

Generic High-Throughput Methods for Multilingual Sentiment Detection

Stefan Gindl, Arno Scharl
Department of New Media Technology
MODUL University Vienna
Vienna, Austria
{stefan.gindl, arno.scharl}@modul.ac.at

Albert Weichselbraun
Department of Information Systems and Operations
Vienna University of Economics and Business
Vienna, Austria
albert.weichselbraun@wu.ac.at

Abstract—Digital ecosystems typically involve a large number of participants from different sectors who generate rapidly growing archives of unstructured text. Measuring the frequency of certain terms to determine the popularity of a topic is comparably straightforward. Detecting sentiment expressed in user-generated electronic content is more challenging, especially in the case of digital ecosystems comprising heterogeneous sets of multilingual documents. This paper describes the use of language-specific grammar patterns and multilingual tagged dictionaries to detect sentiment in German and English document repositories. Digital ecosystems may contain millions of frequently updated documents, requiring sentiment detection methods that maximize throughput. The ideal combination of high-throughput techniques and more accurate (but slower) approaches depends on the specific requirements of an application. To accommodate a wide range of possible applications, this paper presents (i) an adaptive method, balancing accuracy and scalability of multilingual textual sources, (ii) a generic approach for generating language-specific grammar patterns and multilingual tagged dictionaries, and (iii) an extensive evaluation verifying the method's performance based on Amazon product reviews and user evaluations from Sentiment Quiz, a game-with-a-purpose that invites users of the Facebook social networking platform to assess the sentiment of individual sentences.

Keywords – *multilingual sentiment detection; evaluation; grammar patterns; semantic orientation; tagged dictionaries*

I. INTRODUCTION

This paper reports results of RAVEN (Relation Analysis and Visualization for Evolving Networks), a research project that aims to extend the sentiment detection component used in Web portals such as the *US Election 2008 Web Monitor* [16] and the *Media Watch on Climate Change* [8] to multilingual settings (both portals are accessible at www.ecoresearch.net).

This extension needs to address the following main challenges: (i) creating generic, language-independent interfaces for the sentiment detection tool chain, including tokenization, part-of-speech (POS) tagging, applying grammar patterns and querying tagged dictionaries; (ii) revising an existing language detection component in order to ensure its scalability and applicability to large document archives; (iii) using social networking platforms such as Facebook to gather and validate

entries for tagged dictionaries in multiple languages; (iii) evaluating the accuracy of the resulting dictionaries across languages. Processing Web documents from a digital ecosystem's heterogeneous sources (e.g., publication archives of the European Commission or articles of news media in multilingual countries such as Switzerland) requires accurate language detection to classify fragments of text of arbitrary lengths – this classification is important to select the correct tagged dictionary and set of grammar patterns for the subsequent sentiment analysis. Direct quotes and anglicisms further complicate this task and call for adaptive approaches flexible enough to detect embedded text fragments in a language different from the main classification.

II. RELATED WORK

Early work on sentiment detection started with Wiebe's [21] attempt to discriminate positive and negative sentences and Hatzivassiloglou's and McKeown's work [6] on the inheritance of sentiment across adjectives with syntactical relations. In a similar approach, Turney and Littman [20] transfer sentiment values from known paradigmatic terms to unknown terms by using the association measurements *Pointwise Mutual Information* (PMI) and *Latent Semantic Analysis* (LSA). The latter outperformed the former in the chosen test bed. Reviews are a popular source of validation data for sentiment detection algorithms, since they are already labeled by the review author (e.g. star ratings) and are easily available in large quantities. Turney classified reviews, after having identified positive and negative phrases, by measuring the association strength of arbitrary terms with the two terms “excellent” and “poor” [19]. Pang and Lee apply *Support Vector Machines* (SVMs) to guess the number of stars in movie reviews [13]. Popescu and Etzioni also use reviews in [14], but they extract product features judged in the reviews and determine sentiment expressed towards those features, i.e. if the review author has a positive or negative appraisal for that feature. Syntactical features often play a role in sentiment detection as well. Wilson, Wiebe, and Hoffmann explore the helpfulness of certain features in [22] by training different machine learning algorithms on those features. Nicholls and Song examine the impact of four different POS groups (nouns, verbs, adverbs, and adjectives) by training a *Maximum Entropy Classifier* [12].

Many approaches to multilingual sentiment detection rely on machine learning methods. The following summary describes and contrasts previous work in this evolving field.

The early work of Shanahan, Grefenstette, Qu, and Evans [17] gives an initial overview of possible methods to accomplish this task. They present three machine learning approaches for multilingual sentiment detection. The first approach trains a Support Vector Machine model in a certain language. In order to apply that model to other languages, the term vectors with the associated weights are transformed by a machine translation procedure. In the second approach, the machine translation transfers an initial text corpus into several other languages, for which separated classifier models are trained. The third approach dispenses the usage of a machine translation component and uses original text corpora for each language. Similar to the second approach, each language gets its own classifier.

Kim and Hovy [11] present another machine learning approach. A sentiment lexicon for the English language in conjunction with a Maximum Entropy classifier identifies expressed opinions. Two models for applying this method to German e-mail archives are being proposed: (i) translate the messages into English and then apply the classifier; (ii) translate the term of the sentiment lexicon into German and use this newly generated lexicon for sentiment detection.

The work of Boiy and Moens [2] provides a comparison of the efficiency of a Support Vector Machine, Multinomial Naïve Bayes and a Maximum Entropy classifier for English, French and Dutch. Linguistic characteristics such as unigrams or negations serve as features for the training of the classifiers. Balahur and Montoyo [1] use a Support Vector Machine classifier and apply it to English and Spanish customer reviews, using automatically extracted features. Pre-defined context polarity shifters such as negations and modifiers enrich the feature attributes.

Hiroshi et al. [7] refrain from using a trained classifier. They detect sentiment in Japanese texts by translating them into English. A polarity lexicon represents the basis for the extraction of sentiment units, which are multi-term phrases such as “high price”. An aggregation of these sentiment units gives a predication of the overall text sentiment.

Annotated multilingual corpora are important to evaluate precision and recall of the various sentiment detection methods. Evans et al. [3] create such corpora by annotating English, Chinese and Japanese texts on a sentence level.

III. METHOD

A language detection component first identifies all languages used in the input document and forwards monolingual text units and their respective ISO 639-1 language code to the sentiment detection component.

Based on the language code, the sentiment analysis component activates the correct tagged dictionary and set of grammar patterns, computes sentiment values per sentence and per document, and selects a subset of all tagged sentences for verification. These sentences and their sentiment values are forwarded to Sentiment Quiz, which asks ten users to assess the sentiment of a sentence (see Figure 1). Finally, the system compares the

user input with the sentiment values from the sentiment detection component and adjusts the component's tagged dictionary and grammar patterns accordingly.

A. Language Detection

Several standard techniques to detect natural languages are available, usually based on trigrams and common short words [9]. Trigrams compare a document's frequency of three-letter sequences with a language's typical distribution of these same three-letter sequences. Similarly, common short words such as determiners, conjunctions and prepositions help discern a language. As both methods produce comparable results for chunks of text larger than ten words [5], the authors used the computationally lighter short word technique in previous projects.

The short words approach provides maximum throughput, but tends to be too coarse for processing dynamic Web documents [23] and multilingual fragments of text. Adaptively combining more fine-grained methods addresses this problem and allows tagging units of text of arbitrary length: (i) trigrams, comparing the frequencies of three-letter sequences with the distributions of the same three-letter sequences in a particular language, or (ii) the Naïve Bayes algorithm, which evaluates the used vocabulary to determine the text's language.



Figure 1. Sentiment Quiz: Example of a text snippet presented to a player.

Table I reflects the trade-off between accuracy and throughput by contrasting computational effort (CPU Time) with the achieved precision, recall, F-Score and accuracy of the algorithm for five selected languages (English, French, German, Italian, and Spanish). Applying the short words approach yields a speed gain of 275 % compared to the current Bayes implementation and 1650 % compared to the use of trigrams.

TABLE I. EVALUATION OF THE SHORT WORDS (SW), TRIGRAM (T), AND BAYES (B) LANGUAGE DETECTION COMPONENTS ON APPROXIMATELY 6.8 MILLION SENTENCES FROM THE MULTILINGUAL REUTERS CORPUS.

Language (Size)	Method	CPU Time (min)	Performance (%)			
			Precision	Recall	F-Score	Accuracy
English (427723)	SW	24	99.14	94.74	96.89	99.62
	T	400	80.14	100.00	88.98	98.44
	B	67	97.03	99.78	98.38	99.79
French (3534116)	SW	201	99.96	97.76	98.85	98.82
	T	3304	99.9	99.25	99.57	99.56
	B	553	99.82	99.98	99.90	99.89
German (1170228)	SW	66	99.43	99.79	99.61	99.87
	T	1094	100.00	99.66	99.83	99.94
	B	183	99.94	99.95	99.94	99.98
Italian (540826)	SW	31	84.24	99.95	91.43	98.51
	T	506	100.00	99.55	99.78	99.96
	B	85	100.00	99.99	99.99	99.99
Spanish (1137623)	SW	65	99.58	98.88	99.23	99.74
	T	1064	100.00	93.12	96.44	98.85
	B	178	100.00	98.35	99.17	99.72

In its current and extended implementation, the system processes text in the following seven languages: English, French, German, Italian, Spanish, Portuguese, and Russian.

B. Sentiment Detection

The ever increasing size of digital ecosystems and the limits of human cognition are main motivations behind developing methods to automatically determine the sentiment expressed in sentences and documents. Such methods are typically based on the notion that there is a conceptual connection between words and their adjacent text [4]. The semantic orientation towards a target term within a sentence is calculated by measuring the distance d between the target term t and a pre-defined list of sentiment words s known to have positive or negative connotations [15].

For the *US Election 2008 Web Monitor* [16], for example, this list was taken from the tagged dictionary of the General Inquirer containing 4 400 positive and negative sentiment words [18]. Calculated for each sentence separately, the sentiment represents the sum of the tagged words' sentiment charges (SC_s) divided by their distance to the target term. The exponent λ determines the influence of the distance on calculating the semantic orientation:

$$S = \sum \frac{SC_s}{d_{ts}^\lambda} \quad (1)$$

Multilingual applications require a separate tagged dictionary for each language. The size and accuracy of these dictionaries have a strong impact on the validity of this approach. Once a set of terms has been tagged and compiled into a dictionary, reverse lemmatization can be used to extend the dictionary by considering plurals, gerunds, past tense suffixes, and other syntactical variations (e.g. manipulate \rightarrow manipulates, manipulating, manipulated).

C. Sentiment Quiz

The annotated corpus required to evaluate the method was created by means of a Facebook application called "Sentiment Quiz", an application in the tradition of games-with-a-purpose which confronts players with different sentences containing a

target term (target terms in this case were presidential candidates of the US Election 2008). The players assess the sentiment expressed towards a target term in that sentence. Every sentence is rated by a certain number of players, and the average of all ratings determines the overall sentiment charge of the sentence.

D. Grammar Patterns

To improve the results of the simple, scalable method presented in the previous section, this paper suggests an approach with a more detailed and fine-grained handling of syntax. Considering the syntax of a sentence helps increase the accuracy of sentiment detection and to compute sentiment not only for the whole sentence, but for specific entities referenced in the sentence as well (a feature that is not yet addressed in this paper, but will be explored as part of future research efforts).

In natural languages, the impact of a term can be modified by other terms. An example is the term "very", which causes an amplification or intensification of the term it refers to. Related research anticipates the impact of such terms on sentiment terms [10]. Given this evidence, the authors decided to consider three types of modifying terms:

- **Amplifiers:** these terms have an *intensifying* effect on the connected term (e.g., "very", "extremely" or "greatly");
- **Diminishers:** terms with a *weakening* impact on sentiment terms (e.g., "barely", "hardly" or "quite");
- **Negations:** this type of modifier *inverts* the sentiment value of the connected term (e.g., "not", "no" or "never").

In order to deal with these modifiers properly we had to regard their sphere of influence – i.e., how the distance to a target term influences their effect. In some cases, the connected term directly follows the modifier (e.g., "this result is not *good* at all", with *good* as the sentiment term); in other cases modifiers are spatially divided from the connected term (e.g., "he did not achieve such a *good* result"). In this case, there are three terms between the modifier and its connected term. More accurately, the noun phrase "he" interacts with the noun phrase "such a good result" via the verb phrase "did not achieve". The verb phrase itself (with the term "not" as an essential part of it) influences the noun phrase that follows.

It is not ideal to define a constant value for the sphere of a modifier, such as a fixed three-term window. This method would include actually unmodified sentiment terms when the scope has been set too high, or disregard modified ones when set too low. To solve this problem, the authors decided to automatically extract "grammar patterns" typical for a certain modifier. A grammar pattern is a frequently occurring part-of-speech sequence in the vicinity of a modifier. For example, a pattern for the term "not" could be "rb vb in dt nn" (rb = "adverb", vb = "verb", in = "preposition", dt = "determiner" and nn = "noun"), representing sentence snippets such as "do not participate in the Olympics" or "did not win over a majority" (as extracted from an extensive English media corpus).

A maximum pattern length and a list of stop tags for eliminating implausible patterns help refine these extracted patterns. The stop tags include tags for punctuation or conjunctions, which introduce another part of a sentence or combine two

parts, rather than end a grammatical structure. Once extracted, Finite State Machines decide whether a sentiment term is still in the scope of a modifier. Thereby one can easily iterate decision trees representing a term's typical grammar patterns.

A great advantage of the grammar pattern approach is its applicability to multiple languages. The only requirement is a pre-defined set of modifier terms typical for a given language; the system extracts the related patterns from a text corpus in a fully automated way and makes them processable with Finite State Machines. Promising preliminary results for the English language suggest the method's applicability to other languages. Other languages, however, might pose additional challenges – in German, for example, modifiers underlie inflections depending on their case in a sentence. The system therefore has to consider morphological differences of lexically identical terms.

Hand-crafted collections of modifier terms (amplifiers, diminishers, negations) served as seed list. In the case of German, inflected versions of the terms had to be included as well (these terms are direct equivalents of the English seed terms, as far as a direct translation was possible):

- **English:** (i) Amplifiers: *many, more, most, much, very*; (ii) Diminishers: *few, hardly*; (iii) Negations: *absence, lack, never, no, none, not, without*.
- **German:** (i) Amplifiers: *enorm, mehr, mehrere, mehrere, mehreren, mehrerer, sehr, viel, viele, viele, vielen, vieler, vieles, total*; (ii) Diminishers: *kaum, manch, manche, manchem, manchen, mancher, selten, wenig, wenige, wenigen, weniger*; (iii) Negations: *kein, keine, keinem, keinen, keiner, keines, nicht, nie, niemals, ohne*.

E. Domain Ontologies

Advanced sentiment detection systems incorporate domain ontologies to disambiguate terms that can be used in more than one context. Ontologies also allow the assessment of results by matching concepts across languages, and lead to a number of interesting follow-up research questions; e.g., if grammar patterns differ across languages, and whether these differences depend on the type of concept, or on the specific relation between two concepts.

For a proof-of-concept, the environmental domain has been chosen for the following reasons: (i) Due to its complex and dynamic nature, processing environmental knowledge represents a significant research challenge. (ii) There is increasing demand for environmental knowledge among various stakeholders. (iii) Typically, members of environmental communities are intrinsically motivated and can be expected to actively participate in social applications. (iv) The *Media Watch on Climate Change* (www.ecoresearch.net/climate) not only provides a rich archive of domain-specific content, but also a technology platform to analyze and compare results. (v) The availability of GEMET, the *GEneral Multilingual Environmental Thesaurus* (www.eionet.europa.eu/gemet). GEMET has been developed as an indexing, retrieval and control tool for the European Topic Centre on Catalogue of Data Sources and the European Environment Agency. GEMET was conceived as a “general” thesaurus, aimed to define a core of general terminology for the environment. Additional evidence sources to be considered are the *Global Change Master Direc-*

tory (gcmd.nasa.gov) of the NASA Goddard Space Flight Center, as well as the *Semantic Web for Earth and Environmental Terminology* (sweet.jpl.nasa.gov/ontology) of the NASA Jet Propulsion Laboratory.

IV. EVALUATION

The following section presents a formal evaluation based on two different experiments. The first evaluation uses multilingual Amazon product reviews (www.amazon.com) to measure the sentiment detection performance of the three methods (lexical forms, simple negation detection, and advanced grammar patterns) introduced in Section III.B. The second evaluation computes the correlation between the automated techniques and user feedback gathered from the Sentiment Quiz.

A. Amazon Product Reviews

To assess the relative performance of the grammar pattern-based approach, a Web crawler retrieved 2 500 Amazon product reviews. Such reviews have the advantage that their authors already provide information on whether the review is positive or negative when uploading the documents. Including reviews of several product categories (“Electronics”, “Office Products”, “Outdoor Living”, “Photo”, and “Wireless”) helped create a varied, domain-independent sample. As the primary baseline algorithm, a simple method based on lexical forms was chosen, which does not take into consideration any kind of grammar. To compute the sentiment, the system aggregates the sentiment values of positive and negative terms provided by the tagged dictionary (see Section III.B) identified in the text. The second method follows a more complex approach based on lexical forms that invert the sentiment values of the term *directly* preceded by a negation. This baseline helps evaluate the usage of a scalable method to deal with negated expressions within a window of a given character length.

The gathered results are listed in Table II, which compares the baseline algorithm using lexical forms with simple negation detection (inverting the sentiment value of terms that follow a set of negation triggers), and a more complex parsing of grammar patterns. Separating the results for both polarity classes (positive versus negative) documents the strengths and weaknesses of the underlying methods.

TABLE II. SIMPLE NEGATION DETECTION VERSUS MORE ADVANCED GRAMMAR PATTERNS FOR THE ENGLISH AND GERMAN LANGUAGE.

English	Performance (%)			
	Positive Polarity		Negative Polarity	
	Recall	Precision	Recall	Precision
<i>Lexical Forms</i>	78.7	62.7	51.3	72.0
<i>Simple Negation Detection</i>	81.0	64.7	53.9	75.5
<i>Advanced Grammar Patterns</i>	82.0	66.1	56.1	76.9
German				
<i>Lexical Forms</i>	81.8	59.9	23.0	82.7
<i>Simple Negation Detection</i>	83.8	62.2	25.4	87.4
<i>Advanced Grammar Patterns</i>	84.7	61.6	26.6	85.6

The transition from simple lexical forms to the involvement of the negation detection increases both recall and precision. The further improvement through grammar patterns is minor, but represents an important step considering the sheer size of the repository (an improvement in recall of just three percent means that 10 000 additional sentences could be tagged automatically). The sparsity of modifier terms helps explain the modest performance gains. Product reviews tend to be short, and many of them do not contain modifier terms at all.

The handling of modifier terms in the German language turned out to be more difficult. Although using grammar patterns showed a positive impact, the impact is less obvious as compared to English (see Table II). The difference between lexical forms simple negation detection is quite obvious, both in terms of recall and precision. The transition from the simple negation detection approach and the advanced grammar patterns approach, on the other hand, does not show a clear difference. The improvement in recall accompanies a worsening in precision. A further investigation of the modifier terms used will cast a light on the question of whether the assumed usage purpose is correct. A direct translation of modifier terms from one language into another is intuitively obvious, but might be linguistically incorrect.

B. Sentiment Quiz

The second evaluation contrasts the performance of (i) different λ values for lexical forms (see Equation 1), (ii) the simple negation detection, and (iii) advanced grammar patterns using a data set assembled from 1 008 sentences, containing target terms for the US Election 2008. These target terms are the names of political candidates such as Barack Obama and Hillary Clinton. All sentences were manually classified by Sentiment Quiz users, resulting in an average numerical value of the sentiment expressed within a specific sentence.

The evaluation using recall and precision shows the superiority of the simple negation detection and advanced grammar pattern approaches. The authors also measured the efficiency with Pearson's correlation coefficient. This evaluation shows a constant decrease of correlation when increasing the λ exponent (see Table III). Simple negation detection and advanced grammar patterns yielded correlations of 0.48 and 0.49, respectively. In other words, the system achieves the best results when a mathematical distance measure is not invoked (with λ being 0 the distance is disregarded reducing the sentiment computation to a simple word counting).

TABLE III. COMPARISON OF LEXICAL FORMS, SIMPLE NEGATION DETECTION (=SND) AND ADVANCED GRAMMAR PATTERNS (AGP) ON THE SENTIMENT QUIZ DATA.

Method	$\lambda=0$	$\lambda=1$	$\lambda=2$	SND	AGP
Positive Polarity					
<i>Recall (%)</i>	65.2	39.3	23.1	75.7	75.3
<i>Precision (%)</i>	44.2	45.5	46.7	43.6	43.5
Negative Polarity					
<i>Recall (%)</i>	51.5	21.4	10.2	61.4	60.2
<i>Precision (%)</i>	57.2	50.4	43.6	55.9	55.4

C. Discussion

Evaluations based on Amazon product reviews and the assessments of Sentiment Quiz users show significant improvements when comparing the method based on lexical forms with the negation detection. The experiments indicate that advanced grammar patterns might yield better results than the simple negation detection, but the observed effects are not significant.

The differences between the evaluations based on Amazon product reviews and those leveraging the Sentiment Quiz demonstrate that benchmarks obtained from direct user ratings differ from data derived from indirect feedback, which uses, for instance, product ratings to estimate the text's sentiment. This does not come as a surprise, since the sentiment of the text snippet might differ from the author's review classification. This underlines the importance of human evaluators to determine the accuracy of such methods. Applications following the tradition of *games with a purpose*, like the approach presented in this paper, provide an effective chance to gather such evaluations in a cost-effective way.

V. OUTLOOK AND CONCLUSIONS

This paper presents a generic approach for building and evaluating multilingual tagged dictionaries for detecting sentiment in digital ecosystems that are characterized by large archives of unstructured text. It applies these dictionaries using three types of sentiment detection: (i) standard algorithm based on co-occurring lexical forms; (ii) negation detection that inverts the sentiment when encountering a negation trigger; and (iii) grammar patterns that process additional modifiers such as amplifiers and diminishers.

Comparing evaluation results from direct user feedback with data derived from product reviews show considerable differences between the measured precision and recall values. This finding calls for a more detailed evaluation of these deviations, especially in multilingual settings. Future research will investigate the full potential and possible application areas of the proposed grammar pattern approach. To replace the lexical approach (see Section III.B), we aim to generate patterns for the determination of sentiment in a sentence towards a target term (e.g., a person or an organization). Involving relational terms will allow calculating sentiment for different entities that are referenced within a sentence or document. By considering such terms, the algorithm will support simple structures such as "<x> is a good president" and "I support/don't support the president", or more complex structures like "I think/don't think that <x> is a good president". It will also allow distinguishing sentences that express sentiment of a general nature from sentences referencing multiple entities, where relational patterns should be invoked.

ACKNOWLEDGMENT

The presented work is part of the RAVEN research project (Relation Analysis and Visualization for Evolving Networks; www.modul.ac.at/nmt/raven) funded by the Austrian Ministry of Transport, Innovation & Technology (BMVIT) and the Austrian Research Promotion Agency (FFG) within the strategic objective FIT-IT (www.fit-it.at). The authors wish to thank Kate Farrar for proof-reading the manuscript.

REFERENCES

- [1] A. Balahur and A. Montoyo, Multilingual feature-driven opinion extraction and summarization from customer reviews. In: Kapetanios, E., Sugumaran, V., Spiliopoulou, M. (Eds.), NLDB. Vol. 5039 of Lecture Notes in Computer Science, Springer, 2008, pp. 345—346.
- [2] E. Bojy and M.-F. Moens, “A machine learning approach to sentiment analysis in multilingual web texts”. *Information Retrieval*, vol. 12(5), October 2008, pp. 526—558.
- [3] D. K. Evans, L.-W. Ku, Y. Seki, H.-H. Chen, N. Kando, “Opinion analysis across languages: An overview of and observations from the ntcir6 opinion analysis pilot task”. In: WILF '07: Proceedings of the 7th international workshop on Fuzzy Logic and Applications, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 456—463.
- [4] R. Giora, “Discourse coherence and theory of relevance: Stumbling blocks in search of a unified theory”. In: *Journal of Pragmatics*. Vol. 27, 1996, pp. 17—34.
- [5] G. Grefenstette, “Comparing two language identification schemes”. In: 3rd International Conference on Statistical Analysis of Textual Data, Rome, Italy, 1995, pp. 263—268.
- [6] V. Hatzivassiloglou and 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, Morristown, NJ, USA, 1997, pp. 174—181.
- [7] K. Hiroshi, N. Tetsuya, and W. Hideo, “Deeper sentiment analysis using machine translation technology”. In: COLING '04: Proceedings of the 20th international conference on Computational Linguistics, Association for Computational Linguistics, Morristown, NJ, USA, 2004, p. 494.
- [8] A. Hubmann-Haidvogel, A. Scharl, and A. Weichselbraun, “Multiple coordinated views for searching and navigating web content repositories”. *Information Sciences*, 179(12), 2009, pp. 1813—1821.
- [9] D. A. Hull and G. Grefenstette, “Querying across languages: a dictionary-based approach to multilingual information retrieval”. In: SIGIR '96: Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, New York, NY, USA, 1996, pp. 49—57.
- [10] A. Kennedy and D. Inkpen, “Sentiment classification of movie reviews using contextual valence shifters”. *Computational Intelligence*, 22:(2), 2006, pp. 110—125.
- [11] S. M. Kim and E. Hovy, “Identifying and analyzing judgment opinions”. In: Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Association for Computational Linguistics, Morristown, NJ, USA, 2006, pp. 200—207.
- [12] C. Nicholls and F. Song, “Improving sentiment analysis with part-of-speech weighting”. In Machine Learning and Cybernetics, 2009 International Conference on Machine Learning and Cybernetics, July 2009, pp. 1592—1597.
- [13] B. Pang and L. Lee, “Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales”. In EMNLP '02: Proceedings of the ACL-02 conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Morristown, NJ, USA, 2002, pp. 79—86.
- [14] A.-M. Popescu and O. Etzioni, “Extracting product features and opinions from reviews”. In HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Morristown, NJ, USA, 2005, pp. 339—346.
- [15] A. Scharl, I. Pollach, and C. Bauer, „Determining the semantic orientation of web-based corpora”. *Intelligent Data Engineering and Automated Learning*, 4th International Conference, IDEAL, Lecture Notes in Computer Science 2690, 2003, pp. 840—849.
- [16] A. Scharl and A. Weichselbraun, “An automated approach to investigating the online media coverage of us presidential elections”. *Journal of Information Technology & Politics* 5:(1), 2008, pp. 121—132.
- [17] J. G. Shanahan, G. Grefenstette, Y. Qu, and D. A. Evans, “Mining multilingual opinions through classification and translation”. In: AAAI Spring Symposium, 2004, pp. 146—149.
- [18] P. J. Stone, D. C. Dunphy, M. S. Smith, and D. M. Ogilvie, *The general inquirer: A computer approach to content analysis*, 1966.
- [19] P. D. Turney, “Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Morristown, NJ, USA, 2002, pp. 417—424.
- [20] P. D. Turney and M.L. Littman. “Unsupervised learning of semantic orientation from a hundred-billion-word corpus”. Technical report, National Research Council, Institute for Information Technology, 2002.
- [21] J. M. Wiebe, “Tracking point of view in narrative”, *Comput. Linguist.*, 20:(2), 1994, pp. 233—287.
- [22] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann, “Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis”, *Comput. Linguist.*, 35(3), 2009, pp. 399—433.
- [23] A. Xafopoulos, C. Kotropoulos, G. Alpanidis, and I. Pitas, “Language identification in web documents using discrete HMMs”. *Pattern Recognition* 37:(3), March 2004, pp. 583—594.