Компьютерная лингвистика и интеллектуальные технологии

По материалам ежегодной международной конференции «Диалог» (2016)

Выпуск 15

Computational Linguistics and Intellectual Technologies

Proceedings of the Annual International Conference "Dialogue" (2016)

Issue 15

УДК 80/81; 004
ББК 81.1
К63

Программный комитет конференции выражает
искреннюю благодарность Российскому фонду
фундаментальных исследований за финансовую поддержку

Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной международной конференции «Диалог» (Москва, 1–4 июля 2016 г.). Вып. 15 (22). — М.: Изд-во РГГУ, 2016.

Сборник включает 67 докладов международной конференции по компьютерной лингвистике и интеллектуальным технологиям «Диалог 2016», представляющих широкий спектр теоретических и прикладных исследований в области описания естественного языка, моделирования языковых процессов, создания практически применимых компьютерных лингвистических технологий.

Для специалистов в области теоретической и прикладной лингвистики и интеллектуальных технологий.

Содержание

Приглашенные доклады

Основная программа конференции

AN OPINION WORD LEXICON AND A TRAINING DATASET FOR RUSSIAN SENTIMENT ANALYSIS OF SOCIAL MEDIA

Koltsova O. Yu. (ekoltsova@hse.ru)[1],
Alexeeva S. V. (salexeeva@hse.ru)[1,2],
Kolcov S. N. (skoltsov@hse.ru)[1]

[1]National Research Institute Higher School of Economics, St. Petersburg, Russia

[2]St. Petersburg State University, St. Petersburg, Russia

Automatic assessment of sentiment in large text corpora is an important goal in social sciences. This paper describes a methodology and the results of the development of a system for Russian language sentiment analysis that includes: a publicly available sentiment lexicon, a publicly available test collection with sentiment markup and a crowdsourcing website for such markup. The lexicon is aimed at detecting sentiment in user-generated content (blogs, social media) related to social and political issues. Its prototype was formed based on other dictionaries and on the topic modeling performed on a large collection of blog posts. Topic modeling revealed relevant (social and political) topics and as a result—relevant words for the lexicon prototype and relevant texts for the training collection. Each word was assessed by at least three volunteers in the context of three different texts where the word occurred while the texts received their sentiment scores from the same volunteers as well. Both texts and words were scored from −2 (negative) to +2 (positive). Of 7,546 candidate words, 2,753 got non-neutral sentiment scores. The quality of the lexicon was assessed with *SentiStrength* software by comparing human text scores with the scores obtained automatically based on the created lexicon. 93% of texts were classified correctly at the error level of ±1 class, which closely matches the result of *SentiStrength* initial application to the English language tweets. Negative classes were much larger and better predicted. The lexicon and the text collection are publicly available at http://linis-crowd.org.

Key words: sentiment lexicon, web interface, crowdsourcing sentiment markup, Russian blogosphere, livejournal, test collection, topic modeling

1. Introduction

Sentiment analysis (SA) in Russia has so far been focused on polarity detection in customer reviews: this, for instance, can be clearly seen from the content of the Russian Information Retrieval Seminar (*ROMIP*) competition on SA (Chetviorkin et al, 2012; Chetviorkin, Loukachevitch, 2013; Loukachevitch et al, 2015). However,

marketing professionals are not the only potential "consumers" of automatic sentiment analysis techniques. Social scientists get increasingly interested in "online public opinion" on various social and political issues or events, as well as in predicting public reaction to those events with online sentiment data. At the moment, no Russian language sentiment lexicons or machine learning instruments are publicly available (exception: Chetviorkin-Loukashevich dictionary of sentiment-bearing words with undefined polarity for consumer reviews in three domains). As a result, researchers in Russia can only rely on commercial services whose methodologies are never completely disclosed. This is often unacceptable for academic users.

This work seeks to make a first step in the development of freely available resources for the Russian language SA We develop a domain-specific sentiment lexicon and check its quality against the marked-up collection of political and social post fragments written by top bloggers at the most popular Russian blog platform *LiveJournal*. Our sentiment analysis task here is reduced to a relatively simple classification of texts into those with prevailing negative emotions and those with prevailing positive emotions, irrespectively of the object of these sentiments—that is, we do not solve a political support/oppose classification task.

The rest of the paper is organized as follows. We first take a visit on the previous literature. Next, we explain our data collection, sentiment lexicon formation, and the markup procedure. Then, we report word and text assessment results and analyze the quality of the lexicon. Finally, we close the paper with a conclusion.

2. Related work

SA can be conventionally divided into two main approaches (Pang, Lee, 2008; Medhat et al, 2014):

(1) Lexicon-based approach (Taboada et al, 2011). It browses texts for certain words or phrases whose polarity has been predefined, often in relation to the domain of interest. Such thesauri are often supplemented with a set of rules concerning the use of negation or booster words. Some of the well-known limitations of this approach are domain sensitivity and initial lexical insufficiency while its simplicity is one of its main advantages.

(2) Machine learning approach. It uses marked-up text collections (training datasets), as well as feature lists, as information which a mathematical algorithm relies on while classifying other marked-up collections (test sets). Most of such algorithms optimize until the best possible fit with the test set markup is reached. After that, these algorithms are applied to non-marked-up (real world) collections. This more sophisticated approach most often yields better results, however, it is vulnerable for overfitting and requires large marked-up corpora to produce high quality.

In addition, these two approaches work differently for different tasks. For instance, SVM method for the task of binary classification of English-language movie reviews has yielded precision of 86.4% (Pang, Lee 2004), which is particularly

high. At the same time, a lexicon-based approach has been successfully used for sentiment analysis of English language social media with *SentiStrength* system (Thelwall et al, 2010) (for more details see section 4). For the Russian language, during the *ROMIP* SA competition in 2012, the best results in consumer review classification tasks were obtained by machine learning approaches, however, in political news classification, lexicon-based approaches took the lead (Chetviorkin, Loukachevitch, 2013). The competition organizers attribute this latter success to the great diversity of topics (sub-domains) occurring in the news and to the absence of a sufficient training set.

These two conditions are met by user-generated social and political content from blogs and social media, the object of our interest, which is why we have chosen the lexicon-based approach as a first step.

Two main methods of sentiment lexicon generation—manual and semi-automatic—are usually described in literature (Mohammad, Turney, 2013; Taboada et al, 2011). The manual method is a human markup of words into sentiment classes, which can be very reliable when qualified experts are used. Among limitations of this approach are its labor-intensive character (although not more intensive than in the creation of marked-up text collections) and the mentioned above initial insufficiency. The latter means that initially it is hard to think of all potentially sentiment-bearing words without additional methods of their extraction.

This problem is addressed by semi-automatic methods of lexicon generation, notably by bootstrapping techniques (Thelen, Riloff, 2002; Godbole et al, 2007). They start with small lists of words with pre-defined polarity (seed words) and automatically extend them with a number of linguistic instruments. Those include measurement of sematic association between words (Turney, 2002), synonym/antonym dictionaries (Hu, Liu, 2004) or general dictionaries or various pre-existing taxonomies (Esuli, Sebastiani, 2005). Sentiment lexicons developed for other languages are also applied (Mihalcea et al, 2007), although in our experience their usefulness is limited. Sentiment-bearing adverbs may be automatically derived from the respective seed adjectives (Taboada et al, 2011), which is a technique we have borrowed. Chetviorkin and Loukashevitch (2012) offer a methodology of detecting sentiment-bearing words (but not their polarity) for Russian language customer reviews: having manually annotated 18,362 words, they then train a classifier to detect more sentiment-bearing words and show a good quality.

Thus, semi-automatic approaches may solve the problem of labor-intensiveness in manual lexicon construction only partially, while marked-up collections can be a solution only when they emerge without researchers' effort (e.g. consumer reviews). Classification of other types of content in resource-scarce languages faces a cold start problem. In recent years, it is increasingly often addressed with crowdsourcing, both in SA (Hong et al, 2013) and other linguistic tasks (Mohammad, Turney, 2011). Crowdsourcing, as a technique relying on cheap or free labor of a large number of lay persons, brings about its own problems, notably the issue of insufficient quality resulting from the lack of qualification or motivation. Approaches to coping with this are still in their cradle. While Hong et al (2013) suggest to motivate volunteers through gamification, Hsueh et al (2009) develop a number of methods to detect

and discard low-quality assessments. The gold standard in both cases is, however, expert opinion, which itself is prone to individual biases when it comes to the polarity of political texts. In addition, resource-scarce languages may be resource-scarce precisely because crowdsourcing services are unavailable either for technical or financial reasons. We address some of these problems further below.

3. Data collection and markup

3.1. Generating relevant text collection

We extract our collection from our database that includes all posts by top 2,000 LiveJournal bloggers for the period of one year (from March 2013 to March 2014). Earlier we found out that only about a third of those texts may be classified as political or social (Koltsova et al, 2014), hence, we face a problem of retrieving relevant texts. While Hsueh et al (2009) employ manual annotation, this is unfeasible for our collection of around 1.5 million texts, so we adopt a different approach (Koltsova, Shcherbak, 2015). We perform topic modeling, namely Latent Dirichlet Allocation with Gibbs sampling (Steyvers, Griffiths, 2004). It yields results akin to fuzzy clustering, by ascribing each text to each topic out of a predefined number, with a varying probability, based on word co-occurrence. All words are also ascribed to all topics with varying probabilities. When sorted by this probability they form lists that allow fast topic interpretation and labeling by humans. Our *TopicMiner* software (http://linis.hse.ru/soft-linis) was used for all topic modeling procedures.

Our prior experience shows that the optimal number of topics depends most of all on the "size" of topics to be detected (smaller topics demand a larger number). A series of experiments (Bodrunova et al, 2013; Nikolenko et al, 2015) has lead us to choose the number of 300 for the task of retrieving social and political topics. 100 most relevant texts and 200 words of each topic were read by three annotators who have identified 104 social or political topics. The topic was considered relevant if two of the three annotators had chosen it. Intercoder agreement, as expressed by Krippendorf's alpha, is 0.578. Texts with the probability higher than 0.1 in these 104 topics (mean probability = 0.3) were considered relevant and were included into the final working collection which comprised 70,710 posts.

3.2. Selection of potentially sentiment-bearing words

Based on the aforementioned literature, we employed a complex approach to the generation of a proto-lexicon for manual annotation (all details on this approach can be found in Alexeeva et al, 2015) comprising the following elements:

- the list of high-frequency adjectives created by the *Digital Society Lab* (http://digsolab.ru) based on a large collection of Russian language texts from social media, and the list of adverbs automatically derived from the former list;

- Chetviorkin-Loukashevitch lexicon (Chetviorkin, Loukachevitch, 2012) (most of it later discarded);
- Explanatory Dictionary of the Russian Language (Morkovkin, 2003);
- Translation of the free English-language lexicon accompanying *SentiStrength* software (Thelwall et al, 2010);
- 200 most probable words for each of the relevant topics identified by annotators, which was aimed at detecting domain-specific words.

We formed a lexicon of potentially sentiment-bearing words accepting only those that occurred in at least two of the listed sources. The lexicon comprised 9,539 units. However, only 7,546 of them occurred in the texts identified as social or political, and only they were later manually annotated.

3.3. Data markup and evaluation of crowdsourcing results

To avoid some pitfalls of crowdsourcing we have adopted, so to say, a sociological vision of it: our volunteers were not supposed to imitate experts; rather, their contribution was seen as similar to that of respondents in an opinion poll which cannot produce "wrong" answers. For that, we tried to make our sample of assessors as diverse as possible in terms of region, gender, and education. In total, 87 people from 16 cities took part in the assessment.

They worked with our website linis-crowd.org and assessed words' sentiment as expressed in the texts in which they occurred, as well as the prevailing sentiment of the texts themselves, with a five-point scale, from -2 (strong negative), to +2 (strong positive). The texts were to help detect domain-specific word polarity.

Each word was shown with three different texts, one at a time. Each post was cut down to one paragraph since long texts are more likely to include different sentiments. Once each word received three annotations, we went on with further annotation to get more than one assessment for each text. By the time of data analysis, we received 32,437 word annotations and the same number of text annotations (of them, 14 word annotations and 18 text annotations were discarded due to technical errors). In total, the assessors annotated 19,831 texts. Annotated word and text collections are available at http://linis-crowd.org/.

Intercoder agreement in word assessment task (five-class), as expressed by Krippendorf's α, has turned out to be 0.541. To compare, Hong et al (2013) report α as low as 0.11–0.19 for a three-class word sentiment annotation task. Taboada et al (2011: 289) obtain mean pairwise agreement (MPA) of 67.7% in a three-class task of word assessment in customer reviews (NB: α and MPA are not directly comparable). In text annotation task we obtained α=0.278 for all texts and 0.312 for the texts that got non-zero scores (five-class). Hsueh et al (2009) report MPA among Amazon Mechanical Turk annotators to be 35.3% for a four-class task of political blog posts annotation. Ku et al (2006) claim to reach a much higher agreement of 64.7% for four-class blog annotation and 41.2% for news annotation by specially selected assessors. Nevertheless, none of these levels is impressively high. In relation to this, Hsueh and colleagues

(2009) point at the problem of political blogs' ambiguity. We tend to agree that this ambiguity and general lack of societal consensus on the polarity of political issues, not (or at least not only) the lack of quality, cause the low agreement. Therefore, disagreeing individuals cannot be filtered out because they may reflect an important part of the public opinion spectrum. A milder measure of divergence of an annotator's mean score from the global mean allows for a lot of disagreement on individual items. It shows that in our case only 0.5% of all annotations were made by individuals strongly deviating from the global mean.

4. Results

4.1. Word and text assessment results

The majority of words (4,753) were annotated as neutral and therefore excluded from the lexicon. Table 1 shows that although negatively assessed words prevail, positive words have been also detected. At the same time, highly emotional words are quite a few.

Table 1. Distribution of mean scores over words

Mean score (rounded)	Number of words with such score	Share of words with such score, %
−2	225	3
−1	1,666	22
0	4,753	63
1	853	11
2	49	0.6

We have also calculated the variance of scores for each word. Although, as mentioned above, disagreement not necessarily indicates low quality, the usefulness of highly disputable words for sentiment classification of texts is doubtful. Since distance between two neighboring values of scores is one, we have regarded all words with variance =>1 as candidates for being discarded. However, we have found only 153 such words, and most of them looked like sentiment-bearing. In some cases their polarity seemed quite obvious: e.g. gorgeous (*сногсшибательный*), filth (*мерзость*), long-suffering (*многострадальный*), first-class (*первосортный*), while others looked ambiguous: e.g. endless (*бесконечный*), quality (*качество*), and tolerance (*терпимость*). At this stage of the research they have been included in the lexicon with their mean scores (since their number is anyway negligibly small).

The distribution of scores over texts is similar to that of words (see Table 2). Most texts were marked as neutral. Positive class size is obviously insufficient (it relates to the negative class as 1:4.6). The same unbalanced class structure in political blogs is also pointed at by Hsueh et al (2009).

Table 2. Distribution of mean scores over texts

Mean score (rounded)	Number of texts with such score	Share of texts with such score
−2	75	0.4
−1	6,546	34
0	11,760	61
1	1,427	7
2	23	23

4.2. Lexicon quality evaluation

After neutral word filtering and leaving out the words that did not occur in the relevant texts, our lexicon comprised 2,753 items. We installed this lexicon into *SentiStrength* freeware for quality evaluation (Thelwall et al, 2010). All texts were lemmatized with *MyStem2* (Segalovich, 2003) prior to SA. In the default mode, *SentiStrength* ascribes two sentiment scores to each text: one corresponds to the maximal negative word score in the text, and the other—to the maximal positive score. If booster words occur before a given word, the absolute value of its score is increased. If negation is used, the sign of the score is reversed. The integrated text score was then calculated as the mean of the two *SentiStrength* scores.

We defined lexicon quality as the share of correctly detected cases. We first calculated the absolute difference between the rounded mean assessors' score of a given text and the rounded integrated score based on *SentiStrength* results. Then we obtained the share of the exact matches, as well as the share of ±1 class matches, as offered by *SentiStrength* developer (Thelwall et al, 2010). A ±1 class match means that if a text is ascribed to one of the two neighboring classes, the class is considered correctly predicted. In our case the share of ±1 class matches comprises 93.0% which is comparable to Thelwall's results—96.9% (Thelwall et al, 2010). Prediction of the negative classes is better than that of the positive ones (95% and 59% for '−1' and '−2' classes vs. 82% and 19% for '+1' and '+2' classes). As it can also be seen, moderate classes are predicted much better than extreme classes, which are very small, while the dominant '0' class yields 99.6% of ±1 class matches.

SA systems for Russian use different evaluation techniques. The closest to our case was the *ROMIP* SA competition held on texts from political news and from blogs containing customer opinions (Chetviorkin, Loukachevitch, 2013). As sentiment lexicons are domain sensitive, it would be unfair to directly test our lexicon on the texts of a different type and to compare it to the approaches that were developed specially for this type. It would be equally unfair to apply the *ROMIP* methods to our collection. We therefore performed an indirect comparison of the results, using the same methodology of quality evaluation as *ROMIP*. Its best participants in a three-class blog classification task exceeded their baseline by 12–27% in terms

of recall and by 5–29% in terms of precision. In news classification task the respective values were 23–28% and 43–49%. Having converted our data into three classes (positive, negative and neutral), we calculated our baseline, precision and recall (see Table 3).

Table 3. Three-class classification quality

	Recall (macro)	Precision (macro)
Our lexicon	0.43	0.44
Baseline	0.33	0.18
Difference	0.10	0.26

The quality of our lexicon is comparable to that of the *ROMIP* approaches used in the blog classification task and is lower than the quality reached for news. It should be noted that class distribution of the *ROMIP* news collection was much more balanced (Panicheva, 2013) than that of both its blog collection and of our sample. This has made the task of exceeding the baseline more difficult in blog SA. In contrast to most *ROMIP* methods, our lexicon is publicly available and may be improved by the research community.

5. Conclusion and future research

We have presented a lexicon for sentiment analysis of political and social Russian-language blogs. Its quality is comparable to the results obtained for English-language Twitter and for Russian-language blogs with customer opinions. We have also described the results of words and texts annotation based on a crowdsourcing approach. The lexicon and the annotated collection are publicly available at our website linis-crowd.org that allows further crowdsourcing of sentiment markup. This web resource is aimed at the widest research community. While the lexicon can be already used by social scientists, the collection may serve as a benchmark for testing new sentiment instruments. In particular, we are now using it for training machine learning SA algorithms that should help increase the quality of SA. We also plan to improve the lexicon by replicating our research on a collection of blog comments that are potentially much more emotional.

6. Acknowledgements

This work was supported by the Russian Foundation for Humanities, project 'Development of a publicly available database and a crowdsourcing website for testing sentiment analysis instruments', Grant No 14-04-1203.

References

1. *Alexeeva S., Koltsova E., Koltcov S.* (2015) Linis-crowd.org: A lexical resource for Russian sentiment analysis of social media [Linis-crowd.org: lexicheskij resurs dl'a analiza tonal'nosti sotsial'no-politicheskix tekstov], Computational Linguistics and computantional ontologies: Proceedings of the XVIII joint Conference "Internet and modern society (IMS-2015)" [Kompyuternaya lingvistika i vyichislitelnyie ontologii: sbornik nauchnyih statey. Trudyi XVIII ob'edinennoy konferentsii «Internet i sovremennoe obschestvo» (IMS-2015)], St. Peterburg, pp. 25–34.
2. *Bodrunova S., Koltsov S., Koltsova O., Nikolenko S., Shimorina A.* (2013) Interval Semi-Supervised LDA Classifying Needles in a Haystack, Proceeding of the 12th Mexican International Conference on Artificial Intelligence (MICAI 2013) Part I: Advances in Artificial Intelligence and Its Applications, Berlin: Springer Verlag, pp. 265–24.
3. *Chetviorkin I. I., Braslavski P. I., Loukachevitch N. V.* (2012), Sentiment Analysis Track at ROMIP 2011, Proceedings of International Conference Dialog, pp. 739–746.
4. *Chetviorkin I., Loukachevitch N.* (2012) Extraction of Russian Sentiment Lexicon for Product Meta-Domain, Proceedings of COLING 2012: Technical Papers, pp. 593–610.
5. *Chetviorkin I., Loukachevitch N.* (2013) Sentiment Analysis Track at ROMIP 2012, Proceedings of International Conference Dialog, Vol. 2, pp. 40–50.
6. *Esuli A., Sebastiani F.* (2006) SentiWordNet: A publicly available lexical resource for opinion mining, Proceedings of 5th International Conference on Language Resources and Evaluation (LREC), Genoa, pp. 417–422.
7. *Godbole N., Srinivasaiah M., Skiena S.* (2007) Large Scale Sentiment Analysis for News and Blogs, ICWSM'2007, Boulder, Colorado, USA.
8. *Hong Y., Kwak H., Baek Y., Moon S.* (2013) Tower of Babel: a crowdsourcing game building sentiment lexicons for resource-scarce languages, Proceedings of the 22nd International World Wide Web Conference (WWW), pp. 549–556.
9. *Hsueh P., Melville P., Sindhwani V.* (2009) Data quality from crowdsourcing: a study of annotation selection criteria, Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing, Boulder, Colorado, pp. 27–35.
10. *Hu M., Liu B.* (2004) Mining and summarizing customer reviews, Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), Seattle, WA, pp. 168–177.
11. *Koltsova O., Koltcov S., Alexeeva S.* (2014) Do ordinary bloggers really differ from blog celebrities? Proceedings of WebSci ‹14 ACM Web Science Conference, Bloomington, IN, USA, NY: ACM, pp. 166–170.
12. *Koltsova O., Shcherbak A.* (2015) 'LiveJournal Libra!': The political blogosphere and voting preferences in Russia in 2011–2012, New Media and Society, vol. 17, no. 10, pp. 1715–1732.
13. *Ku L.-W., Liang Y.-T., Chen H.-H.* (2006) Opinion Extraction, Summarization and Tracking in News and Blog Corpora, Proceedings of the AAAI-CAAW›06.

14. *Loukachevitch N. V., Blinov P. D., Kotelnikov E. V., Rubtsova Y. V., Ivanov V. V., Tutubalina E.* (2015) SentiRuEval: testing object-oriented sentiment analysis systems in Russian, Proceedings of International Conference Dialog, Vol. 2.
15. *Medhat W., Hassan A., Korashy H.* (2014) Sentiment analysis algorithms and applications: a survey, Ain Shams Engineering Journal, Vol. 5, Issue 4, pp. 1093–1113.
16. *Mihalcea R., Banea C., Wiebe J.* (2007) Learning multilingual subjective language via cross-lingual projections, Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic, pp. 976–983.
17. *Mohammad S, Dorr B., Hirst G., Turney P.* (2011) Measuring degrees of semantic opposition, Technical report, National Research Council Canada.
18. *Mohammad S. M., Turney, P. D.* (2013), Crowdsourcing a word-emotion association lexicon, Computational Intelligence, Vol. 29 no. 3, pp. 436–465.
19. *Morkovkin V. V.* (2003) Explanatory dictionary of Russian language: structural words: prepositions, conjunctions, particles, interjections, parentheses, pronouns, numbers, connections [Ob'jasnitelnyj slovar' russkogo jazyka: Structurnyje slova: predlogi, sojuzy, chastitsy, mezhdometija, vvodnyje slova, mestoimenija, chislitelnyje, svjazannyje slova], Astrel, Moscow.
20. *Nikolenko S., Koltcov S., Koltsova O.* (2015) Topic Modeling for Qualitative Studies. Journal of Information Science (R&R).
21. *Pang B., Lee L.* (2004) A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts, 42nd Meeting of the Association for Computational Linguistics[C] (ACL-04), pp. 271–278.
22. *Pang B., Lee L.* (2008) Opinion Mining and Sentiment Analysis, Foundations and Trends in Information Retrieval, Vol. 2, no. 1–2, pp. 1–135.
23. *Panicheva P.* (2013) ATEX: a rule-based sentiment analysis system processing texts in various topics [Sistema sentimentnogo analiza ATEX osnovannaya na pravilah pri obrabotke tekstov razlichnyh tematik], Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2013" [Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Trudy Mezhdunarodnoy Konferentsii "Dialog 2013"], pp. 101–113.
24. *Segalovich I.* (2003) A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search engine, Proceedings of MLMTA'2003, pp. 273–280.
25. *Steyvers M., Griffiths T.* (2004) Finding scientific topics, Proceedings of the National Academy of Sciences, Vol. 101, no. Suppl. 1, pp. 5228–5235.
26. *Taboada M., Brooke J., Tofiloski M., Voll K., Stede M.* (2011) Lexicon-Based Methods for Sentiment Analysis, Computational Linguistics, Vol. 37, no. 2, pp. 267–307.
27. *Thelen M., Riloff E.* (2002) A Bootstrapping Method for Learning Semantic Lexicons using Extraction Pattern Contexts, Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002), pp. 214–221.
28. *Thelwall M., Buckley K., Paltoglou G., Cai D., Kappas A.* (2010) A. Sentiment strength detection in short informal text, Journal of the American Society for Information Science and Technology, Vol. 61, no. 12, pp. 2544–2558.
29. *Turney P.* (2002) Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, Proceedings of 40th Meeting of the Association for Computational Linguistics, Philadelphia, PA, pp. 417–424.