The synergistic paradigm for profiling authorship of blogosphere content by profession

Maria Komova and Alina Petrushka

Lviv Polytechnic National University, S. Bandery st. 12, Lviv, 79013, Ukraine

Abstract

The article studies the identification of the authorship of blogs by correlating the bloggers' professional affiliation with the content features of their blogs. The research aims to develop a consolidated synergistic paradigm of authorship profiling based on the bloggers' professional affiliation. The research methodology involves determining the synergistic paradigm of the blogosphere by the topic, the type of author, the level of generalization of information, the nature of iconic records, genre, emotional coloring, and language of posts; forming the consolidated synergistic paradigm of authorship profiling by profession. Interdependence in the system "picture of the bloggers' world - content" resulting from the synergy of all social and communication factors is considered by the method of postulated extrapolation. The scientific novelty of the work is the consolidated synergistic paradigm of authorship profiling in online editions by bloggers' professions using the postulated extrapolation of documentary information in the blogosphere. It is a communication model of the correlation between the picture of the bloggers' world and the content of blogs. The authorship profiling phenomenon is considered a formalized detailing of information interaction as a complex, open system characterized by structural, functional, coordination, channel, and semantic disclosures. The consolidated synergistic paradigm of authorship profiling demonstrates the correlation of such a component of the bloggers' picture of the world as their professional affiliation with the characteristics of their blogs. Each professional group of bloggers has its model of a consolidated synergistic paradigm of authorship profiling, which summarizes characteristics from all profiling features.

Keywords

Authorship Profiling, Blogosphere, Postulated Extrapolation, Synergistic Paradigm, Professional Affiliation

1. Introduction

The spread of information technologies and the growth of mass access to information via the Internet cause global communication changes in social life. Social networking sites have become an important channel of social communication, a platform for exchanging social information in any format on any topic. They opened vast multimedia opportunities for expressing meanings - knowledge, emotions - with the help of texts, graphics, sounds, and emojis. The rapid spread of information technology and cyberspace has transformed the nature of human identity from physical to virtual.

Social media Twitter, Facebook, Instagram, LinkedIn, blogs are not only universal disseminators of information, but also the most used sources of disinformation, programmed and targeted psychological influence, platforms for carrying out critical cyber attacks. The same networks are a medium for distributing anonymous messages with malicious purposes: Internet fraud, impersonation, identity theft, use of fake profiles in social networks, and plagiarism.

CMiGIN 2022: 2nd International Conference on Conflict Management in Global Information Networks, November 30, 2022, Kyiv, Ukraine
EMAIL: mariia.v.komova@lpnu.ua (M. Komova); alina.i.petrushka@lpnu.ua (A. Petrushka)
ORCID: 0000-0002-4115-3690 (M. Komova); 0000-0002-8769-4599 (A. Petrushka)

The massive scale of misuse of social media platforms is unfolding to psychologically influence the media on users' cognitive and behavioral positions. Dissemination of false information in the media is a powerful social communication technology, as it affects social perception, the formation of public opinion, and the determination of the electoral behavior of voters. Incorrectly quoting the statements of political opponents serves as a weapon that destroys their reputation. Easy access to the distribution and consumption of information causes the successful application of various special social communication technologies to influence the electoral opinion of users. Spreaders of fake news use social media platforms to exploit people's ignorance and lack of critical thinking for their purposes.

The complex infrastructure of the use of information technologies is widely integrated into the activities of modern institutions. Its functioning is threatened by critical cyber-attacks that can penetrate information systems without hindrance. Cybercrime is associated with such phenomena as the leakage of critical information, the spread of fake messages, cyberbullying, and cloud encryption.

The need to ensure information security in the sphere of management and business, as well as other spheres of social life, has actualized scientific research aimed at the unambiguous identification of a cybercriminal. Innovative methods of combating and identifying offenders have opened up the possibility of bringing them to justice in critical cyber security breaches.

Detecting unreliable information in social networks is a significant legal, political, moral, and ethical problem due to the difficulty connecting some information with known and reliable subjects. Determining the origin of sources can help society fight against unverified, incomplete, false information. Author profiling is a technology for identifying the demographic characteristics of authors in social media.

2. Recent research and publications

2.1. Identification of relevant information

Automated analysis of Twitter users' tweets from Arab countries aims to establish opinions on COVID-19 vaccines. Diachronic analysis with a sampling interval of 4 months was applied. Practical (medical) aspect: to establish the most popular vaccine in Arab countries and identify the reasons for people's reluctance to vaccinate. The second methodological aspect concerns the development of a model to detect vaccine-related tweets: tagged and untagged with prominent virus hashtags. For this, various natural language processing methods were applied based on data obtained from 1,098,376 unique tweets. The detection of vaccine-related tweets is solved as a binary classification problem. Using a statistical method - logical regression - allows the identification of marked and unmarked tweets. For the identification of vaccine-related tweets, the logistic regression model shows the highest accuracy of 0,82. The results of the analysis of attitudes to Covid and vaccination can be used during population vaccination, vaccine advertising [1].

The study of the most popular tourist destinations in Granada and their perception was based on data from 235,755 tweets on Twitter and 90,725 posts on Instagram. The authors classify tourist sentiment from messages in English and Spanish using different methods, including deep learning models. The best test results were obtained by using a bidirectional encoder. For Spanish texts, Google developed BERT, a transformer-based machine learning method for natural language preprocessing. Tweeteval was used for the multi-aspect classification of English texts. A Spanish-Tourism-BERT model is proposed to find the most popular tourist destinations and classify their perception using hashtags and negative sentiment markers for each destination. It allows the revealing both positive and negative perceptions of tourist objects with the help of their identification. Studying the most popular tourist destinations provides useful analytical information for improving the quality of tourist services and formulating optimal marketing strategies [2].

Huge volumes of data, particularly in textual, unstructured, and structured forms, are present on these social networks. Because these data have frequently been used in cybercrimes like cyberterrorism, cyberbullying, etc., extracting information from these data has now become a significant challenge in order to protect the privacy of the data [3]. Information from authors’ profiles in social networks, reactions and comments of other users can serve as an effective means of identifying fake news. In particular, the researchers propose a multi-step method that will make it possible to obtain judgments about the fakeness of a message based on the prediction of the feedback position [4]. The results of

another study confirm the effectiveness of identifying misinformation in social networks based on the analysis of users' communication interaction. The researchers propose their own Conversational Sentiment Analysis Model method for analyzing the emotional coloring of dialogues in the Twitter microblog [5].

In contrast, the results of comparing the effectiveness of different machine learning classifier algorithms for the analysis of messages from Twitter demonstrated the advantages of the Neural Network Classifier' algorithm, which demonstrates the highest accuracy [6].

The semantic component, which is not considered in the main methods, can be used as an additional functional tool for identifying relevant information. The hypothesis is that concepts and relationships between these concepts are correlated with relevant and irrelevant information. It can increase the performance of classifiers. The semantic approach is applied within the SVM classifier. This model is tested on different collections of Twitter profiles. The results show the effectiveness of the integration of the semantic component in the categorization of Twitter profiles [7].

2.2. Profiling the authorship of messages in social media

High identification accuracy is characteristic of authorship profiling methods to identify persons involved in non-compliance with the principles of academic integrity, producing plagiarism, committing cybercrimes, and spreading fake news.

In researching text identification models on Twitter (plagiarism identification), an urgent problem arises in establishing the effectiveness of using one model when analyzing to study similar cases. These models reduce the number of features without significantly changing the efficiency level. Data show that reducing the set of functions to 300 does not decrease efficiency. The analysis reveals specific terms that clearly distinguish the two genders [8].

Researchers developed a system for the unambiguous, most accurate, and fast identification of criminals in cyberspace. This technology involves analyzing the tweets of different users globally, identifying cybercriminals, and providing this information to the police for authorship identification. During analytical research, various online and offline databases are used. Texts from different Twitter users are used for intellectual analysis. A comparative analysis of modern research methods and software provides results regarding evaluating the effectiveness of various methods. Combined methods for text analysis involve the analysis of textures, algorithms, and polygraphs. These new technologies demonstrate high efficiency and will be used in future technologies as a tool against cybercrime [9].

In information war conditions, identifying fake news spreaders becomes especially relevant. Automatic identification of fake news distributors is based on machine learning methods. Various linguistic, personalized, and stylistic features and embedded words are extracted from Twitter to create a model with the help of PAN@CLEF Profiling Fake News Spreaders [10].

The identification of social network users is driven by the need to identify individuals who register multiple accounts and use them to publish fake messages to undermine the primary purpose of the social network. Existing methods for solving the problem of user identification by multiple identifiers are based on statistical data analysis and limited use of deep semantic information in messages. A deep learning method is proposed for determining semantic relationships at the document and user levels. Identifying the stylistic features of the text and establishing similarities in different messages makes it possible to identify their authorship [11].

With the help of methods of determining the anonymous authorship of texts in social media, the author of a written text is identified from a group of suspected authors. Users regularly use social media platforms Twitter, Facebook, and Instagram to share information. Twitter occupies a dominant position, forwarding millions of user messages daily. Identifying the author of micro texts is difficult due to the small amount of suspicious text. The synthesis-based convolutional neural network model consists of feature extraction and classification. Three different types of features are extracted from the original tweets. Three different deep learning-based methods (capsule, LSTM, and GRU) are used to obtain different sets of features. These functions compare and detect hidden signs of authorship. Softmax is used to predict class labels. Maps for different models illustrate text fragments for authorship prediction. A standard Twitter data set is used to evaluate the performance of the developed systems. Experimental evaluation shows that the proposed synthesis-based network can outperform previous methods [12].

Identification of the most likely author of social network publications is relevant for fields such as literature, law, cyber security, forensics, and plagiarism identification. Various models of automated natural language processing (NLP) are used. A system for identifying the author of literary articles based on a convolutional neural network (CNN) is proposed. This system allows it to identify different writing styles by visualizing writing patterns. Experiments confirmed the achievement of maximum accuracy of 93.58%. The system outperformed standard manual methods [13].

The need to identify the authorship of unreliable information spread in social networks is characteristic of the sphere of the political life of society. For this purpose, the method of associating certain information's content characteristics with a particular politician's activities is used. To determine authorship in social networks, particularly on Twitter, a metric based on compression – Normalized Compression Distance (NCD) is applied to compare the author's text with other authors' texts. The methodology works with an accuracy of 80.3% in a scenario with six different policies [14].

Correctness and relatively high accuracy in predicting authorship are achieved by using identification methods based on one of the demographic characteristics of social media users: gender, age, and profession.

Researchers of the content of social networks often experience difficulties determining the age of authors of published texts. Twitter users do not publish information about their age, limiting the possibility of profiling authors. Accurate information about age groups can be useful in marketing research, which provides knowledge about the characteristics of purchasing priorities. Linguistic studies show that users of different ages are distinguished by the use of excellent vocabulary and grammatical forms and various graphic symbols (smileys, emojis, icons). Typical emojis used by users of different age groups in their messages are set. With the help of methods of text analysis and artificial intelligence, it was established that the type and number of graphic symbols used in tweets indicate the age group of their authors [15].

Content theft and spoofing have become widespread in news texts, social media posts, and emails. Criminals operating in cyberspace must be identified. For profiling and searching for an anonymous suspect, essential attributes are gender, age, language, dialect in the region, and personality. A neural architecture is proposed for author gender determination on multimodal Twitter data. Bidirectional GRU is used to learn the coded representation of the text part of the tweet, and ResNet-50 is used to establish features from the images. An integrated author profile model was formed to predict the gender of a Twitter user by a combination of text and image representations. Experimental results show that the model achieved an accuracy of 84.03%. The system shows the writing patterns of men and women [16].

The methodology for establishing a correlation between bloggers' picture of the world and publishing activity is based on the application of the traditional authorship profiling method. The proposed method of postulated extrapolation of the blogosphere expands the scope of the application of the authorship profiling method. Author profiles established when examining a particular data set can be extrapolated to any other data set. The method of postulated extrapolation is used to prove the postulate: bloggers' picture of the world correlates with the features of their publishing activity. The method of postulated extrapolation can be applied to study the bloggers' picture of the world in any blogosphere according to various demographic and socio-political criteria: professional composition, gender, educational level, and religious, political, and cultural preferences. The use of the method of postulated extrapolation of documentary information in the blogosphere makes it possible to substantiate the statement that features characterize a specific professional group of bloggers for content filling of blogs [17]. The use of methods of authorship profiling and postulated extrapolation actualizes further comparative studies [18].

Technological opportunities have emerged for user profiling with the advent of personalized online services. Information about the gender of the author of the message has an essential role in ensuring the activity of law enforcement agencies. Information about the user's gender is unavailable to other users regarding anonymity and confidentiality. Female and male users have differences in the lexical and graphic means they use in messages. Different models of emotion-aware multimodal gender prediction are explored to determine the user's gender using their text posts. Emotional cues in multimodal posts that include text and images help predict a user's gender. The PAN 2018 dataset is enriched with emotion labels. Various multitasking architectures have been developed for gender prediction. The gender prediction results on the PAN 2018 test dataset show that the multimodal system (with text and

image) with emotion support is more effective than the unimodal model (only with text or only with image) [19].

Authorship profiling with mixed-code prediction involves analyzing data based on socio-demographic features in various combinations: gender, age, profession, political ideology, marital status, and educational level.

The profiling of political authors to identify gender, profession, and political ideology in social networks is based on deep learning architecture Spanish BERT and RoBERTa. The system adequacy level is 90% [20].

Implementing the project "PoliticEs: Spanish Author Profiling for Political Ideology" involves determining the political ideology, gender, and profession of the user of a message in social networks in Spanish. TF-IDF is applied to pre-prepared SentencePieces and custom tokens obtained by encapsulating named entities. Deep models in Spanish using manually selected feature classes was built [21].

To profile the authors, terminological analysis is offered by identifying the most frequently used terms and their functions in the data set. A negative pattern is established: the accuracy of profile prediction remains the same due to irrelevant and redundant terms in the dictionary set. Researchers searched for feature selection algorithms to determine the most informative characteristics of terms and avoid redundant features. Documents are represented as vectors using these essential information functions. A new algorithm for selecting features is proposed based on dividing terms into different categories. Various machine learning algorithms are used to evaluate the effectiveness of the proposed model. It achieved high accuracy in predicting age and gender [22].

In the field of author profiling, an important aspect is the identification of bullying in social networks. In particular, the researchers carried out a thorough study of the role classification of participants in the communication process during bullying, a correlation was made between the position of students and their role in cyberbullying [23, 24]

2.3. Ethical problems of authorship identification

The stylometric method of identifying the author based on the stylistic features of the text is widely used in historical research or to establish copyright. However, there is a caveat to its use in the context of privacy and protection of personal data on the Internet. Assessing the potential risks and consequences of using stylometry methods is important. A model of automated human identification using stylometric methods is presented. The risks regarding the preservation or violation of confidentiality and protection of personal data related to the use of stylometry were analyzed in the context of evaluating the effectiveness of stylometric identification [25].

Automated processing of natural language and analysis of social media actualize the issue of moral and ethical aspects of using the received information. The classification of moral principles applied in the analysis of textual data depends on the text and the author. A comparison of traditional and new text classifiers based on language models in English and Portuguese was made. Classification of moral principles depends on lexical information. Different models may be more suitable for a specific task [26].

Thus, the methodology of text identification and profiling of the authorship of messages in social media is a relevant and thematically branched segment of linguistic, computer, and information research (Fig. 1).

Research on authorship profiling is characterized by intensive development, a wide range of practical applications, and a high level of application efficiency. Attention is drawn to the dynamism, diversity, and innovation of research methods: logical regression method, deep learning method, machine learning method, semantic analysis, intellectual text analysis, terminological text analysis, convolutional neural network model, multimodal gender prediction method with consideration of emotions, multimodal identification system method the gender of an anonymous user, a method of associating specific characteristics of certain information with the activities of a particular politician, a metric based on compression (Normalized Compression Distance, NCD), a conceptual method of vectorization. As identifiers of categorical groups, writing patterns, graphic symbols, text, and images and their various combinations are used.

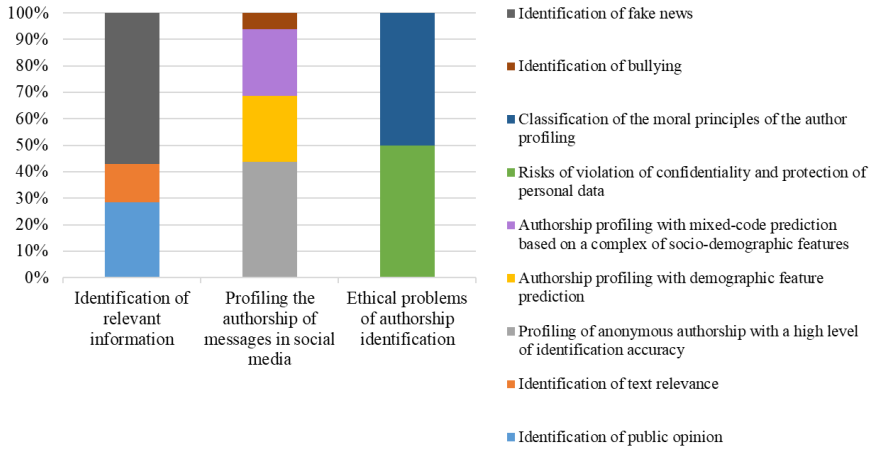

Figure 1: Research topics on profiling the authorship of messages in social media

Authorship profiling methods are widely used in many areas: forensic analysis, security, marketing, education, reputation management, prediction of fake profiles, sentiment analysis, detection of sources of disinformation, automatic adjustment of customer service communication, psychographic analysis of text indicating individual and social behavior of a politician.

The traditional technique of predicting author profiles involves identifying stylistic features inherent in the texts of different authors. Researchers identify many stylistic features, but more is needed to predict authors' profiles accurately.

For automated information classification, researchers use a wide range of methods for selecting statistical characteristics and machine learning. Intellectual analysis of text authorship is a complex technological task. Artificial intelligence technologies are used to identify, protect, recognize, create, extract, and document digital evidence, which can then be used as proof of illegal actions against social network users or to analyze critical data.

Profiling of the author takes place in two directions: analysis of the content of social networks (texts and images created by users); classification of authors by demographic classes (age, gender, language, country) according to the characteristics of the content they created.

The analysis of the texts published by the authors is used for specific and unambiguous identification of the authorship or definition of the category of the author's profile. Specific identification of content authorship is gaining popularity. Identification and coverage of various target groups that actively use social networks is part of the toolkit of political influence. Developing systems capable of automatically obtaining this information is of considerable interest. Improves the accuracy of predicting feature usage profiles based on content: words and n-grams of words with the highest frequency of use, part-of-speech tags, and symbols. Various machine learning models are offered to determine the target category (logistic regression, decision tree, k nearest neighbors, support vector machine, naive Bayes, neural networks, and random forest). Various types of signs are used (service words, n-grams of letters), leading to many stylistic markers.

Author profile identification technologies are used to analyze texts created in English and other national languages. Profiling of authorship of texts written in Indian and Persian languages is developing.

The question of the effectiveness of the models of unique identification of authorship or categorical profiling of authors is a key issue during their experimental testing based on processing data from social networks.

3. Results

3.1. A synergistic paradigm of authorship profiling by topic

Censor.NET blogs are heterogeneous by thematic (Komova, 2020). They are multi-disciplinary by topics since bloggers consider various political, military-political, socioeconomic, and humanitarian problems in their relationship, in a cause-and-effect aspect. The topics of the posts can relate to both

the blogger's professional sphere and current socio-political and socioeconomic issues. A systematic, integrated approach to the coverage of topics is a common feature of the posts. Only some blogs can be conditionally classified as industry blogs.

The study of the documentary information of the blogosphere, created on the Censor.NET blog platform, with the help of postulated extrapolation, involves the formation of a synergistic paradigm of authorship profiling in three projections:

- topic of posting → professional group;
- topic with the highest productivity → professional group;
- professional group → topic of posting.

3.1.1. Posting topic → professional group

The general thematic direction of the Censor.NET blogosphere is established by identifying the topics of the posts, grouping them by subject areas, and generalizing and systematizing topics. The accepted selection criteria ensure equality of conditions and comparability of the research results: the object of the research is the five last posts of the top-5 bloggers with the highest number of posts from each group. Thus, the topic of the posts was studied by 25 posts from each of the 16 groups of bloggers. In total, the topics of 400 posts were explored. The systematization of the obtained results allows for determining the general subject area profile of the Censor.NET blogosphere, the thematic structure of the posts of top bloggers, and the ranked representativeness of topics in top bloggers' posting (as of July 1, 2019).

The thematic range of posts shows that the general subject area profile of the blogosphere on Censor.NET is defined by posts on military-political, security, political, legal, socioeconomic, and humanitarian issues.

Dominant positions in the blogosphere Censor.NET, by quantitative indicators, occupies five industry topics that are most actively discussed in the blogosphere (covered in more than 80% of posts):

- "The Russian-Ukrainian war" is covered in 88 posts (22%) of the vast majority of top blogger groups: government officials at various levels, military personnel, cultural figures, journalists, affiliated experts, politicians, political scientists, personalities, foreign bloggers; most frequently discussed topics: annexation of Crimea, occupation, and deoccupation of Donbas; personnel of the Armed Forces (soldiers, volunteers, volunteers, captured soldiers, sailors, awards, memory of fallen soldiers);
- "Political system" is highlighted in 77 posts (19%) of the following groups of top bloggers: deputies of various levels, government officials of various levels, educators and scientists, cultural figures, unaffiliated experts, political scientists, clergy; most frequently discussed topics: activity of political parties, public and political organizations; church activities (local church, tomos, inter-denominational conflicts);
- "Law enforcement system" is highlighted in 68 posts (17%) of the following groups of top bloggers: deputies of various levels, judges and lawyers, entrepreneurs and bank employees, educators and scientists, journalists, and politicians; most frequently discussed topics: anti-corruption (political and economic corruption, electronic declaration); reform of the judicial system;
- "Humanitarian, information sphere" is highlighted in 56 posts (14%) of the following groups of top bloggers: military personnel, educators and scientists, affiliated experts, non-affiliated experts, and clergy; most frequently discussed topics: national identity, preservation of cultural heritage (language policy of the state, historical memory, genocide, Holodomor, UPA, worldview formation, self-awareness, self-affirmation, psychological portrait of the nation), information security (information war, hybrid war, decommunization, anti-Ukrainian propaganda, expansion of the "Russian world");
- "Economy" is covered in 36 posts (9%) of the following groups of top bloggers: judges, lawyers; entrepreneurs, bank employees; doctors; most frequently discussed topics: economic reform (development of industries and transport, land reform, investment policy, informatization, electronic government).

The most popular topics are health care (25 posts), social sphere (13 posts), Ukraine and the world (12 posts), foreign policy (10 posts), Armed Forces (9 posts), education, and science (6 posts) (Fig. 2).

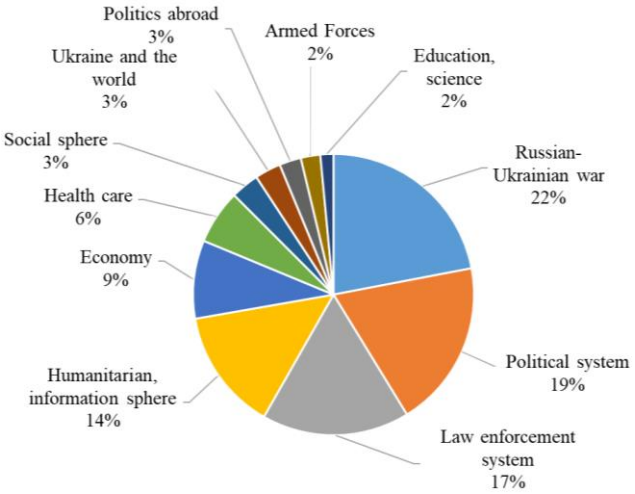

Figure 2: Thematic structure of top bloggers' posts on Censor.NET

Comparing the topics of posts with the highest productivity and the group of top bloggers allows us to visualize their relationship in the synergistic paradigm of authorship profiling in the projection "posting topic → professional group" (Fig. 3a–3d).

The productivity of posting on political topics is divided between categories of bloggers almost in half. It is dominant for political scientists, clergy, government officials, deputies, unaffiliated experts, cultural figures, educators, and scientists and marginal for foreign bloggers, affiliated experts, journalists, military personnel, politicians, entrepreneurs, lawyers, and personalities. Medical workers are not interested in politics.

Posting on economic topics demonstrates that economic issues are relevant for two categories: entrepreneurs and lawyers. Medical workers, government officials, foreign bloggers, politicians, political scientists, affiliated experts, journalists, deputies, and cultural figures show little interest in covering economic issues. Unaffiliated experts, military personnel, clergy, educators, scientists, and personalities do not create posts on economic topics (Fig. 3a).

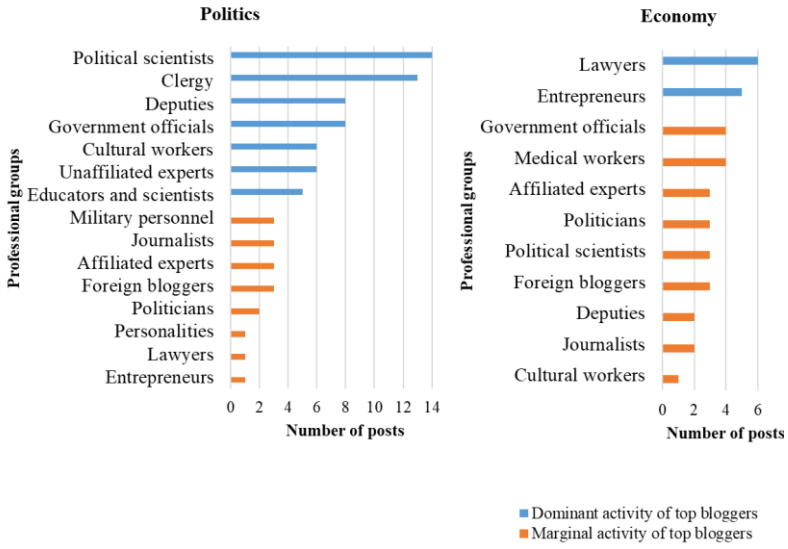

Figure 3a: Authorship profiling in the "posting topic → professional group" projection

Posting on humanitarian topics is a priority for a relatively small group of bloggers: unaffiliated and affiliated experts, clergy, and military personnel. At the same time, cultural figures, politicians, journalists, entrepreneurs, foreign bloggers, lawyers, educators and scientists, deputies, personalities, and political scientists create a few humanitarian posts. Government officials and medical workers show no interest in this topic. The activities of the law enforcement system are highlighted in the posts of all

categories of bloggers, except for the clergy and medical workers. This topic is dominant among lawyers, entrepreneurs, politicians, deputies, journalists, educators, scientists, and government officials.

The topic of the activities of the law enforcement system is marginal among foreign bloggers, unaffiliated and affiliated experts, cultural figures, personalities, political scientists, and military personnel. Marginal positions are occupied by posts on general issues of the development of the Armed Forces of Ukraine by foreign bloggers, journalists, politicians, unaffiliated experts, entrepreneurs, lawyers, and deputies (Fig. 3b).

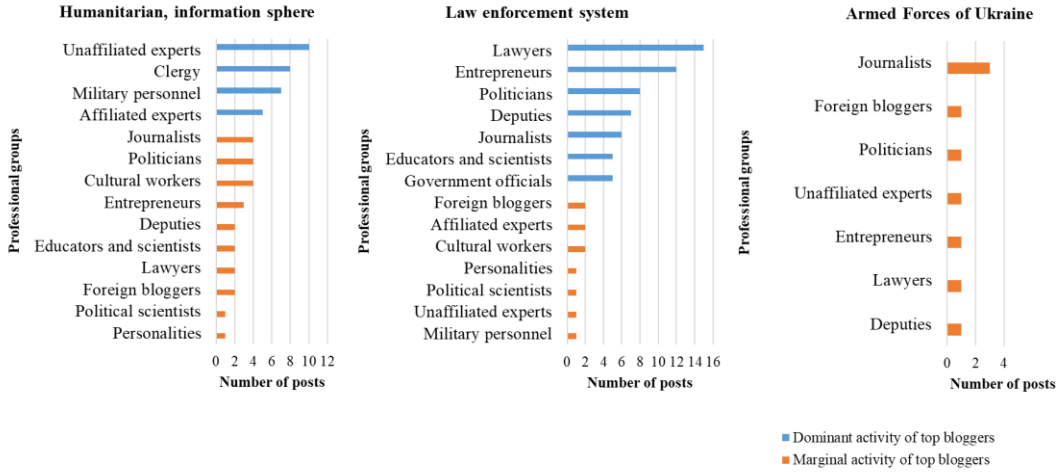

Figure 3b: Authorship profiling in the "posting topic → professional group" projection

The topic of the Russian-Ukrainian war worries all categories of bloggers except for medical workers. It is a dominant topic for most bloggers: personalities, military personnel, foreign bloggers, cultural figures, government officials, unaffiliated and affiliated experts, political scientists, and politicians. This topic occupies marginal positions among the clergy, journalists, entrepreneurs, educators and scientists, deputies, and lawyers.

The topic of the Russian-Ukrainian war worries all categories of bloggers except for medical workers. It is a dominant topic for most bloggers: personalities, military personnel, foreign bloggers, cultural figures, government officials, unaffiliated and affiliated experts, political scientists, and politicians. This topic occupies marginal positions among the clergy, journalists, entrepreneurs, educators and scientists, deputies, and lawyers (Fig. 3c).

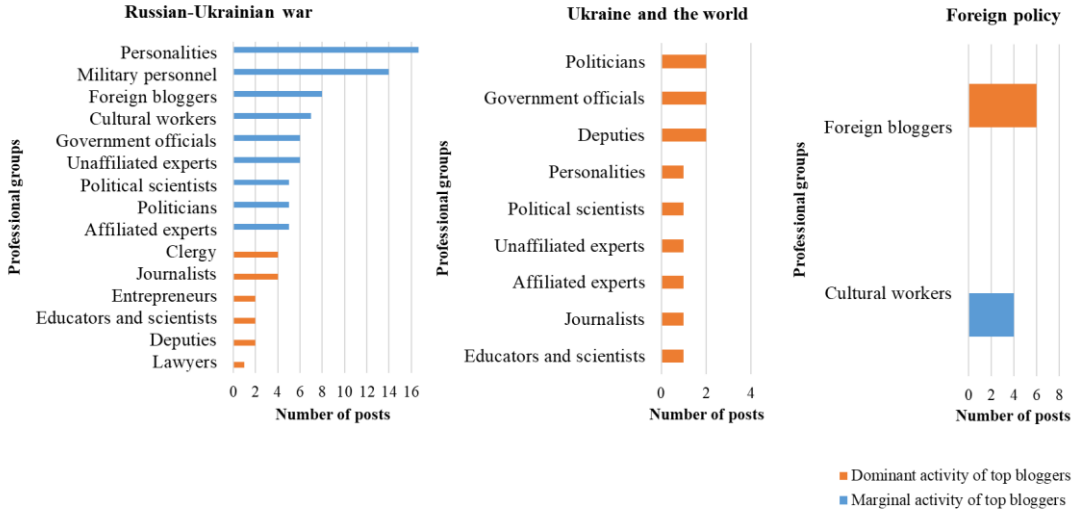

Figure 3c: Authorship profiling in the "posting topic → professional group" projection

Ukrainian society's educational, scientific, medical, and social spheres occupy marginal positions in bloggers' posts. Issues of education and science are covered only by educators and scientists, journalists, cultural figures, and deputies. The number of posts is minimal. Concerns about the provision of health care are expressed by medical workers, which create a significant number of posts (17 posts out of 25 investigated). Personalities, affiliated experts, educators, scientists, and journalists also write on popular medical topics.

The social sphere is also reflected in the posts of an extremely limited circle of bloggers. Affiliated experts, medical workers, lawyers, educators, scientists, and entrepreneurs create 1-4 posts from 25 researched within each category (Fig. 3d).

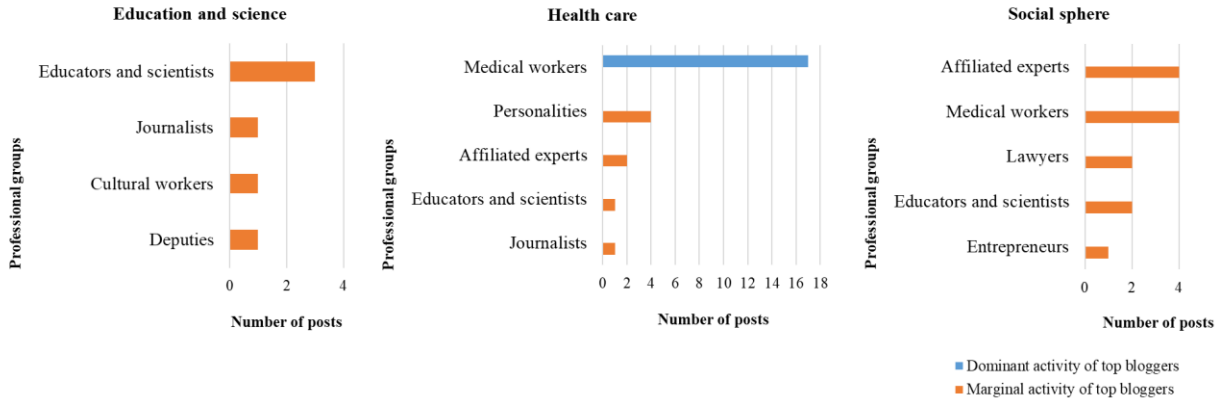

Figure 3d: Authorship profiling in the "posting topic → professional group" projection

3.1.2. Posting topic → professional group

The topics with the highest posting productivity fully represent modern public discourse: 44 posts – anti-corruption; 39 posts – annexation of Crimea, occupation, and deoccupation of Donbas; 38 posts – activities of political parties; 25 posts – economic reforms; 24 posts – national identity (Fig. 4).

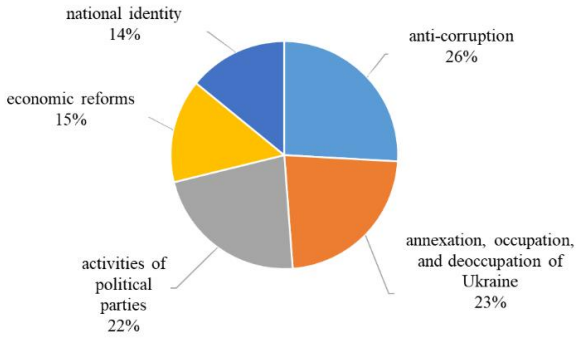

Figure 4: Topics with the highest posting in Censor.NET

Significantly, posts on topics with the highest posting productivity are created by bloggers from the vast majority of professional groups. These bloggers belong to groups 9-13. Entrepreneurs and bank employees most actively consider the topic "Anti-corruption". The topic "Annexation of Crimea, occupation and deoccupation of Donbas" is most actively considered by a group of personalities, which are bloggers who did not specify information about their professional affiliation or belonging to civil society. The topic "Activity of political parties" is most actively considered by political scientists. The topic of "Economic reforms" is most actively considered by medical workers. The topic "National identity" is most actively considered by clergy representatives.

A comparison of such indicators as the distribution and dominance of a particular topic in the posts of top bloggers reveals a pattern: the most common topics are also those that are most often included in the dominant group (Fig. 5).

The priority group in terms of spread and dominance includes the following topics: Russian-Ukrainian war (spread – 15 / dominance – 8), politics (15/7), law enforcement system (14/6), humanitarian and information sphere (12/5), economy (11/1). This observation proves the chosen methodology's correctness and the conclusions' validity.

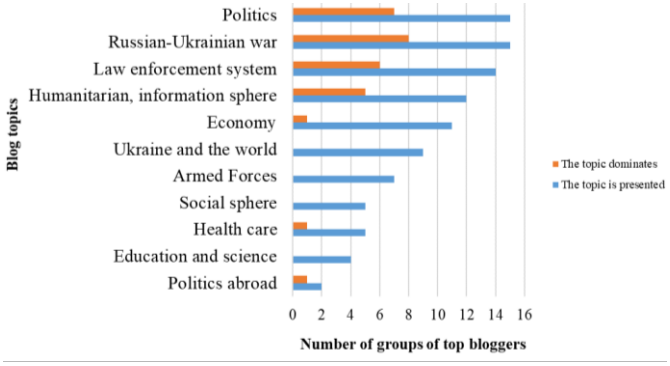

Figure 5: Spread and dominance of topics in groups of top bloggers

Comparing topics with the highest productivity and a group of top bloggers allows us to visualize their relationship in the synergistic paradigm of authorship profiling (Fig. 6).

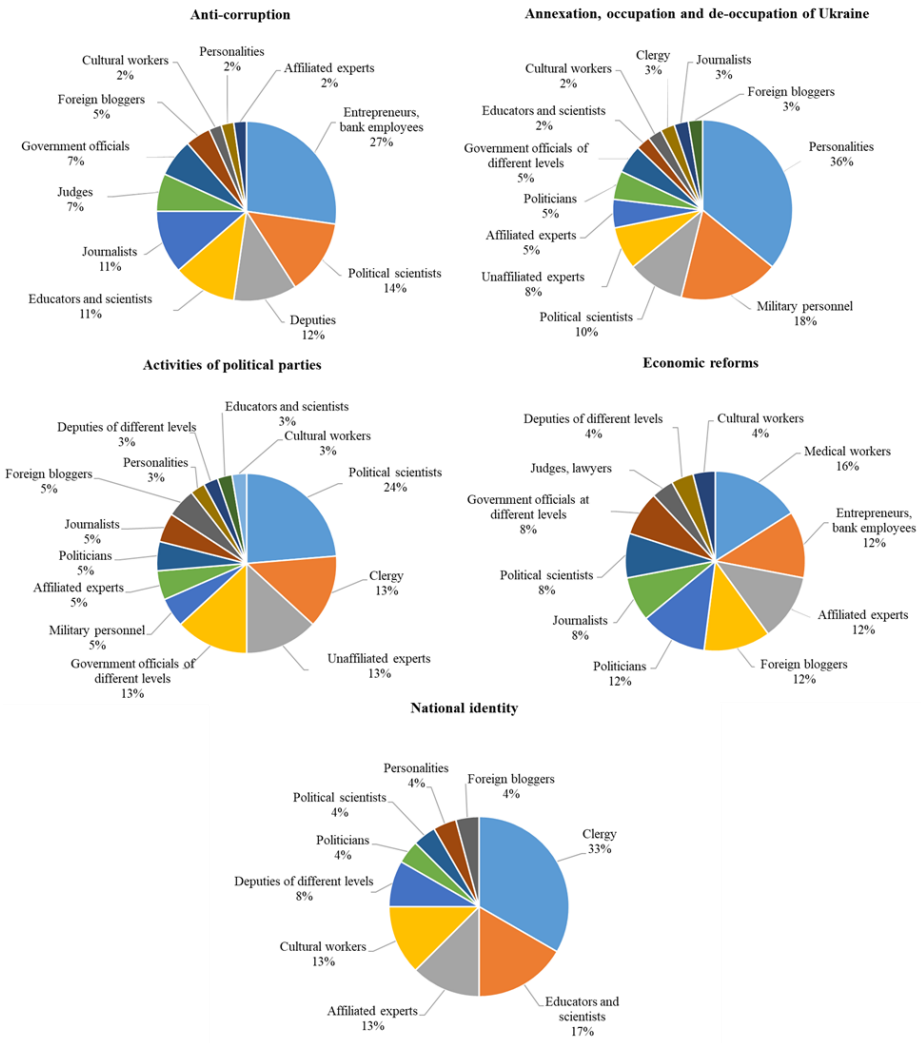

Figure 6: Authorship profiling by topics in the projection "topic with the highest productivity → professional group"

3.1.3. Posting topic → professional group

A synergistic paradigm of authorship profiling in the "professional group → topic of posting" projection is formed by classifying posts according to the thematic distribution of posts by top bloggers of Censor. NET. We determine the following quantitative and qualitative indicators within each of the 16 groups of top bloggers:

- number of posts covering a particular topic;
- dominant and marginal themes;
- correlation of indicators of distribution and dominance of topics in posts by groups of top bloggers.

This information forms the basis of a synergistic paradigm. Different thematic focuses characterize the content of different professional groups of top bloggers.

The synergistic paradigm of authorship profiling in the projection "professional group → topic of posting" shows the regularity that each of the groups of bloggers has 1-3 dominant topics, and the rest of the topics are on the margins (Fig. 7a–7f).

For deputies, the dominant topics are the political and law enforcement systems, and marginal topics are Ukraine and the world, the Russian-Ukrainian war, the humanitarian and information sphere, the economy, Armed Forces, education, and science.For government officials, the dominant topics are the political and law enforcement systems and the Russian-Ukrainian war, and marginal topics are the economy, Ukraine, and the world. For judges and lawyers, the dominant topics are the law enforcement system and the economy, and marginal topics are the social sphere, the Russian-Ukrainian war, the Armed Forces, and the political system (Fig. 7a).

Figure 7a: Authorship profiling in the “professional group → posting topic” projection

The dominant topics for bank employees and entrepreneurs are the law enforcement system and the economy. The marginal topics for this bloggers' group are the Russian-Ukrainian war, the humanitarian and information sphere, the economy, Armed Forces, the social sphere, and the political system. For medical workers, the dominant topic is health care, and marginal topics are the economy and social sphere (Fig. 7b).

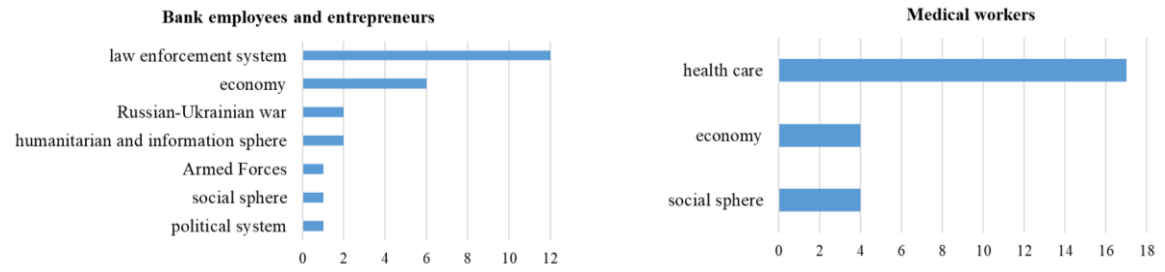

Figure 7b: Authorship profiling in the “professional group → posting topic” projection

For military personnel, the dominant topics are the Russian-Ukrainian war, the humanitarian and information spheres, and the marginal topics are political and law enforcement systems and Armed Forces. The dominant topics for foreign bloggers are the Russian-Ukrainian war and foreign policy. The marginal topics of this bloggers' group are the economy, political and law enforcement systems, humanitarian and informational sphere, and Armed Forces. For personalities, the dominant topics are

the Russian-Ukrainian war, and the marginal topics are health care, Ukraine and the world, political and law enforcement systems, humanitarian and information sphere (Fig. 7c).

Figure 7c: Authorship profiling in the "professional group → posting topic" projection

For educators and scientists, the dominant topics are the humanitarian and information spheres and political and law enforcement systems. The marginal topics for educators and scientists are education and science, the Russian-Ukrainian war, the social sphere, Ukraine and the world, and health care. For cultural figures, the dominant topics are the Russian-Ukrainian war and the political system. The marginal topics for cultural figures are the humanitarian and information sphere, foreign policy, law enforcement system, education and science, and economy (Fig. 7d).

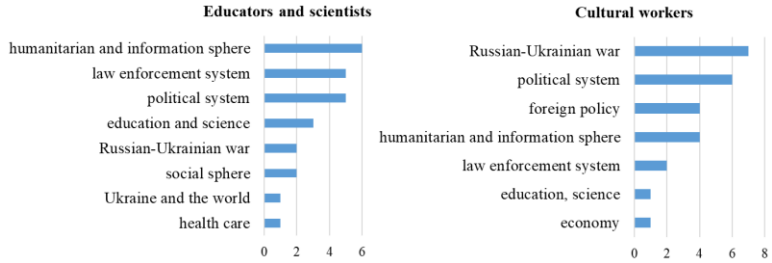

Figure 7d: Authorship profiling in the "professional group → posting topic" projection

Russian-Ukrainian war and the humanitarian and information spheres. The topics of the social sphere, economy, political and law enforcement systems, health care, Ukraine and the world are marginal for affiliated expert. The dominant topics for unaffiliated experts are the humanitarian and information sphere, the Russian-Ukrainian war, and the political system. The marginal topics for this group of bliggers are Ukraine and the world, the Armed Forces, and the law enforcement system. For journalists, the dominant topic is the law enforcement system. The marginal topics for journalists are the Russian-Ukrainian war, the humanitarian and information sphere, the Armed Forces, the political system, the economy, Ukraine and the world, health care, education and science (Fig. 7e).

Figure 7e: Authorship profiling in the "professional group → posting topic" projection

For politicians and activists, the dominant topics are the law enforcement system and the Russian-Ukrainian war. The politicians' marginal topics are the humanitarian and information sphere, economy, Ukraine and the world, political system, and Armed Forces. For political scientists, the dominant topics are the political system and the Russian-Ukrainian war. The marginal topics for political scientists are the economy, Ukraine and the world, the law enforcement system, humanitarian and information

sphere. For the clergy, the dominant topics are the political system, the humanitarian and information spheres, and the marginal topic is the Russian-Ukrainian war (Fig. 7f).

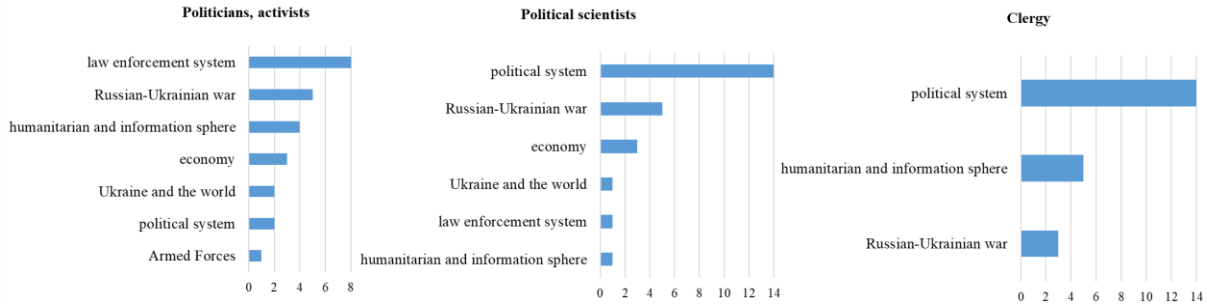

Figure 7f: Authorship profiling in the "professional group → posting topic" projection

Thus, the synergistic paradigm of authorship profiling in the projection "professional group of top bloggers → topic of posting" reflects the dominant and marginal subject area topics on which bloggers of a particular professional group write.

3.2. A synergistic paradigm of authorship profiling by author type

The Censor.NET blogosphere is a collective blog that a group of bloggers keeps according to the owner's rules. The blogosphere consists of 527 blogs of individuals and 21 blogs of various public volunteer, veteran, and environmental organizations.

These organizations are institutions of civil society:

- organizations for research and implementation of state policy: ISER (Institute for Social & Economic Research), "Police of Chernihiv region";
- human rights organizations: "Cassations and Appeals" (National Center for Human Rights);
- veteran, volunteer and paramilitary organizations: People's Project (People's Project (All-Ukrainian volunteer center)), Seni Cup (sports society for people with a physical disability), HELP (Public organization "Forpost"), Research and Production Enterprise "Temp-3000", School of military divers, Volunteer Union, Union wives and mothers of fighters ATO members, The 'Return Alive' Foundation, Victory Sisters Foundation, WWU Heart of a Warrior, National Corps, Luhansk Partisans;
- environmental organizations: WWF in Ukraine, World Wildlife Fund, Center for Environmental Initiatives Ecoaction, ECO Patrol.

Thus, the synergistic paradigm of profiling authorship in the Censor.NET blogosphere by authorship in the projection "professional group → authorship" is manifested in the complete belonging of blogs of all groups to the category of collective blogs. The bloggers kept blogs on the Censor.NET online platform according to the owner's rules.

3.3. A synergistic paradigm of authorship profiling by the level of generalization of information

During the formation of a synergistic paradigm of authorship profiling by the level of generalization of information, we identify author, monitoring, and quotation blogs (Fig. 8).

Author blogs that contain the original author's text (277 posts) predominate in the Censor.NET blogosphere. It allows the blogosphere to be an interesting source of primary information often cited, particularly during trolling [27]. The high number of author blogs confirms that the blogosphere provides a vast opportunity for self-expression, the realization of one's creative ideas, and the disclosure of the blogger's point of view.

The blogosphere testifies that the document's role is significant in modern society. Formulation of certain statements using documented argumentation is the core of monitoring blogs (108 posts). The main content in these blogs presents comments about other sites and blogs with links on them. Entrepreneurs and unaffiliated experts are the dominant authors of monitoring blogs.

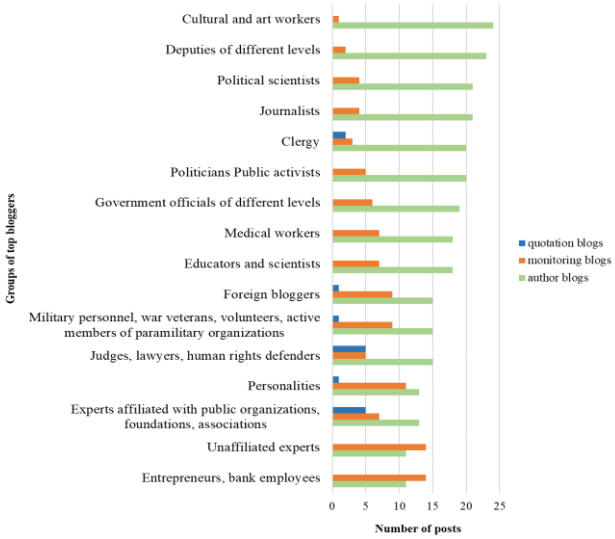

Figure 8: Authorship profiling in the "professional group → level of generalization of information" projection

The segment of quotation blogs, the main content of which is quotes from other blogs with links to them, is small (15 posts). They are represented in the following groups: judges and lawyers, military personnel, affiliated experts, clergy, and foreign bloggers. Quotation blogs are not dominant in any of the professional groups.

3.4. A synergistic paradigm of authorship profiling according to the nature of sign

The Censor.NET blogosphere vividly testifies to the creative palette of bloggers. The predominant type of sign system is a text blog, in which the main content is text (194 posts). However, the content of text blogs is widely supplemented with multimedia. The text is accompanied by photos in 169 posts, photos and videos in 17 posts.

Posts in which photos accompany the text are created by entrepreneurs and bank employees, military personnel, journalists, unaffiliated experts, political scientists, and personalities. In these groups, this type of information recording is dominant. The analyzed group of posts by top bloggers includes photoblogs (8 posts) and vlogs (12 posts). However, they are in marginal positions (Fig. 9).

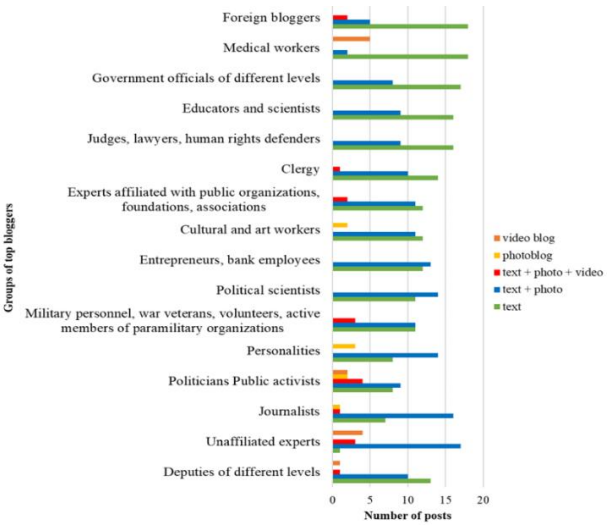

Figure 9: Authorship profiling in the “professional group → nature of iconic records” projection

A cursory look at the blogosphere shows the presence of podcasts, the main content of which are audio files (specially dictated interviews, lectures belonging to the oral genre), and music blogs, the main content of which are musical works. However, they go beyond the analyzed group of posts of top bloggers.

3.5. A synergistic paradigm of authorship profiling by post genres

During the formation of a synergistic paradigm of authorship profiling by genre of posts, a relatively wide variety of genres was revealed: diary entries, sketches, memoirs, essays, and literary works. The Censor.NET blogosphere demonstrates the active growth of blogs, which can be considered analytical sketches, the topic of which are actual facts and social events.

Out of 400 analyzed posts of top bloggers, 315 posts belong to this category. Posting in the form of analytical sketches is characteristic of all professional groups. The most active in posting analytical essays are unaffiliated experts and political scientists (25 posts each), affiliated experts, judges, and entrepreneurs (24 posts each) (Fig. 10a).

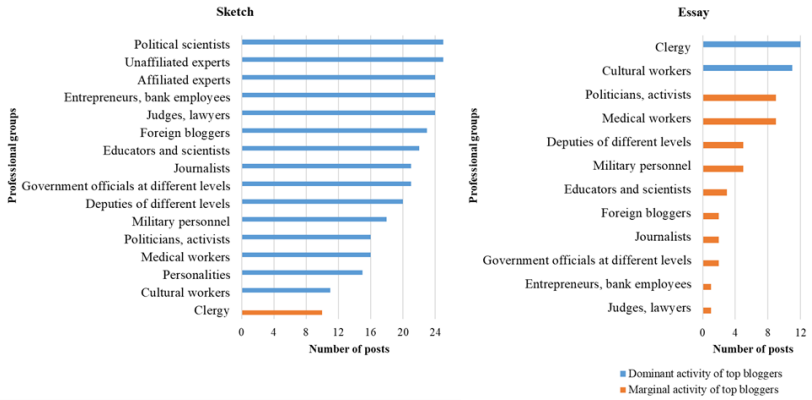

Figure 10a: Authorship profiling in the "professional group → post genre" projection

Extended, aphoristic, metaphorical essays with thoughts and emotional impressions on a specific occasion or issue are common in the blogosphere (62 posts).

Memories of bloggers with meaningful facts of their own lives in the past are on the margins. We highlight such a type as a literary blog, the content of which is a literary work (poem, play) (2 posts). Clergypersons and cultural figures mostly write essays and literary works. Memoirs and diaries are mostly written by personalities (Fig. 10b).

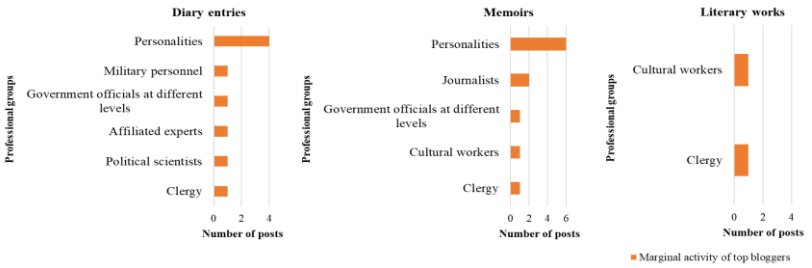

Figure 10b: Authorship profiling in the "professional group → post genre" projection

3.6. A synergistic paradigm of authorship profiling according to the nature of sign

As a result of the formation of a synergistic paradigm of authorship profiling by the emotional coloring of blog posts, it is possible to identify the state of social balance and the level of meeting the

needs of bloggers from different professional groups. The blogosphere mainly contains critical posts highlighting harmful, undesirable societal phenomena that cause negative emotions (206 posts). The most critical experts are unaffiliated experts, personalities, and military personnel (19, 18, and 17 posts, respectively). Government officials of all levels (16 posts), clergy (14 posts), foreign bloggers (13 posts), judges, and lawyers (11 posts) write positive posts that highlight desirable and necessary phenomena for society and evoke positive emotions. Neutral posts do not have pronounced positive or negative features and demonstrate the lowest representation. Posting on neutral topics is the domain of political scientists, judges, and military personnel (7 posts each) (Fig. 11).

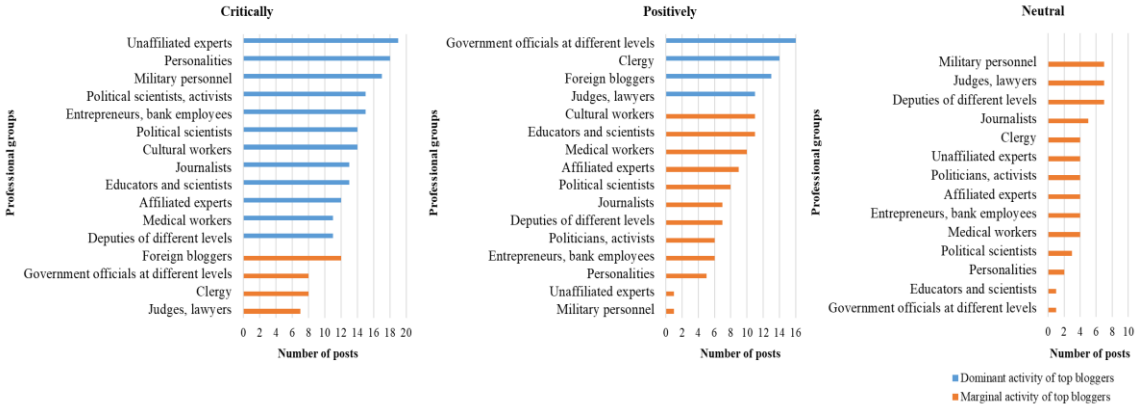

Figure 11: Authorship profiling in the "professional group → emotional coloring of posts" projection

This group of blogs also includes those that contain signs of trolling. The Censor.NET blogosphere has a substantial percentage of well-known, recognizable bloggers. An impersonal form of a troll account without a personal photo, with signs of fake accounts or intentionally provocative accounts, with minimal personal information and an insignificant number of own entries, with a predominance of reposts about "enemies", will not contribute to the achievement of the troll's goal: to draw attention to himself and provoke an emotional response with emotional arguments, to involve readers in long and useless discussions, to intensify the conflict, to put into doubt values. Therefore, on such a resource as Censor.NET, subtle forms of trolling are more successfully used, which are based on the psychology of personal and collective behavior, and exploit moral and ethical norms accepted in society. From this point of view, the group "Personalities" is of interest, although we find posts with signs of trolling in other groups.

"Connoisseur" model. Traits of trolling: brevity; indisputability; emotionlessness and unequivocal statement from the position of a "connoisseur"; manipulation of factual material; violation of the unity of the complex of objective data about the fact and its interpretation in the system "the date of the event - the place of the event - the actual presence of the event itself - persons involved in events". For example, one of the posts on the blog of M. Moskalova, containing a small amount of the author's text and two links. A discussion ensued between the author, which included numerous new references to support the argument for the original claim, and commentators to refute this claim [28].

"Speculator" model. Traits of trolling: opposition of two realities, which do not exist on the same plane, or two problems, the solution of which is not on the same field of argumentation; the use of a significant social, moral and ethical problem for leveling another issue, unimportant for the troll. Thus, in one of posts by Yu. Andreiev, the author contrasts the Independence Day parade with the need to provide for children with cancer, which is accompanied by emotional appeals to the president, and requests to the public for help [29].

"Pseudo friend" model. Traits of trolling: extensive logical reasoning; multifaceted argumentation for the affirmation of certain realities that do not correspond to the interests and beliefs of the blogger, and in the conclusions - emasculation of the core of the same reality. Thus, Yu. Kasianov examines the issue of Ukraine's exit from the Russian protectorate. In the conclusions, he highlights his views on the ideals of a true Ukrainian nation - "without flags, embroideries, torchlight processions and the law on languages, but with a deep-rooted attraction to personal freedom and state independence." Thus, it opposes individual freedom and national culture [30].

Trolling as a provocation of an emotional reaction of a specific group of readers, the public in general, is actualized in current conditions when the populism of politicians and public activists determines the electoral priorities of society when the existing base of public discourse is significantly reduced. According to the British researcher of modern media P. Pomerantsev, modern society lives in a post-factual world, when the emotional setting is important, not the facts. Emotional trolling, fueled by conspiracy theories ("as reported by well-informed sources"), has a significant impact on audience segments that do not naturally think critically [31].

3.7. A synergistic paradigm of authorship profiling by the language of posts

Blogs created on the Censor.NET platform are primarily in Ukrainian (Fig. 12).

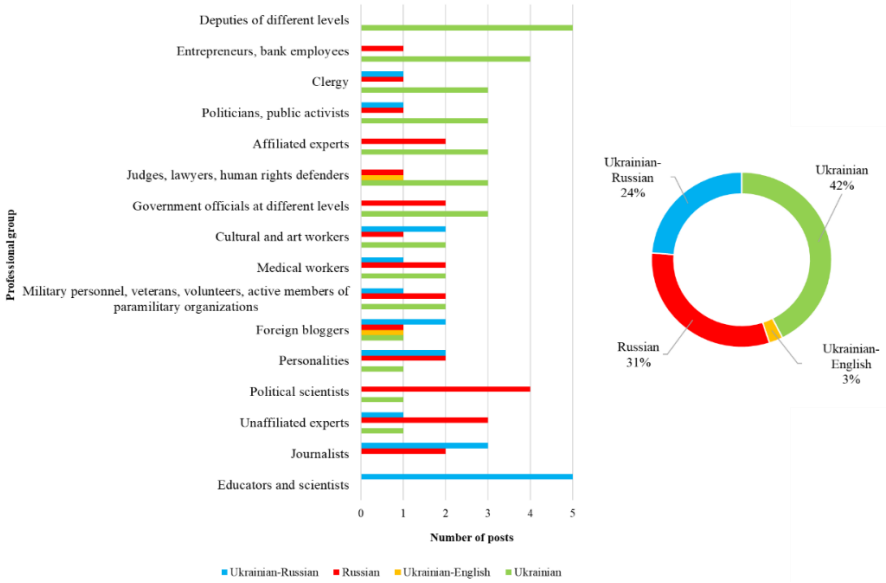

Figure 12: Authorship profiling in the "professional group → language of the post" projection

Ukrainian-language blogs constitute a shaky majority - 290 blogs (53%) against Russian-language and bilingual – 258 blogs (47%). However, general information about the linguistic characteristics of blogs, in general, does not reflect the accurate picture since the linguistic environment also depends on the indicators of posting activity. In the posts created in the group of top bloggers, there is a change in the language environment. An active group of top bloggers creates content in Ukrainian (in combination with Ukrainian-English) in 36 blogs out of 80 blogs and in Russian (in combination with Ukrainian-Russian) in 44 blogs. Since Ukrainian-language posts are scattered, we define the Russian-speaking environment as the sum of blogs written only in Russian (25 blogs) and Russian and Ukrainian (19 blogs). Ukrainian-English blogs are also few (2 blogs). Thus, in the posts of top bloggers, compared to the general indicators of Censor.NET, the percentage of Ukrainian-language content decreases from 53% to 45%, while Russian-language content, on the contrary, increases – from 47% to 55%. The Censor.NET blogosphere is mostly Russian-language content.

3.8. Consolidated synergistic paradigm of authorship profiling

The consolidated synergistic paradigm of authorship profiling is a generalized and formalized representation of the content characteristics of posts created by a specific professional group of bloggers, taking into account the following characteristics: topic, type of authorship, level of generalization of information, the nature of sigh records, the genre of posts, emotional coloring of posts, language of posts (table 1).

Table 1
Consolidated synergistic paradigm of authorship profiling in Censor.NET

Group of bloggers	Topic	Authorship	The level of generalization of information	Sign system	Genre	Emotional coloring	Language
Deputies	Political system, law enforcement system	Collective blog	Author	Text	Sketch	Critically	UA
Government officials	Political system, Russian-Ukrainian war	Collective blog	Author	Text	Sketch	Positively	UA
Lawyers	Law enforcement system, economy	Collective blog	Author	Text	Sketch	Positively	UA
Entrepreneurs	Law enforcement system, economy	Collective blog	Monitoring	Text, text + photo	Sketch	Critically	UA
Military personnel	Russian-Ukrainian war, humanitarian and information sphere	Collective blog	Author	Text, text + photo	Sketch	Critically	UA& RU
Educators, scientists	Humanitarian and information sphere, political system, the law enforcement system	Collective blog	Author	Text	Sketch	Critically	UA or RU
Medical workers	Health care, social sphere, economy	Collective blog	Author	Text	Sketch	Critically	UA& RU
Cultural workers	Russian-Ukrainian war, political system	Collective blog	Author	Text	Sketch	Critically	UA& {UA or RU}
Journalists	Law enforcement system, humanitarian and information sphere, Russian-Ukrainian war	Collective blog	Author	Tex + photo	Essay	Critically	UA& RU
Affiliated experts	Russian-Ukrainian war, humanitarian and information sphere	Collective blog	Author	Text	Sketch	Critically	UA
Unaffiliated experts	Humanitarian and information sphere, political system, Russian-Ukrainian war	Collective blog	Monitoring	Text + photo	Sketch	Critically	RU

Group of bloggers	Topic	Authorship	The level of generalization of information	Sign system	Genre	Emotional coloring	Language
Politicians	Law enforcement system, Russian-Ukrainian war	Collective blog	Author	Text	Sketch	Critically	UA
Political scientists	Political system, Russian-Ukrainian war	Collective blog	Author	Text + photo	Sketch	Critically	RU

4. Conclusion

The consolidated synergistic paradigm of authorship profiling is a communication model of the correlation between the bloggers' picture of the world and the content of blogs. Applying the postulated extrapolation method to create new knowledge allows us to consider the interdependence in the system "bloggers' picture of the world - content" as a result of the synergy of all socio-communication factors. We consider the phenomenon of the synergistic paradigm of authorship profiling as a formalized detailing of information interaction. The concept of the synergistic paradigm of authorship profiling is included in the context of the interpretation of information interaction as a complex, open system, which is characterized by structural, functional, coordination, channel, and semantic manifestations.

In the synergistic paradigm of authorship profiling, the correlation of such a component of the bloggers' picture of the world as the professional affiliation of bloggers with the characteristics of their blogs is modeled. Each professional group of bloggers has its model of a consolidated synergistic paradigm of authorship profiling, which summarizes characteristics from all profiling features.

Thus, the synergistic paradigm of authorship profiling shows that the topic of state security and national identity is a defining marker in the blogs of most professional groups. The popularity of political and humanitarian topics testifies to the relevance but unrealization of political expectations, social ideals, and the hidden resources and potential of social activity. The absence among the priority posts of economic, educational, scientific, scientific, and technical topics indicates the exclusion of the citizen, and the general public from achieving sustainable development of society, economic growth, improving welfare and social protection, and raising socio-economic standards of living. In line with applying a special method of postulated extrapolation of documentary information, using the tools of content analysis, lexical-semantic analysis, a generalized axiological and worldview portrait of bloggers was modeled by studying marked vocabulary. The technologies of using marked vocabulary in social networks have become widespread due to the massive potential to influence blogger-leaders on other users' worldviews and behavioral positions. The identification of marked vocabulary makes it possible to identify trends in social attitudes, systemic connections between external (behavioral) and internal (cognitive) patterns, and the role of social networks in their modeling.

The basis for building a model of a synergistic paradigm of authorship profiling can be the dependencies between any other component of the world picture (gender affiliation, educational level, religious, political, and cultural preferences) and the features of the content, as well as the publication activity of bloggers, feedback characteristics with bloggers.

5. References

[1] W. Aljedaani, I. Abuhaimed, F. Rustam, M. W. Mkaouer, A. Ouni, and I. Jenhani, Automatically detecting and understanding the perception of COVID-19 vaccination: a middle east case study, Social Network Analysis and Mining 12(1) (2022). doi: 10.1007/s13278-022-00946-0.

[2] M. S. Viñán-Ludeña and L. M. de Campos, Discovering a tourism destination with social media data: BERT-based sentiment analysis, Journal of Hospitality and Tourism Technology 13(5) (2022) 907–921. doi: 10.1108/JHTT-09-2021-0259.
[3] 3. M. N. Omri, F. Fkih, Dynamic Editing Distance-based Extracting Relevant Information Approach from Social Networks, *IJCNIS* 4 (6) (2022) 1–13. doi: 10.5815/ijcnis.2022.06.01.
[4] 4. S. K. Kiran, M. Shashi, K. B. Madhuri, Multi-stage Transfer Learning for Fake News Detection Using AWD-LSTM Network, *IJITCS* 14 (5) (2022) 58–69. doi: 10.5815/ijitcs.2022.05.05.
[5] 5. I. Fadhli, L. Hlaoua, M. N. Omri, Sentiment Analysis CSAM Model to Discover Pertinent Conversations in Twitter Microblogs, *IJCNIS* 14 (5) (2022) 28–46. doi: 10.5815/ijcnis.2022.05.03.
[6] 6. G. Mostafa, I. Ahmed, M. S. Junayed, Investigation of Different Machine Learning Algorithms to Determine Human Sentiment Using Twitter Data, *IJITCS* 13 (2) (2021) 38–48. doi: 10.5815/ijitcs.2021.02.04.
[7] O. Mabrouk, L. Hlaoua, M. N. Omri, Exploiting ontology information in fuzzy SVM social media profile classification, Applied Intelligence 51(6) (2021) 3757–3774. doi: 10.1007/s10489-020-01939-2.
[8] C. Ikae, J. Savoy, Gender identification on Twitter, Journal of the Association for Information Science and Technology 73(1) (2022) 58–69. doi: 10.1002/asi.24541.
[9] N. Pasieka, V. Sheketa, M. Kulynych, Yu. Romanyshyn, S. Chupakhina, and O. Khytrova, Machine Training for Intelligent Analysis of Text for the Identification of the Author, CEUR Workshop Proceedings 3187 (2021) 58–68. URL: https://ceur-ws.org/Vol-3187/paper6.pdf.
[10] R. Agarwal, S. Gupta, and N. Chatterjee, Profiling Fake News Spreaders on Twitter: A Clickbait and Linguistic Feature Based Scheme, in: P. Rosso, V. Basile, R. Martínez, E. Métais, and F. Meziane (Eds.), Natural Language Processing and Information Systems, volume 13286 of Lecture Notes in Computer Science. Springer, Cham, 2022, pp. 345–357. doi: 10.1007/978-3-031-08473-7_32.
[11] X. Du, S. Chen, Z. Liu, and J. Wang, Multiple userids identification with deep learning, Expert Systems with Applications 207 (2022). doi: 10.1016/j.eswa.2022.117924.
[12] C. Suman, R. Kumar, S. Saha, and P. Bhattacharyya, 'Authorship Attribution Using Capsule-Based Fusion Approach', in: E. Métais, F. Meziane, H. Horacek, and E. Kapetanios (Eds.), Natural Language Processing and Information Systems, volume 12801 of Lecture Notes in Computer Science. Springer, Cham, 2021, pp. 289–300. doi: 10.1007/978-3-030-80599-9_26.
[13] A. Dhar et al., Author Identification from Literary Articles with Visual Features: A Case Study with Bangla Documents, Future Internet 14(10) (2022). doi: 10.3390/fi14100272.
[14] S. P. Muñoz, C. Oliva, L. F. Lago-Fernández, and D. Arroyo, Advancing the Use of Information Compression Distances in Authorship Attribution, in: F. Spezzano, A. Amaral, D. Ceolin, L. Fazio, and E. Serra (Eds.), Disinformation in Open Online Media, Lecture Notes in Computer Science. Springer-VerlagBerlin, Heidelberg, 2022, pp. 114–122. doi: 10.1007/978-3-031-18253-2_8.
[15] A. Majkowska, K. Migdał-Najman, K. Najman, and K. Raca, Graphic characters as Twitter age group identifiers, in: Modern classification and data analysis: methodology and applications to micro- and macroeconomic problems, Studies in classification, data analysis, and knowledge organization. Springer,Cham, 2022, pp. 275–288. doi: 10.1007/978-3-031-10190-8.
[16] C. Suman, R. S. Chaudhary, S. Saha, and P. Bhattacharyya, An attention based multi-modal gender identification system for social media users, Multimed Tools Appl 81(19) (2022) 27033–27055. doi: 10.1007/s11042-021-11256-6.
[17] M. Komova, Method of postulative extrapolation of documental information in the blogosphere', Library Science. Record Studies. Informology 1 (2021) 50–57. doi: 10.32461/2409-9805.1.2021.229855.
[18] M. Komova, Thematic filling of the blogosphere: dynamics and perspectives, Scientific notes of V. I. Vernadsky Taurida National University. Series: Philology. Journalism 2 (2020) 127–136. doi: https://doi.org/10.32838/2663-6069/2020.2-4/20.

[19] C. Suman, A. Singh, S. Saha, and P. Bhattacharyya, Development of Multi-task Models for Emotion-Aware Gender Prediction, in: Proceedings of 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 2022, pp. 1–8. doi: 10.1109/IJCNN55064.2022.9892404.
[20] S. S. Carrasco and R. C. Rosillo, LosCalis at PoliticEs 2022: Political Author Profiling using BETO and MarIA, CEUR Workshop Proceedings 3202 (2022).URL: https://ceur-ws.org/Vol-3202/politices-paper1.pdf.
[21] A. Manea and L. P. Dinu, UniRetro at PoliticEs@IberLef 2022: Political Ideology Profiling using Language Models, CEUR Workshop Proceedings 3202 (2022). URL: https://ceur-ws.org/Vol-3202/politices-paper8.pdf
[22] V. S. Reddy, V. K. Prasad, D. N. M. Rao, and S. C. Satapathy, Intelligent Systems and Sustainable Computing: Proceedings of ICISSC 2021. Springer Nature, 2022.
[23] L. Pomytkina, Y. Podkopaieva, K. Hordiienko, Peculiarities of Manifestation of Student Youth' Roles and Positions in the Cyberbullying Process, IJMECS 13 (6) (2021) 1–10. doi: 10.5815/ijmecs.2021.06.01.
[24] Z. Avkurova, S. Gnatyuk, B. Abduraimova, S. Fedushko, Y. Syerov, and O. Trach, Models for early web-attacks detection and intruders identification based on fuzzy logic. Procedia Computer Science, vol. 198, pp. 694–699, Jan. 2022, doi: 10.1016/j.procs.2021.12.308.
[25] A. Patergianakis and K. Limniotis, Privacy Issues in Stylometric Methods, Cryptography 6(2) (2022). doi: 10.3390/cryptography6020017.
[26] A. G. J. Lan and I. Paraboni, Text- and author-dependent moral foundations classification, New Review of Hypermedia and Multimedia 28(1–2) (2022) 18–38. doi: 10.1080/13614568.2022.2092655.
[27] Information war against Ukraine. Identification of Russian "trolls", hostile sites and groups - RESEARCH!', Carthago delenda est (c), 2016. URL: https://streebogblog.wordpress.com/2016/10/04/інформаційна-війна-проти-україни-іде/.
[28] M. Moskalova, The prohibition to open fire was given by Viktor Muzhenko in November 2015 [Zapriet otkryvat ohon dal Viktor Muzhenko v noiabrie 2015], Censor.NET, 2019. URL: https://censor.net/ru/blogs/3131550/zapret_otkryvat_ogon_dal_viktor_mujenko_v_noyabre_2015.
[29] Y. Andreiev, Volodymyr Groysman promises to find money for the parade. At this time, children with cancer die without treatment [Volodymyr Hroisman obitsiaie znaity hroshi na parad. V tsei chas pomyraiut onkokhvori dity bez likuvannia], Censor.NET, 2019. URL: https://censor.net/ru/blogs/3137185/volodimir_groyisman_obtsya_znayiti_grosh_na_parad_v_tseyi_chas_pomirayut_onkohvor_dti_bez_lkuvannya.
[30] Y. Kasianov, Give up or fight [Voievat nielzia sdavatsia], 2019. URL: https://censor.net/ru/blogs/3131417/voevat_nelzya_sdavatsya.
[31] P. Pomerantsev, Nothing is True and Everything is Possible, Reprint edition. PublicAffairs, 2015.