

ISSN 2708-2032
e-ISSN 2708-2040



**INTERNATIONAL
UNIVERSITY**

**INTERNATIONAL
JOURNAL OF INFORMATION
& COMMUNICATION TECHNOLOGIES**

**Volume 2, Issue 4
March 2021**

ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН
MINISTRY OF EDUCATION AND SCIENCE OF THE REPUBLIC OF KAZAKHSTAN



**INTERNATIONAL JOURNAL OF
INFORMATION AND COMMUNICATION
TECHNOLOGIES**

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ
ИНФОРМАЦИОННЫХ И
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

**ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ
КОММУНИКАЦИЯЛЫҚ
ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ**

Том 2, Выпуск 8
December 2021

Главный редактор – Ректор АО МУИТ,
к.ф.-м.н.
Хикметов А.К.

Заместитель главного редактора –
Проректор по НИМД, PhD, ассоц. профессор
Дайнеко Е.А.

Отв. секретарь – Директор департамента по науке, к.т.н., ассоц. профессор
Ипалакова М.Т.

ЧЛЕНЫ РЕДКОЛЛЕГИИ:

Отельбаев М.О., д.ф.-м.н., профессор, АО «МУИТ», Рысбайулы Б., д.ф.-м.н., профессор, АО «МУИТ», Синчев Б.К., д.т.н., профессор, АО «МУИТ», Дузбаев Н.Т., PhD, проректор по ЦИИ, АО «МУИТ», Сейлова Н.А., к.т.н., декан ФКТК, АО «МУИТ», Мухамедиева А.Г., к.э.н., декан ФЦТ, АО «МУИТ», Ыдырыс А., PhD, заведующий кафедрой «МКМ», АО «МУИТ», Саксенбаева Ж.С., к.т.н., заведующий кафедрой «ИС», АО «МУИТ», Шильдибеков Е.Ж., PhD, заведующий кафедрой «ЭиБ», АО «МУИТ», Аманжолова С.Т., к.т.н., заведующий кафедрой «КБ», АО «МУИТ», Ниязгулова А.А., к.ф.н., заведующий кафедрой «МиИК», АО «МУИТ», Айтмагамбетов А.З., к.т.н., профессор, АО «МУИТ», Джоламанова Б.Д., ассоциированный профессор, АО «МУИТ», Разак А., PhD, профессор, АО «МУИТ», Алмисреб А.А., PhD, ассоциированный профессор, АО «МУИТ», Мохамед А.Н., PhD, ассоциированный профессор, АО «МУИТ», Prof. Young Im Cho, PhD, Gachon University (South Korea), Prof. Michele Pagano, PhD, University of Pisa (Italy), Tadeusz Wallas, PhD, D.Litt., Adam Mickiewicz University in Poznań (Poland), Тихвинский В.О., д.э.н., профессор, МГУСИ (Россия), Масалович А., к.ф.-м.н., Президент Консорциума Инфорус (Россия), Lucio Tommaso De Paolis, Research Director of the Augmented and Virtual Laboratory (AVR Lab), Department of Engineering for Innovation, University of Salento (Italy), Prof. Liz Bacon, Deputy Principal and Deputy Vice-Chancellor, Abertay University (Great Britain).

Издание зарегистрировано Министерством информации и общественного развития Республики Казахстан. Свидетельство о постановке на учет No KZ82VPY00020475 от 20.02.2020 г.

Журнал зарегистрирован в Международном центре по регистрации сериальных изданий ISSN (ЮНЕСКО, г. Париж, Франция)

Выходит 4 раза в год.

УЧРЕДИТЕЛЬ:

АО «Международный университет информационных технологий»

ISSN2708-2032 (print)
ISSN2708-2040 (online)

СОДЕРЖАНИЕ

ИНФОКОММУНИКАЦИОННЫЕ СЕТИ И КИБЕРБЕЗОПАСНОСТЬ

- Кожухметова Б.А., Губский Д.С., Дайнеко Е.А., Ипалакова М.Т.*
Численно-математическое моделирование современных устройств СВЧ и КВЧ диапазонов на примере микрополоскового резонатора.....6
- Мубаракова С.Р., Аманжолова С.Т., Ускенбаева Р.К.*
Актуальность кибербезопасности в современном мире.....12
- Разак А., Әділ А.Ж., Аманжолова С.Т.*
Новый инструмент для обнаружения взлома Wi-Fi на основе технологии блокчейн.....18

ЦИФРОВЫЕ ТЕХНОЛОГИИ В ЭКОНОМИКЕ И МЕНЕДЖМЕНТЕ

- Аукен В.М.*
Анализ взаимодействия государственных доходов и аудита.....38
- Бердыкулова Г.М.*
Методология преподавания экономических дисциплин в цифровую эру.....42

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

- Элле В.Ж., Мелисова Л.Т., Куандыков А.А., Куатбаева А.А., Аманбайқызы З.*
Свойства реальных бизнес-процессов с точки зрения проектирования.....49
- Кошимбай А.Б., Молдагулова А.Н.*
Исследование метода анализа и обработки данных социальных сетей с целью определения тональности.....55
- Базарбеков И.М., Шарипов Б.Ж.*
разработка бизнес-процесса для получения онлайн услуг в организации образования62
- Жунусов Д.О., Алиаскаров С.Ж.*
метод классификации текстов на основе алгоритмов машинного обучения.....69

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

- Синчев Б.*
О полиномиальной разрешимости класса np-complete.....75

CONTENTS

INFORMATION AND COMMUNICATION NETWORKS, CYBERSECURITY

- Kozhakhmetova B.A., Gubsky D.S., Daineko Y.A., Ipalakova M.T.*
Numerical and mathematical modeling of modern devices of UHF and EHF bands on the example of a microstrip resonator.....6
- Mubarakova S.R., Amanzholova S.T., Uskenbayeva R.K.*
Relevance of cybersecurity in the modern world.....12
- Razaque A., Adil A. Zh., Amanzholova S.T., Valiyev B.B.*
Blockchain technology-featured novel air-cracking tool for wi-fi hacking detection.....18

DIGITAL TECHNOLOGIES IN ECONOMICS AND MANAGEMENT

- Auken V.M.*
Interaction analysis of government revenue and audit.....38
- Berdykulova G.M.*
Methodology of teaching the economic disciplines in digital era.....42

INTELLIGENT SYSTEMS

- Elle V., Melissova L., Kuandykov A.A., Kuatbayeva A.A., Amanbaikyzy Z.*
Properties of real business processes from a design point of view.....49
- Koshimbay A.B., Moldagulova A.N.*
Research method of analyzing and processing social network data in order to determine the tonality.....55
- Bazarbekov I.M., Sharipov B.Zh.*
development of a business process for obtaining online services in the organization of education62
- Zhunissova D.O., Aliaskarov S.Zh.*
method for text classification based on machine learning algorithms69

MATHEMATICAL AND COMPUTER MODELING

- Sinchev B.*
On polynomial decision of class NP-complete.....75

МАЗМҰНЫ

АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ЖЕЛІЛЕР, КИБЕРҚАУІПСІЗДІК

Кожяхметова Б.А., Губский Д.С., Дайнеко Е.А., Ипалакова М.Т.

Микрожолқты резонатор мысалында АЖЖ және ЕЖЖ диапазондарының заманауи құрылғыларын сандық-математикалық үлгілеу.....6

Мубаракова С.Р., Аманжолова С.Т., Ускенбаева Р.К.

Қазіргі әлемдегі кибер қауіпсіздіктің өзектілігі.....12

Разак А., Әділ А.Ж., Аманжолова С.Т.

Блокчейн технологиясына негізделген Wi-Fi хакерін анықтаудың жаңа құралы.....18

ЭКОНОМИКАДАҒЫ ЖӘНЕ МЕНЕДЖМЕНТТЕГІ ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАР

Аукен В.М.

Мемлекеттік кірістер және аудиттің өзара әсерлері.....38

Бердіқұлова Ғ.М.

Цифрлық дәуірде экономиканы оқыту әдістемесі.....42

ИНТЕЛЛЕКТУАЛДЫ ЖҮЙЕЛЕР

Элле В.Ж., Мелисова Л.Т., Қуандықов А.А., Қуатбаева А.А., Аманбайқызы З.

Жобалау тұрғысынан нақты бизнес-процестердің қасиеттері.....49

Көшімбай А.Б., Молдагулова А.Н.

Тоналдылықты анықтау мақсатында әлеуметтік желілердің деректерін талдау және өңдеу әдісін зерттеу.....55

Базарбеков И.М., Шарипов Б.Ж.

Білім беру ұйымында онлайн қызмет көрсету үшін бизнес-процесін дамыту.....62

Жунусов Д.О., Алиаскаров С.Ж.

Машинналық оқыту алгоритмдері негізінде мәтіндер классификациясының әдісі.....69

МАТЕМАТИКАЛЫҚ ЖӘНЕ КОМПЬЮТЕРЛІК МОДЕЛЬДЕУ

Синчев Б.

NP-complete сыныптың полиномиялық шешімі туралы.....75

Koshimbay A.B., Moldagulova A.N.

International Information Technology University, Almaty, Kazakhstan

RESEARCH METHOD OF ANALYZING AND PROCESSING SOCIAL NETWORK DATA IN ORDER TO DETERMINE THE TONALITY

Abstract. A wide spread of social online services and the advancement of Big Data technologies poses a challenge to utilize data from social media in numerous circles. Nowadays, the «social listening» and substance examination advances pick up ubiquity in Data Science. The sentiment analysis of the text is one of the especially important tasks in the field of natural language processing. It is used in different spheres. This article discusses the main methods of identifying emotions in text data and analyzes the current achievements in the field of computer analysis of emotions in text data. At the moment, there are many unresolved problems in the field of automatic text analysis to determine the emotional coloring of the vocabulary in social media texts.

Keywords: sentiment-analysis, tonality of text, text vectorization, machine learning, support vector machine, tone dictionary

Introduction

The majority of people nowadays can't imagine their lives without social media. According to the research results, social media platforms such as Facebook, Instagram, Twitter, and VKontakte have become an indispensable part of everyone's life during the last decade. A significant number of businesses are establishing themselves on social media platforms. Consumer relationships can be maintained, requests and feedback can be responded to rapidly, and marketing campaigns can be run through social media. As a consequence, social media have a significant impact on the socialization of modern individuals. The rise of social media has signified not just the transfer to new data sources, but also the possibility for each user to be there. Now, if a user comes across something interesting, he or she can put it in his or her profile [1].

Quite a substantial shift in the way information is distributed offers enormous possibilities for individuals and society as a whole. In general, the online world transports a massive amount of data, including user-to-user communication. It should be emphasized that major media firms use public evaluation based on the number of likes, postings, and comments to explore public opinion. In social networks, there is a massive amount of data. For instance, Facebook, which had 2.23 billion active monthly users at the end of summer 2018, posts over half a million comments per minute and over 100,000 photos per day. Every day, Twitter's audience sends out more than half a billion tweets, amounting to roughly 200 billion messages every year. This is a massive amount of unstructured data that has never been seen before. Likes and dislikes ratings may potentially become public tools for researching user opinions in networks [1]. Hence, there is a need for a tool that can automatically extract relevant information from publications, distinguish reviews from advertisements, and determine, among other things, the users' attitudes toward the topic of interest. By processing such data, mass media companies can improve the quality of their content, identify their target audience, and assess the attitude to materials in a comprehensive perspective [2].

Social media posts come in a wide range of content formats. Video (YouTube), images (Instagram, Pinterest), and text (other social media platforms are the most common types of material (Twitter, Facebook). This report will concentrate primarily on textual information and the methods for processing it, specifically, based on the semantic analysis of texts [2].

The purpose of the article is to give an overall overview of various methods for monitoring and evaluating social media opinion in order to perform an in-depth analysis of unstructured data and extract negativity and threat from text arrays. The ability to recognize emotionally colored vocabulary and analyze the user's appraisal of the product can be determined based on the tone of textual messages [2].

It was previously impossible to find an automated solution to this issue. Presently, computer linguistics' functionalities enable the extraction of information from texts using computer technology and accurate models. Assessment of the emotional coloring of a text is one of the objectives of this research (text tone analysis, content analysis, sentiment analysis). As a result of the research, there has been defined a variety of open text sources that present data about people's perspectives on a variety of topics. There is a need for greater research in the field of text tone analysis in order to obtain more complicated statistical data [3].

Sentiment analysis of texts is defined as a set of approaches for automatically selecting an emotionally colored language in writings and assessing the writers' emotional reactions (opinions) to the items mentioned in the text. Sentiment analysis is utilized in marketing research, audience loyalty tracking for various themes and brands, and other applications [3].

The study of emotions in textual data seems to be applicable to a variety of fields, including evaluation of the properties of products and services based on online customer feedback, analysis of unfavorable emotions in messages, prediction of stock market trends, the political environment based on news streams. In order to analyze such a large amount of data, many methods for automatically determining the tone of the text have been introduced in recent years, which will be reviewed in this article.

Materials and research methods

Text tonality determination is a difficult task whose outcome is highly dependent on the context, study region, and amount of text data. There are a few fundamental techniques for determining a text's tone. The three most widely acknowledged ways to determine the tone of a sentence can be divided into several categories. (Figure 1)

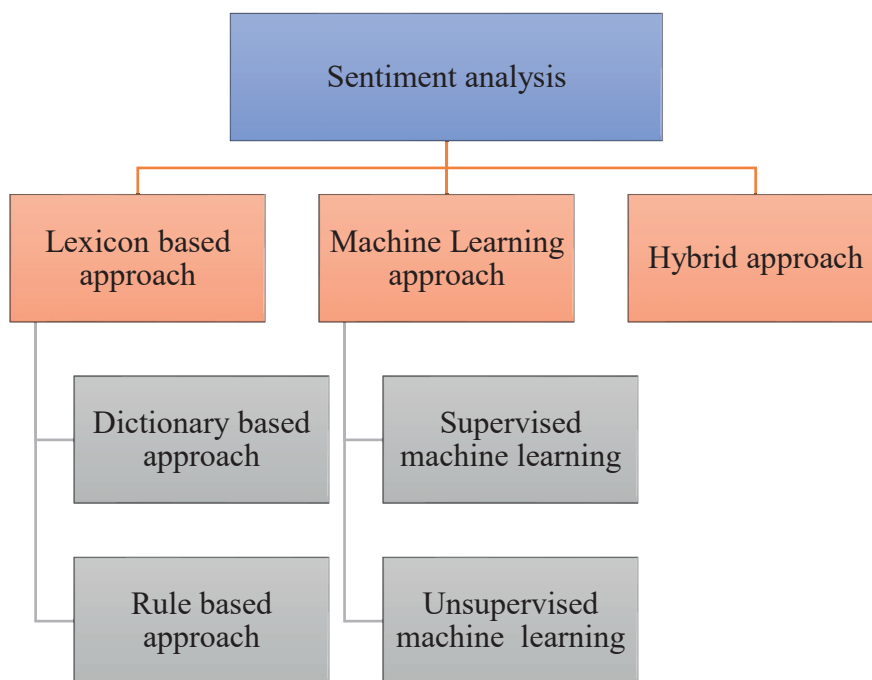


Figure 1– Classification of approaches to sentiment analysis

1. Methods based on rules and dictionaries

The conversational style of speech is common in text messages published on social media. The inclusion of slang, changed words, typos, and other elements that make it difficult to interpret the text are the main challenges in analyzing users' emotions. The text quality is strongly influenced by the social network in question. Also, it should be borne in mind that subjective and objective texts should be differentiated. A subjective text, as opposed to the objective one, contains information on an event without the author's personal perspective. It is subjective assessments that are of interest, as they allow of the extraction of a certain author's viewpoint. The majority of objective articles are informational in nature, hence they are removed from the bulk of the material under study [4].

Rule-based approaches typically specify a set of rules in a particular scripting language that denotes subjectivity, polarity of opinion. This approach entails the creation of numerous rules, based on "if-then" rules. For instance, if the particle "not" appears before a positive adjective, the construction is considered to be negative. This strategy also includes the use of tone dictionaries, in which the words examined belong to a specific category (positive, negative, neutral, and so on). The rules accept a variety of inputs, including standard NLP (NeuroLinguistic Programming) approaches like stemming, tokenization, voice tagging, and syntactic analysis. Two sets of polarized terms, such as negative words like «awful», «worse», «ugly», and positive words like "good," "best," and "beautiful," are a fundamental illustration of these techniques [4].

The text is given. The number of favorable words in the text is counted. There is also a count of bad terms.

When the number of positive words in a text exceeds the number of negative ones, the positive emotion is the polar opposite of the negative feeling. Otherwise, the result is neutral. This approach is extremely basic because it ignores the word combinations in the sequence. The more advanced processing analysis is feasible; however, such systems are extremely difficult to construct, because they require additional support for new terminology and language, despite the fact that new rules are more difficult to enforce. Furthermore, adding new rules may result in unfavorable outcomes because such systems necessitate a large investment in manual configuration and adherence to them [4].

The second linguistic approach is using tone dictionaries. The main concept of this method is to create tonal glossaries, which are collections of words and emotional states with a numerical positive or negative tonal score. When applying this method, linguists conduct the so-called rule-based tone analysis. The outcome of this study is a collection of rules (also known as a lexicon or sentiment lexicon analysis) that classify words as positive or negative, along with the corresponding intensity measure [5].

Furthermore, attribution of this or that text to a specific tone class is calculated on the basis of all collected weights of the text terms. The arithmetic mean of the weights is most commonly employed for this, with the sum of the weights or artificial neural networks being utilized in rare circumstances [5]. The text is regarded as positive if the number of positive words exceeds the number of negative ones, and vice versa. Clearly, this method is inefficient. It necessitates the creation of a big word dictionary, which must be updated on a regular basis. Furthermore, the fact that the number of positive words outnumbers the number of negative words is not a trustworthy standard by which the work might be evaluated properly [5].

2. Machine learning approach.

Supervised machine learning algorithms

A supervised machine learning approach necessitates a training set of texts that are marked up inside an emotional space and used to build a statistical or probabilistic classifier. A set of training samples is required to solve the sentiment analysis problem. A set of pre-labeled review texts is provided for supervised learning. A pair of feature vectors, which is a representation of a single text, and the tonality of the text constitute an individual instance of this set. The marked set of texts (the training sample) is evaluated, and a statistical pattern is developed for use in classifying new input vectors in supervised machine learning algorithms. There are several algorithm classifiers based on the supervised machine learning approach, such as linear classifier, SVM, decision tree classifier, Naive Bayesian classifier, etc. The Naive Bayesian classifier and the support vector machine approach are the most widely used methods in the field of tone analysis [6].

A Naive Bayesian classifier (NB) is a probabilistic classifier that relies on the Bayes theorem and assumes class independence. The Naive Bayes approach ensures that no relationship exists between the system's various parameters. It is particularly useful for huge data sets and, despite its simplicity, outperforms other more complicated algorithms in terms of consistency.

Table 1 represents the advantages and disadvantages of using Naïve Bayes classifier: [6]

Table 1 - Table of advantages and disadvantages of the Naive Bayes classifier

Benefits	Drawbacks
The Naive Bayesian classification method is simple and quick to implement.	One of the most significant drawbacks of the Naïve Bayesian classification is strong independence of features, as in actual life it is quite difficult to have a set of features that are totally independent of one another.
It will converge more quickly than such discriminative models as regression models.	Another issue with Naive Bayesian classification is that it has a "zero-frequency" feature, which implies that if a classifier has a category but is not observed in the training data set, then the Naive Bayesian model would give it as a zero probability and will be unusable for prediction.
It requires less training data.	
It's extremely scalable because the number of predictors and data points scales linearly.	
It has the ability to produce probabilistic predictions and can work with both continuous and discrete data.	

Support vector machines (SVMs) are linear classifiers. Regression predicts a continuous value, while linear classification predicts a label or a group. The method's basic idea is to create a hyperplane that isolates sample items as well as possible. SVM classifies input data by identifying a hyperplane that separates classes in n-dimensional spaces. One of the benefits of SVM is its versatility, as it can solve problems using a variety of kernel functions. SVM classifiers also have a high level of accuracy and can handle enormous data sets. Because SVM classifiers only use a fraction of training points, they require extremely little memory. However, they have long learning times, so in practice, they are not suitable for large datasets. Another disadvantage is that SVM classifiers do not work well with overlapping classes. [6]

Unsupervised machine learning

In contrast to the methods described above, the unsupervised learning method determines relations and patterns between objects without labeled data. Such methods include Gaussian mixture and k-nearest neighbor models.

K-means is an algorithm that finds k training examples within the shortest distance to the provided sample. The class of the object of interest will be the most common among k objects. To implement it, the algorithm needs a training sample of marked reviews. It was necessary to calculate the distance between the vector of this review and vectors from the training sample in order to establish the class of review from the test sample and determine the minimum distance between k items in the training sample (k is given by the expert or chosen according to efficiency estimates). The input vector's class is the one where more than half of the nearby k vectors are members. As an advantage, the k-mean algorithm has high accuracy, insensitivity to outliers, and no assumptions about data entry. Yet, it has drawbacks such as high temporal complexity and high spatial complexity. Once the sample is unbalanced, one class's sample size is quite huge while the sample size of the other classes is quite tiny. If a sample is entered, then the class with the largest sample size is the one with the most K adjacent values, which causes issues in classification [7].

3. Hybrid approach

The techniques that combine some of the methods outlined above are known as hybrid methods. A hybrid method employed in learning models by A.C. Koenig and E. Brill included the method based on tone dictionaries and the method of reference vectors. Using this strategy, the authors attained a learning accuracy of 72 % [8].

Results and discussion

According to the reviewed articles and works, scientists generally merge approaches to achieve the best outcomes. V.G. Vasiliev, S. Davydov, and M.V. Khudyakova in their works use a linguistic strategy complemented with machine learning approaches to adjust individual classification rules through training [9]. Numerous studies have demonstrated that integrating linguistics to overcome the tonality problem produces beneficial consequences. Rule-based algorithms provide more accurate results than machine learning approaches that apply the statistics and probability theory because their operation is intimately tied to the meaning of words. However, as previously indicated, the linguistic method has some significant limitations. While comparing different algorithms for tonality analysis, it must be highlighted that rule-based methods produce more accurate outcomes than machine learning approaches. It can be asserted that each approach is unique, and the combination of different approaches results in an increased accuracy of the training model. Table 2 shows a comparison of methodologies based on the primary criteria for selecting a tonality analysis method.

Table 2- Comparison of tonality analysis methods

	Accuracy	Automation	Training data	Easy to apply	Applicability in commercial systems
Rule based approach	High accuracy	automated	No data required	-	+
Dictionary based approach	Not unique	Automated within the same subject area	Data required	+	-
Supervised machine learning	Moderate accuracy	Automated	Data required	+/-	+
Unsupervised machine learning	Low accuracy	Automated	No data required	+	+

Tang and colleagues (Tang et al.) claim that: [10]

«To differentiate and evaluate comments on Internet reviews, most of the available methods rely on the natural language processing algorithms. However, while comments on Internet evaluations are less formal than those on news stories or magazine articles, they still demand poor accuracy. Many of the sentences in the books contain grammatical faults as well as unknown components not found in dictionaries.»

Using the sentence syntax method, the researcher was able to reach the best level of accuracy. All the aforementioned algorithms seem to be unable to function with word order, since saving the words necessitates the creation of a text array in the form of a matrix, with each row representing a vectorized word. However, only convolutional neural networks are capable of using it. Convolutional neural networks, on the other hand, are only implemented when the maximum classification accuracy is required. This approach is rarely applied in reality, but it improves the model's accuracy by 2% on average when compared to other classifiers [11].

M.V. Chernyshevich in his work analyzes the main existing types of classifying the tone of opinions and offers his opinion scale, which operates with both absolute and comparative evaluations. He conducted a study of the emotional tone of readers' comments. In his research, 38 newspaper articles were selected and a library of readers' comments was compiled. The study discovered that negative remarks heavily dominated in the Russian-language comments (59.3%), whereas neutral comments made up the largest group in the English-language comments (46.1 %). Both Russian and English-language comments had the lowest percentage of good remarks, although Russian-language comments had twice as many of them as the English-language comments (17.7% and 8.4 %, respectively). The researchers note that Russian-speaking users frequently employ negative evaluative language, as well as a lot of irony, sarcasm, and insults directed towards the state and other commentators, whereas comments on English-language articles have a neutral tone [12].

In constructing an application for assessing the tone of messages from social networks, Bobyakova D.A. used several approaches: supervised machine learning, specifically the Naïve Bayesian classifier, and the dictionary-based approach. In the beginning, the frequency dictionary included 100 terms with the highest frequency of words in Russian. Furthermore, the majority of the terms in the compiled dictionary turned out to be pronouns, connectives, and prepositions, resulting in a vocabulary of only 29 items. The tests were performed on a Twitter database of 100,000 texts. To pre-process the texts, the author removed stop words, links, hashtags, and words with the highest frequency of occurrence. Bigrams and unigrams were used to show the documents. This method demonstrated 86.6 % training accuracy and 89.1 % recall [13].

In general, the automatic text tonality analysis is a sufficiently impartial and effective method that can be used successfully in both sociological and linguistic research.

Conclusion

This article investigates the textual features of messages in public networks in the context of developing ways to assess the emotional coloring of opinions and discusses strategies to analyze text messages.

Natural language analysis tasks are gaining popularity as the volume of unstructured textual data grows. Modern techniques can be used to handle problems of this type, given the availability of open machine learning libraries. The number of conferences in the field of sentiment analysis is growing year by year, and so does the number of publications in Russian and other foreign languages on text analysis. According to the Google Academy, about 700 works on the sentimental analysis of Russian-language texts were published in 2018, as opposed to 8,500 works on the English-language texts. It should be noted that, based on the above material, we can determine that the researchers obtained the accuracy of the tonal analysis of Russian-language texts of about 80%, and in English-language texts, the accuracy reaches 96% [14].

The existence of numerous works on the topic of sentiment analysis indicates that this topic is relevant today and is in demand in many areas, such as the economic market, politics, marketing, etc. The approach proposed in the work can be used for marketing, sociological and political research. It also allows monitoring the loyalty of the audience to a particular topic or brand, which gives the management an opportunity to make timely decisions. However, as the analytical study demonstrates, sentimental analysis algorithms for Russian-language texts are less developed than those for foreign-language texts.

The investigation of various approaches of tonality analysis by different experts assisted in the selection of the appropriate method of analysis for developing a system for analyzing Russian-language comments. At the moment, a training sample for training a multi-class classifier of Russian-language Internet texts is being developed and the maximum weighted average f1- score, which is the average of accuracy values, is approaching 50%. The hybrid methods consist of supervised machine learning and a bag-of-words rule-based approaches. This

model uses a hybrid approach to analyze the emotional coloring of the text tone. More work needs to be done to search and develop the optimal method for training a model for Russian language text messages. Thus, it can be concluded that there is a high demand in the modern world for automatic sentiment analysis of texts, as well as an increase of its application possibilities.

REFERENCES

1. Kolmogorova A.V. (2019) The use of texts of the "Internet revelation" genre in the context of solving the problems of sentiment analysis. [Bulletin of NGU. Series Linguistics and Intercultural Communication] 3.71-82.
2. Moussa, M., Mohamed, E. & Haggag, M. (2018) A survey on opinion summarization techniques for social media. *Future Computing and Informatics Journal*. 3 (1). 82–109 PP.
3. Boudad, N. et al. (2018) Sentiment analysis in Arabic: A review of the literature. *Ain Shams Engineering Journal*. 9 (4). 2479–2490.
4. Taboada, Maite & Brooke, Julian & Tofiloski, Milan & Voll, Kimberly & Stede, Manfred. (2011). *Lexicon-Based Methods for Sentiment Analysis*. *Computational Linguistics*. 37. 267-307.
5. Kaur, Fatehjeet & Bhatia, Rekha. (2016). Sentiment Analyzing by Dictionary based Approach. *International Journal of Computer Applications*. 152. 32-34.
6. Kolmogorova A.V., Kalinin A.A., Malikova A.V. (2018) Linguistic principles and methods of computational linguistics for solving problems of sentiment analysis of Russian texts. [Series: Actual problems of philology and pedagogical linguistics] 1, 139–148
7. Yousefpour A., Ibrahim R., Hamed H.A. (2017) Ordinal-based and frequency-based integration of feature selection methods for sentiment analysis. [Expert Systems with Applications] Vol. 75. 80–93 PP.
8. Khudyakova M.V., Davydov S., Vasilyev V.G. (2017) Classification of feedback of users using fragment rules.
9. Poetze F., Ebster C., Strauss C. (2018) Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts. [Proceedings of the 9th International Conference on Ambient Systems, Networks and Technologies (ANT)]. Vol. 130,660–666 PP.
10. Pang B., Lee L. (2018) Opinion mining and sentiment analysis. [Foundations and Trends in Information Retrieval]. Vol. 2, 1-135P.
11. Stieglitz S., Mirbabaie M., Ross B., Neuberger C. (2018) Social media analytics – Challenges in topic discovery, data collection, and data preparation. [International Journal of Information Management]. Vol. 39,156–168 PP.
12. Chernyshevich M.V. (2018) Sentiment classification for the task of automatic text sentiment analysis. *Vestnik Uchenye zapiski UO VGU im. P.M. Masherova* 28, 136–140
13. D.A. Bobyakova (2017) Development of an application for analyzing the sentiment of texts from social networks. [Final qualification work of ITMO University, St. Petersburg, Russia] 2017 – 74 P.
14. K. Archipenko, I. Kozlov., Y. Trofimovich, K. Skornyakov., A. Gomzin., D. Turdakov (2017) Comparison of neural network architectures for the analysis of Russian tweets. [Proceedings of the International Conference on Computational Linguistics and Intelligent Technologies.] URL: <http://www.dialog-21.ru/media/3380/arkhipenkoetal.pdf>

Көшімбай А.Б., Молдагулова А.Н.

Тоналдылықты анықтау мақсатында әлеуметтік желілердің деректерін талдау және өңдеу әдісін зерттеу

Аңдатпа. Әлеуметтік онлайн-сервистердің кең таралуы және Үлкен Деректер технологияларының дамуы әлеуметтік желілердегі деректерді әртүрлі салаларда қолдануға деген қызығушылықтарды артыруда. Қазіргі таңда контент анализ және "әлеуметтік желілерді мониторинг" технологиялары танымалдылыққа ие болуда. Атап айтқанда, мәтіннің тоналдылығын талдау табиғи тілді өңдеу саласындағы маңызды міндеттердің бірі болып табылады. Ол әртүрлі салаларда қолданылады. Ұсынылған мақалада мәтіндік деректердегі эмоцияларды сәйкестендірудің негізгі әдістері қарастырылады. Мәтіндік деректердегі эмоцияларды компьютерлік талдау саласындағы қазіргі жетістіктер талданды. Қазіргі уақытта әлеуметтік медиа мәтіндерінің лексикасының эмоционалды түсін анықтау үшін автоматты талдау саласында көптеген шешілмеген мәселелер бар.

Түйін сөздер: әлеуметтік желілерді мониторинг, комментарийлерге анализ жасау, әлеуметтік көңіл-күй, пайдаланушының қабылдауын бағалау, машиналық әдіспен оқыту

Кошимбай А.Б. Молдагулова А.Н

Исследование метода анализа и обработки данных социальных сетей с целью определения тональности

Аннотация. Быстрое распространение общественных онлайн-сервисов и эволюция технологий «Больших данных» инициировали внимание к применению сведений из общественных сетей во всевозможных секторах экономики. На сегодняшний момент, известность технологии завоевывают, технологии как «прогноз социальных сетей» (social listening) и контент анализа. В особенности анализ тональности текста является одной из важных задач в области обработки естественного языка. Необходимо подчеркнуть, что данная технология применяется в разных областях. В предоставленной статье рассматриваются основные методы идентификации эмоций в текстовых данных. Исследованы и проанализированы существующие достижения в области компьютерного анализа эмоций в текстах. В результате исследования, на данный момент существует множество нерешенных проблем в области автоматического анализа для определения эмоциональной окраски текстов в социальных сетях.

Ключевые слова: анализ тональности текста, сентимент-анализ, классификатор, метод опорных векторов, тональный словарь

Авторлар туралы ақпарат

Көшімбай Айасем Бекжанқызы – Халықаралық ақпараттық технологиялар университетінің «Ақпараттық жүйелер» кафедрасының магистранты.

Молдагулова Айман Николаевна – PhD, Халықаралық ақпараттық технологиялар университетінің «Ақпараттық жүйелер» кафедрасының қауымдастырылған профессоры.

Сведения об авторах

Кошимбай Айасем Бекжанқызы – магистрант кафедры «Информационные системы» Международного университета информационных технологий.

Молдагулова Айман Николаевна – PhD, ассоц. профессор кафедры «Информационные системы» Международного университета информационных технологий.

About the authors

Aiasem B. Koshimbay – Master's student, Department of Information Systems, International University of Information Technology.

Aiman N. Moldagulova – PhD, Assoc. Professor, Department of Information Systems, International University of Information Technology.

INTERNATIONAL JOURNAL OF INFORMATION AND
COMMUNICATION TECHNOLOGIES

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ
КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

Ответственный за выпуск	Есбергенов Досым Бектенович
Редакторы	Медведев Евгений Юрьевич
Компьютерная верстка и дизайн	Жадыранова Гульнур Даутбековна

Редакция журнала не несет ответственности за
недостоверные сведения в статье и
неточную информацию по цитируемой литературе

Подписано в печать 15.12.2021 г.
Тираж 500 экз. Формат 60x84 1/16. Бумага тип.
Уч.-изд.л. 6.5. Заказ №170