ECAI 2010
H. Coelho et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-606-5-511

Opinion Question Answering: Towards a Unified Approach

Alexandra Balahur and **Ester Boldrini** and **Andrés Montoyo** and **Patricio Martínez-Barco**[1]

Abstract. Nowadays, the Web contains large amounts of heterogeneous (factual and opinionated) data, which is becoming equally important for users to access. The need to efficiently manage this information leads to the necessity of building automatic systems that efficiently process it. In this paper, we propose and evaluate a series of techniques whose aim is to improve the performance of an Opinion Question Answering (OQA) system. We include additional resources and processes with the objective of limiting the sources of errors in the different stages involved - question analysis, answer retrieval and filtering, answer re-ranking. We propose new elements that are significant in these stages and show that their use improves the performance of the system. We conclude that the suggested techniques help to influences the task in a positive manner.

1 INTRODUCTION

The State of the Blogosphere 2009 survey published by Technorati[2] demonstrates that there is a growing influence of the blogosphere on subjects ranging from business to politics and to the way information travels through communities. In a year when revolutions and elections were organized by blogs, users are blogging more than ever, and the State of the Blogosphere is so strong that the attitudes held by bloggers don't differ very much by age or gender, or even across geographies. The same report shows that, in contrast to the general idea about bloggers, each day it is more and more the professionals who decide to use this means of communication, thus contradicting the common belief about the predominance of an informal editing [1]. Due to the growing interest in this type of text, the subjective content of the Web is constantly increasing and reflecting people's opinion about a wide range of topics. [7]. The abovementioned Technorati research demonstrates that self-expression and sharing expertise through opinion continue to be the primary motivations for bloggers. They describe significant, positive impacts on their personal lives, as well as their business. Moreover, blogs represent an important source of real-time, unbiased information that is useful to many applications. However, in order to exploit the content of this subjective information, its processing must be automated. The Natural Language Processing (NLP) task dealing with the treatment of subjective data is called Sentiment Analysis (SA). Users must be able to efficiently access this data, through queries or questions. While techniques to retrieve objective information have been widely studied and implemented, opinion-related tasks still represent an important challenge. Therefore, the aim of our research is to study and implement appropriate methods for the task of QA in the context of opinion. In the research, this setting is known as Opinion Question Answering (OQA).

2 MOTIVATION AND CONTRIBUTION

Although research in opinion-related tasks has gained importance in the past years, there are still many aspects that require analysis and improvements. This is especially true for approaches that combine subjectivity analysis with other NLP tasks (e.g. QA, summarization). The TAC 2008 Opinion Pilot task, as well as the subsequent research performed on the competition data, have shown that answering opinionated questions and summarising subjective information are significantly different from the same tasks in the context of factual data. Therefore, the first motivation of our work is the need to identify and explore the challenges raised by opinion Question Answering (OQA), as opposed to the factual QA. To this aim, we analyse the improvements that can be brought in the different stages of the OQA process - question treatment, opinion retrieval, opinion analysis, topic detection; in the context of these subtasks, we study the contribution of discourse analysis, through the application of coreference resolution. This is motivated by the conclusions drawn by previous studies [1]; in this context, our aim is to verify if the inclusion of such resources and tools - whose performance is most of the times not optimal - affects the performance of the system in a positive or negative manner. Our contribution to this respect is the identification of the challenges related to OQA as opposed to traditional QA. In consequence, a further contribution resides in adding the appropriate methods, tools and resources to resolve the identified issues. In order to test the relevance of each tool, resource and technique, we will carry out a separate, as well as global evaluation. Finally, our work is motivated by the fact that although previous approaches have shown that opinion questions have much longer answers than factual questions, the work done in OQA so far has only considered a sentence-level approach. To this respect, our further contribution resides in employing two different types of retrieval - at a sentence level and at the level of 3 snippets. It is important to mention that in the context of OQA, considering longer snippets does not necessarily mean that there is a higher chance to encounter the answer (due to more information being present). The need to retrieve longer snippets of text is motivated by the fact the in some of the cases, one opinion is expressed through two or three consecutive sentences (e.g. "They estimated a growth in employment by 3%. Obviously, this didn't happen.").

3 RELATED WORK

Most of the state of the art in QA systems is focused on the needs of answering factual questions. For this type of systems, advanced

[1] University of Alicante, Spain, email: {abalahur, eboldrini, montoyo, patricio@dlsi.ua.es}
[2] http://technorati.com/blogging/feature/state-of-the-blogosphere-2009/

methods exist to process questions, retrieve appropriate answers and perform answer extraction [12]. However, due to the recent growth in the volume of subjective information present on the Internet and the great impact it has on the daily lives of people all over the world, treating factual information is no longer sufficient. Opinionated questions revealing answers about people's opinions have long as well as complex answers, which in most of the cases are not easily retrievable, as they extend over more than one sentence or document. Moreover, opinion questions usually require the retrieval of sentences that are relevant to the query, but, at the same time, have a specifically required polarity (e.g. "Why do people enjoy Starbucks coffee?" requires a positive opinions on Starbucks coffee and "What reasons do people give for not liking Starbucks coffee?" requires the OQA system to retrieve the negative sentiments expressed on the Starbucks coffee). Due to these characteristics, traditional QA systems, developed for objective queries, are not entirely suitable to retrieve answers to opinion questions [17]. Research focused on building factoid QA systems has a long tradition, however, it is only recently that studies have started to focus on the creation and development of opinion QA systems. Example of this can be [16] who took advantage of opinion summarization to support Multi-Perspective QA system, aiming at extracting opinion-oriented information of a question. [23] separated opinions from facts and summarized them as answer to opinion questions. [9] identified opinion holders, which are frequently asked in opinion questions. Recently, [21] consider the use of convolution kernels for opinion source detection. Moreover, [8] proposed a method to identify strong clues of subjectivity on adjectives, which can be used as discriminative features for fact versus opinion classification. Apart from these studies, specialized competitions for systems dealing with opinion retrieval and QA have been organized in the past few years. The TAC 2008 Opinion Pilot track proposed a mixed setting of factoid and opinion questions. It is interesting to note that most of the participating systems only adapted their factual QA systems to overcome the newly introduced difficulties related to opinion mining and polarity classification. The Alyssa system [14] was added new components for the classification of the polarity of the question and of the extracted answer snippet, using a Support Vector Machines (SVM) on opinion mining corpora. [19] performs query analysis to detect the polarity of the question using defined rules and filters opinion based on Nave Bayes. Other relevant competition focused on the treatment of subjective data is the NTCIR MOAT (Multilingual Opinion Analysis Test Collection). The approaches taken by the participants in this task are relevant to the process of opinion retrieval, which is the first step performed by an OQA system. High-scoring systems employed specific techniques to accommodate the opinion retrieval difficulties. For example, [24] used an almost unsupervised approach applied to two of the subtasks: opinionated sentence and topic relevance detection.[13] applied a sequential tagging approach at the token level and used the learned token labels in the sentence level classification task and their formal run submission was is trained on MPQA [20]. [10] competed for the traditional Chinese using the supervised lexicon-based and machine learning approaches. [22] propose an OM system which extracts topic-related opinions at a sentence level and identifies their target with the purpose of selectig relevant opinions. After having analysed the aforementioned contributions and the performance of the systems implemented, we conclude that this task requires the use of additional tools and methods, focused on the specific aspects related to the treatment of opinion (source, target, polarity) and joint topic-opinion analysis.

4 CORPORA

In order to conduct our experiments, we employed two collections of blog posts: EmotiBlog [5] and the TAC 2008 Opinion Pilot test collection, which is part of the Blog06 corpus. It contains the documents that include the answers to the opinion questions given on 25 targets. EmotiBlog is composed of blog posts in English extracted form the Web. As a consequence, it represents a genuine example of this textual genre. EmotiBlog is a monothematic corpus about the Kyoto Protocol, annotated with the improved version of EmotiBlog [4].

OM is a very complex task due to the high variability of language. Thus, the improvement brought by EmotiBlog is an annotation model for a fine-grained detection of subjective speech. From the first version of the model, the components that were not statistically relevant have been eliminated. The elements which compose the improved version of the model are presented in Table 1.

Table 1. EmotiBlog structure

Elements	Description
Obj.Speech	Confidence, comment, source, target.
Subj. speech	Confidence, comment, level, emotion, phenomenon, polarity, source and target.
Adjectives/Adverbs	Confidence, comment, level, emotion, phenomenon, modifier/not, polarity, source and target.
Verbs/ Names	Confidence, comment, level, emotion, phenomenon, polarity, mode, source and target.
Anaphora	Confidence, comment, type, source and target.
Capital letter/ Punctuation	Confidence, comment, level, emotion, phenomenon, polarity, source and target.
Phenomenon	Confidence, comment, type: collocation, saying, slang, title, and rhetoric.
Reader/Author Interpretation (obj.)	Confidece, comment, level, emotion, phenomenon, polarity, source and target.
Emotions	Confidence, comment, accept, anger, anticipation, anxiety, appreciation, bad, bewilderment, comfort, compassion.

The first distinction consists in separating between objective and subjective speech. Subsequently, the corresponding elements have to be specified. In case of an objective sentence its source - the author of the opinion - and target - the object on which the opinion is expressed - (when necessary also the level of confidence of the annotator and a comment) have to be inserted. Dealing with subjective sentences, adjectives, adverbs, punctuation, names can be marked; we also annotate cases of anaphora at a cross-document level (to interpret the storyline) and the sentence type (simple sentence, title but also saying or collocation). We decided to underline cases of saying and collocations since they have a relevant subjective charge and they are strictly dependent on cultures, personal background and other socially-related aspects. We believe that annotating such phenomena could improve the comprehension of subjectivity. It is worth mentioning that we annotate them with a single label due to the fact that their meaning is obtained from the overall expression and not by the single words it contains. Finally, the Reader and the Writer interpretation can be marked in objective sentences. They are extremely relevant because they are employed to interpret correctly an apparent objective discourse, but which in reality represents an factual argument for a given opinion (e.g. "During his mandate, unemployment rates increased from 12% to 15%."). The first is useful to extract what

is the interpretation of the reader and the second to understand the background of the reader (i.e. "The tragedy happened in 1989" instead of saying "The earthquake happened in 1989"). The questions whose answers are annotated with the EmotiBlog annotation schema are the subset of opinion questions in English presented in [1]. They are shown in Table 2.

Table 2. Questions over the EmotiBlog corpus

Number	Question
2	What motivates people's negative opinions on the Kyoto Protocol?.
5	What are the reasons for the success of the Kyoto Protocol?
6	What arguments do people bring for their criticism of media as far as the Kyoto Protocol is concerned?
7	Why do people criticize Richard Branson?
11	What negative opinions do people have on Hilary Benn?
12	Why do Americans praise Al Gore's attitude towards the Kyoto protocol?
15	What alternative environmental friendly resources do people suggest to use instead of gas en the future?
16	Is Arnold Schwarzenegger pro or against the reduction of CO2 emissions?
18	What improvements are proposed to the Kyoto Protocol?
19	What is Bush accused of as far as political measures are concerned?
20	What initiative of an international body is thought to be a good continuation for the Kyoto Protocol?

The main difference between the two corpora employed is that Emotiblog contains only posts on the Kyoto Protocol and the TAC 2008 corpus is composed by documents on a multitude of topics. Therefore, different techniques must be adjusted in order to properly treat each of them. For example, in order to retrieve the target of the discourse, we can employ a Semantic Role Labeling (SRL) system and it would be more relevant for the TAC corpus since it presents a great variety of topics. On the other hand, a Semantic Role (SR) approach for EmotiBlog would be more effective to highlight all the subtopics related to it. The TAC posts are more focused on the main topic, while EmotiBlog post are about the Kyoto protocol, but they are about all the aspects and satellite topics of this main topic. Regarding the source of the discourse, both corpora have the same situation. Normally a post is written by an author but there are also many cases of multiple source, for example when the author of the post cites another person's discourse. As a consequence we need a correference resolution system at both cross- and intra-document. The cross-document level would be extremely relevant for detecting and understanding all the discourse and how each author of each posts answers to the previous ones and the internal correference is needed for this possible mixture of sources and for understanding all the correferent elements in general.

5 EXPERIMENTS

5.1 Question analysis

In order to be able to extract the correct answer to opinion questions, different elements must be considered. As stated in [1] we need to determine both the expected answer type of the question - as in the case of factoid ones - as well as new elements - such as expected polarity type. However, opinions are directional - i.e., they suppose the existence of a source and a target to which they are addressed. Thus, we introduce two new elements in the question analysis - expected source (ES) and expected target (ET). These two elements are selected by applying SRL and choosing the source as the agent in the sentence and the direct object (patient) as the target of the opinion. The expected answer type (EAT) (i.e opinion or other) is determined using machine learning using SVM, by taking into account the interrogation formula, the subjectivity of the verb and the presence of polarity words in the target semantic role. In the case of expected opinionated answers, we also compute the expected polarity type (EPT) - by applying opinion mining on the affirmative version of the question (e.g. "Why do people like George Clooney?" - "People like George Clooney for X"). This is done through a series of patterns that we previously defined in [2].

5.2 Candidate snippet retrieval

In the answer retrieval stage, we employ two strategies:

1. Using the JIRS (JAVA Information Retrieval System) IR engine (Gómez et al., 2007) to find relevant snippets. JIRS retrieves passages (of the desired length), based on searching the question structures (n-grams) instead of the keywords, and comparing them.
2. Using the "Yahoo" search engine to retrieve the first 20 documents that are most related to the query. Subsequently, we apply Latent Semantic Analysis on the retrieved documents and extract the words that are most related to the topic. Finally, we expand the query using words that are very similar to the topic and retrieve snippets that contain at least one of them and the ET.

5.3 Polarity and topic-polarity classification of snippets

In order to determine the correct answers from the collection of retrieved snippets, we must filter only the candidates that have the same polarity as the question EPT. For polarity detection, we use a combined system employing SVM machine learning on unigram and bigram features trained on the NTCIR MOAT 7 data and an unsupervised lexicon-based system. The latter uses the Opinion Finder lexicon to filter out subjective sentences - that contain more than two subjective words or a subjective word and a valence shifter (obtained from the General Inquirer resource). Subsequently, it accounts for the presence of opinionated words from four different lexicons - Micro WordNet [6], WNAffect [18], Emotion Triggers [3] and General Inquirer [15]. For the joint topic-polarity analysis, we first employ Latent Semantic Analysis to determine the words that are strongly associated with the topic, as described in Section 4.2. Consequently, we compute the polarity of the sentences that contain at least one topic word and the question target. We also employed JavaRAP[3] to resolve the correference elements. JavaRAP is an implementation of the classic Resolution of Anaphora Procedure (RAP) given by Lappin and Leass (1994). It resolves third person pronouns, lexical anaphors, and identifies pleonastic pronouns.

5.4 Filtering using Semantic Roles

Finally, answers are filtered using Semantic Roles [11]. Since blog entries implicitly belong to some author, present through first person markables (I, my etc.). Besides that, we intuitively consider that

[3] http://www.comp.nus.edu.sg/ qiul/NLPTools/JavaRAP.html

sources and targets of opionins should have the agent or patient semantic roles in the opinionated sentences (e.g. "George Clooney is a terrific actor", "I really hate George Clooney"). In order to test this intuition, we filter for the next processing step all snippets that have the required target and source as agent or patient. It is probable that sources and targets are also found in other semantic roles. However, due to the fact that no special method is priorly employed for source and target determination, we consider this to be a fair approximation for the present study.

6 Evaluation and Discussion

We evaluate our approaches on both the EmotiBlog question collection, as well as the TAC 2008 Opinion Pilot test set. We compare them against the performance of the [1] system and the best scoring system (as far as F-measure is concerned) in the TAC 2008 task. For both the TAC 2008 and EmotiBlog sets of questions, we employ the Semantic Roles system in sentiment analysis and determine the ES, ET and EPT. Subsequently, for each of the two corpora, we retrieve 1-phrase and 3-phrase snippets. The retrieval of the of the EmotiBlog candidate snippets is done using query expansion with LSA and filtering according to the ET. Further on, we apply sentiment analysis (SA) using the approach described in Section 5.3 and select only the snippets whose polarity is the same as the determined question EPT. The results are presented in Table 3.

The retrieval of the TAC 2008 1-phrase and 3-phrase candidate snippets was done using JIRS. Subsequently, we performed different evaluations, in order to assess the impact of using different resources and tools. Since the TAC 2008 had a limit of the output of 7000 characters, in order to compute a comparable F-measure, at the end of each processing chain, we only considered the snippets for the 1-phrase retrieval and for the 3-phases one until this limit was reached.

1. In the first evaluation, we only apply the sentiment analysis tool and select the snippets that have the same polarity as the question EPT and the ET is found in the snippet. The ET at this step is only the target directly mentioned in the question. (e.g. Why do people like George Clooney? - here, the target is George Clooney).
2. In the second step, we add the result of the LSA process to filter for further processing only the snippets from 1. containing the words related to the topic.
3. In the third step, we filter the results in 2. by applying the Semantic Roles (SR) system and setting the condition that the ET and ES are the agent or the patient of the snippet. It is important to mention that in this step, ET and ES can also be terms that are discovered to be closely-related to the topic using LSA (e.g. George Clooney - actor).

The results are computed using the pyramid nugget score proposed in the TAC 2008 Opinion Pilot competition. The results are shown in Table 4.

From the results obtained, we can draw the following conclusions. Firstly, the hypothesis that OQA requires the retrieval of longer snippets was confirmed by the improved results, both in the case of EmotiBlog, as well as the TAC 2008 corpus. In the case of EmotiBlog, where the topic remains relatively constant along the posts, judging opinion and correctly classifying the sentiment it contains can be more accurately done in a larger context, from which the target can be also extracted - see Table 3. This fact was confirmed by the experiments on the TAC 2008 corpus, where it was seen that the 1-sentence passages that were retrieved were not enough to capture the meaning of the opinion expressed; additional context was thus needed and obtained from the use of a larger span of text. Moreover, in some cases, 3-sentences-long snippets contain information that cannot be otherwise extracted, as it contains anaphoric references. Secondly, the results presented in Table 4 show that opinion questions require the joint topic-sentiment analysis (in the first two settings, where LSA was not employed, the system performs worse than in the remaining configurations); analysisng the improvements in F-measure, we can see that the use of topic-related words for computing the affect value influences the results in a positive manner. Thirdly, another conclusion that we can draw is that target and source detection is a relevant step at the time of answer filtering, not only helping in the more accurate retrieval of answers, but also at placing at the top of the retrieval the relevant results. Nonetheless, as we can see from the relatively low improvement in the results, much remains to be done in order to appropriately tackle OQA. As seen in the results, there are still questions for which no answer is found (e.g. 18). This is due to the fact that its treatment requires the use of inference techniques that are presently unavailable (i.e. define terms such as "improvement"). Finally, from the analysis of the errors, we could see that even though some tools are in theory useful and should produce higher improvements - such as SR - their performance in reality does not produce drastically higher results. There are many cases in which we obtain sentences with pronouns and if we do not have at our disposal the all document we cannot understand to whom it is referring. In order to solve this problem JavaRAP substitutes pronouns or correference elements in general, with their corresponding antecedent. After having analysed the results we detected some cases of mistakes that are repeated through our answer corpus. It is well known that correference resolution is a challenging task, above all due to the high variability and employment of cultural elements in language; the world knowledge is extremely difficult for a system to deduce. From preliminary results obtained using JavaRap for co-reference resolution, we also noticed that the performance of the OQA lowered, although theoretically it should have improved. As a consequence we detected some errors that are recurrent and that are due to these abovementioned reasons. The first one is that in many cases the system replaces the correference element with the first possible antecedent (the nearest one). It coincides in genre and number but it is not the correct one. There are also other cases in which JavaRAP takes the next possible antecedent and thus threats the anaphora as a cataphora. Another interesting case is when we obtain a sentence in which we have to replace two different correferences and the system labels them with the same antecedent. Those are examples of mistakes that are extremely difficult to solve. A possible solution for dealing with such a problem is the retrieval of larger spans of text, as proposed. In this case, we can subsequently split the paragraphs into smaller chunks of text and treat the snippets as separate opinion units, but considering the context in which they appear ("interpolate" for target, source and polarity detection).

Table 4. Results for TAC 2008 question set

System	F-measure
Best TAC	0.534
Worst TAC	0.101
JIRS + SA+ET (1 phrase)	0.377
JIRS + SA+ET (3 phrases)	0.431
JIRS + SA+ET+LSA (1 phrase)	0.489
JIRS + SA+ET+LSA (3 phrases)	0.505
JIRS + SA+ET+LSA+SR (1 phrase)	0. 533
JIRS + SA+ET+LSA+SR (3 phrases)	0.571

Table 3. Results for questions over EmotiBlog

Q. No	No.A.	Baseline (Balahur et al., 2009)				1 phrase +ET+SA				3 phrases +ET+SA			
		@1	@5	@10	@50	@1	@5	@10	@50	@1	@5	@10	@20
2	5	0	2	3	4	1	2	3	4	1	2	3	4
5	11	0	0	0	0	0	2	2	2	1	2	3	4
6	2	0	0	1	2	1	1	2	2	0	1	2	2
7	5	0	0	1	3	1	1	1	3	0	2	2	4
11	2	1	1	1	1	0	0	0	0	0	0	0	1
12	3	0	1	1	1	0	1	2	3	0	0	1	2
15	1	0	0	1	1	0	0	1	1	1	1	1	1
16	6	1	4	4	4	0	1	1	2	1	2	2	6
18	1	0	0	0	0	0	0	0	0	0	0	0	0
19	27	1	5	6	18	0	1	1	2	0	1	1	1
20	4	0	0	0	0	0	0	1	1	0	0	1	2

When we analysed the results obtained we confirmed some of our initial hypotesis. The first one is that, when retrieving 1 sentence it is more frequent to have a sentenced with the direct topic the question asks involved. Some time the answer can be correct or other times incorrect, but when we retrieve 3 sentences the result is slightly different. Having a larger quantity of information there are many cases in which we have a sentences not directly related to the topic but referring to a subtopic of this. On the one hand, the systems have more options to retrieve the sentence and moreover to obtain a more complete information but at the other hand, there is also he risk of including not relevant information. Another conclusion we can draw form the results obtained is the SR detection. It can b eemployed for both source and subject. Regarding the topic of EmotiBlog the SR recogniser performed better because the subject is the same and all the subtopic are related, while for the TAC it is more useful even of less effective since we have a huge variety of topics with their corresponding suptopics. Regarding the source of the discourse, the SR system performed in an equivalent way but with low results. We believe an effective coreference resolution system must be used because it would be an added value to the task.

7 CONCLUSIONS AND FUTURE WORK

In this paper, we presented and evaluated different methods and techniques, as well as their combined use, for improving the task of QA in the context of opinion. From the evaluations performed using different NLP resources and tools, we concluded on the usefulness of joint topic-sentiment analysis, as well as the target and source identification. We have also shown that by retrieving longer answers, the results have improved. As far as corefernce resolution is concerned, our experiments have demonstrated that the use of existing systems does not lead to improved results. The problem of finding references can however be partially surpassed by retrieving larger spans of text and judging opinion in this extended context. From the results obtained, we concluded that opinion QA requires the development of appropriate strategies. Future work includes the employment of external knowledge sources that refine the semantics of queries ad also the use of a temporality system. We believe this would be extremely useful in order to discriminate opinions and classify them regarding to their date. This would be employed in case of questions such as: "What was the peopleś opinion about Mc Donaldś in 1980?" or "How does the opinion about Bush changed during its mandate?". As it can be seen, in order to find the correct answers, the retrieval system should discriminate between different periods and subsequently filter only the relevant information.

ACKNOWLEDGEMENTS

This paper has been partially supported by The Spanish Ministry of Science and Education of the Spanish Government (Ministerio de Ciencia e Innovacin - Gobierno de España) grant no. TIN2009-13391-C04-01, and Conselleria d'Educacin - Generalitat Valenciana (grant no. PROMETEO/2009/119 and ACOMP/2010/288. We would also like to thank the referees for their very useful comments.

REFERENCES

[1] A. Balahur, E. Boldrini, A. Montoyo, and P. Martínez-Barco, 'Opinion and generic question answering systems: a performance analysis', in *Proceedings of the ACL-IJCNLP 2009*, (2009).

[2] A. Balahur, E. Lloret, O. Ferrandez, A. Montoyo, M. Palomar, and R. Muñoz, 'The dlsiuaes team's participation in the tac 2008 tracks', in *Proceedings of the Text Analysis Conference 2008 Workshop*, (2008).

[3] A. Balahur and A. Montoyo, 'Applying a culture dependent emotion triggers database for text valence and emotion classification', *Procesamiento del Lenguaje Natural*, **40**(40), (2008).

[4] E. Boldrini, A. Balahur, P. Martínez-Barco, and A. Montoyo, 'Emotiblog: a fine-grained model for emotion detection in non-traditional textual genres', in *Proceedings of WOMSA 2009*, (2009).

[5] E. Boldrini, A. Balahur, P. Martínez-Barco, and A. Montoyo, 'Emotiblog: an annotation scheme for emotion detection and analysis in non-traditional textual genres', in *Proceedings of the 5th International Conference on Data Mining (DMIN 2009)*, (2009).

[6] S. Cerini, V. Compagnoni, A. Demontis, M. Formentelli, and G. Gandini, 'Micro-wnop: A gold standard for the evaluation of auto-matically compiled lexical resources for opinion mining', (2007).

[7] H. Cui, V. Mittal, and M. Datar, 'Comparative Experiments on Sentiment Classification for Online Product Reviews', in *Proceedings of the 21st National Conference on Artificial Intelligence AAAI 2006*, (2006).

[8] V. Hatzivassiloglou and K. McKeown, 'Predicting the semantic orientation of adjectives', in *Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL*, (1997).

[9] S.M. Kim and E. Hovy, 'Determining the Sentiment of Opinions', in *Proceedings of COLING 2004*, (2004).

[10] B. Lu, B.K. Tsou, and O.Y. Kwong, 'Supervised approaches and ensemble techniques for chinese opinion analysis at ntcir-7', in *Proceedings of NTCIR-7 Workshop meeting*, (2008).

[11] P. Moreda, *Los Roles Semanticos en la Tecnologia del Lenguaje Humano: Anotacion y Aplicacion*, Doctoral Thesis, 2008.

[12] A. Moschitti, S. Quarteroni, R. Basili, and S. Manandhar, 'Exploiting syntactic and shallow semantic kernels for question/answer classifi-

cation', in *Proceedings of the 45th Conference of the Association for Computational Linguistics (ACL) 2007*, (2007).
[13] L. Qu, C. Toprak, N. Jakob, and I. Gurevych, 'Sentnece level subjectivity and sentiment analysis experiments in ntcir-7 moat challenge', in *Proceedings of NTCIR-7 Workshop meeting*, (2008).
[14] D. Shen, J. Leidner, A. Merkel, and D. Klakow, 'The alyssa system at trec qa 2007: Do we need blog06?', in *Proceedings of The Sixteenth Text Retrieval Conference*, (2007).
[15] P.J. Stone, D. C. Dumphy, M. S. Smith, and D. M. Ogilvie, *The General Inquirer: A Computer Approach to Content Analysis*, The MIT Press, 1966.
[16] V. Stoyanov, C. Cardie, D. Litman, and J. Wiebe, 'Evaluating an opinion annotation scheme using a new multi-perspective question and answer corpus', in *AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, (2004).
[17] V. Stoyanov, C. Cardie, and J. Wiebe, 'Multi-perspective question answering using the opqa corpus', in *Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*, (2005).
[18] C. Strapparava and A. Valitutti, 'Wordnet-affect: an affective extension of wordnet', in *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, (2004).
[19] V. Varma, P. Pingali, R. Katragadda, S. Krishna, S. Ganesh, K. Sarvabhotla, H. Garapati, H. Gopisetty, K. Reddy, and R. Bharadwaj, 'Iiit hyderabad at tac 2008', in *Proceedings of Text Analysis Conference, at the joint annual meeting of TAC and TREC, 2008*, (2008).
[20] J. Wiebe, T. Wilson, and C. Cardie, 'Annotating expressions of opinions and emotions in language', in *Language Resources and Evaluation*, volume 39, (2005).
[21] M. Wiegand and D. Klakow, 'Convolution kernels for opinion holder extraction', in *Proceedings of NAACL 2010*, (2010).
[22] K. Youngho, K. Seongchan, and M. Sung-Hyon, 'Extracting topic-related opinions and their targets in ntcir-7', in *Proceedings of NTCIR-7 Workshop meeting*, (2008).
[23] D. Yu and V. Hatzivassiloglou, 'Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences', in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (2003).
[24] T. Zabibalov and J. Carroll, 'Almost-unsupervised cross-language opinion analysis at ntcis-7', in *Proceedings of NTCIR-7 Workshop meeting*, (2008).