Research of the estimated emotional components for the content analysis

A Iskhakova\textsuperscript{1,2}, A Iskhakov\textsuperscript{1} and R Meshcheryakov\textsuperscript{1}

\textsuperscript{1} Cyberphysical systems laboratory, V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, 65 Profsoyuznaya street, Moscow 117997, Russia
\textsuperscript{2} Information systems security department, Tomsk State University of Control Systems and Radioelectronics, 40 prospekt Lenina, Tomsk, 634050, Russia

E-mail: shumskaya.ao@gmail.com

Abstract. The research focuses on the automation of intellectual processing of Internet content for the timely detection and blocking of content operating destructively on the user’s mind. It deals with subjects of current interest (such as pressure upon mentality of the user, incitation, cyberbullying, incitement of the masses, etc.) and connects several scientific directions. This article shows results of the research in the field of text-based Internet content analysis. Numerical indicators of the estimated emotional components for content with signs of aggression and mental pressure upon the user are shown in work. The obtained results of the lexical analysis of electronic texts demonstrate the presence of a natural increase in the proportion of use of negatively coloured lexemes in the texts with signs of aggression. Also, the structure of the socio-cyberphysical system of analyzing Internet content with a description of key blocks and elements is presented.

1. Introduction

Virtual ways of communication and informatization are actively developing in the last decades. Today it is possible to say that they are not only widespread but also replaced other forms of communication and information exchange in many spheres of life. According to the statistical data of analytical agency «We Are Social» for January 2018, more than 4 billion users, that is 53 \% of all population of the Earth, are using the Internet in one or another way.

Development of virtual ways of communication, popularization of social networks and other forms of communication through the Internet services practically move social interaction on the other level. Thereby, eventually it becomes more noticeable and indisputable for each person. Virtual tools allow us to considerably simplify a lot of processes in our society. The problem of the use of means of communication to influence the society is not new; however, the Internet provides excessively wide opportunities for that process. For many users, it is the most priority platform for communication, a tool providing the possibility to receive up-to-date information on various phenomena and events. In this connection, the risks of Internet misuse by malefactors are increasingly dangerous.

The increasing level of communication and informing via text allows a malefactor to use vulnerability and trustfulness of a person to achieve improper purposes: from intrusive spam and propagation of unreliable information to purposeful propaganda of various ideas and fraud. Anonymity is a characteristic feature of the Internet, which aggravates the extensiveness of the range of problems both for the user and for the owner of an information system used for communication.
The authors consider an actual problem of security provision of individual, group and mass consciousness in the Internet environment at the expense of the development of methods to automatically detect destructive materials that will form a basis of an embedded socio-cyberphysical system. This system will be distinguished from the classical concept by involving a person and society.

1.1. Relevance

Counteractions and reactions of the manifestations of aggression, extremism, suppression and humiliation of human dignity in the virtual environment is a relevant task today. The problem of safe use of the virtual environment by a person arises for the following reasons:

- the high popularity of various virtual portals, social networks;
- expansion of their functionality and availability to all segments of the population;
- anonymity (or an opportunity to act from another name);
- emergence of a set of fraudulent and criminal schemes.

The specified Internet space problems are widely discussed at various levels: by political and public figures, and, for example, by teachers and psychologists. The information resource centers with a possibility of psychological support for people whose interests were touched in these ways are created in many countries of the world (including ones based at the educational, public, public institutions and also on the Internet). Nevertheless, it is obvious that there is a need for predictive work in the direction for protecting an Internet environment and its “cleaning” from harmful and destructive content. The existing developments including ones that are introduced in Internet browsers and many web resources, unfortunately, have no sufficient sensitivity to such kind of content and do not show the sufficient level of efficiency. Besides, methods of formation and spread of the content are also constantly being modified and improved. These facts complicate effective search and detecting of harmful and malicious content. That is why the problem is particularly acute today.

The relevance of a subject for Russia is confirmed by its mentions in two important documents: “Bases of state policy in the field of the international information security until 2020” (2013) and “The main directions of scientific research in the field of ensuring information security” (2017). These documents describe the plan of policy, scientific researches and methods of ensuring information security of the population of the country. Therefore, the fact of mentioning and understanding at the state level of the problem of destructive impact on the population through the Internet is very significant.

The variety of research and publications in the described and related fields of both Russian and foreign scientists indicates that in the last five years a lot of studies has been conducted on this direction, and the considered subject is important for scientific community and for society in general.

1.2. The dangers of the virtual environment

The research in the field of Internet content analysis is actual last years. It is due to increasing and increasingly stronger tendency of a person to interact with virtual reality and involve a user into a virtual medium. The existing investigations are of interdisciplinary character, as the given area is socially and linguistically oriented and also have medical, mathematical as well as engineering aspects.

According to statistical investigations [1] unlawful acts in a virtual medium in different countries of the world are directly proportional to the number of Internet users which is steadily growing for the last decade as well as the degree of their involvement. The increasing functionality of the Web-resources also promotes the increase in the activity of malefactors: a modern person is not foreign to revealing personal data, trusting certain websites, mass-media, information channels, undertaking financial transactions online. The pressure upon a person frequently occurs through the influence on sensitive topics: injustice, politics, social movements, financial questions, illnesses, difficult life situations, etc. Unlike obvious crimes (theft, fraud, defamation), when a user can address law enforcement agencies to protect his or her interests, the influence on a person trough propaganda, humiliation, aggression cannot be formally observed in the most cases, so there are no authentic data on statistics of the real scale of the problem [2].

The concept of destructive impact on the identity and mentality of the user of the virtual environment
is rather wide and many-sided. In order to concretize it, it is necessary to differentiate its some components. Interaction through the Internet network does not include direct contact (a meeting of participants of communication). It means that active (physical) aggression towards the user is excluded [3], however the virtual environment allows to carry out other forms of aggression and destructive influence. Within the conducted research the concept of destructive impact through the Internet content affects the area of impact on the personality from the psychological point of view for the purpose of manipulation with his opinion, actions, imposing of the ideas and acts, deception. In addition, there is one more component of manifestation of aggression in the virtual environment that is obviously allocated – persecution, “bullying” or “cyberbullying”. This kind of negative impact is directed to insult the personality, humiliation, deliberate or accidental bringing to mental disorders. The similar phenomena which are indirectly relating to a subject of the conducted research are not considered further as subject of the analysis.

Thus, in this work the concept of destructive impact of Internet content is understood as such forms and methods of deliberately distributed materials which are designed to have an impact on opinion of the user, to change his mood and a condition, to convince of certain facts, relationships of cause and effect. It implies verbal, passive and indirect effects on the network user with elements of aggression, heightened emotionality, mood management, etc. [4]

Problems of security provision of individual, group and mass consciousness, including protection from destructive influence of information and from the use of information technologies for terrorism propaganda are selected as one of the basic directions of scientific research in the field of provision of information security in the Russian Federation (approved by the Security Council of the Russian Federation, August, 31st 2017). Owing to universal propagation and observed progress in the field of Internet technologies, the given medium is most actively used for above mentioned unlawful acts in the form of information influence on the population.

The development of the socio-cyberphysical systems that is intellectual control systems, a person along with his or her actions and knowledge being a part of them, is an actual research area. Socio-cyberphysical systems represent the technology of the future; they will unite the knowledge of a person and also an ability of faster and effective analysis, self-training, and decision-making based on the application of computing means and intellectual algorithms.

1.3. Review

The virtual medium of social interaction is widely investigated for the solution of some practical problems. The majority of known methods and algorithms in the given area are directed on revealing of certain features of the Internet content, and also on classification or characteristic of the user. Thus, the majority of investigations are applied and are directed to the attraction of an audience to promoted products, to earning a profit etc.

The review of the papers on the analysis of the Internet content in the Russian and foreign praxis showed that there are some approaches to the analysis of Web-resource content and determination of the elements negatively influencing the user. At the same time many researches and the offered methods are closely connected with the used language and are poorly transferable to other language. That can be explained by use of the certain language lexical and semantic features in each separate case.

In [5-7] there are reviews of the features of Russian-speaking texts, considering filtration of harmful content. The databases of sets of lemmas of words which are processed by means of a grammatical dictionary are considered as language signs of negative influence on a user, including aggression, suppression, extremist elements. The authors use an approach based on keywords of the special dictionary and the passive subject heading list. Advantages of such methods of a research of content can be considered simplicity of a configuration and application, but disadvantages are the high value of type I errors and the need to clarify the settings for each separate case.

In [8, 9] the authors mentioned the fact of development of an author's corpus of texts of extremist nature. This allowed them to reach efficiency more than 60% in definition of text content of extremist
orientation, including short messages were investigated. It should be noted that short texts are an important component of similar content.

The paper [10] reveals the high quality of the techniques on texts of certain subjects but simultaneously point out the problems connected to errors, the ones in the lexical analysis among others. This problem is common and is also mentioned in [11]. It is connected with the fact that insufficient study of a lexical component or use of irrelevant sets of characteristics leads to high indications of type I errors and as a result impossibility to apply a method in actual practice.

Thus, it is possible to conclude that currently the problem of the intellectual analysis of the Internet content is significant, extremely urgent and is rapidly developing. A considerable number of printed publications, including those in international conferences on these problems, is an evidence of the needs of the society for the solution of problems and challenges that have arisen along with the close implementation of the virtual world in the real life of mankind. At the same time, the development in the given area is of extensive character and is directed, first of all, on the solution of separate applied problems, commercial ones among them. In this regard, there is a necessity of integration of investigation results in the field of Internet content processing, detecting dangerous phenomena in the virtual medium, estimation of influence on users and development of software in order to create a uniform socio-cyberphysical system for data control processes.

2. Materials and methods

2.1. Approaches to the problem solution

The problems of text attribution generally refer to not only the determination of the text author but also the determination of various features of the text as well as its author. Despite a variety of such features and a considerable number of applied problems, generally, the attribution boils down to the classification of texts by one or several features. The problem of classification of e-mails is in detail considered in papers [12, 13], where special attention is given to the problem of detecting the automatically generated texts. On the basis of received in [13] calculations and regularities, it is possible to draw a conclusion on perspectivity of the approach based on detecting and calculating the values of content features that to different degrees manifest themselves for the chosen classes.

When solving the attribution problems, the text is represented in the form of a structured set of values of the selected features. The text can be defined in the form of the following form t:

\[ t = (a_1; a_2; \ldots; a_{m-1}; a_m), \]

\[ a_i = p_i(t), \]

\[ p_1, p_2, \ldots, p_{m-1}, p_m \in P, \]

where t - is a certain form of the investigated text presented in the form of values of its features;

\( p_i \) – is the feature accepted for classification of texts, which value refers to the image-vector t of the investigated text, at \( i = 1 \ldots m \);

\( a_i \) – is the value of the feature \( p_i \) calculated uniformly for the all sample, at \( i = 1 \ldots m \);

\( P \) – is the set of features of the texts used for the classification of texts;

\( m \) – is the number of features of the texts used for classification of texts, the capacity of the set P.

If we understand the investigated objects (texts) as points in the multidimensional features space, the classification problem is formulated as an object reference to one of the classes expressed by "a condensation of points" representing a homogeneous subset of objects in space.

The primary goal when using such an approach is to detect steady dependences between an available qualitative feature of the text and measured numerical values. The objective of this research consisted in the analysis of two classes of content: usual texts with neutral emotional colouring and texts with elevated emotionality with an aggression component on the examples of papers [14, 15]. As a practical experience of research show, the text characteristics which can be used for the automation of content
classification process are not always obvious differences of the text which a user or even an expert can notice. In the present paper, it is offered to carry out a lexical analysis by the estimation of the text emotionality level on an example of consideration of an evaluation component in lexemes of text content fragments.

2.2. Research on the text properties

For an estimation of the presence of negative components in a text material (Internet content) the dictionary of estimated words and expressions of Russian “RuSentiLex” (Russian Sentiment Lexicon) in the version of 2017 [16] was used. All lexical items and their values are considered in the dictionary from three points of view:

- the polarity of a word expressing an estimation component: positive, negative or neutral (attributing of pairs of polarities are also possible);
- the tonality source: directly expressed estimation, emotion or connotation;
- the voice-frequency distinctions between values of a polysemic word (if the value of the parameter changes depending on the word meaning).

To estimate the display of emotional components in materials, including sharply negative, sharply positive, aggressive, offensive etc., we analysed 100 texts with the length from 1500 to 17000 characters divided in advance into classes on neutral (usual) and with signs of aggressive presentation of information (propaganda, imposition). The comparison of the emotional component of the content was conducted on the basis of the following quantitative features characterising the presence of estimation components:

- \( d_n \) – is the general part of the lexemes the value of which has no estimation component (“neutral”):

\[
d_n = \frac{n_n}{N},
\]

where \( n_n \) – is the total of the lexemes the value of which has no estimation component (“neutral”);

\( N \) – is the number of elementary lexemes in the text.

- \( d_{pos} \), \( d_{neg} \) – is the part of the lexemes the value of which has an estimation component, “positive” and “negative” accordingly:

\[
d_{pos} = \frac{n_{pos}}{N},
\]

\[
d_{neg} = \frac{n_{neg}}{N},
\]

- \( d_e \) – is the estimation of the text emotionality level, indicating the presence of a general emotional colouring of the content, equal to the part of the lexemes which value has an estimated component:

\[
d_e = d_{neg} + d_{pos}.
\]

3. Results and discussion

The numerical values of the text features for 10 copies from each class of materials are brought in Table 1 for the set of usual texts with neutral emotional colouring (class 1) and in Table 2 for the set of texts with the elevated emotionality with an aggression component (class 2). The data are averaged over the indicator of 1000. The tables also show additional indicators: 1 - an average of elementary lexemes for 1000 characters of the text.
### Table 1. Numerical values of features for a set of usual texts with neutral emotional colouring.

| No. | Class 1 |   |   |   |   |
|-----|---------|---|---|---|---|
| 1   | 154.847 | 0.432 | 0.112 | 0.456 | 0.544 |
| 2   | 140.104 | 0.533 | 0.23 | 0.237 | 0.763 |
| 3   | 125.613 | 0.425 | 0.214 | 0.361 | 0.639 |
| 4   | 162.566 | 0.411 | 0.168 | 0.421 | 0.579 |
| 5   | 151.193 | 0.38 | 0.206 | 0.414 | 0.586 |
| 6   | 173.102 | 0.306 | 0.135 | 0.559 | 0.441 |
| 7   | 165.266 | 0.314 | 0.178 | 0.508 | 0.492 |
| 8   | 135.885 | 0.416 | 0.269 | 0.315 | 0.685 |
| 9   | 146.941 | 0.43 | 0.301 | 0.269 | 0.731 |
| 10  | 138.187 | 0.422 | 0.276 | 0.302 | 0.698 |
|     | 141.213 | 0.469 | 0.157 | 0.374 | 0.626 |

### Table 2. Numerical values of features for a set of texts with elevated emotionality with an aggression component.

| No. | Class 2 |   |   |   |   |
|-----|---------|---|---|---|---|
| 1   | 135.811 | 0.368 | 0.351 | 0.281 | 0.719 |
| 2   | 151.628 | 0.338 | 0.409 | 0.253 | 0.747 |
| 3   | 122.181 | 0.249 | 0.531 | 0.22 | 0.78 |
| 4   | 138.219 | 0.345 | 0.448 | 0.207 | 0.793 |
| 5   | 161.205 | 0.402 | 0.462 | 0.136 | 0.864 |
| 6   | 177.109 | 0.394 | 0.423 | 0.183 | 0.817 |
| 7   | 141.513 | 0.289 | 0.449 | 0.262 | 0.738 |
| 8   | 150.389 | 0.352 | 0.465 | 0.183 | 0.817 |
| 9   | 134.708 | 0.378 | 0.521 | 0.101 | 0.899 |
| 10  | 154.721 | 0.245 | 0.498 | 0.257 | 0.743 |
|     | 162.565 | 0.354 | 0.412 | 0.234 | 0.766 |

The calculations results are also presented graphically in figure 1 (a-e) in the form of box plot that is a convenient way of visually interpreting groups of numerical data through their quartiles. There are two groups of data (two classes of texts), so each diagram also has two graphics.
Figure 1. Box plots of the obtained numerical data of the studied characteristics (a-e).
As you can see in the graphic interpretation, the numerical values of several features considerably differ for the two selected classes of material. Thus, it is possible to confirm the hypothesis that the Internet content with the expressed destructive influence on the personality of a user has a higher rate of emotionally coloured lexemes. For required text materials the rate of the lexemes with the value has a “negative” estimation component is on the average 2 times as high compared to usual texts and makes 35-53%.

4. Socio-cyberphysical system for determining the destructive impact on the user
Intelligent analysis of electronic text messages and other Internet content involves close integration of the special algorithms, knowledge bases, network resources, computing power, and the user directly. A conceptual structure of a destructive content detection socio-cyberphysical system that is proposed by authors is presented in figure 2. It includes many elements and the user himself. This scheme initially explains the idea that this system, artificial in its origin, has a positive effect on a person, minimizes the negative impact coming from the Internet, and increases the efficiency of human interaction with the virtual environment.

Figure 2. The structure of the socio-cyberphysical system for Internet content analysis.
The key blocks of the proposed socio-cyberphysical system of analyzing Internet content are:

- multi-agent crawler subsystem;
- subsystem of search and primary categorization of resources;
- decision making subsystem;
- knowledge base.

One of the major key components of the system is a separate independent subsystem of search and primary categorization of resources. In fact, the given module represents a multi-agent topic-focused crawler system designed for independent web browsing. The operation of agents is based on the principles of the web crawler operation of global search systems (search of resources in a strict sequence, indexation of new pages, adding of unknown websites in a database) except for features of specialised algorithms to monitor the web space. Thus, the agents are not engaged in the content analysis but rather transmit the selected references in the data storage system, considering the separation into the corresponding categories (references to images, video series, sound files etc.). The ranging subsystem represents a set of program modules and stored procedures realised by means of a by a relational database control system. The objective of this module is to select the determination of weight (popularity) of the investigated content according to meta-characteristics of the server software and application of popular data on resource indexation. The operation of the subsystem of decision-making is based on the above-mentioned approach and also an author's methods and algorithms for detecting potentially dangerous content. As a linguistic laboratory, an interdisciplinary group of experts-analysts is understood who are capable to objectively estimate the results of cluster analysis, could supervise process key parameters (the size of time intervals of supervision, proximity measure of clusterization, the number of clusters in a cluster structure etc.) Besides, the given system implies interfaces for any interested user (realisation of plug-ins for web browsers and open API).

5. Conclusion
The development of progressive technologies in the field of the intellectual analysis of the Internet content will allow us to increase the efficiency of monitoring and operative blocking of the materials possessing signs of the destructive effect on the mentality of a person. The relevance of a problem of detecting of signs of destructive impact of the Internet user is confirmed not only by results of sociological researches, but also by a set of publications and projects, directly and indirectly affecting this problem area. It is established that a problem of identification of destructive impact on mentality, behavior and point of view of the person in the virtual environment is possible to refer to weakly formalized tasks. The main reason is fact that variety of the methods are used for generation and spread such content is too high. Besides, they are continuously being modified and combined by malefactors. The attempt to formalize and classify methods of destructive impact on Internet users with the indication of methods for automation of their definition is given in article.

In the given paper, an actual approach in the field of the intellectual analysis of electronic text-based materials for detecting of the content possessing destructive effect on the condition, mood, outlook and mind of a user is presented. There were obtained the results of the lexical analysis of electronic texts according to indicators of the existence of emotional components attesting to the presence and natural increase of the use rate of negatively coloured lexemes in required texts. The structure of an envisaged socio-cyberphysical system of the Internet content analysis with the description of key blocks and elements is presented.

6. Acknowledgments
The reported study was funded by RFBR according to the research project № 18-29-22104.

References
[1] Komarov A A 2016 Legal Science and Law Enforcement Practic 1(35) 172–177
[2] Batoev V B 2016 The bulletin of Voronezh institute of the Ministry of internal affairs of Russia 2
37–43

[3] Bykov I A, Balakhonskaya L V, Gladchenko I A and Balakhonsky V V 2018 2018 IEEE Communication Strategies in Digital Society Workshop (ComSDS) (St. Petersburg) pp 12–14

[4] Serkina N N and Lomakina E A 2015 Information security and issues of prevention of cyber extremism among young people: materials of the university conference (Magnitogorsk) pp 219-224

[5] Iskhakova A O 2016 Proceedings of Tomsk State University of Control Systems and Radioelectronics 3 65–69

[6] Zlokazov K. V. 2015 Political Linguistics Journal 1 244-251

[7] Devyatkin D A, Kusnezova Y M, Shvets A V and Chudova N V 2014 Artificial intelligence and decision making 2 27-40

[8] Pazel'skaya A G and Solovyev A N 2011 Komp'yeturnaya lingvistika i intellektual'nye tekhnologii: Po materialam ezhegodnoj Mezhdunarodnoj konferencii «Dialog» (Bekasovo). 10(17) 574–586

[9] Nakano T, Suda T, Okaie Y and Moore M J 2016 2016 IEEE Tenth International Conference on Semantic Computing (ICSC) (Laguna Hills, CA) pp 337–341

[10] Weir G R S, Santos E D, Cartwright B and Frank R 2016 2016 IEEE International Conference on Cybercrime and Computer Forensic (ICCCF) (Vancouver, BC) pp 1–3

[11] Korzh R, Peleshchysyn A, Syerov Y and Fedusenko S 2016 2016 XIth International Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT) (Lviv) pp 77–79

[12] Iskahkova A and Meshcheryalov R 2017 2017 Second Russia and Pacific Conference on Computer Technology and Applications (RPC) pp 85–89

[13] Shumskaya A O 2016 SPIIRAS Proceedings 6 (49) 124–140

[14] Tian Y. 2010 2010 Second International Conference on Multimedia and Information Technology, Kaifeng pp 123–126

[15] Repina E A and Belyanin V P 2016 The effectiveness of mass communications. Social and linguistic aspects pp 164–171

[16] Loukachevitch N 2016 Proceedings of Language Resources and Evaluation Conference LREC-2016 pp 1171–1176