and generate a formal query. Some works used a modified sort of CG to analyse sentiment for targets ([23]).

In target-dependent sentiment analysis, domain ontology is a very important knowledge-base to indentify entities, features and opinion ([32]). Ontology is also used in building and restructuring CGs of text ([29], [31]). In our system, we developed a general ontology and domain ontologies for certain industries. An ontology will capture information and sentiment terms related to objects and their attributes.

In this paper we introduce to a semantics-oriented method to classify sentiment at clause level ([34]), which focuses on target-dependent approach ([33]). To assign correctly sentiment for given targets, we use ontology and conceptual graph together combinate with sentiment operational semantic rule which will be discuss detail later. This paper is organized as follows. Section 2 recalls some preliminaries concept. Section 3 introduces our sentiment analysis model. Section 4 presents the architecture of our SAS (Sentiment Analysis System). Section 5 discusses some experimental results. Finally, Section 6 concludes the paper.

Tung Thanh Nguyen
Hochiminh City University of Technology
Hochiminh City, Vietnam

Tho Thanh Quan
Hochiminh City University of Technology
Hochiminh City, Vietnam

II. Preliminaries

A. *Conceptual Graph*

Definition 1 (Basic Conceptual Graph). A basic conceptual graph (BG) defined over a vocabulary $V = (T_c, T_r, I)$, is a 4-tuple $G = (C, R, E, l)$ satisfying the following conditions:

- (C, R, E, l) is a finite, undirected and bipartite multigraph called the underlying graph of G, denoted $graph(G)$. C is the concept node set, R is the relation node set (the node set of G is $N = C \cup R$). E is the family of edges.
- l is a labeling function of the nodes and edges of $graph(G)$ that satisfies: (i) A relation node r is labeled by $l(r) \in T_r$. $l(r)$ is also called the type of r and is denoted by $type(r)$; (ii) The degree of a relation node r is equal to the arity of $type(r)$ and (iii) Edges incident to a relation node r are totally ordered and they are labeled from 1 to $arity(type(r))$. ([38])

Example 1. Fig. 1shows a CG representingthe sentence "*I like smartphone A*". The CG consists two concept nodes ("I" is a subject, "smart phone A" is an object) and one relation node ("like").

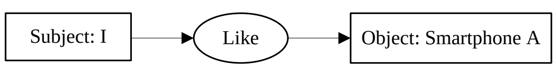


Figure 1. Aconceptual graph of sentence "I like smartphone A"

In fact, many sentences are written in complicated structures including clauses connected by conjunctions. In this case, complicated sentences can be expressed in an extended form of CG, known as *nested conceptual graph* (NCG).

Definition 2 (Nested Conceptual Graph). An elementary NCG G is obtained from a normal BG H by adding a third field to the label of each concept node c equal to **. The set of boxes of G is $boxes(G) = \{H\}$ and the complete concept node set of G is $X_G = C_H$. A trivial bijection exists between elementary NCGs and normal CGs (when no ambiguity occurs we do not distinguish between them).

- Let H be an NCG and D an elementary NCG. The graph G obtained from H by substituting D for the third field ** of a concept node c in X_H is an NBG. $boxes(G) = boxes(H) \cup boxes(D)$, and $X_G = X_H \cup X_D$.
- The third field of any concept node $c \in X_G$ is called the description of c and is denoted $Descr(c)$. ([38])

Example 2. Fig.2 shows a NCG of sentence "*I like smartphone A but I buy smartphone C of Brand B*". The NCG is represented as a tree with the root is relation node "*But*" and two branches are two clauses of the sentence.

B. *Ontology*

Building a CG from natural language sentences requires a domain knowledge which can help recognize concepts as well as relationship of concepts in a sentences. The most important domain knowledge is ontology. A domain ontology will capture information in a certain industry including brands, product lines, paticular products, features, typical sentiment terms and interrelationships of those.

A CG of a sentence can be built in two steps (for more detail, refer[29]):

- Extract sentiment phrases from the sentence according to patterns ([35], [36]). A sentiment term may be a verb, an adjective or a special type of word which is typical for a feature.
- Recognize concepts in the sentence and restructure CG according to ontology schema.

In the following examples, we use the concept of T-Box and A-Box to describe the ontologies in Fig. 3. Basically, T-Box captures the relations between concepts and A-Box describes individuals (or instances) of concepts.

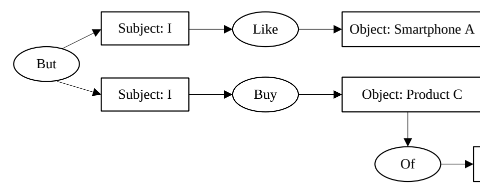
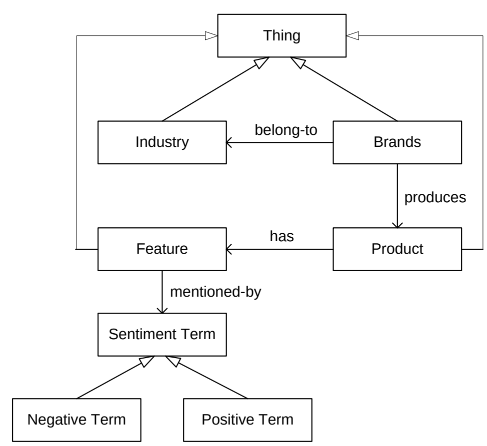
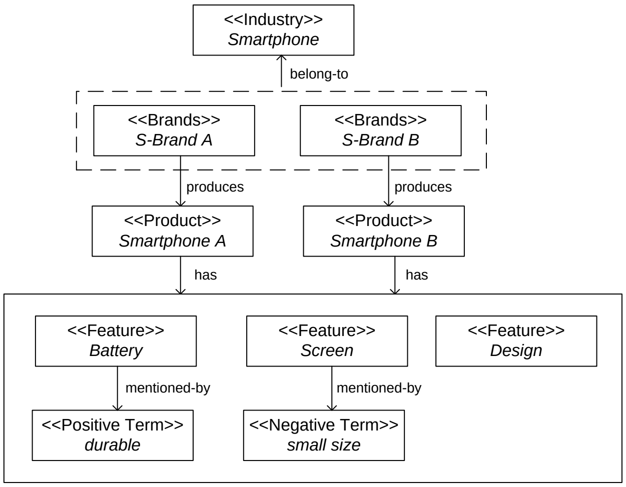


Figure 2. A nested conceptual graph of sentence "I like smartphone A but I buy smartphone C of Brand B"



(a) The T-Box



(b)

The A-Box

Figure 3. An example of Industry Ontology

Example 3. Fig. 3 gives an example of an Industry Ontology OS. The T-Box shows that in an Industry, there may have some Brands. Each Brand can produce many Products and each Product has various Features. All of these concepts are subconcepts of Thing in the Generic Ontology, i.e. they can be mentioned by positive or negative terms. Apparently, this T-Box is shared by all of Industry Ontologies.

The A-Box of this Industry Ontology shows that in fact this ontology describes concepts in the industry of smartphones. There are two brands, namely *S-Brand A* and *S-Brand B*, which respectively produce products of *Smartphone A* and *Smartphone B*. A smartphone product has some certain features, such as *Battery* and *Screen*. While inherited sentiment terms from the concept of *Thing*, this *Industry Ontology* also introduces some new domain-specific sentiment terms, such that *durable* is a positive term with regard to *Battery*, whereas *small size* is a negative term to *Screen*.

III. Sentiment Analysis Model

Definition 1.2 (Sentiment analysis model). A sentiment model is a 3-tuple of (C, O, Δ) where:

- C is a nested conceptual graph which is built from a text.
- O is a domain ontology.
- Δ is a transition of $node(C) \times Eva_O \rightarrow node'(C) \times Eva_O{}'$, where Eva_O is a valuation of each concept in Ointo {*positive*, *negative*,*neutral, unknown*}.

Each pair $(node(c), Eva_O)$ is called a *state*. Sentiment value of concepts in CG will be calculated when the system travels over states. Basically, the system selects a starting state, travelling order and termination condition according to *operational sentiment rules*define by human expert. Table I illustrates some operational semantic rules which will be used in the following examples.

Example 4. We revisite the CG in Fig.2. At the beginning, the sentiment value of concepts are initialized with *unknown*, whereas *Brand A* can be infered from domain ontology as a producer of *Smartphone A*:

$$(\text{null}, \text{Eva}_{O_1} = \{\text{brand A: unknown}, \text{smartphone A: unknown}, \text{brand B: unknown}, \text{product C: unknown}\})$$

When processing node "*buy*", sentiment valuesare inferred and assignedinto corresponding concepts as follows:

$$(\text{buy}, \text{Eva}_{O_2} = \{\text{brand A: unknown}, \text{smartphone A: unknown}, \text{brand B: positive}, \text{product C: positive}, \})$$

In the next state, at the node "*but*", according to the operational sentiment rule, all concepts in front of "but" will be assign negation of sentiment value of the clause behind "but". Beside that, "but" is also a terminal node so the system will close after sentiment analysis at node "but" finish.

$$(\text{but}, \text{Eva}_{O_2}) = \{\text{brand A: negative}, \quad \text{smartphone A: negative}, \text{brand B: positive}, \text{product C: positive}\}$$

$$(TER, Eva_{O_2}) = \{\text{brand A: negative}, \quad \text{smartphone A: negative}, \text{brand B: positive}, \text{product C: positive}\}$$

TABLE I. EXEMPLAR SENTIMENT RULES

Rule	Operation	Description
Buy(A,B)	Concept(B) = positive	In "*A buy B*" clause, all concepts behind "buy" will be positive.
Like(A,B)	Concept(B) = positive	In "*A like B*" clause, all concepts behind "*like*" will be positive.
Hate(A,B)	Concept(B) = positive	In "*A hate B*" clause, all concepts behind "*hate*" will be negative.
But(A,B)	(Concept(A) = ¬ Concept(B)) ∧ terminate	In "*A but B*" clause, concepts in front of "but" will be assigned negation of sentiment value of clause behind "*but*" and terminate sentiment processing.
Defeat(A,B)	Concept(A) = positive, Concept(B) = negative	In "*A defeat B*" clause, concepts in front of "*defeat*" will be assigned positive and concepts behind "*defeat*" will be assigned negative.

IV. Sentiment AnalysisSystem

Fig. 4 presents our sentiment analysis system (SAS). To be adapted from a generic semantic search model into the specific sentiment search problem, our SAS framework is enhanced with the following features:

Generic Ontology and Industry Ontologies. In SAS, a common *Generic Ontology* is constructed, which captures domain-independent sentiment terms. Apart from that, a set of *domain ontologies* are also developed. Since our framework is specifically intended for social monitoring systems, which are often used to monitor brands and product in various industries like *Smartphones* or *Babycares*, each domain ontology should capture information in a certain industry. Thus, we regard our domain ontologies as *Industry Ontologies*. We also develop an incremental strategy to automatically enrich an Industry Ontology from the Generic Ontology and a set of training documents, which is introduced as an external module of *Incremental Update* in SAS.

Query processing. The query processing in SAS is merely keyword-based. The submitted keywords are names of industry, brands or products that users want to observe sentiment opinions from the public. Firstly, traditional keyword-based techniques are applied in SAS to retrieve relevant documents. Then sentiment analysis is performed to infer the sentiment opinion of the retrieved data.

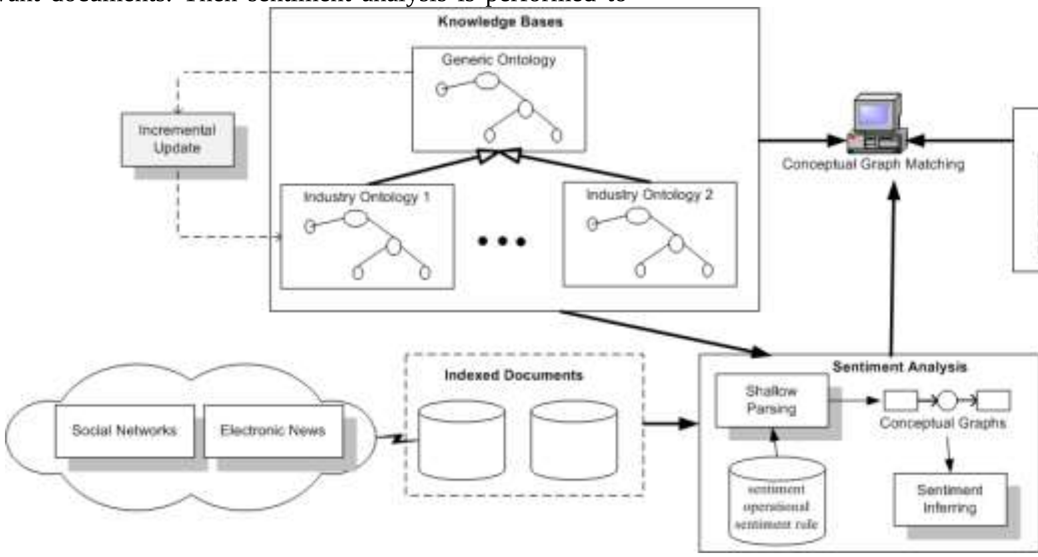


Figure 4. The SAS Architecture

V. Experiments

Among eight datasets, there were 2 dataset of generic domains and the remaining 6 datasets belonging to various industries. For generic domains, we developed a *Generic Ontology* capturing 767 sentiment words and phrases; among which there are 375 positive phrases and 392 negative phrases.

For each specific industry, we constructed a corresponding *Industry Ontology*. The concepts captured in each Industry Ontology are brands and products in the corresponding industry. For example, for *Smartphone Ontology*, the brands included*Apple, HTC,* etc. and the products included *iPhone 5s*, *iPad Air*, *HTC One Max*, *HTC Sensation*, etc. Thus, the datasets and ontologies reflected a real demand of brand managers who always wanted to observe the opinion of users regarding their products.

We then measured the accuracy of our sentiment classification approach. As compared to traditional framework of generic semantic search, we enhanced our sentiment search by introducing our Sentiment Phrasing Rules and the enrichment of Industry Ontology. Therefore, we did compare the performance of various sentiment analysis strategies as follows.

- CSS-FULL: we applied our full CSS framework.
- CSS-GEN: we only used Generic Ontology in the CSS framework
- CSS-NO-RULES: we did not use Sentiment Phrasing Rules in the CSS framework.
- SVM: we used SVM for sentiment classification, as this technique was employed by various related works.

Fig.5 presents the accuracy percentage when we applied those analysis strategies on the collected datasets. It can be observed that in generic domains like *Amway* or *Mobifone*, the accuracy performances of CSS-FULL and CSS-GEN were more or less the same. However, in specific domains, where corresponding Industry Ontologies were constructed properly, CSS-FULL outperformed all of other strategies.

It is notable that SVM could compete with CSS-GEN in domains where neutral data were dominant, e.g. *Mobifone*, *Beer* or *Banking*. It can be explained that in neutral data, the occurrence frequencies of sentiment phrases were not high, thus SVM can demonstrated its capabilities of recognizing irrelevant samples (i.e. recognizing samples without sentiment opinions). However, when the occurrences of sentiment terms become large, SVM achieved poor performance due to the complexity of the language structures, which can negate the sentiment meaning. This property was also reflected via the fact that CSS-NO-RULES and SVM virtually achieved the same performance in all datasets.

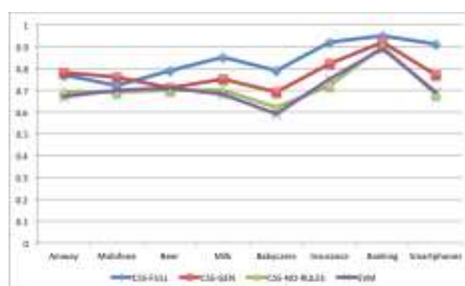


Figure 5. Accuracy performance of sentiment analysis strategies

VI. Conclusion

In this paper, we present a hybrid approach for sentiment analysis. In one hand, we apply the classical technique of using conceptual graphs and domain ontology for shallow parsing non-complex (but non-trivial) natural sentences. In the other hand, we introduce the concept of sentiment analysis model with enhances conceptual graph with operational sentiment rules. By doing so, we can combine the powerful means of shallow parsing by CG, domain knowledge of ontology and expertise knowledge of sentiment rules. The experiments have shown that our approach outperforms the most popular technique of SVN in this area.

Acknowledgement

This work is supported by the research project B0212-20-02TD funded by Vietnam National University – Ho Chi Minh City. We are also grateful to YouNet Media for supporting real datasets for our experiments.

References

[1] Nasukawa, T, & Yi, J, "Sentiment Analysis: Capturing Favorability Using Natural Language Processing", Proceedings of the 2nd International Conference on KnowledgeCapture, pp. 70-77, 2003.

[2] Dave, K, Lawrence, S, & Pennock, M, "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews", Proceedings of the 12thInternational Conference on World Wide Web, pp. 519-528, 2003.

[3] Padmaja, S, Fatima, S, "Opinion Mining and Sentiment Analysis - An Assessment of Peoples' Belief: A Survey,ternational Journal of Ad hoc", Sensor & Ubiquitous Computing, pp. 21–33, 2013.

[4] Melville, P, Gryc, W, & Lawrence, R. D, "Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification", Proceedings of the 15th ACM SIGKDDInternational Conference on Knowledge Discovery and Data Mining, pp. 1275-1284, 2009.

[5] Xia, R, Zong, C, & Li, S, "Ensemble of Feature Sets and Classification Algorithms for Sentiment Classification", Information Sciences, pp. 1138-1152, 2011.

[6] Zhang, Z, Ye, Q, Zhang, Z, & Li, Y, "Sentiment Classification of Internet Restaurant Reviews Written in Cantonese", Expert Systems with Applications, pp. 7674-7682, 2011.

[7] Tan, S, & Zhang, J, "An Empirical Study of Sentiment Analysis for Chinese Documents", Expert Systems with Applications, pp. 2622-2629, 2008.

[8] Ye, Q, Zhang, Z, & Law, R, "Sentiment Classification of Online Reviews to Travel Destinations by Supervised Machine Learning Approaches", Expert Systems with Applications, pp. 6527-6535, 2009.

[9] Prabowo, R, & Thelwall, M, "Sentiment Analysis: A Combined Approach", Journal ofInformetrics, pp. 143-157, 2009.

[10] Xu, K, Liao, S. S, Li, J, & Song, Y, "Mining Comparative Opinions from Customer Reviews for Competitive Intelligence", Decision support systems, pp. 743-754, 2011.

[11] Go, A, Bhayani, R, & Huang, L, "Twitter Sentiment Classification Using Distant Supervision", CS224N Project Report, Stanford, pp. 1-12, 2009.

[12] Gindl, S, & Liegl, J, "Evaluation of Different Sentiment Detection Methods for Polarity Classification on Web-Based Reviews", Computational Aspects of Affectual andEmotional Interaction, 2008.

[13] Shimada, K, & Endo, T, "Seeing Several Stars: A Rating Inference Task for A Document Containing Several Evaluation Criteria", Advances in Knowledge Discovery andData Mining, pp. 1006-1014, 2008.

[14] Yu, H, & Hatzivassiloglou, V, "Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences", Proceedings of the2003 Conference on Empirical Methods in Natural Language Processing, pp. 129-136, 2003.

[15] Lin, D, "Automatic Retrieval and Clustering of Similar Words", Proceedings of the17th International Conference on Computational Linguistics, vol. 2, pp. 768-774, 1998.

[16] McDonald, R, Hannan, K, Neylon, T, Wells, M, & Reynar, J, "Structured Models for Fine-to-Coarse Sentiment Analysis", Annual Meeting-Association for ComputationalLinguistics, 2007.

[17] Täckström, O, & McDonald, R, "Discovering Fine-Grained Sentiment with Latent Variable Structured Prediction Models", Advances in Information Retrieval, pp. 368-374, 2011.

[18] Narayanan, Liu & Choudhary, "Sentiment Analysis of Conditional Sentences", Proceedings of the 2009 Conference on Empirical Methods in Natural LanguageProcessing, vol. 1, pp. 180-189, 2009.

[19] Asher, N, Benamara, F, & Mathieu, Y, "Distilling Opinion in Discourse: A Preliminary Study", Proceedings of the International Conference on ComputationalLinguistics, pp. 7-10, 2008.

[20] Somasundaran, S, Ruppenhofer, J, & Wiebe, J, "Discourse Level Opinion Relations: An Annotation Study", Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue, pp. 129-137, 2008.

[21] Zhou, L, Li, B, Gao, W, Wei, Z, & Wong, K. F, "Unsupervised Discovery of Discourse Relations for Eliminating Intra-Sentence Polarity Ambiguities", Proceedings ofthe Conference on Empirical Methods in Natural Language Processing, pp. 162-171, 2011.

[22] Zirn, Niepert, Stuckenschmidt, & Strube, "Fine-Grained Sentiment Analysis with Structural Features", Proceedings of the 5th International Joint Conferenceon Natural Language Processing, pp. 336-344, 2011.

[23] Guang Qiu, Bing Liu, Jiajun Bu, Chun Chen, "Opinion Word Expansion and Target Extraction through Double Propagation," Computational Linguistics, Vol. 37, No. 1, Pages 9-27, March 2011.

[24] Mei, Q, Ling, X, Wondra, M, Su, H, & Zhai, C, "Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs", Proceedings of the 16th International Conference onWorld Wide Web, pp. 171 180, 2007.

[25] Li, F, Huang, M, & Zhu, X, "Sentiment Analysis with Global Topics and Local Dependency", Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, 2010.

[26] Zhao, Jiang, Yan,& Li, X, "Jointly Modeling Aspects and Opinions with a MaxEnt-LDA Hybrid", Proceedings of the 2010 Conference on Empirical Methods inNatural Language Processing, pp. 56-65, 2010.

[27] Sauper, C, Haghighi, A, & Barzilay, R, "Content Models with Attitude", Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, 2011.

[28] Mukherjee, A, &Liu, B, "Aspect Extraction through Semi-Supervised Modeling", Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, vol. 1, pp. 339-348, 2012.

[29] Tho Thanh Quan, Siu Cheung Hui, "Ontology-based Natural Query Retrieval using Conceptual Graphs," 10th Pacific Rim International Conference on Artificial Intelligence proceedings, pp. 15-19, 2008.

[30] Tru H. Cao, Hai T. Do, Bao T.N. Pham, Tuyen N. Huynh and Duy Q.Vu, "Conceptual Graphs for Knowledge Querying in VN-KIM", 13th International Conference on Conceptual Structures, pp. 27-40, 2005.

[31] Tru H. Cao, Truong D. Cao, Thang, "A Robust Ontology-Based Method for Translating Natural Language Queries to Conceptual Graphs," 3rd Asian Semantic Web Conference proceeding, pp. 479-492, 2008.

[32] Lili Zhao, Chunping Li, "Ontology Based Opinion Mining for Movie Reviews" Third International Conference, KSEM 2009 proceeeding, pp 204-214, November, 2009.

[33] Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, Tiejun Zhao, "Target-dependent Twitter Sentiment Classification", Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 151–160, June, 2011.

[34] Theresa Wilson, Janyce Wiebe, Rebecca Hwa, "Just how mad are you? Finding strong and weak opinion clauses", in Proceedings of AAAI, pp. 761–769, 2004. (Extended version in Computational Intelligence, vol. 22, no. 2, pp. 73–99, 2006.

[35] Peter D. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews", Proceedings of the 40th Annual Meeting on Association for Computational Linguistics,pages 417-424, 2002.

[36] Basant Agarwal, Vijay Kumar Sharma, and Namita Mittal, "Sentiment Classification of Review Documents using Phrase Patterns",Advances in Computing, Communications and Informatics (ICACCI), pages 1577 – 1580, 2013.

[37] Maite Taboada ,Julian Brooke ,Milan Tofiloski ,Kimberly Voll ,Manfred Stede, "Lexicon-Based Methods for Sentiment Analysis", MIT Press Journals Vol. 37 No. 2, Pages 267-307, 2011.

[38] Michel Chein and Marie-Laure Mugnier, "Graph-based Knowledge Representation - Computational Foundations of Conceptual Graphs", Springer, 2009.