A content analysis software system for efficient monitoring and detection of hate speech in online media

Yuliya Krylova-Grek[1,2], Oleksandr Burov[3,4]

[1]*National University of "Kyiv-Mohyla Academy", 2 Hryhoriya Skovorody Str., Kyiv, 04655, Ukraine*
[2]*Uppsala University, Gamla torget 3, 753 20 Uppsala, Sweden*
[3]*Institute for Digitalisation of Education, 9 M. Berlinskoho Str., Kyiv, 04060, Ukraine*
[4]*University of Vienna, 5 Liebiggasse, Vienna, 1010, Austria*

Abstract

This paper presents the results of interdisciplinary project that is a combination of computer program and psycholinguistic approach to media study. In the research we presented the programs that can be used for monitoring and analysis of media content to identify hate speech at its early stage. The aims of research were the following: 1) develop content analysis program for monitoring Russian media outlets; 2) apply the psycholinguistic approach for identifying hidden and manipulative hate speech. In the research there were used two types of content-analysis: quantitative and qualitative. Quantitative content analysis was conducted with computer program that was developed to select publication that could have contained hate speech. For qualitative content analysis the psycholinguistic method of text analysis was used. The method applies for identification methods and tolls that journalists use to incriminate hidden and manipulative hate speech. It is hypothesized that programs of content-analysis help to optimize work and makes it less time-consuming and more effective for analyst, journalists and other specialists who involved into media study. Methods. Quantitative content analysis, psycholinguistic method of qualitative content-analysis. Quantitative content analysis was developed with Python programming language. The publications were selected according to the key words, periods of search (month) and the name of outlet. The list of key words includes words that are used in media for discrimination, dehumanization, and marginalization of objects of hate. Implementation such a program helped to reduce time of monitoring of media outlets. The qualitative content-analysis was conducted with the authors' psycholinguistic method of text analysis that can be applied for analyzing media texts. The programs of content analysis were applied within the project "Hate Speech in Online Media Publicizing Events in Crimea". The results were published in a data analysis report on spreading the hate speech in the Russian language media communicating the armed Ukraine – Russia conflict and events related to it in Crimea on a regular base (December 2020 – May 2021). The research showed that the content analysis programs used in the project are useful tools for systematizing and processing data in humanities research and can be used by a wide range of specialist who have deal with collection and processing of information (media, communication, human rights and so on).

Keywords
content-analysis, media, hate speech, text

1. Introduction

Nowadays, people deal with large amounts of information from different online resources. According to the Digital 2020 Global Overview, the average person spends 6 hours and 43 minutes online per day, or approximately 40% of their time [1]. To navigate a large amount of information, a person needs critical thinking and analysis skills, which are considered to be among the priorities in the 21st century [2].

Examining large amounts of textual data is a time-consuming and labour-intensive process that can be simplified by using computer programs that allow quicker and more efficient processing of information and receive reliable quantitative results [3]. Content analysis is a qualitative and quantitative method of analysing the content of documents in order to identify and measure various facts and trends reflected in these documents. The peculiarity of programs are that the documents are studied in their social context through content analysis. It can be used both as a primary research method and in combination

with other methods (e.g. in studies of media performance, in classifying responses to open-ended questionnaires). Unlike other research methods, content analysis of a text is characterised by the fact that its procedure involves counting the frequency and volume of references to certain semantic units of the text under study. The quantitative characteristics of the text obtained in this way make it possible to draw conclusions about the qualitative content of the documents.

At the same time, if information analysis is part of a professional activity, a specialist needs additional special skills for monitoring, processing, and analysing data. For example, a such specialist as media analyst, PR, journalist etc. has not only to collect and analyse information, but also to monitor sources to find a certain type of information. Therefore, it would be advisable to introduce special courses in the curriculum to teach the basics of big data analysis for future humanities professionals whose work involves processing and analysing information. In this context, teaching content analysis skills allows reducing time and conduct research, proceed data and analysis information more effectively. For example, for civil activists such skills allow effectively identify texts that contain manipulation, disinformation and hate speech. In our study, we are talking about journalists, human right defenders, and other specialties related to media analysis and processing.

In the context of our work, we will consider an example of the content analysis software application for conducting quantitative and qualitative research of the online media space. The study was conducted as part of a project involved a technical specialist, a psycholinguist, the human right activists, and journalists. The task of the project was to monitor Russian media to observe the situation before the military aggression, because we consider media as an important tool for understanding the vectors of strategic communication in the country. The results of the study (2014–2021) revealed that the Russian media actively used covert and manipulative hate speech aimed at creating a dehumanised, marginalised, and demonised image of Ukrainians as speakers of the Ukrainian language and culture. Hidden and manipulative hate speech was a signal that should have drawn public attention to further discrimination and genocide against the targets of hate.

As we see, the importance of monitoring hate speech at its initial stage is stated in documents of international organisations and research by scholars. For example, the organization "Anti-Defamation League" use "Pyramid of hate" that illustrates the prevalence of bias, hate and oppression in society and demonstrates the progressive intensification of hate: from supporting each next step to genocide. The lowest level "bias attitude" includes stereotypes, insensitive remarks, microaggression and so on. The upper levels "bias motivated violence" and "genocide" include act of violence and annihilate an entire people [4]. The stages in the bottom act as a base for mass atrocities can be consider as an early stage of hate speech that can be detected with the help of content analysis.

In the 2017 annual report of the Council of Europe highlights the danger of the dissemination and amplification of poor-quality information in media that is divided into three types: mis-, dis-, and mal-information, which are differed based on the dimensions of harm and falseness: mis-information is when false information is shared, but no harm is meant; dis-information is when false information is knowingly shared to cause harm, and mal-information is when genuine information is shared to cause harm. The report stressed that hate speech is used to cause harm and discriminate against a group of people on religion, race, and other grounds: "…people are often targeted because of their personal history or affiliations. While the information can sometimes be based on reality (for example targeting someone based on their religion) the information is being used strategically to cause harm" [5].

In recent years, there has been considerable interest in issues of the negative influence of mass media in humanities (psychology, linguistics, sociology, etc.), IT, and Computational linguistics.

In linguistics, much work on the potential of speech propaganda and manipulation has been carried out by Bulyigina and Shmelev [6], Elswah and Howard [7], Aronson and McGlone [8], Soloviova [9]. Numerous studies have been published on the distortion of the meaning of concepts in media texts (McGlone et al. [10], Shmelev [11], Burridge and Allan [12], Vakaliuk et al. [13], Pilkevych et al. [14] and others).

There is a vast amount of literature on specific features of media representations of war and military conflicts (Pocheptsov [15, 16], Pack [17], Kamalipour [18], Galtung [19], Dawes [20]). In particular, Dawes [20] points out that since speech shapes our perception of reality, the age of information can be

called the age of manipulation.

In computational linguistics, the study and detection of hate speech explore by using natural language processing. For example, Schmidt and Wiegand [21] studied the ways of the automatic detection of hate speech. All surveyed methods include common features that are usually used in the computer program to identify hate speech, such as a set of negative words or expressions, using various complex features using ("dependency parse information", "features modeling specific linguistic constructs", "meta-information" and so on). At the same time the authors stressed that in most cases, computer program results can't be considered as full, because "they are only evaluated on individual data sets most of which are not publicly available" [21]. Taking into consideration the weak features of computer analysis of hate speech, we think that the best result researchers can receive if they combine computer and human inspection.

2. Aims

The aims: 1) develop content analysis program for monitoring Russian media outlets; 2) apply the psycholinguistic approach for identifying hidden and manipulative hate speech.

3. Theoretical background

Content analysis is a quantitative and qualitative method of analysing content (text, pictures etc.) in order to identify and measure various facts and trends reflected in these materials. The purpose of quantitative content-analysis is to fins valid resources and to collect certain data The purpose of qualitative content analysis is to organize and elicit meaning from the collected data and to draw realistic conclusions from it. The utilization of content analysis in Humanities and Social Sciences are described by Dey [22], Lester et al. [23], Bengtsson [24]. The authors pay attention to qualitative content analysis and stress the advantages of utilizing capabilities of computer (applications, programs, soft) for gathering and proceeding data. However, the authors didn't consider the question of media content-analysis and didn't propose any special programs that can help to gather and measure hate speech in online media.

The peculiarity of content analysis of a media text is its direct correlation with external circumstances, such as the socio-political situation. Content analysis is based on counting the occurrence of determined components in the analysed information materials, supplemented by identifying statistical relationships and analysing the structural links between them.

The need to use content analysis for analysis of large text arrays was caused by the development of mass communications in the late nineteenth and early twentieth centuries. The content-analysis was used for analytical research of texts in the media outlets. This method can be used as the main method, for example, to analyse the political orientation of a publication, or as an auxiliary or control method with other methods, for example, to measure media effectiveness. Content-analysis can also be applied alongside with other methods, for example linguistic and psycholinguistic, for evaluating the effectiveness of a media outlet.

One of the main founders of the research procedure is the sociologist Lasswell [25], who was the first to examine the impact of the media on the worldview of the population during the First World War. He chose the texts of newspapers, bulletins, and other information messages as sources, identified key themes, statements and social models on which propaganda was based. Based on the data obtained, he drew conclusions about the strategic goals of the countries of a conflict [25]. The method of content analysis was implemented the during second world war for identifying if an American newspaper consists pro-Nazi texts. As a result, the newspaper was closed, and the method was widely recognised and began to be actively used to analyse media products [26].

The classical model of content analysis proposed by Lasswell [27] is the following: 1) depending on the purpose of the study, the text is divided into parts, each of which was subsequently subjected to analysis, then the results were compared and summarised; 2) key units are selected for text analysis.

The keywords may vary depending on the research objective, for example, a word-symbol such as the name of a leader, the name of a country, the name of an ideology, etc. As a result of the analysis of information messages, we should get answers to the main 5 questions: who transmits the information, what is the information about, how the information is disseminated, what audience the information is intended for, what effect the message has [27].

Lasswell's research became the basis for the development of further procedures for analysis of media texts and answering a variety of questions, ranging from the peculiarities of the worldview of the average citizen to the ideological orientation of a particular media organ [22].

Today, various types of content analysis are used both for analysing large data sets and for analysing a single text. Because of a large demand of such programs. IT industry is actively developing software to improve functionality and customise content-analysis programs to analyse big text data, which parses text from unformatted content and unstructured data from social media, news reports, surveys, etc. to provide practical information such as the mentioning frequency of a certain brand, personality event, or counting particular words in the texts (e.g., programs for semantic content analysis such as Semantrum, Nvivo, MAXQDA, Yoshikoder, Advego, SentiStrength, Nvivo, etc.). The brief characteristics of some of these programmes are given below.

ADVEGO is a free programme for semantic analysis of texts. It performs text statistics by the number of words, determines the most frequently used words and their number, as well as words and phrases that are part of the semantic core [28].

Yoshikoder is a free programme that allows you to find online texts by a given phrase or expression. This programme allows you to examine keywords-in-context, and perform basic content analyses, in any language. At the same time, in our research we need to use some specific keywords that are not included in the dictionary [29].

Programme SentiStrength which refers to the programme for Sentiment Analysis or Opinion Mining. The programme is set up to search for evaluative and emotive vocabulary in the text according to a pre-compiled dictionary, which is divided into groups: "negative" and "positive" vocabulary, which is placed on the scale of emotional intensity of the statement [30].

MaxQDA, NVivo software can be applied in a range of sectors: from social science and education to healthcare and business. These programs can be used to analyse data from interviews, surveys, field notes, web pages, and journal articles. These programmes are similar in terms of functionality and operation [31, 32].

4. Methods

In the research the following methods were used: quantitative content analysis, psycholinguistic method of qualitative content-analysis. Quantitative content analysis was developed with Python programming language. The publications were selected according to the key words, periods of search (month) and the name of outlet. The list of key words includes words that are used in media for discrimination, dehumanization, and marginalisation of objects of hate. Implementation such a program helped to reduce time of monitoring of media outlets. The qualitative content-analysis was conducted with the psycholinguistic method of text analysis that is the authors' method that can be applied for analysing media texts.

4.1. Quantitative content analysis

The study outlines the approaches and opportunities for interdisciplinary interaction of humanities and computer sciences. The research of hate speech was done with collaboration with the Crimean Human Rights Group (Siedova and Krylova-Grek [33]). The objective of the research was to identify hidden and manipulative hate speech in a number of Russian media outlets. A content analysis programme was used for selecting media texts based on selected keywords. The study was conducted in 2014-2022 and consisted of three stages: basic monitoring; quantitative content analysis; qualitative content-analysis; conclusions.

I. Basic monitoring. At this stage, a list of media outlets has been made. The study investigated 11 online media outlets with audiences of over 1 million readers per month.

II. Quantitative content analysis. This stage involves 1) monitoring of the selected media outlets; 2) identification of groups in relation to which negative characteristics are applied; identification of keywords used to create a negative image of the selected groups; 4) texts selection.

III. Qualitative content analysis. At this stage, we applied the author's psycholinguistic analysis of the selected texts.

For conducting quantitative content-analysis, a special program was developed by analyst and technical specialists of the Crimean Human Rights Group O. Sedov (now the program is under the process of patent). The program allows to optimize the teamwork by surveying a large array of information to find a pool of texts that can incite xenophobia and hatred.

The program sets up the following parameters for hate speech content analysis:

1. Online resource name. We specified the names of news agencies that specialize in current news. The initial cohort includes the nine most popular online media, whose traffic ranges from 1 million to 15 million visitors monthly.

2. Keywords. We entered words that usually accompany texts with hate speech. To single out keywords, we developed Hate dictionary based on the careful analysis of publications in selected media. Monitoring and word selection took place in 2017-2018. The dictionary includes more than 400 words. It should be noted the dictionary is constantly supplemented and changed due to the emergence of new narratives, concepts, and words.

3. Search period. We set the time interval: date and year. In our research, the most appropriate search period that can provide us with the required amount of data is one month. For example, from 1 of August to 31 of August.

The content analysis program searched the selected websites to find in which hate speech regarding such ethnic and social groups might be used: Ukrainians, Crimean Tatars, Jews, Residents of Crimea and Donbas Russians, Activists and Journalists, Euromaidan Participants, LGBTQ groups. Most of the hate words referred to Ukrainians.

Key words included new created lexicon, archetypes of the second world war and other words that marginalized some ethnic and social group. Studies found that articles featuring hate speech often use terms and concepts that are not recorded in the official dictionaries of the Russian language. Some of these words are 'made up', created purposefully to incite hatred. In some cases, such words are formed by combining parts of words denoting nationality and obscene lexical units, for example, "kriptobanderovtsy" (a compound word made up of "crypto" and "Banderite"), "natsgady" (a compound word made up of a shortened form of Nazi/Nationalists that sounds similar and is used by Russian media as a play on words; and the word 'gady' that is similar to 'foul people' in English (skunk, despicable), "khokhlodauny" (a compound word made up of 'khokhly' (see above) and the Russian word for Down's syndrome sufferers ([douny] /n., pl.(ukr.)/).

Apart from other things, we considered a well-known vocabulary available in glossaries, and with negative connotations, that was used by journalists in the investigated media in relation to objects of hatred, ridicule, marginalisation, and so on. Most such words go back to of WWII archetypes, and common negative stereotypes. For example, "fascist", "fascism", "nazi". Moreover, based on the phonetic similarity of the words "Nazi" and "nationalism" in Ukrainian and Russian, journalists use the word [natsist] instead of [natsionalist] (Nazi, nationalist).

Among other words are those that humiliate and marginalise another people's language by distorting the phonetic sound of words for sarcastic or mocking reasons, and bracketing the phonetic spelling of words (the transliteration of Ukrainian words into Russian in terms of our study), which contextually, in the publications, is sarcastic and expresses contempt for the Ukrainian language and its speakers (e.g., "svidomyye", "nezalezna").

The study materials were monitored and sampled using a content analysis program: texts were selected using set key words and word combinations. Such words and word combinations had been

collected by the project's monitoring group during preliminary studies into the hate content of Russian-language online media and public social networks. Nine Russian-language sites were searched. The Russian language vocabulary used in relation to the main ethnic groups, as well as to the most vulnerable social groups of the population, who currently live in the territory occupied by the Russian Federation, was added to the list of key words and word combinations.

As an example, we delve into the interface of the hate speech content analysis program in greater detail. First and foremost, it's important to highlight the distinctive features of the program's terminology: "query words" refer to the specific words we input into the search based on our Dictionary of Hate. On the other hand, "keywords" are the words that most frequently appear in the text of a publication. For instance, let's take the online platform Politnavigator as an example. We selected the keywords for September 2019 (figure 1).
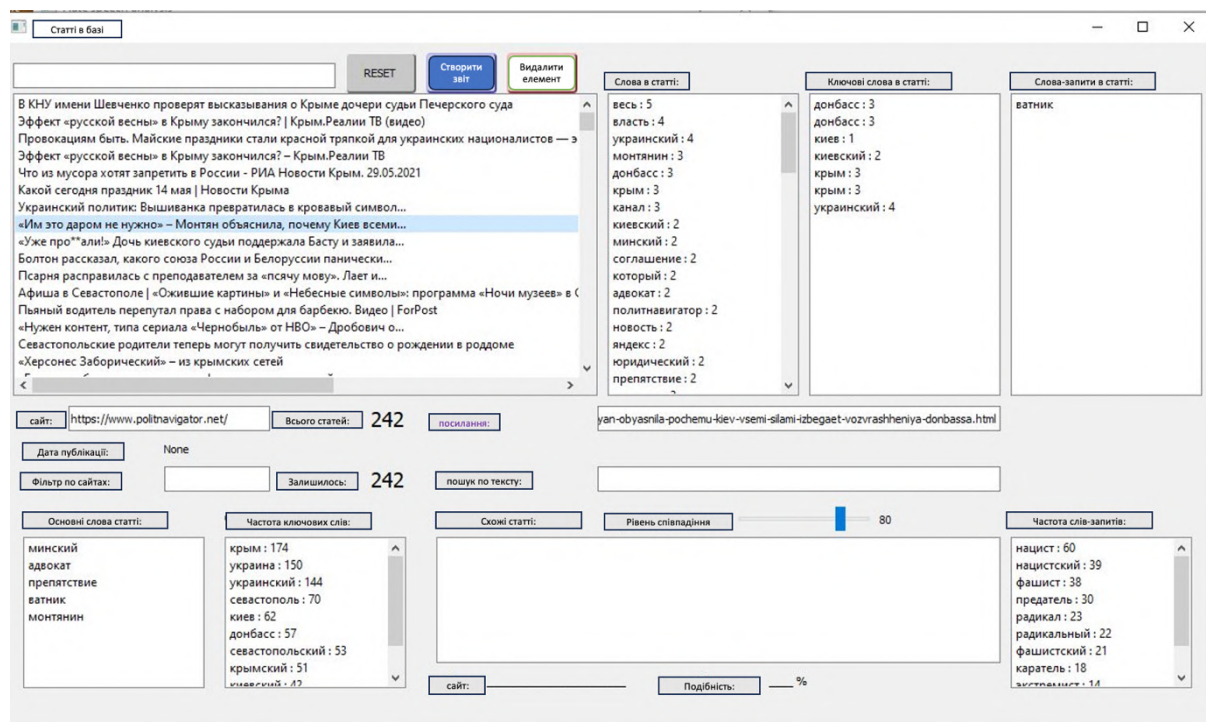
Figure 1: The interface of content-analysis program.

The program's interface presented us with the following information:

- the total number of articles containing the selected keywords (both in their titles and within the text);
- a list of these articles along with corresponding links;
- a list of the keywords identified within the articles;
- the frequency of each keyword's appearance within the articles;
- the number of query words used.

Upon hovering over a specific article, the interface provided the subsequent details:

- a direct link to the respective article;
- a hyperlink directing to the media outlet's website where the article is published;
- the keywords featured within the article;
- the frequency and number of the keywords;
- the frequency and number of the query words.

This interface design allows for a thorough examination of hate speech content, aiding in the analysis and understanding of its prevalence and distribution across various publications.

The content analysis program was used to electronically process the content of the selected websites for the indicated period regarding abovementioned key words and word combinations. The sample of publications produced by the content analysis program was subject to a psycholinguistic analysis exercise. The content was divided into two groups based on the exercise results: publications featuring hate speech, and publications with other types of manipulations.

4.2. The psycholinguistic analysis

The texts selected through content analysis underwent psycholinguistic analysis to distinguish between those containing hate speech and those that did not. The content analysis program selected all texts that contain key words, but not all the texts consists hate speech. The reasons are related to the algorithm for configuring the content analysis programme which scans the page together with comments and other information. Therefore, the reasons for the errors are justified by the following factors: 1) comments under the text included hate vocabulary. As the research aimed to analyse the products of the media specialists' activities, the comments were not taken into account and such texts were attributed to error. Among other things, comments can be a product distributed by bots or specifically hired people, which requires additional technical methods for their analysis; 2) texts in which keywords have a direct meaning; for example, the word "fascists" used in the text give a factual retrospective to the military events of the Second World War. At the same time, there were only a few such texts (2%) that were removed from the list for further analysis.

Subsequently, the texts containing hate speech were scrutinized with psycholinguistic analysis to identify the methods and techniques employed by journalists in these publications [34]. The analysis is a part of an innovative author's methodology, which assists in identifying both direct and manipulative hate speech that does not contain direct insults and calls for gender, racial or religious intolerance, but forms a negative attitude towards certain groups and individuals. The psycholinguistic analysis was performed manually, as only a professional can assess sarcasm, infer indirect meanings of words, decode and elucidate the significance of newly created words that do not exist in the dictionary.

After having conducted the textual analysis, hate speech was divided into three types, which are characterized by the specific linguistic and graphic tools used in the publication:

Type 1: direct hate speech;
Type 2: indirect (hidden) hate speech;
Type 3: manipulative hate speech.

5. Results and discussion

The monitoring period of online media are from December 1, 2020, to May 31, 2021. We think that and the amount of content for this period is enough to validate the study outcomes. The project monitoring group obtained 1,284 publications when keyword-based electronic content sampling had been completed. The content was processed according to the psycholinguistic method of media text analysis, and the result was that 560 publications featuring hate speech elements selected from the entire content were received by the project analytical experts. These include 16 texts with Hate Speech Type One; 341 texts with Hate Speech Type Two; and 203 texts with Hate Speech Type Three [34].

Type 1. Direct use of hate speech is characterised by: the use of obscenities, direct insults and calls for violence.
Type 2. Indirect use of hate speech is characterised by: marginalisation of the other party with usage of offensive ethnonyms, polarisation, dividing a society into in-group and out-group, generalisation (attributing one case to an entire group), negative sarcasm and irony, the use of archetypes and stereotypes that develop a certain world view and attitude, the creation of new concepts with negative connotations.

Type 3. Manipulative hate speech is characterised by: substituting the meanings of concepts that create negative associations; using fake; use the opinion of biased or 'pseudo-experts' citing (people who have little or no experience or knowledge about the problem on which they comment); distorting and misinterpreting historical facts; justifying aggression or violence against a target group;enhancing informative messages with non-linguistic means (photograph, pictures); using manipulative titles that does not match, or distorts, the information presented in the text of the article.

The study investigated eleven Russian language online media, including: five news websites, with news items on the situation in Crimea largely dominating the content, the audiences of over 1 million readers per month, and a minimum 25% share of Ukrainian readers; three Russian news websites regularly writing about Crimea and Donbas, with audiences of over 1 million readers per month, and a minimum 25% share of Ukrainian readers; two news websites, with largely dominating content that describes the situation in Crimea, financed out of the Russian Federation budget; and the official website of the "government" of the Pravitel'stvo Kryma (figure 2). The analysis of the auditoriums at selected sites was conducted using SimilarWeb, a platform that offers insights into global digital traffic [35].
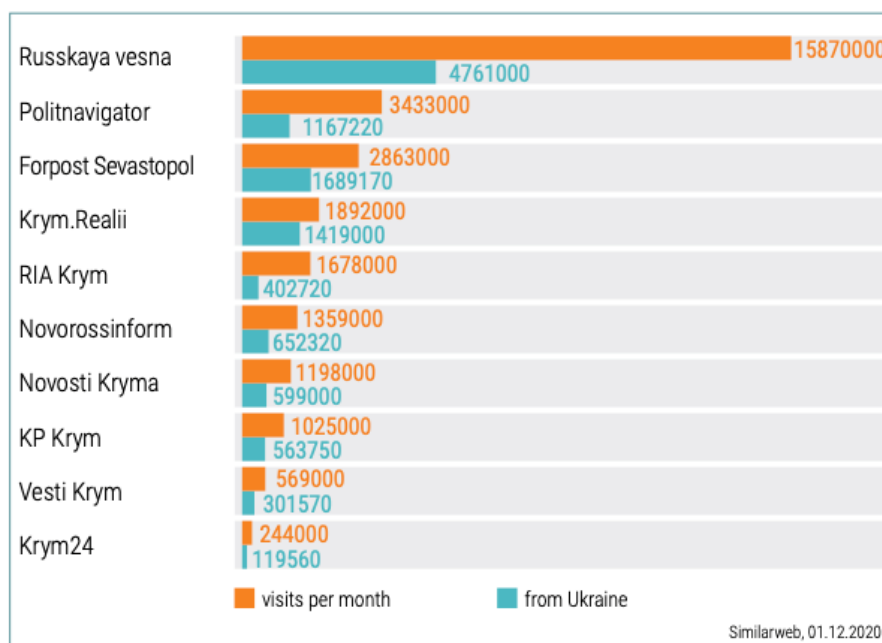
Figure 2: Audience of studied websites.

The reason for including texts that did not contain hate speech in the search results is as follows: a) certain articles did contain words from the dictionary, but they corresponded to their direct meaning, for example, the word "down" was used to refer to people with Down syndrome; b) the program counted words and expressions contained in the comments under the publication. The second aspect was the main reason for the large percentage of inappropriate texts.

6. Conclusion

The effectiveness of using a quantitative content analysis programme to monitor hate speech is determined by the following criteria:

1) the amount of time a specialist spends on selecting texts: a specialist spends up to 15 minutes on selecting articles;

2) search criteria: in the software, you can set the search period, the source to be analysed, and a fairly large number of keywords that can be entered, for example, in our study we identified about 400 words related to hate speech;

As a result, reducing the time spent on quantitative analysis allows you to spend more time on qualitative analysis of texts. Another advantage of the quantitative-qualitative approach is that, unlike other search engines and programs, it has better functionality for selecting texts, reduces the time for text selection, and when combined with psycholinguistic analysis of texts, allows us to go beyond simple mathematical calculations and analyse the psycholinguistic methods and techniques used by journalists to influence the minds of the audience.

Acknowledgments

References

[1] Digital 2020, Global digital overview, Technical Report, Hootsuite, 2020. URL: https://media.rbcdn.ru/media/reports/Digital_2020.pdf.

[2] The Future of Jobs Report 2020, Technical Report, The World Economic Forum, 2020. URL: https://www3.weforum.org/docs/WEF_Future_of_Jobs_2020.pdf.

[3] B. J.-Y. Guédé, From Content Analysis to Content Analysis of Digital Social Networks, in: 2022 International Conference on International Studies in Social Sciences and Humanities (CISOC 2022), Atlantis Press, 2022, pp. 235–248. doi:10.2991/978-2-494069-25-1_23.

[4] Anti-Defamation League, Pyramid of Hate, 2021. URL: https://www.adl.org/sites/default/files/pyramid-of-hate-web-english_1.pdf.

[5] C. Wardle, H. Derakhshan, Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking, Technical Report, Council of Europe, 2017. URL: https://rm.coe.int/information-disorder-report-november-2017/1680764666.

[6] T. Bulyigina, A. Shmelev, Linguistic conceptualization of the world (based on the material of Russian grammar), Moscow, 1997.

[7] M. Elswah, P. N. Howard, "Anything that Causes Chaos": The Organizational Behavior of Russia Today (RT), Journal of Communication 70 (2020) 623–645. doi:10.1093/joc/jqaa027.

[8] J. Aronson, M. S. McGlone, Stereotype and social identity threat, in: T. D. Nelson (Ed.), Handbook of prejudice, stereotyping, and discrimination, Psychology Press, New York, 2009, pp. 153–178.

[9] T. Soloviova, The features of the learning of political linguistics at higher education institutions, Educational Dimension 3 (2020) 217–232. doi:10.31812/educdim.v55i0.3944.

[10] M. S. McGlone, G. Beck, A. Pfiester, Contamination and Camouflage in Euphemisms, Communication Monographs 73 (2006) 261–282. doi:10.1080/03637750600794296.

[11] D. N. Shmelev, Euphemism [evfemizm], in: Y. N. Karaulov (Ed.), Russkij yazyk. Enciklopediya [Russian language. Encyclopedia], 2 ed., Bolshaya ros. encikljpediya. Drofa, Moscow, 2008.

[12] K. Burridge, K. Allan, Euphemism & dysphemism: Language used as shield and weapon, Oxford University Press, 1991.

[13] T. Vakaliuk, I. Pilkevych, D. Fedorchuk, V. Osadchyi, A. Tokar, O. Naumchak, Methodology of monitoring negative psychological influences in online media, Educational Technology Quarterly 2022 (2022) 143–151. doi:10.55056/etq.1.

[14] I. A. Pilkevych, D. L. Fedorchuk, M. P. Romanchuk, O. M. Naumchak, Approach to the fake news detection using the graph neural networks, Journal of Edge Computing 2 (2023) 24–36. doi:10.55056/jec.592.

[15] G. G. Pocheptsov, Theory of communication, Refl-book, Moscow, 2001.

[16] G. G. Pocheptsov, (Des)information, Palivoda, Kiev, 2019.

[17] J. L. Pack, Book Review: The Language of War by Steve Thorne, 2006. London and New York: Routledge, pp. vii+ 104 ISBN: 0 415 35868 X (pbk), Language and Literature 18 (2009) 408–410. doi:10.1177/09639470090180041201.

[18] Y. R. Kamalipour, Language, media and war: Manipulating public perceptions, Media 3 (2010) 87.

[19] J. Galtung, Language and War: Is There a Connection?, Current Research on Peace and Violence 10 (1987) 2–6. URL: https://www.jstor.org/stable/40725052.

[20] J. Dawes, The language of war: literature and culture in the US from the Civil War through World War II, Harvard University Press, 2002. doi:10.4159/9780674030268-intro.

[21] A. Schmidt, M. Wiegand, A Survey on Hate Speech Detection using Natural Language Processing, in: L.-W. Ku, C.-T. Li (Eds.), Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media, Association for Computational Linguistics, Valencia, Spain, 2017, pp. 1–10. doi:10.18653/v1/W17-1101.

[22] I. Dey, Qualitative data analysis: A user friendly guide for social scientists, Routledge, 2003. doi:10.4324/9780203412497.

[23] J. N. Lester, Y. Cho, C. R. Lochmiller, Learning to do qualitative data analysis: A starting point, Human resource development review 19 (2020) 94–106. doi:10.1177/1534484320903890.

[24] M. Bengtsson, How to plan and perform a qualitative study using content analysis, NursingPlus Open 2 (2016) 8–14. doi:10.1016/j.npls.2016.01.001.

[25] H. D. Lasswell, The theory of political propaganda, American political science review 21 (1927) 627–631. doi:10.2307/1945515.

[26] N. V. Kostenko, V. F. Ivanov, Experience of Content Analysis: Models and Practices, Centre for Free Press, 2005.

[27] H. D. Lasswell, The structure and function of communication in society, The communication of ideas 37 (1948) 136–139.

[28] Advego, Semantic Text Analysis for SEO, 2024. URL: https://advego.com/text/seo/.

[29] W. Lowe, Yoshikoder: Cross-platform multilingual content analysis, 2015. URL: https://yoshikoder.org.

[30] Y. Z. Hung, Python 3 wrapper for SentiStrength, 2021. URL: https://github.com/zhunhung/Python-SentiStrength.

[31] Nvivo, 2024. URL: https://help-nv.qsrinternational.com/12/win/v12.1.112-d3ea61/Content/cases/cases.htm.

[32] MAXQDA | Official Site | All-In-One Tool for Qualitative Data Analysis, 2024. URL: https://www.maxqda.com/.

[33] I. Siedova, Y. Krylova-Grek, Hate Speech in Online Media Publicizing Events in Crimea: a data analysis report on spreading the hate speech in the Russian language media communicating the armed Ukraine – Russia conflict and events related to it in Crimea on a regular base (December 2020 – May 2021), Technical Report, Kyiv, 2022. URL: https://crimeahrg.org/wp-content/uploads/2023/04/hate-spe.pdf.

[34] Y. Krylova-Grek, Psycholinguistic approach to the analysis of manipulative and indirect hate speech in media, East European Journal of Psycholinguistics 9 (2022) 82–97. doi:10.29038/eejpl.2022.9.2.kry.

[35] Similarweb, 2024. URL: https://www.similarweb.com/.