U.P.B. Sci. Bull., Series C, Vol. 78, Iss. 3, 2016 ISSN 2286-3540

RESEARCH CHALLENGES IN OPINION MINING FROM A NATURAL LANGUAGE PROCESSING PERSPECTIVE

Dumitru-Clementin CERCEL[1], Ştefan TRĂUŞAN-MATU[2]

In this paper, we focus on the Natural Language Processing (NLP) techniques that influence the precision of the opinion mining results. We analyze the challenges in opinion mining from a NLP perspective in order to describe a method with a better precision of the results. In this way, we select the different NLP techniques that can be used in opinion mining, and then we evaluate their impact on the opinion mining task. We used in our experiments a set of forum data. Experimental results show that by applying NLP techniques such as coreference resolution, negation, and stemming, it is possible to improve the opinion mining performance compared to a basic NLP method which does not use these three techniques.

Keywords: Opinion Mining, POS Tagging, Dependency Parsing, Coreference Resolution, Valence Shifters, Stemming

1. Introduction

Opinion mining is a research field of great interest because many applications are based on opinion mining, including opinion summarization [1], opinion propagation [2], [3] and sentiment prediction [4]. Opinion mining is a research field that can apply Natural Language Processing (NLP) techniques. The main distinction between NLP and opinion mining is that NLP deals with the whole semantics of a text, whereas opinion mining is limited to some semantic aspects of a text, such as the target entity, the opinions about the target entity, the positive, negative, or neutral sentiment of the opinions [5, p. 13].

Esuli and Sebastiani [6] indicate that opinion mining implies two steps. The process of opinion mining begins with a binary classification step in order to distinguish between objective and subjective sentences. Only the subjective sentences are handled further. Thus, for each subjective sentence, the type of polarity is determined: positive or negative. Then, the algorithm finds the sentiment strength of the opinion words: weak sentiment, mild sentiment, or strong sentiment. Although Esuli and Sebastiani consider that *opinion* and

[1] Teaching Assistant, Faculty of Automatic Control and Computers, University POLITEHNICA of Bucharest, Romania, e-mail: clementin.cercel@gmail.com
[2] Professor and Researcher, Faculty of Automatic Control and Computers, University POLITEHNICA of Bucharest, and Research Institute for Artificial Intelligence of the Romanian Academy, Romania, e-mail: trausan@cs.pub.ro

subjectivity are total synonyms, their approach does not extract all the opinions expressed in texts because subjectivity is not equivalent to existence of opinions. Opinion mining and subjectivity analysis are partially similar tasks. There are situations in which a subjective sentence does not imply a positive or negative opinion about something (e.g. "I think that he went home") [5, p. 27]. Similarly, an objective sentence can express sentiments. An example for illustrating this last idea is the following sentence: "The earphone broke in two days" [5, p. 27]. This is an objective sentence, but it implies a negative implicit opinion.

The approach that we use to detect and characterize the opinions from texts is based on NLP techniques and consists from three steps. Firstly, we focus on two basic preprocessing techniques applicable to a text containing opinions: tokenization and Part-of-Speech (POS) tagging. Subsequently, we consider the dependency relations between words as a method used to determine opinion words. Then, we apply a technique intensively studied in the NLP field, coreference resolution. This technique should be considered in opinion mining because otherwise "opinion information will be lost" or "opinions may be assigned to wrong entities" [7]. Finally, we employ two NLP techniques that influence the opinion word sentiment detection step in an opinion mining process. Here, we apply the stemming technique to find opinion words in the opinion lexicons. Moreover, in the case in which an opinion word is modified by a negation word, then the polarity of the opinion word changes.

We evaluate the proposed method on a part of the Internet Argument Corpus [8] on which we make own annotations. The results show that all these particular NLP techniques improve the opinion mining process.

The rest of the paper is divided into the following sections. In Section 2, we review the main methods used for opinion mining. Section 3 describes our approach for opinion mining. Experimental results and discussions are presented in Section 4. In the last section we will discuss the conclusions of this research.

2. State of the Art

Various opinion mining methods have been developed for interpreting the textual data which are in a permanent increase. Typically, the opinion mining methods [9], [10], [11], [12] has focused on dependency parsing. In this paper, we have extended these previous works and have introduced additional NLP techniques such as coreference resolution and negation. The computational treatment of coreference resolution improves in precision the opinion word extraction step. We also improve the opinion word sentiment detection process by incorporating the negation of opinion words.

Methods of identifying the sentiment expressed by an opinion word can be classified into: machine learning methods, corpus-based methods, and opinion lexicon-based methods. The main approaches of machine learning as supervised and unsupervised techniques can be used to classify sentiments [13], [14].

The opinion lexicon-based and corpus-based methods often use seed sets of opinion words. More exactly, given two small seed sets whose orientation we know a priori (i.e. each set has positive and negative opinion words, respectively), we will extend these initial seed sets. As regards the opinion lexicon-based methods [15], in order to infer how strongly or weakly a new opinion word is semantically related to the opinion words from the seed sets, the measures that take into account the semantic relations between words in hierarchies such as WordNet [16] are used. As regards the corpus-based methods [17], the new opinion words are extracted from a domain corpus.

3. Proposed Method

In what follows, we consider that the opinion mining process involves three steps for a given set of text documents and a given target entity: firstly, a text preprocessing step is necessary to conveniently prepare the input data, secondly, identifying opinion words about the target entity in documents (the opinion word extraction step), and finally, determining the sentiment of the opinion words, which can be positive, negative, or neutral (the opinion word sentiment detection step). Although specific methods can be used to solve each of the three steps, we take into account that there are some NLP techniques that influence the results of the opinion mining process. In Figure 1, we outline these NLP techniques for each step.

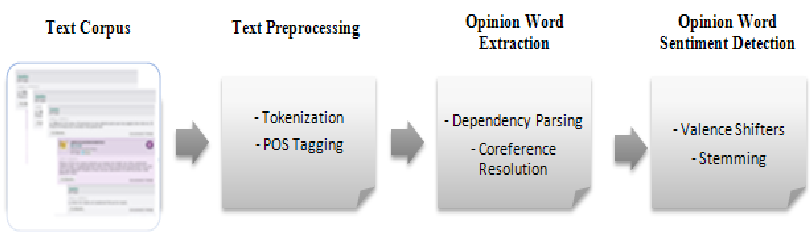

Fig. 1. Overview for our approach to opinion mining using NLP techniques. On the top of the rectangles we mentioned the main steps of the opinion mining process, and inside the rectangles we enumerated the NLP techniques specific to each step.

3.1. Text Preprocessing

Text preprocessing techniques must be adapted for the type of application that implements them. For example, in opinion mining algorithms "we need to avoid removing the syntax-containing words" while text preprocessing techniques are addressed [18, p. 389].

Tokenization is defined by Manning and Schütze [19, p. 124] as a step of preprocessing that consists in splitting "the input text into units called tokens" such as words, numbers, or punctuation marks. Thus, a token is a sequence of characters such as letters, numbers, punctuation marks, or any sign that has a particular meaning. The tokenization process is complex since a sentence may contain syntactically ambiguous forms such as abbreviations (e.g. "Ph.D."), dates (e.g. "13-08-86"), acronyms (e.g. "ACM"), numbers, or punctuation marks.

POS tagging [20], [21] is the task of assigning the corresponding part of speech to each word in an input text. However, may be ambiguities in the POS tagging process. These ambiguities arise because a word can have multiple morphological values depending on context. Many methods have been proposed in order to improve the accuracy of results.

The sentences from the text documents are preprocessed (i.e. tokenization and POS tagging) by using Stanford CoreNLP [22]. Stanford CoreNLP is a complex package and provides support for many NLP tasks.

3.2. Opinion Word Extraction

The goal of this step is to identify opinion words about target entities in a given text document. In our approach, we consider that a target entity is syntactically defined by a noun term and opinion words express opinions about these noun terms and can be one of the following parts of speech: adjective, adverb, or verb.

Performing a morphological and syntactic analysis of a sentence, we can establish dependency relations between its words. Marneffe et al. [23] developed a set of Stanford dependency relations between words in a sentence. Each Stanford dependency relation is a binary grammatical relation between two words. One word is called the *governor* word (also called *head* word) and the other is called the *dependent* word. Formally, a Stanford dependency relation is represented in the form *abbreviated relation name (governor*, *dependent)*, where the dependency relation exists between the *governor* word and the *dependent* word.

From an opinion mining perspective, Stanford dependency relations between words in a sentence have importance because there are some dependency relations that can be used to extract opinions. More concretely, these Stanford dependency relations for opinion mining directly or indirectly establish a syntactic dependency relation between a noun term and the opinion word about this noun term. In Table 1, we summarized the relevant Stanford dependency relations [23] used by us to identify pairs in the form of (*noun term*, *opinion word*) from sentences.

Table 1

The set of dependency relations used for opinion extraction

Dependency Relations	Description
dobj (“direct object”)	The opinion word is a verb or a complement following a copular verb
nsubj (“nominal subject”)	The opinion word is a complement following a copular verb
amod (“adjectival modifier”)	The opinion word is the adjectival modifier of a noun
advmod (“adverbial modifier”)	The opinion word is an adverbial modifier of a verb
acomp (“adjectival complement”)	The opinion word is the adverbial complement of a verb
xcomp (“open clausal complement”)	The opinion word is the complement following a copular verb

Algorithm 1 extracts pairs in the form of (*noun term*, *opinion word*) from a given sentence by using the Stanford dependency relations described above. For each dependency relation, we apply its definition.

Algorithm 1: Identifying Stanford Dependency Relations for Opinion Mining

Input: s – sentence

Output: Ω – a set of pairs (noun term, opinion word)

$\Omega \leftarrow \emptyset$
R $\leftarrow$ Parse(s) // the set $\{rel_j(h_k, d_l)\}_{j,k,l\in\mathbb{N}}$ of all dependency relations
  between the words from the sentence *s*;
for each dependency relation $rel_j(h_k, d_l)$ in R **do**
  if rel_j = “**dobj**” and **checkVerb**(h_k) and **checkNoun**(d_l) **then**
    $\Omega \leftarrow \Omega \cup (h_k, d_l)$
  end if
  if rel_j = “**nsubj**” and **checkAdjective**(h_k) and **checkNoun**(d_l) **then**
    $\Omega \leftarrow \Omega \cup (h_k, d_l)$
  end if
  if rel_j = “**amod**” and **checkAdjective**(h_l) and **checkNoun**(d_k) **then**
    $\Omega \leftarrow \Omega \cup (d_l, h_k)$
  end if
  if rel_j = “**advmod**” and **checkVerb**(d_l) and there is $rel_{jj}(h_{kk}, d_{ll})$ so that
      rel_{jj} = “**rcmod**” and $d_k = h_{kk}$ **then**
    $\Omega \leftarrow \Omega \cup (h_l, d_{ll})$
  end if

if rel_j = "**acomp**" and **checkVerb**(h_k) and there is $rel_{jj}(h_{kk}, d_{ll})$ so that
    rel_{jj} = "**rcmod**" and $h_k = d_{ll}$ **then**
  $\Omega \leftarrow \Omega \cup (d_l, h_{kk})$
end if
if rel_j = "**xcomp**" and **checkVerb**($h_{i,k}$) and there is $rel_{jj}(h_{kk}, d_{ll})$ so that
    rel_{jj} = "**ccomp**" and $h_k = d_{ll}$ **then**
  $\Omega \leftarrow \Omega \cup (h_{kk}, d_l)$
end if
end for

The task of coreference resolution implies the identification of "noun phrases (NPs) that refer to the same entity in a text" [24]. For instance, in the review "I bought a Canon S500 camera yesterday. It looked beautiful. I took a few photos last night. They were amazing", the noun "It" refers to the object "Canon S500 camera", and the pronoun "They" refers to the attribute "photos" [7]. We resolve dependency parsing and coreference by applying NLP tools from the Stanford CoreNLP package.

3.3. Opinion Word Sentiment Detection

After the step of extraction of opinion words from sentences, the next step is to establish the sentiment expressed by each opinion word. We take into account two NLP techniques that influence the opinion word sentiment detection step such as valence shifters and stemming.

Valence shifters are words or expressions that intensify, diminish, or shift the polarity (from positive to negative or vice versa) of opinion words in a sentence. For our purpose we consider negation, as a category of valence shifters. The use of negation changes the polarity of opinion words, from positive into negative. For example, in the sentence "I don't like the snow", the verb "like" has a negative polarity because of the negation "not". Negation can appear in many ways: negative words (e.g. "no", "not", "never", "none", "nobody", "no more", "nowhere"), prefixes (e.g. "un-"), or suffixes (e.g. "-less") [5, p. 32]. To identify negation in sentences, we take into consideration the Stanford dependency relation called *neg* ("negation modifier").

In this paper, we adopt an opinion lexicon-based approach to identify the sentiment expressed by each opinion word. However, there is an issue when the opinion words with the same root refers to a certain inflected form or derived opinion word, but the opinion lexicon contains other morphological forms or different opinion words of the same lexeme. A solution can be given by stemming algorithms that are studied in the NLP area. In a definition given in the NLP literature, stemming means "a process that strips off affixes and leaves you with a stem" [19, p. 132]. In inflectional languages, besides some invariable words (i.e. without inflectional forms), words usually have the following structure in a sentence: a stem, which carries the lexical sense, plus inflectional affixes, which have not an independent sense. After removing the inflectional affix remains the stem. We also work with Stanford CoreNLP to perform stemming on opinion words.

Opinion lexicons represent important resources to identify the sentiment of opinion words. There are available several opinion lexicons that enable us to classify opinion words into one of the following sentiment categories: positive, negative, or neutral. In this way, we classify the opinion words by applying the algorithms presented in [25]. These algorithms use the following four opinions lexicons:

- *SentiWordNet* version 3.0 [26] is an opinion lexicon in which the opinion words are classified into three sentiment categories: positive, negative, and neutral. For this purpose, an automatic annotation of all the synsets contained in the WordNet 3.0 lexical database is performed, so that each synset *s* receives three real sentiment scores *Pos(s)*, *Neg(s)*, and *Obj(s)* depending on the respective positive, negative, and neutral sentiment of its words. Finally, this opinion lexicon contains 155,287 words from 117,659 synsets.
- The *Micro-WNOp* [27] lexicon is composed of a subset of WordNet synsets, more precisely 1,105 synsets, which represent 1,960 distinct words. Unlike SentiWordNet, in the Micro-WNOp lexicon, the annotation of the sentiments of the synsets is made manually. Every Micro-WNOp synset was annotated with two numerical scores: one score indicates the grade of positivity of its component words, and another score indicates the grade of negativity of the component words of the same synset.
- The MPQA subjectivity lexicon [28] comprises 8,221 subjective clues defined as single-word clues and which have "subjective usages". Of the final subjective words, 33.1% have positive sentiment, 59.7% have negative sentiment, 0.3% have both positive and negative sentiment, and 6.9% have neutral sentiment.
- Bing Liu's opinion lexicon [29] contains 6,786 opinion words. Opinion words in this lexicon are divided into two sets according to their sentiment: a set of opinion words with positive sentiment and another set of opinion words with negative sentiment. The opinion lexicon created by Bing Liu has been semi-automatically generated.

4. Experimental Results

In this section, we analyze our experiments, including the data set and the evaluation metrics.

4.1. Dataset

We performed experiments on forum data from the Internet Argument Corpus (IAC) created by Walker et al. [8] to evaluate the effectiveness of the method presented in this paper. This corpus contains forum threads collected from the website *4forums*[3]. IAC is a dataset freely available on the Internet and each forum thread is saved in the JSON[4] format.

In all the below-presented experiments, we used seven forum threads from the Internet Argument Corpus. The selected forum threads have been annotated by two human annotators. These have annotated each forum thread from two points of view. Firstly, they are extracted all occurrences of opinion words. Then they are asked to label each opinion word with the corresponding sentiment from the following variants: positive, neutral, or negative sentiment. Finally, the general agreement between human annotators is 85.39%.

4.2. Results

First, we evaluate the effectiveness of the opinion word extraction step. In this experiment, we compare two methods: dependency parsing (depend) and dependency parsing with coreference resolution (depend + coref). We use a set of well-known evaluation metrics from information retrieval such as precision, recall and F-measure [30] to quantify the performance of this step. In Table 2, we show a comparison between the two methods. All these results indicate the effectiveness of usage of the dependency parsing with coreference resolution in the opinion word extraction step.

Table 2

Performance of opinion word extraction

Method	Precision	Recall	F-measure
Depend	80.56%	74.23%	77.27%
depend + coref	85.12%	78.43%	81.64%

Next, we present the results for the opinion word sentiment detection step. For this purpose, we compare the performance of the following three methods: the opinion lexicon-based method but without considering stemming (lexicon), the opinion lexicon-based method with stemming (lexicon + stem), the opinion lexicon-based method with both stemming and negation (lexicon + stem + neg). We use the precision measure to evaluate the effectiveness of this step. We present all these results in Table 3.

[3] http://www.4forums.com/
[4] http://json.org/

Table 3

Performance of opinion word sentiment detection

Method	SentiWordNet	Micro-WNOp	MPQA Subjectivity Lexicon	Bing Liu's Lexicon
Lexicon	53.25%	34.24%	51.48%	49.97%
lexicon + stem	74.9%	48.16%	72.41%	70.28%
lexicon + stem + neg	78.77%	52.03%	76.28%	74.15%

A comparison between these results indicate that the performance of this step using the opinion lexicon-based method with stemming is higher than using only the opinion lexicon-based method by 21.65% precision for SentiWordNet. This result indicates that English is an inflectional language and individuals use in forum data the inflected or derived opinion words.

Then we see, by incorporating negation into the opinion lexicon-based method with stemming, performance is further improved. The opinion lexicon-based method with both stemming and negation outperforms both the only opinion lexicon-based method and the opinion lexicon-based method considering stemming with precision equal to 25.52% and 3.87%, respectively, for SentiWordNet. This last result shows that negation has an important role in sentiment classification. This may be due to the fact that its frequency of usage in texts is high.

As resulting from Table 3, the low results are obtained for the Micro-WNOp lexicon because it does not contain enough synsets, but only one subset of English words. Thus, many opinion words are not in the Micro-WNOp lexicon.

The reason for choosing several opinion lexicons to perform tests is that there is no opinion lexicon to return the exact sentiment of opinion words by taking into consideration both the context and domain in which these opinion words are used.

We thus can observe that the NLP techniques play an important role for each step in the proposed opinion mining process. Moreover, the three steps being successive, not solving one of them leads to the propagation of errors in the following steps. Although intensively studied in the NLP research community, these NLP techniques are not completely solved.

5. Conclusions

In this paper we address the research challenges of the opinion mining problem from a NLP perspective. We previously presented an opinion mining method which consists of specific steps: (1) a preprocessing step to conveniently prepare the input data; (2) a step of extraction of opinion words about target entities; (3) a sentiment classification step of each opinion word. Even if each of the three steps can be implemented by means of specific methods, we use NLP techniques in each step.

Experimental results show the effectiveness of the proposed methods both in opinion word extraction and opinion word sentiment detection steps. The experimental results for the opinion word extraction step show that by combining dependency parsing with coreference resolution improves the results of the baseline method containing only dependency parsing by 4.56% precision. We also note that for the opinion word sentiment detection step, the opinion lexicon-based method with both stemming and negation achieves the best precision of 78.77%.

As future work we intend to incorporate and other NLP techniques into the opinion mining method presented in this paper like opinion-word sentiment disambiguation. Moreover, there are and other categories of valence shifters: ironic sentences, presuppositional items such as the adverbs "barely" and "hardly", modifiers that enhance or weaken the strength of the sentiment of opinion words, modal verbs used to express opinions such as possibility or necessity, or negation [31].

REFERENCES

[1] *A. Balahur, M.A. Kabadjov, J. Steinberger, R. Steinberger, A. Montoyo,* Summarizing Opinions in Blog Threads, in: PACLIC, 2009, pp. 606-613.
[2] *D.-C. Cercel, S. Trausan-Matu*, Opinion Propagation in Online Social Networks: A Survey, in: Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14), ACM, Thessaloniki, Greece, 2014, pp. 1-10.
[3] *D.-C. Cercel, S. Trausan-Matu*, User-Level Opinion Propagation Analysis in Discussion Forum Threads, in: G. Agre, P. Hitzler, A. Krisnadhi, S. Kuznetsov (Eds.) Artificial Intelligence: Methodology, Systems, and Applications, Springer International Publishing, 2014, pp. 25-36.
[4] *J. Kim, J.-B. Yoo, H. Lim, H. Qiu, Z. Kozareva, A. Galstyan*, Sentiment Prediction Using Collaborative Filtering, in: ICWSM, 2013.
[5] *B. Liu*, Sentiment Analysis and Opinion Mining, Morgan & Claypool, 2012.
[6] *A. Esuli, F. Sebastiani,* Determining the semantic orientation of terms through gloss classification, in: Proceedings of the 14th ACM international conference on Information and knowledge management, ACM, Bremen, Germany, 2005, pp. 617-624.
[7] *X. Ding, B. Liu*, Resolving object and attribute coreference in opinion mining, in: Proceedings of the 23rd International Conference on Computational Linguistics, Association for Computational Linguistics, Beijing, China, 2010, pp. 268-276.
[8] *M.A. Walker, J.E.F. Tree, P. Anand, R. Abbott, J. King*, A Corpus for Research on Deliberation and Debate, in: N. Calzolari, K. Choukri, T. Declerck, M.U. Dogan, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis (Eds.) LREC, European Language Resources Association (ELRA), 2012, pp. 812-817.
[9] *R. Johansson, A. Moschitti*, Syntactic and semantic structure for opinion expression detection, in: Proceedings of the Fourteenth Conference on Computational Natural Language Learning, Association for Computational Linguistics, 2010, pp. 67-76.
[10] *Y. Wu, Q. Zhang, X. Huang, L. Wu*, Phrase dependency parsing for opinion mining, in: Proceedings of the 2009 Conference on Empirical Methods in Natural Language

Processing: Volume 3-Volume 3, Association for Computational Linguistics, 2009, pp. 1533-1541.
[11] *M.S. Almeida, C. Pinto, H. Figueira, P. Mendes, A.F. Martins*, Aligning opinions: Cross-lingual opinion mining with dependencies, in: Proc. of the Annual Meeting of the Association for Computational Linguistics, 2015.
[12] *M. Zhang, W. Che, Y. Shao, T. Liu*, Improve Chinese Semantic Dependency Parsing via Syntactic Dependency Parsing, in: Asian Language Processing (IALP), 2012 International Conference on, IEEE, 2012, pp. 53-56.
[13] *A. Bermingham, A.F. Smeaton*, Classifying sentiment in microblogs: is brevity an advantage?, in: Proceedings of the 19th ACM international conference on Information and knowledge management, ACM, 2010, pp. 1833-1836.
[14] *R. Xia, C. Zong, S. Li*, Ensemble of feature sets and classification algorithms for sentiment classification, Publisher, City, 2011.
[15] *J. Kamps, M. Marx, R.J. Mokken, M. de Rijke,* Using WordNet to measure semantic orientation of adjectives, in: LREC, 2004.
[16] *G.A. Miller*, WordNet: a lexical database for English, Publisher, City, 1995.
[17] *V. Hatzivassiloglou, K.R. McKeown*, Predicting the semantic orientation of adjectives, in: Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, Association for Computational Linguistics, Madrid, Spain, 1997, pp. 174-181.
[18] *C.C. Aggarwal, C. Zhai*, Mining Text Data, Springer, 2012.
[19] *C.D. Manning, H. Schütze*, Foundations of statistical natural language processing, MIT Press, 1999.
[20] *D.-C. Cercel, Ş. Trăuşan-Matu*, POS tagger bazat pe modelul HMM de ordinul doi, Publisher, City, 2012.
[21] *D.-C. Cercel, S. Trausan-Matu*, A POS Tagger analysed in collaboration environments and literary texts, in: Proceedings of EGC Conference, Toulouse, France, 2013, pp. 287-292.
[22] *C.D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S.J. Bethard, D. McClosky*, The Stanford CoreNLP natural language processing toolkit, in: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 2014, pp. 55-60.
[23] *M.-C. de Marneffe, C. Manning*, Stanford typed dependencies manual, in, Technical report, Stanford University, 2008.
[24] *V. Stoyanov, C. Cardie, N. Gilbert, E. Riloff, D. Buttler, D. Hysom*, Coreference resolution with reconcile, in: Proceedings of the ACL 2010 Conference Short Papers, Association for Computational Linguistics, Uppsala, Sweden, 2010, pp. 156-161.
[25] *D.C. Cercel, S. Trausan-Matu*, Opinion Influence Analysis in Online Forum Threads, in: Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 2014, pp. 228-235.
[26] *A.E.S. Baccianella, F. Sebastiani*, SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining, in: Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), European Language Resources Association (ELRA), Valletta, Malta, 2010.
[27] *S. Cerini, V. Compagnoni, A. Demontis, M. Formentelli, G. Gandini*, Micro-WNOp: A gold standard for the evaluation of automatically compiled lexical resources for opinion mining, in: A. Sans (Ed.) Language resources and linguistic theory: Typology, second language acquisition, English linguistics, Franco Angeli Editore, 2007.
[28] *T. Wilson, J. Wiebe, P. Hoffmann*, Recognizing contextual polarity in phrase-level sentiment analysis, in: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Vancouver, British Columbia, Canada, 2005, pp. 347-354.

[29] *M. Hu, B. Liu*, Mining and summarizing customer reviews, in: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, Seattle, WA, USA, 2004, pp. 168-177.
[30] *S. Robertson*, Evaluation in information retrieval, in: Lectures on information retrieval, Springer, 2001, pp. 81-92.
[31] *L. Polanyi, A. Zaenen*, Contextual Valence Shifters, in: J. Shanahan, Y. Qu, J. Wiebe (Eds.) Computing Attitude and Affect in Text: Theory and Applications, Springer Netherlands, 2006, pp. 1-10.