Journal of Theoretical and Applied Information Technology
31st January 2017. Vol.95. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ASPECT-BASED OPINION SUMMARIZATION : A SURVEY

[1]WARIH MAHARANI, [2]DWI H. WIDYANTORO, [3]MASAYU L. KHODRA
[1]School of Electrical Engineering and Informatics
Institut Teknologi Bandung , West Java, Indonesia
E-mail: [1]wmaharani@telkomuniversity.ac.id, [2]dwi@stei.itb.ac.id, [3]masayu@stei.itb.ac.id

ABSTRACT

With the growing popularity of Web 2.0, social media is becoming the largest source of information. Because of the huge number of unstructured reviews, it is impossible to summarize all this information manually. Therefore, efficient computational methods are needed for mining and summarizing the reviews to produce a representative summarization. We presents a detailed and systematic overview of the last update in the aspect-based opinion summarization, including state-of-the-arts and methods that widely used in aspect-based opinion summarization. This paper also describes a comparison of several methods in summary generation, including text-based and visual-based opinion summarization. Finally, this paper presents some research opportunities and challenges in aspect-based opinion summarization.

Keywords: *Web 2.0; Social Media; Opinion Summarization; Aspect-Based Summarization.*

1. INTRODUCTION

With the explosive growth of social media, people can share their opinions easily in their blogs, microblogs, comments, forum discussions and social network sites. Nowadays, people tend to explore user reviews and discussions in a forum on the e-commerce websites before they buy a product. However, due to the number of user reviews, it is difficult for the customer to find a proper review in accordance to user needs. It encourages research on opinion summarization [1, 14, 24, 32]. Opinion summarization can be viewed as a multi-document summarization [22]. The main purpose of opinion summarization is to help customers in finding opinions easily and help in decision-making. Organization and companies can leverage user reviews so that they do not need to conduct surveys of customer satisfaction. It can be used for their business development.

In general, there are two approaches in opinion summarization: traditional-based summarization and aspects-based summarization [22]. The method used in traditional-based summarization similar with the text summarization methods. It produces an opinion summarization by extracting some important sentences. Whereas aspect-based summarization involves the aspect and sentiments about them. Moreover, it is a quantitative summarization, which means that it involves the number of aspects and opinion pairs. Aspect-based opinion summarization can be divided into three main subtasks i.e. (1) aspect and opinion extraction, (2) sentiment classification and (3) opinion summarization [13]. Aspect and opinion extraction is the most widely discussed in many studies in the last decade, because this is the first subtask that can affect the next subtask. However, the survey paper that specifically discusses on opinion summarization is still limited, especially research on aspect-based opinion summarization. Whereas current product reviews on the web are extensive growing, especially coupled with the development of sentiment analysis studies. Therefore, we present the most recent or state-of-the-art papers in aspect-based opinion summarization. Kim [16] provided a comprehensive opinion summarization, starting with state of the art, techniques and approaches used in opinion summarization. It is also presented a summary generation techniques and visualization technique as well as performance evaluation. Meanwhile, Liu discussed the fundamental of opinion mining and presented a comprehensive survey in opinion mining and sentiment analysis [24]. Nevertheless, these surveys are more focused on a general opinion summarization studies, which are limited to articles conducted before 2012. The previous survey papers presented a general opinion summarization studies. They concluded that there

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

are many kinds of techniques have been proposed, including aspect extraction methods and sentiment classification methods. However, there is no survey paper that analyzed the visual representations of aspect-based opinion summarization. In addition, there have been no survey papers that specifically discuss in detail about the aspect-based opinion summarization. The purpose of this paper is to fill the existing gap, by providing a detailed review and different insight about aspect-based opinion summarization. Therefore, in this paper, we focus on aspect-based opinion summarization with detailed review, starting with aspect extraction, sentiment classification and summary generation. This paper also describes a comparison of several methods for summary generation. The contribution of this paper is the comparison among different summary generation techniques, elaborated the effectiveness in different domains and languages. We also present a brief description of some visual representations of opinion summarization.

The paper is organized as follows: Section 2 presents the general opinion summarization; Section 3 describes in detail the aspect-based opinion summarization approach and comparative analysis of the methods used in previous studies, specifically various visualization of opinion summarization; finally, Section 4 concludes the paper and challenges as well as opportunities for further research.

2. OPINION SUMMARIZATION

According to Liu [22, 23], opinion is a quintuple (o_j, f_{jk}, so_{ijkl}, h_i, t_l), where o_j is an object, f_{jk} is an aspect of objects o_j, so_{ijkl} is the opinion on aspect f_{jk} of the objects o_j, h_i is a person who expresses the opinion (opinion holder) and t_l is the time when the opinions is expressed by h_i [22, 23]. Sentiment so_{ijk} generally categorized into positive, negative and neutral opinion. Based on this definition, all information in the quintuple associated with each other to generate the structure of the target and opinion. In general, there are several points that differentiate text summarization and opinion summarization. These differences can be seen in Table 1 [29, 32].

2.1 Traditional-based Opinion Summarization

Traditional-based opinion summarization does not focus on aspects of a particular object. Some researchers used the extractive and abstractive approach as in text summarization [43]. Traditional-based opinion summarization will produce a summary that does not consider the aspects and its opinions, but it will extract the most important sentences in the document.

Wang exploited traditional approach to summarize data phone conversation [10]. This approach selects the most important sentences from data phone conversations. The disadvantage of this approach is that it is not considered the objects, aspects and its opinions, so they may produce a summary which is not related to objects, aspects and its opinions. It may be useless for users if they want to know about the pros and cons of an aspect and its opinions.

Table 1: Text and Opinion Summarization

Criteria	Text Summarization	Opinion Summarization
Focus	Extraction of important words	Generally focused on the objects, aspects and their opinion (o_j, f_{jk}, s_{oijkl}). h_i, t_l is not compulsory
Structure	Ordered time	More structured with objects segmentation, aspects and their opinion Not necessarily sequential time, except for summarization with timeline-based approach
Source	News articles, scientific papers	Blog, forum products and services, reviews
Output	Text	Text or visual
The Level of Compression	1% - 30%	Not considered

2.2 Aspect-based Opinion Summarization

Aspect-based opinion summarization involves objects, aspects, and followed by opinion [22, 23]. Some domains widely used in this research are electronic product reviews [13, 14], movie reviews [54], hotel services reviews, restaurant reviews and reviews from travel agency [51], as well as specifics topics such as phone conversations or blogs. According to Figure 2, there are 3 basic tasks in aspect-based opinion summarization: aspect extraction, sentiment classification and opinion summaries generation. Each task can be performed with several approaches such as machine learning and natural language processing.

The purpose of aspect extraction is to identify and extract topics, objects, aspects and opinions from a document. Some researchers performed a feature extraction task to generate features such as term frequency (number of occurrences of the term in the document), term co-occurrence (a feature that appears together, such as unigram/n-gram), POS (Part of Speech), opinion term (words that express positive or negative opinion), negation

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and syntactic dependency. While the sentiment classification determines aspects, sentences or document into positive, negative and neutral polarity. The third task is the summary generation which generates a document summary based on previous tasks.

3. ASPECT-BASED SUMMARIZATION

In the aspect-based opinion summarization, opinions will be categorized based on objects, aspects and its opinions, with a different level such as aspect level, sentence level or document level. The representations of the aspect-based summary can be text or visual-based. Aspect-based opinion summarization covers both object, aspect and its opinion polarity. This paper focuses on summary generation in aspect-based opinion summarization. But since aspect-based opinion summarization involves aspect extraction and sentiment classification tasks, therefore in this paper we will discuss many methods which developed in each of these subtasks at a glance.

In general, there are several approaches that can be performed in aspect-based opinion summarization as shown in Figure 1.

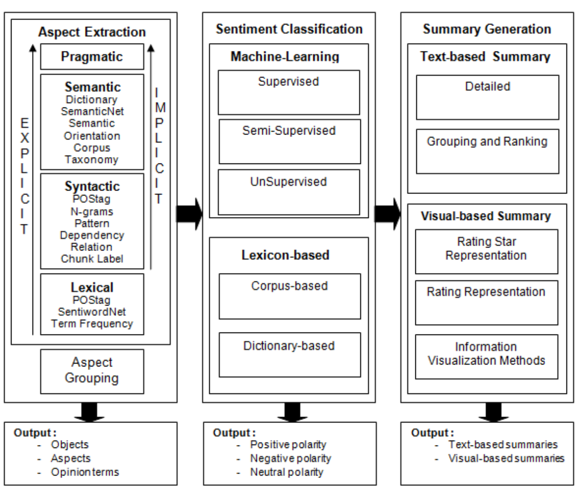


Figure 1: The main tasks in Aspect-based Opinion Summarization

There are several approaches that can be used to generate a summary based on previous research. Opinion summaries can be represented in text-based summary and visual-based summary. In the next section, this paper will focus on summary generation methods, beginning with a general discussion of aspect extraction and sentiment classification.

3.1 Aspect Extraction

In general, there are two types of aspects contained in data review, namely explicit aspect expression and implicit aspect expression. Explicit aspect extraction has been widely explored in many researches. However, still limited work has been done in extracting implicit aspects. Extraction can also be done at some levels of languages, such as lexical, syntactic, semantic and pragmatic. Various techniques have been used for aspect extraction, i.e frequency-based [13, 19, 22, 25, 30, 34], lexicon-based [8, 50], syntactic-based [7, 17, 18, 37, 42, 54] and learning-based [15, 49].

Aspect extraction was first studied by Hu & Liu [13]. They extracted aspects using association rule mining to find frequent item set [13, 14]. The idea is that they extract nouns/noun phrases, which is widely mentioned in the data review as candidate-aspect. This method is quiet simple and effective. Although this method is simple and quietly effective, however, this method has not been able to find implicit features and low-frequency aspect. However, this method has been developed in many researches [19, 25, 30, 34]. Popescu [34] evaluated the extracted noun phrase by using a PMI score. Ku [19] used a TFIDF scheme by using other frequency-based approach. Moghaddam [30] improved the frequency-based approach with an additional pattern rule to remove noise terms.

The drawback of the frequency-based method was improved by using the lexicon-based approach [8]. Ding [8] proposed a holistic lexicon-based approach to solving the previous problem, which improves the frequency-based method in [13, 14]. Instead of looking at the current sentence alone, this approach exploits external information and evidence in other sentences and other reviews, and some linguistic conventions in natural language expressions to infer orientations of opinion words. No prior domain knowledge or user inputs are needed. This approach has been proven that the method can extract the aspect which has a small frequency because the search is based on a noun closest to the opinion words. This approach became widely developed in subsequent studies such as Zhang [50].

The lexicon-based method depends on opinion lexicon being used. Therefore, many research develop syntactic-based approach, which is more simple and effective. Syntactic-based approach utilized the dependency relationships and rule-based pattern to extract list of aspect candidates. Syntactic-based approach was first developed by Turney, by utilizing POStag pattern [42]. Then,

Journal of Theoretical and Applied Information Technology
31st January 2017. Vol.95. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Zhuang [54] expanded the method by using dependency relationships generated by a dependency parser MINIPAR. Wiebe [38, 46] and Kobayashi [17, 18] presented a similar approach. And then Qiu developed Double propagation method, which is simple and effective in generating high recall [36, 37].

In addition to these three methods, learning-based approach also been developed in recent years because it promises a great performance [2, 26, 49]. Zhai [49] applied semi-supervised learning with Expectation Maximization algorithm that used sharing word and lexical similarity in WordNet to label data set automatically. Lu [26] used another approach to build a context-aware sentiment lexicon that utilized multiple sources such as a general-purpose sentiment lexicon, content-dependent sentiment lexicon, sentiment ratings, WordNet and linguistic heuristic rule: "and", "but" and negation [26]. The construction of sentiment lexicon started from some seeds that consist predetermined aspects. While Bross addressed the context-sensitive problems by using a taxonomy products and WordNet [2]. In general, machine learning approach can produce a relatively high performance in a particular domain.

3.2 Sentiment Classification

This task classifies sentences into positive, negative and neutral opinion. In general, sentiment classification can be done by three approaches: supervised [32, 51], semi-supervised [5, 12, 49] and unsupervised [10, 42].

In supervised learning, some researchers usually used Naïve Bayes [47, 51], Maximum Entropy [28, 47], Artificial neural Network (ANN), and Support Vector Machine (SVM) [28, 39, 51]. Pang [31] proposed supervised learning to classify data reviews into positive and negative opinion using Naïve Bayes, Maximum Entropy and SVM. Pang [31] used n-gram as subjectivity clues to classify polarity. Sheng [3] and Zhu [52] applied ANN to classify subjective opinions into positive and negative opinion, which used adjective and adverb term as features. While Zhu [52] used unigram and rule-based learning with back propagation learning algorithm to classify opinions. Ye [48] also used Naïve Bayes, SVM and N-grams to classify travel reviews.

Table 2 compares several research that used supervised learning algorithm to classify user review. The comparison used "*" , "***" and "*****" value to represent the method's performance in each research.

Table 2: Supervised LearningMethods

Data format	NB	ME	SVM	ANN
Free format (Cornell dataset) [41]	*	***	*****	-
Pros, cons & detail reviews [65]	*****	-	***	-
Free format [61]	***	-	*****	-
Free format (Cornell dataset) [47]	-	-	*****	-
Free format (Cornell dataset) [24]	-	-	*****	-
Free format [52]	***	-	*****	-
Free format (Cornell dataset) [66]	-	-	***	*****
Pros, cons & detail [60]reviews	*	***	*****	-
Free format [36]	***	**	*****	-

In general, SVM performed the best performance compared to Naïve Bayes, Maximum Entropy and ANN [28, 35, 39, 47, 48, 51]. However, Naïve Bayes produced better performance than SVM, for smaller dimension dataset [51]. The problems in applying supervised learning are data annotation effort for training examples and difficult to scale up to a large number of application domain. Therefore, some research using unsupervised learning to overcome this problem [10, 11, 20].

Turney applied unsupervised learning to classify user reviews into the recommended or not recommended [42]. They classified opinion polarity based on some syntactic pattern and the average value of the phrase semantic orientation. Semantic orientation calculated using Pointwise Mutual Information (PMI). Unsupervised learning approach has major advantages, which does not need annotation data for training examples. In general, the opinion summarization involving very large datasets, so it will take a great effort to manually annotate the dataset. Unsupervised learning approach can offer a solution in dealing with these problems. Thus, unsupervised learning has shown a quite well performance in a large-scale of domain. However, the performance generated by unsupervised approach is not as good as the supervised learning [9, 33, 34, 42].

Semi-supervised learning is performed to avoid some issues in supervised and unsupervised. Several research-based approach used opinion lexicon to classify user reviews [13, 19, 55]. Gamon [9] used semi-supervised learning to classify user reviews. The seed set used in this research initially with 10 terms, which consisted of 4 positive terms (good, excellent, love, happy) and 6 negative terms (bad, lousy, terrible, hate, said,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

unreliable). Number of negative terms more than positive terms because the classification performance on negative sentences lower than positive sentences. Seed set used to create a sentiment vocabulary of unlabeled data. Li [20] modified K-Means with 3 strategies which used TF-IDF weighting, voting mechanism, and term scoring mechanism. The purpose of TF-IDF weighting is to improve the accuracy, the voting mechanism is made to produce a more stable cluster, while term scoring mechanism is to improve the performance by combining the WordNet and clustering results [20].

3.3 Summary Generation

3.3.1 Text-based Opinion Summarization

Several aspect-based opinion summarization generate a text-based summary in accompanied by statistical data for each object/aspect. In addition, a summary accompanied by the number of sentences and their opinion, which is based on details aspect. Each pairs of aspects and their opinions are grouped based on positive and negative opinions [13] as shown in Figure 2.

```
Digital Camera:
  Aspect: GENERAL
     Positive:   100 sentences
     Negative:    36 sentences
  Aspect: Lens
     Positive:   150 sentences
     Negative:   77 sentences
  Aspect: Battery
     Positive:   180 sentences
     Negative:    96 sentences
…….
```

```
Digital Camera:
+ GENERAL (136)
- Battery (276)
  The battery life is long (180)
  The battery was not available anywhere (10)
  ……….
+ Lens  (227)
  ……….
```

Figure 2: Text-based Opinion Summarization

In general, text-based summary generation implemented through ranking stage based on the weight of opinion sentence [19, 33, 34]. Sentences with the highest weight will be displayed as the result of a summary. It is to overcome the weaknesses of the statistical approach, because of the details review that should be displayed. Sentences with the highest weight used as the most representative sentences from the document. This representation is suitable if it involves a large dataset, because it shows the most representative opinion sentences, more concise and effective for users because it focuses on the object/aspect and its opinion [10, 11, 43].

Das [6] used Information Retrieval (IR) based technique to identify the most "informed" sentences from any cluster and it can be termed as IR based cluster center for that particular cluster. Meanwhile Zhu proposed a graph-based method to identify informative sentences [53]. They formulated the informative sentence selection problem in opinion summarization as a community leader detection problem in social computing [53].

3.3.2 Visual-based Opinion Summarization

In addition to text-based summaries, opinion summaries can be represented in a visual form. Several previous studies generate visual-based summaries, such as ratings, stars, bars, values or other information visualization [19, 24, 27, 40]. Visual representation can provide a snapshot of the opinion more interesting, such as Opinion Observer [21]. Visual representation make the summaries quicker and easier to read than text representation.

The visualization will affects users' perceptions. Some well-known commercial website reviews present a product summarization with various visualizations, such as in Amazon, Yelp, TripAdvisor, Rotten Tomatoes, IMDb and eBay. They used several visualizations such as thumbs up or thumbs down, positive and negative signs, star rating and rating meter. Moreover, there are also different rating systems accompanied by an interesting polarity visualization, such on RottenTomatoes website.

Visualizations can related to the rating scales. Visualization such as design, shape, colors and font are features of a scale that are not necessarily essential, but help to give a visualization its "look and feel" [4, 41]. Users will use visual cues to aid them in their interpretation of response scale items. Information processing is related to human memory. A long term memory is formed by past interactions and experiences, which needs to be considered when designing the layout of a visualization.

Colin Ware, Director of the Data Visualization Research Lab at the University of New Hampshire stated that a human can distinguish differences in line length, shape orientation, and color (hue) readily without significant processing effort; these are referred to as "preattentive attributes" [45]. These attributes are categorized into : (1) form, which consists of orientation, line length, line

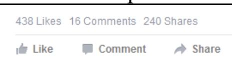
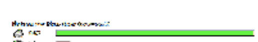
Journal of Theoretical and Applied Information Technology
31st January 2017. Vol.95. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

width, size, shape, curvature, added marks and enclosure, (2) color, which consists of intensity and hue, and (3) spatial position or 2-D position. Our eyes can catch these attributes immediately when we look at a visualization. They can be perceived in less than 10 milliseconds, even before we make a conscious effort to notice them. These attributes come into play when we analyze any visualization. Position and length attribute can used to perceive quantitative data with precision. The other attributes are useful for perceiving other types of data such as categorical, or relational data. For example, we can easily find the longest bar in bar chart, as it calls on the preattentive attribute of length.

Visual-based opinion summarization can be categorized as either quantitative or content based [44]. In this paper, we focus on the quantitative based visualization. This paper will divide the visualization into five categories: numbers, symbols, bars, rating meter and others. Table 3 provides several quantitative-based rating visualizations based on its preattentive attribute.

Table 3: Categories of Visualization

Categories	Visualization	Example of Visualization
A. Numbers	i. Number of user's likes	 www.facebook.com
	ii. Number in 10 rating scales	 www.tv.com
	iii. Number in 30 rating scales	 www.zagat.com
	iv. Number in 100 rating scales	 www.metacritic.com
B. Symbols	i. Thumbs up/down	 www.vanno.com
	ii. Plus/minus sign	
	iii. Smileys	 www.happy-or-not.com
	iv. Stars	 www.starburstmagazine.com
	v. Bullets	 www.tripadvisor.com
C. Bars	i. Bars in 5 rating scales	
	ii. Bars in 10 rating scales	
	iii. Numbers of user's likes in bars	
D.Rating meter	i. Rating meter	
E. Others	i. Tomatoes icon : form (shape), color (hue)	www.rottentomatoes.com

Journal of Theoretical and Applied Information Technology
31st January 2017. Vol.95. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. SUMMARY AND OPEN RESEARCH ISSUES

Along with an explosive growth of the social media, as well as the user reviews, thus the importance of opinion summarization research. The main tasks in opinion summarization include aspect extraction, sentiment classification and summary generation. Aspect extraction and sentiment classification commonly used both machine learning based and natural language based approach. As for the summary generation task will depend on focus, goals and users need. If a user is intended to focus on a shorter summary representation, clear and attractive, it is more appropriate to use visual-based approach. Whereas if a user is intended to find out all the details along with the growing number of opinion, it is more suitable to use the text-based approach with a statistical result. However, in this study we do not discuss the methods which used to generate the value rating that has been discussed in the summary generation. Despite of previous research, the current aspect-based opinion summarization research still have many limitations that can be improved for future studies. There are open research issues that can be develop for future research:

4.1 Open-domain opinion summarization

Most of existing studies work on a domain-spesific research, so how to develop an open-domain opinion summarization?

4.2 Context-based opinion summarization

Currents research leads to understand the sentence's context, so there are opportunities for further research to extend those research with several improvements using natural language processing techniques.

4.3 Big data analysis

The amount of dataset reviews on the internet leads to Big Data Analysis problems, so it can be opportunities in Big Data Analysis.

4.4 Personalized opinion summarization

Generation of summaries can be done with several approaches, but most existing studies have not focused on customized summary generation techniques with focus, goals and user needs. This could open up opportunities for future research that focus on the development of summaries generation according to specific's needs of users, which leads to the personalization of opinion summarization.

REFERENCES:

[1] Beineke, P. et al. 2004. Exploring sentiment summarization. *AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications (AAAI tech report SS-04-07)* (2004).

[2] Bross, J. and Ehrig, H. 2013. Automatic construction of domain and aspect specific sentiment lexicons for customer review mining. *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management* (2013), 1077–1086.

[3] Chen, L.-S. et al. 2011. A neural network based approach for sentiment classification in the blogosphere. *Journal of Informetrics*. 5, 2 (2011), 313–322.

[4] Couper, M.P. et al. 2004. What they see is what we get response options for web surveys. *Social science computer review*. 22, 1 (2004), 111–127.

[5] Dalal, M.K. and Zaveri, M.A. Semi-Supervised Learning based Opinion Summarization and Classification for Online Product.

[6] Das, A. and Bandyopadhyay, S. 2010. Topic-based Bengali opinion summarization. *Proceedings of the 23rd International Conference on Computational Linguistics: Posters* (2010), 232–240.

[7] Deng, L. and Wiebe, J. 2015. Mpqa 3.0: An entity/event-level sentiment corpus. *Conference of the North American Chapter of the Association of Computational Linguistics: Human Language Technologies* (2015).

[8] Ding, X. et al. 2008. A holistic lexicon-based approach to opinion mining. *Proceedings of the 2008 International Conference on Web Search and Data Mining* (2008), 231–240.

[9] Gamon, M. and Aue, A. 2005. Automatic identification of sentiment vocabulary: exploiting low association with known sentiment terms. *Proceedings of the ACL Workshop on Feature Engineering for Machine Learning in Natural Language Processing* (2005), 57–64.

[10] Ganesan, K. et al. 2010. Opinosis: a graph-based approach to abstractive summarization of highly redundant opinions. *Proceedings of the 23rd International Conference on Computational Linguistics* (2010), 340–348.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[11] Ganesan, K. and Zhai, C. 2012. FindiLike: preference driven entity search. *Proceedings of the 21st international conference companion on World Wide Web* (2012), 345–348.

[12] Goldberg, A.B. and Zhu, X. 2006. Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization. *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing* (2006), 45–52.

[13] Hu, M. and Liu, B. 2004. Mining and summarizing customer reviews. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (2004), 168–177.

[14] Hu, M. and Liu, B. 2006. Opinion extraction and summarization on the web. *AAAI* (2006), 1621–1624.

[15] Jin, W. et al. 2009. A novel lexicalized HMM-based learning framework for web opinion mining. *Proceedings of the 26th Annual International Conference on Machine Learning* (2009), 465–472.

[16] Kim, H.D. et al. 2011. Comprehensive review of opinion summarization. (2011).

[17] Kobayashi, N. et al. 2005. Collecting evaluative expressions for opinion extraction. *Natural Language Processing--IJCNLP 2004*. Springer. 596–605.

[18] Kobayashi, N. et al. 2007. Opinion mining from web documents: Extraction and structurization. *Information and Media Technologies*. 2, 1 (2007), 326–337.

[19] Ku, L.-W. et al. 2006. Opinion Extraction, Summarization and Tracking in News and Blog Corpora. *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs* (2006), 100–107.

[20] Li, G. and Liu, F. 2010. A clustering-based approach on sentiment analysis. *Intelligent Systems and Knowledge Engineering (ISKE), 2010 International Conference on* (2010), 331–337.

[21] Liu, B. et al. 2005. Opinion observer: analyzing and comparing opinions on the web. *Proceedings of the 14th international conference on World Wide Web* (2005), 342–351.

[22] Liu, B. 2012. Sentiment analysis and opinion mining: synthesis lectures on human language technologies. *Morgan & Claypool Publishers*. (2012).

[23] Liu, B. 2010. Sentiment analysis and subjectivity. *Handbook of natural language processing*. 2, (2010), 627–666.

[24] Liu, S. et al. 2009. Interactive, topic-based visual text summarization and analysis. *Proceedings of the 18th ACM conference on Information and knowledge management* (2009), 543–552.

[25] Long, C. et al. 2010. A review selection approach for accurate feature rating estimation. *Proceedings of the 23rd International Conference on Computational Linguistics: Posters* (2010), 766–774.

[26] Lu, B. et al. 2011. Multi-aspect sentiment analysis with topic models. *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on* (2011), 81–88.

[27] Lu, Y. et al. 2009. Rated aspect summarization of short comments. *Proceedings of the 18th international conference on World wide web* (2009), 131–140.

[28] Maharani, W. 2013. Microblogging sentiment analysis with lexical based and machine learning approaches. *Information and Communication Technology (ICoICT), 2013 International Conference of* (2013), 439–443.

[29] Mani, I. and Maybury, M.T. 1999. *Advances in automatic text summarization*. MIT Press.

[30] Moghaddam, S. and Ester, M. 2010. Opinion digger: an unsupervised opinion miner from unstructured product reviews. *Proceedings of the 19th ACM international conference on Information and knowledge management* (2010), 1825–1828.

[31] Pang, B. et al. 2002. Thumbs up?: sentiment classification using machine learning techniques. *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10* (2002), 79–86.

[32] Pang, B. and Lee, L. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*. 2, 1–2 (2008), 1–135.

[33] Popescu, A.-M. et al. 2005. OPINE: Extracting product features and opinions from reviews. *Proceedings of HLT/EMNLP on interactive demonstrations* (2005), 32–33.

[34] Popescu, A. and Etzioni, O. 2004. Extracting Product Features and Opinions from Reviews. (2004).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[35] Prabowo, R. and Thelwall, M. 2009. Sentiment analysis: A combined approach. *Journal of Informetrics*. 3, 2 (2009), 143–157.

[36] Qiu, G. et al. 2007. Expanding Domain Sentiment Lexicon through Double Propagation College of Computer Science Department of Computer Science University of Illinois at Chicago. (2007).

[37] Qiu, G. et al. 2011. Opinion word expansion and target extraction through double propagation. *Computational linguistics*. 37, 1 (2011), 9–27.

[38] Riloff, E. and Wiebe, J. 2003. Learning extraction patterns for subjective expressions. *Proceedings of the 2003 conference on Empirical methods in natural language processing* (2003), 105–112.

[39] Tan, S. and Zhang, J. 2008. An empirical study of sentiment analysis for chinese documents. *Expert Systems with Applications*. 34, 4 (2008), 2622–2629.

[40] Titov, I. and McDonald, R. 2008. A joint model of text and aspect ratings for sentiment summarization. *Urbana*. 51, (2008), 61801.

[41] Tourangeau, R. et al. 2004. Spacing, position, and order interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly*. 68, 3 (2004), 368–393.

[42] Turney, P.D. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. *Proceedings of the 40th annual meeting on association for computational linguistics* (2002), 417–424.

[43] Wang, D. and Liu, Y. 2011. A Pilot Study of Opinion Summarization in Conversations. *ACL* (2011), 331–339.

[44] Wang, J. et al. 2013. Clustered layout word cloud for user generated review. *Yelp Challenge, Virginia Polytechnic Institute and State University*. (2013).

[45] Ware, C. 2012. *Information visualization: perception for design*. Elsevier.

[46] Wiebe, J. 2000. Learning subjective adjectives from corpora. *AAAI/IAAI* (2000), 735–740.

[47] Xia, R. et al. 2011. Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences*. 181, 6 (2011), 1138–1152.

[48] Ye, Q. et al. 2009. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications*. 36, 3 (2009), 6527–6535.

[49] Zhai, Z. et al. 2011. Clustering product features for opinion mining. *Proceedings of the fourth ACM international conference on Web search and data mining* (2011), 347–354.

[50] Zhang, L. and Liu, B. 2011. Extracting Resource Terms for Sentiment Analysis. *IJCNLP* (2011), 1171–1179.

[51] Zhang, Z. et al. 2011. Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications*. 38, 6 (2011), 7674–7682.

[52] Zhu, J. et al. 2010. Sentiment classification using the theory of ANNs. *The Journal of China Universities of Posts and Telecommunications*. 17, (2010), 58–62.

[53] Zhu, L. et al. 2013. Graph-based informative-sentence selection for opinion summarization. *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on* (2013), 408–412.

[54] Zhuang, L. et al. 2006. Movie review mining and summarization. *Proceedings of the 15th ACM international conference on Information and knowledge management - CIKM '06*. (2006), 43.

[55] Zhuang, L. et al. 2006. Movie review mining and summarization. *Proceedings of the 15th ACM international conference on Information and knowledge management* (2006), 43–50.