Mining consumers' opinions on the web

Andreas Auinger [a], Martin Fischer [b]

[a] *FH OÖ Studienbetriebs GmbH, Wehrgrabengasse 1-3, 4400 Steyr, AUSTRIA*

[b] KTM-Sportmotorcycles AG., Stallhofnerstraße 3, A-5230 Mattighofen, AUSTRIA

ABSTRACT

Comparing consumer opinion concerning one's own products and those of the competitors to find their strengths and weaknesses is a crucial activity for marketing specialists in the production industry to overcome the requirements of marketing intelligence and product benchmarking. Hence, web-forums, blogs and product review websites provide valuable findings and discussions that record public opinion. Therefore, a huge variety of opinions and commentary about consumer products is woven into the web, which offers a new opportunity for companies to understand and respond to the consumer by analyzing this raw feedback. This paper presents an approach that combines results and understandings from several procedures to encounter the challenge of opinion mining. The proposed architecture includes a wide variety of state-of-the-art text mining and natural language processing techniques. Furthermore, the key elements of applications for mining large volumes of textual data for marketing intelligence are reasoned: a suite of powerful mining, visualization technologies and an interactive analysis environment that allows for rapid generation and testing of hypothesis. The concluding results show that recent technologies look promising, but are still far from a semantically correct textual understanding. Furthermore, this paper presents the results of a proof-of-concept of Text Mining with SPSS software. It is argued that SPSS Text Mining cannot meet the requirements to perform opinion mining as is being requested by market research at the moment.

Contact: andreas.auinger@fh-steyr.at

1 INTRODUCTION

Marketing places emphasis on the identification and subsequent satisfaction of consumer needs. To examine consumer needs and to implement effective marketing strategies and plans aimed at satisfying those needs, marketing managers need relevant, current information about consumers, competitors and other forces in the marketplace. Surprisingly, textual data on the web and in a company's own databases, which constitute a significant part of consumer information, have been widely ignored by companies in the past. However, the web especially has acquired immense value as an actively evolving repository of knowledge for market research. Recent web users of web 2.0 don't just use the web, they contribute content actively in applications like web-forums, weblogs or product review sites. Therefore, modern companies have to take the challenge to utilize the mass of user-generated content to get additional market insights. Text Mining has great potential to overcome these current deficiencies. Unfortunately, research in the field of natural language processing (NLP) encounter a range of difficulties due to the sophisticated nature of human language. In comparison to well structured numerical or categorical data, user generated content is usually made for other human users, and lacks computer-readable structures.

Moreover, the area of opinion mining encounters the additional problem of text classification, which is orthogonal to the usual task of Text Mining. In traditional text classification [cf. 5,16] the focus is on topic identification using statistical techniques, whereas the for opinion mining needed sentiment classification focuses on the assessment of writer's sentiment toward the topic. Emotions and opinions are expressed in a subtle manner, and therefore cannot be satisfactory analyzed with statistical, keyword-based methods. Turney [14] concludes these circumstances with the following statement: "The whole is not necessarily the sum of the parts". This paper is organized as follows. Firstly, a general overview of the identified opinion mining strategies is given. Secondly, the proposed opinion mining system architecture is presented. Subsequently, a brief proof-of-concept demonstrates the capabilities of SPSS Text Mining in terms of opinion mining. Finally, the last section concludes with lessons learned and gives ideas for further investigations.

2 OPINION MINING STRATEGIES

Opinion mining approaches either operate at the level of documents, i.e. sentences, or consider each clause that implies e.g. a product feature in combination with an expressed opinion (opinion mining at the clause/feature level). Document classification [12;14] tasks include picking out articles and classifying reviews as "positive" or "negative". A very common sentence level task is to classify sentences as "subjective" or "objective". However, for many applications, just identifying opinionated sentences may not be sufficient, since it is common to find two or more opinions in a single sentence, or to find a sentence containing opinions as well as facts. For an information extraction system trying to distinguish between positive and negative statements, it is crucial to be able to identify the particular clauses that contain opinions and to determine their semantic orientation.

3 TECHNOLOGY PROSPECTUS

Figure 1 gives an architectural overview of the proposed opinion mining system.

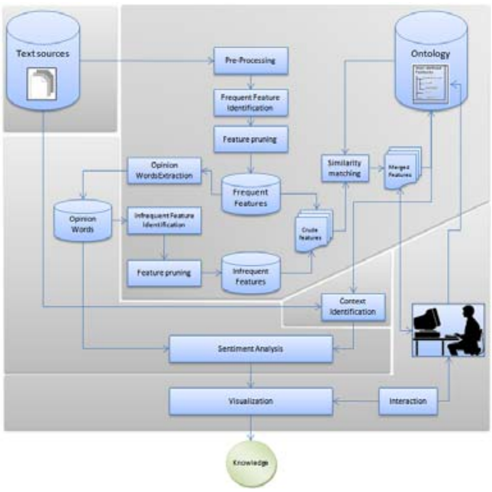

Figure 1: Architecture of the proposed opinion mining system (partly adopted from [1])

After text material is crawled through by adequate routines [2], product features are being identified and matched with the product ontology. In the next step the system analyzes available context information in order to recognize the correct sentiment orientation of context-dependent opinion words and to assign opinions correctly to the product features. The most sophisticated part of the proposed opinion mining system is the sentiment analysis. To encounter the difficulties of statistical based methods, the utility of linguistic rules [4] to obtain the sentiment orientation of consumers' statements are proposed. This section concludes with several suggestions about how results can be intuitively and meaningful visualized.

3.1 Ontology

An ontology is commonly is a description model for a specific domain, basically applicable to explain its complex structures and coherences. For the purpose of an opinion mining system, the ontology is used as a framework for the ensuing analysis of context information and sentiment analysis. In such a hierarchical structure two simple counters determine, whether product features are being seen as positive or negative. Assumed that we summarize the expressed opinions of a particular motorcycle model (KTM RC8), the summary looks like the following:

```
model RC8:
   chassis:
       positive: 126      <individual statements>
       negative: 78       <individual statements>
   design:
       positive: 111      <individual statements>
       negative: 34       <individual statements>
```

Chassis and design are opinion features for motorbikes. There are 126 consumer statements that express positive opinions about the chassis, and 78 that express negative opinions. <individual statements> points to the specific positive (or negative) comments. To create such a product ontology, you could also add all products and their components manually with an ontology editor (e.g. Protégé[1]). This could result in a very accurate ontology, but would probably not cover the features, the creator hasn't thought about and would be very time-consuming. Hu et al.[7] developed an unsupervised approach which is fully automated and operates on the principles of word co-occurrence identification to overcome this deficiency. Carenini et al. [1] identified quality limits for this method: (i) identified product features coincide in their meaning (redundancy) and (ii) there is no information about hierarchical coherences. They introduce an improved method for feature extraction which includes user-specific prior knowledge of the evaluated entity. The task of feature extraction is turned into one of term similarity matching by mapping crude (learned) features into a user-defined product ontology. This hybrid approach combines the advantages of both mentioned techniques. From the unsupervised techniques it inherits most of its portability. Prior knowledge and organization is mapped in form of user-defined features (UDF).

The system maps the output of the unsupervised feature extraction process to the UDF ontology and thereby eliminates redundancies and provides conceptual organization interactively and is user guided. After designing a set of UDF, and producing a set of crude features (CF), similarity matching is performed to map the CF to the UDF. The result is a set of merged features (MF). This process has several potential benefits. Firstly, grouping similar or identical features under the same UDF reduces redundancy. Secondly, hierarchical relationships between features are introduced and can be exploited in organizing and presenting the extracted information. A third benefit is that such information is framed in a way that the user envisions the product to be described and reviewed. Once the MF hierarchy is produced, the procedure intends for an interactive user-guided revision process where the user can scan

[1] Protègè Ontology Editor, Stanford University, CA/USA.

the intermediate output for errors or omissions. Any misplaced CF can be corrected and if the user notices a class of CF features that were inappropriately mapped, he/she can modify the UDF to accommodate for those features and rerun the similarity-matching step. In this way the UDF is not only reusable, it is adaptable as well.

3.2 Context Information

A key element of an opinion mining system is the correct identification of the discussion topic. In "common" language it is rather unusual to mention an opinion and the corresponding object all together in a sentence. The following extract from a motorcycle web-forum post illustrates the importance of correct context identification:

> *"[…] I didn't have problems with brakes and clutch. The reverse is true - the brake system sparked me. At the beginning not to cuttingly, but if it comes down to it, the brakes work perfectly. […] Up to now, I couldn't completely rev up the engine, but so far, torque and sound are persuading. […]"*

Obviously the mentioned components "brakes", "engine" and "sound" are evaluated positively, but it is not evident to which motorcycle model this post refers to. Without this information, consumers' feedback can't be analyzed in a structured way. Therefore, global and local context have to be considered. The obtained information is not only of interest for the maintenance of the product ontology, but also for the ensuing sentiment analysis. The overall subject of conversation is defined as "global context". At a first glance, the determination of the global context looks predestined for a classical text categorization system. Such a system segments all analyzed documents into predefined categories on the basis of occurring words. The category "RC8|brakes", for instance, is characterized by a surpassing number of words like "brake calliper, brake, RC8, dashpot, etc.". And in this way, all articles or posts in which these words frequently occur are allocated to that category. Problems could emerge if we consider that the categories "RC8|Brake" and "SuperDuke|Brake" differ only in terms of their product label. Their components (brakes, engine, design, etc.) are more or less the same. If there are not enough product specific expressions in the analyzed clause, the categorization routine cannot assign the correct category. However, an opinion mining system's endeavour is to compare different products, and to point out exactly what consumers like and dislike. Therefore, a classical text categorization strategy is unrewarding. Fortunately, common online communication media already provide an inherent structure. In the case of web-forums, the overall discussion topic is defined by the particular thread title, which we further define as "global context".

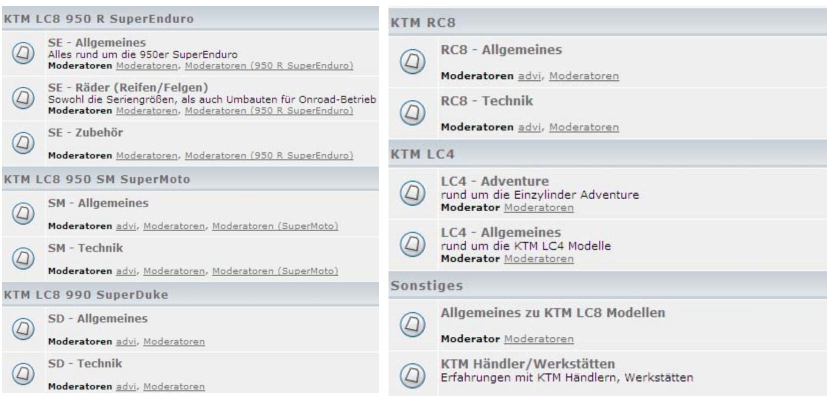

Figure 2: Example of a web-forum content structure.

Figure 2 illustrates an example of a typical web-forum content structure. A similar content structure can also be discovered on product review sites. Weblogs are different in terms of structure, but the global context of weblog articles can also be identified based on their title.
Frequently, the global topic of a discussion contribution is not the sole topic. In the context of our motorcycle examples, an article about the “RC8” (global context) can also mention other models, too. Therefore, the direct environment (local context) of the particular object has to be examined.

3.3 Sentiment Analysis

Several techniques use opinion expressing words such as “great”, “amazing”, “poor”, “bad”, etc., to decide the orientation of an opinion on a product feature. Turney [14] applies an unsupervised learning technique based on mutual information between document phrases and the words “excellent” and “poor” to find indicative words of opinions for classification. Pang et al [12] examine several supervised machine learning methods for sentiment classification of movie reviews. They show that classifiers perform well on whole reviews or articles, but poorly at a clause level. Other related works on sentiment classification and opinions discovery include [2; 7; 8; 10; 13; 15].

Although the orientations of these words are obvious, the orientations of many other words depend on the context. For example, the word “small” can indicate a positive or a negative opinion on a product feature depending on the feature. Ding and Liu [4] present an approach, which infers the orientations of opinions on e.g. a product feature using context information in combination with several linguistic rules:

1. ***Intra-sentence conjunction rule:*** The example sentence “*The engine power is high*” does not clearly distinguish whether “high” intends a positive or a negative opinion. The algorithm identifies whether “high” is clarified as positive (or negative) in any other statement such as: “This motorcycle’s chassis is *great* and it has a *high* engine power”. From this sentence, we can discover that “high” is positive for “engine power” because it is conjoined with the positive word “great”.
2. ***Pseudo intra-sentence conjunction rule:*** Natural language allows one to drop explicit conjunction words, like “and”. Let us use the example sentence *“the engine power is high”* again. Assuming that another product reviewer has written following sentence: “This motorcycle’s engine power is high, which is *great*”. The sentence indicates that the semantic orientation of “high” for “engine power” is positive due to “great”, although no explicit conjunction word like “and” is used.
3. ***Inter-sentence conjunction rule:*** The conjunction rule of preceding points can also be extended to neighboring sentences. The idea is that it is common in natural language to express the same opinion in a few consecutive sentences. Opinion changes are indicated by words such as “but”, “however”, etc. Following sentences are therefore considered to be natural: “The chassis is *great*. The engine power is *high*.” or “The chassis is *great*. *However*, the engine power is *low*.” On the contrary, the following phrase is not natural: “The chassis is *great*. The engine power is *low*.”

The polarity of context-dependent opinion words (e.g. “high”, “low”, etc.) can be determined, because of the fact that the polarity of context independent words (e.g. “great”, “cheapish”, etc.) is known. By the means of mentioned linguistic rules, we can infer in our example that “high” is positive and “low” is negative in the context of “engine power”. However, it has to be remarked that for rules 1 and 2, it may occur that in the product reviews or post the same opinion word for the same feature has conflicting orientations. This circumstance appears if authors have different opinions about a product feature, e.g.: “The chassis of this motorcycle is very hard, which I [don’t] like.” If more people indicate that “hard” is positive (or negative) for a chassis, the algorithm treats it as positive (or negative).

Synonym and Antonym Rule: If a context-dependant word is determined as positive (or negative) in context for a specific product feature, its synonyms are also considered positive (or negative, and its antonyms are considered negative (or positive). For the last example, we know that “high” is positive for “engine power”. Then we also know that “low” is negative for “engine power”.

The described algorithm works iteratively. During each iteration step, the opinion orientations of some opinion word can infer the orientations of some other opinion words. The learned polarities (extracted from the opinion words), in accordance to context information, are recorded in the respective ontology entry.

3.4 Visualization

In order to obtain the desired market knowledge and to be able to understand the gained data, the relevant information has to be presented in an intuitive and meaningful manner to market analyst. A number of metrics that provide a high-level summary of the relevant, multidimensional web content to facilitate top-down exploration of data have been identified [6].

1.1.1. Buzz Count

Buzss count is a simple count of the number (or percentage) of messages, articles or other documents that contain specific contents. Figure 3 illustrates the visualization approach proposed by [6].

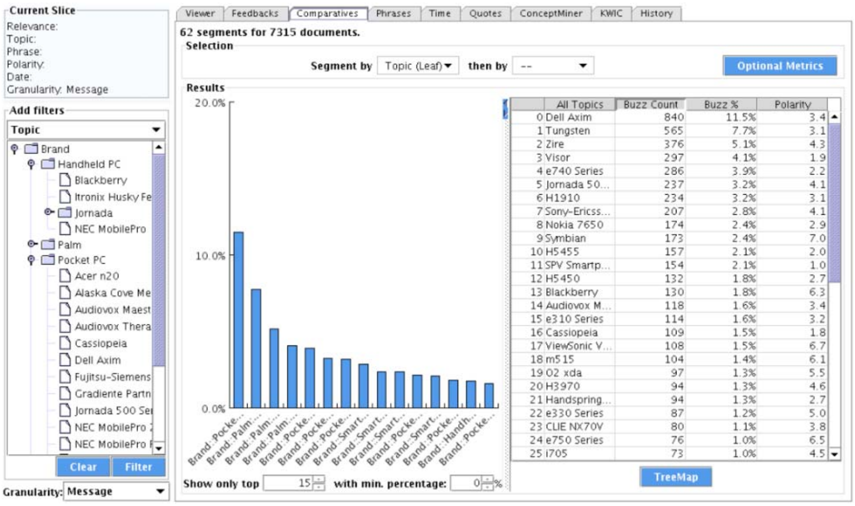

Figure 3: GUI with selected Buzz Count view [6].

SentimentMetrics.com[2] enhances this visualization approach with the ability to monitor the media presence of words (e.g. products and their features) in a chronological way.

---

[2] http://sentimentmetrics.com/ (last visited: August 2008)

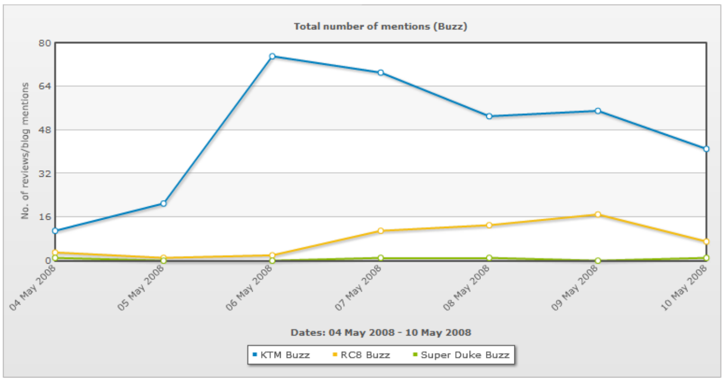

Figure 4: Buzz Count GUI approach by SentimentMetrics.

1.1.2. Sentiment Polarity

Sentiment polarity describes a score, representing the overall sentiment expressed about a topic, e.g. a brand or intersection of topics, products or product features. With the aid of this metric, a market analyst can clearly draw conclusions about the strengths and weaknesses of his own and competitive products. Figure 5 exemplarily illustrates the sentiment visualization approach by Liu et al [11]. It compares consumer opinions of two digital cameras along different feature dimensions, e.g. picture, battery, zoom, etc.

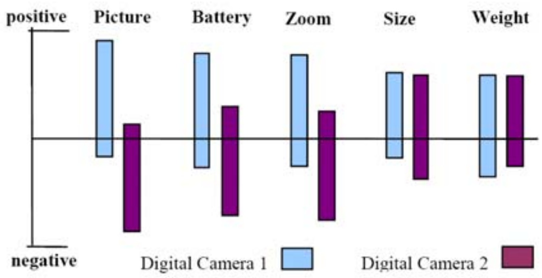

Figure 5: Visual comparison of consumer opinions.

Figure 5 illustrates that digital camera 1 is a superior camera in the minds of consumers and most consumers have negative opinions about the picture quality, battery and zoom of digital camera 2. The visualization enables the user to clearly see how the cameras compare with each other along each feature dimension.

1.1.3. Source Dispersion

Source dispersion measures the spread of a discussion concerning a particular topic. High values indicate that many people are talking about a particular topic, where low values indicate that a discussion is being focused on by a small group of people. SentimentMetrics.com[3] uses a simple pie chart to indicate the comparison (compare **Error! Reference source not found.**). These metrics serve two purposes. Firstly, they define a starting point for top-down exploration. Secondly, they provide dashboard-

[3] URL: http://sentimentmetrics.com/ (last visited: August 2008)

style summary statistics that can be disseminated within an organization, tracked over time and monitored for improvement or directionality.

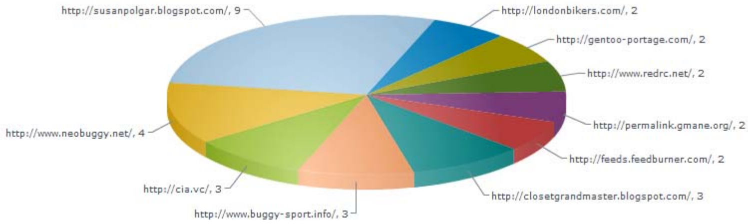

Figure 6: Simple pie chart to visualize source dispersion.[4]

4 SPSS TEXT MINING FOR MARKET RESEARCH

In order to proof the practical operability of opinion mining for market research, we took a closer look at the Text Mining solution of SPSS Inc.[5], one of the market-leading predictive analytics software providers. SPSS Text Mining, an add-on for SPSS Clementine (Data Mining suite), is a classical text mining application. It is available in several languages, and ought to be able to solve classical text mining issues like categorization and clustering, and also multi-linguistic sentiment analysis tasks which we concentrate on here in this article. The main objectives of the test were to clarify following issues:

- Opinion mining at the product feature level with respect to the context of the statement
- Intuitive visualization of results
- Handling of different languages/countries

4.1 Results

The requested analysis level was only able to be partly fulfilled by SPSS Text Mining during our test. It works compliantly if enough text material is provided, and consequently the pattern recognition routines work properly. The user can determine if consumers are evaluating the motorcycle model "RC8" positively or negatively. In the next step, we tried to identify reasons (product features) for these opinions. Unfortunately, the system architecture is too raw to identify explicit product features. The software is rather designed to analyze text material on a document level on the basis of the called *bag-of-feature* [5] approach. The strategy of this approach is to evaluate text documents based on the sole occurrence or absence of words. Grammar, word order and dependencies between words are completely ignored. However, opinion mining aims to identify expressed statements with respect to the characteristics of natural language.

The next issue investigated was the context-identification, one of the main challenges in opinion mining. In SPSS Text Mining, the semantic of words is managed with different libraries. It provides predefined libraries with a lot of common words. One library lists all positive words, the other one the negative ones, for instance. The user has also the possibility to adjust and to create new libraries. One word can be allocated to only one library. In natural language, however, it often happens that a word

[4] URL: http://sentimentmetrics.com/ (last visited: August 2008).

[5] URL: http://www.spss.com/text_mining_for_clementine

has different meanings, which can not be mapped to SPSS Text Mining. For example, “hot” is normally considered to be positive in a design context, on the other side it might indicate a negative characteristic in coherence to an engine or exhaust system context. Due to the fact that SPSS Text Mining can only assign a word to one library, the interpretation of statements is inaccurate.

The visualization of results is generally available with SPSS Text Mining. However, the application isn’t able to integrate a timeline to the analysis. For that reason, the validity of the visualization is very limited, and in the context of market research more or less useless. Furthermore, we couldn’t indicate the possibility to constrain the analysis in terms of a geographical area. That feature would be helpful to estimate how consumers in different areas/countries of the world evaluate products and their features.

In summary, it seems that the operational area of SPSS Text Mining is rather the organization and classification of whole documents than the analysis of individual clauses, which is requested by the challenge of opinion mining.

5 CONCLUSIONS AND OUTLOOK

Online discussion, in the form of web-forums, weblogs and review websites, represent a valuable opportunity for many types of analyses. In this article, we proposed an opinion mining system, which transfers unstructured free text to a structured or semi-structured summary. The automation of this process in the context of e-Commerce and e-Business, assists market researchers and product designers to better understand consumer needs and to facilitate enterprise information management. Through this work, we have addressed the challenge of information overload facing product designers by providing a theoretical end-to-end opinion mining system that gathers specific types of online content and delivers analytics based on classification, NLP, phrase finding and other mining technologies in a marketing intelligence application. Furthermore, it is being claimed that beside these Text Mining techniques, a usable interface is crucial to enable visual analysis. Nevertheless, although our architecture is a collocation of several approved methods; its practical functionality has to be examined in a further work.

Some outstanding problems still remain rather unnoticed and require further investigation efforts. Hu et al. [8] plan to improve and refine their techniques to deal with pronoun resolution and the determination of the strength of opinions. Furthermore, it is noticed that although some consumer comments regarding product features cannot be labelled as either positive or negative, they are still valuable. For example, the following two sentences are extracted from Hu’s corpus [8]:

1: The phone’s sound quality is great.
2: The most important thing for me is sound quality.

Both sentences discuss the product feature sound quality. Unlike the first sentence, the second one does not offer any attitude orientation, yet it does provide valuable information for designers about which features consumers are really concerned about. Such neutral comments and suggestions are currently not being considered in the method of opinion mining.

The analysis of directly expressed opinions is one way to gather information. Another possibility is the analysis of comparative sentences. A comparison is not concerned with an object in isolation. Instead, it compares the object with others. Direct comparisons are one of the most convincing ways of evaluation, which may even be more important than opinions on each individual object. Jindal et al. [9] present a supervised learning approach to identify and interpret comparative sentences from text documents.

Ideally, opinion mining ought to be able to address also fairly sophisticated issues too – identifying the object of sentiment, detecting mixed and overlapping sentiments in a text, identifying and dealing with sarcasm, etc. Certainly, we are still far away from this kind of textual understanding, but hopefully our proposed opinion mining system architecture is a step in the right direction.

REFERENCES

[1] Carenini, G. / Ng, R. / Zwart, E.: Extracting Knowledge from Evaluative Text. University of British Columbia, 2005.

[2] Chakrabarti, S. (2007): Mining the Web. San Francisco (USA)

[3] Dave, K. / Lawrence, S. / Pennock, D. (2003): Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. In: WWW2003.

[4] Ding, X. / Liu, B. (2007): the Utility of Linguistiv Ruls in Opinion Mining. In: SIGIR’07.

[5] Feldman, R. / Sanger, J. (2007): The Text Mining Handbook. Adavanced Approaches in Analyzing Unstructured Data. Cambridge, New York (USA).

[6] Glance, N. / Hurst, M. / Nigam, K. (2005): Deriving Marketing Intelligence from Online Discussion. In: KDD’05.

[7] Hu, M. / Liu, B. (2004a): Mining Opinion Features in Customer Reviews. American Association for Artificial Intelligence.

[8] Hu, M. / Liu, B. (2004b): Mining and Summarizing Customer Reviews. In: KDD ’04.

[9] Jindal, N. / Liu, B. (2006): Identifying comparative sentences in Text Documents. University of Illinois at Chicago.

[10] Kim, S. / Hovy, E. (2004): Determining the Sentiment of Opinions. University of Southern California.

[11] Liu, B. / Hu, M. / Cheng, J. (2005): Opinion Observer: Analyzing and Comparing Opinions on the Web. In: WWW2005.

[12] Pang, B. / Lee, L. / Vaithyanathan, S. (2002): Thumbs up? Sentiment Classification using Machine Learning Techniques. In: Proceedings of EMNLP, pp. 79-86.

[13] Popescu, A-M / Etzioni. O. (2005): Extracting product features and opinions from reviews. EMNLP-05.

[14] Turney, P. (2002): Thumbs Up or Thumbs Down? Sentiment Orientation Applied to Unsupervised Classification of Reviews. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 417-424.

[15] Wilson, T. / Wiebe, J. / Hwa, R. (2004): Just how mad are you? Finding strong and weak opinion clauses. University of Pittsburgh.

[16] Zanasi, A. (2007): Text Mining and its Applications to Intelligence, CRM and Knowledge Management, Southhampton/Boston.