Mathematical and algorithmic support for the early detection process automation of potentially dangerous content in Internet resources

E A Schwarzkopf1, O N Choporov, K A Razinkin1, V G Yurasov and Alexander N Mazalov2

1Voronezh State Technical University, 14, Moscow ave., Voronezh, 394026, Russian Federation
2Tambov State Technical University, 116, Sovetskaya str., Tambov, 392000, Russian Federation

E-mail: csit@bk.ru

Abstract. A mathematical formalization of procedures for the process of potentially dangerous content early detection by the degree of user responsiveness to the content is proposed. In this regard, a methodology for determining the list of resources scanned during monitoring and ranking them according to the degree of user involvement is recommended. The corresponding average daily metrics of commentability, approval and replication in terms of the content unit of the resource under study are proposed. Registration of potentially dangerous content is planned to be carried out in comparison with regular indicators of dynamics popularity acceleration for analyzed content unit. The proposed software can be successfully used in monitoring the Internet space in the context of struggling with destructive content.

1. Introduction
The informational confrontation in social networks today has become an objective reality, which makes it necessary to conduct comprehensive research in this direction [1-6]. One of the most effective tools to do this is the risk analysis methodology [7-12]. Such an approach often underlies the search and assessment of the danger of destructive content [13-15].

However, when it is necessary to organize operational counteraction to resonant content, methodological support should be modernized in the context of organizing early detection of such content. Obviously, this procedure must be automated. To do this, it is necessary to create an appropriate mathematical basis [16-18]. In this regard, the following tasks should be solved:

- determination of the list of Internet resources to be scanned when searching for resonant content;
- parameters of the identified content need to be quickly measured and already at the initial stage of its perception by users of the resource, compare them with the regular indicators of the usual posts of the resource under study;
- in the future, potentially dangerous content must be deleted or counter-content generated that mitigates the consequences of distribution.
The tasks formulated in the work actually reduce the automation cycle of the search for potentially dangerous content to the following (figure 1) repeating sequence of procedures.

In it, first of all, a bypass of Internet resources is realized with increased susceptibility (reactivity) of users to the content coming to it [19-20]. Here, the formalization of the ranking of these resources and the formation of a list of those based on the computing power of the monitoring system are objectively necessary. In the future (figure 1) a thematic analysis (on a monitoring task), the system will catch content with signs of destructiveness. At the next stage, measuring the parameters for identifying content will allow analysts to focus on potentially dangerous content already at the start of an abnormally high growth in their popularity (the phenomenon of resonant content).

Figure 1. Potentially Hazardous Content Detection Algorithm.
This will be followed by a management decision, which provides for at least two scenarios: the elimination (blocking) of inappropriate content or its rapid generation of antipodes that mitigate the effects of the analyzed content on the Internet audience. The key factor here is early detection and responsiveness (the information life of content on the network is very limited).

2. Definition of the list of scanned Internet resources

Below are three parameters that characterize the replicability, approval and commenting of content on the Internet resources:

- the replicability of R can be defined for the resource as the ratio of the number of reposts and the number of content (on average per day);
- approval L, by analogy, you can calculate the ratio of the average daily values (resource) of the number of likes and content;
- commentability C, in turn, will be the ratio of the average number of comments and contents averaged over the day.

In this case, it seems possible to go into three-dimensional space and use vector estimates. The demand for content in the studied resource can be described by the vector $\mathbf{K}$. Obviously, for many resources, we got a whole space of vectors.

Accordingly, for many resources (say, within the framework of online partnerships of single network), an integral vector can be constructed as the sum of the vectors of each of them.

As a numerical estimate of the demand, we can propose the calculation of the vector module $\|\mathbf{K}\| = \sqrt{R^2 + L^2 + C^2}$.

In vector content analysis, preference was given to modules of resource vectors as characteristics of the danger and / or positivity of the content circulating in the studied resource (daily, but the measurement period can theoretically be longer based on monitoring data). However, this is not the only quantity characterizing the vector. There are also angles:

- LR – angle between the R axis and the vector $\mathbf{K}$;
- LL – angle between the L axis and the vector $\mathbf{K}$;
- LC – angle between the C axis and the vector $\mathbf{K}$.

From the point of view of vector content analysis, angles can be very useful for characterizing the audience of the investigated resource. So, in the case when LL<450 and, accordingly, LC>450, it can be confidently stated that this audience is more inclined to approve of the content offered to its attention, rather than commenting (discussing) their content call this option the “dominant of approval”.

Based on the foregoing, a list of resources scanned during monitoring can be built and ranked.

3. Responsiveness metrics of users of Internet resources

When considering the functional features of Internet resources, you can highlight the daily parameters that are in the public domain and can be reduced to the metrics of the susceptibility of users of the resource.

Based on this, metrics are supposed, the essence of which is that they involve the determination of the degree:

- virality of content circulating in a resource as a ratio of the power of the response sets of its users and the content views per day;
- the generation activity of users of the resource in the form of a ratio of the power of the multitude of posts set in the resource and unique and other users marked per day;
- resource user activity as a ratio of the power of multiple visits and users (by category) per day;
the involvement of users of the resource in the processes of analysis, distribution and commenting on its contents through the ratio of the power sets of user and content reactions per day.

The list of the above metrics can be further expanded in view of the greater variability of the characteristics of interest, but from the point of view of monitoring, the degree of involvement seems to be the most informative, which we will consider in more detail below.

To assess the degree of involvement of users of the studied Internet resource in the processes of analysis (evaluation) of distribution and discussion (commenting) viral (we assume that the property of content virility activates these processes, that is, encourages the user to like, make reposts and formulate comments on a particular post) content appropriate to apply the average daily measurements that can be obtained in the public domain of many social networks and sites.

Such a calculation is necessary for the subsequent risk analysis of the processes of perception and distribution of viral content.

In this regard, it is proposed to use the average daily number of reactions, content, views and visits necessary for subsequent measurements of the risk of involvement of resource users in the above-mentioned processes of diffusion of viral content in the Internet.

At the same time, it is possible to evaluate the activity of users of the resource and the virality of the content circulating in them, which forms an additional picture of the popularity of the resource in the context of the generation and consumption of content by its users. In view of the foregoing, further it is necessary to develop integral criteria for ranking the resources necessary for content monitoring of Internet spaces. It is fundamentally necessary to assess the expected damage. In this regard, it seems possible to use the degree of involvement of users of the resource, which is determined in terms of one unit of content of the resource above. From here, multiplying its value by the number of contents (for the study period) we get the total number of reactions to viral contents, that is, in fact, the amount of negative consequences.

4. Conclusion

Thus, for the studied resources, it is possible to obtain averaged estimates of the danger of the distribution of inappropriate content and analyze the obtained visualizations of the popularity of the resource and its directivity through three-dimensional space.

As an operational assessment of the regionality of the studied content at the stage of its launch into the Internet space, it may be recommended to calculate the second derivative of the reactivity dynamics $S''(t)$ of the resource users for this content. If this indicator exceeds the regular (average daily) $S_0$ levels, such content can be considered as potentially dangerous with all the ensuing actions. E.g. $S''(t) \gg S \rightarrow 0$, where $S(t)=R(t)+L(t)+C(t)$ is the basis for such a response by the administrator, moderator, monitor.

In practical terms, the above calculations would be appropriate to tabulate for Internet resources visited by regional users. To normalize the visibility indicator, it is also recommended to take into account the number of users in the studied Internet resource, which will be a divisor to the average rating of approval for the resource.

Thus, for the studied resources, it is possible to obtain averaged estimates of the danger of the distribution of inappropriate content and analyze the obtained visualizations of the popularity of the resource and its directivity through three-dimensional space.

These indicators can be used to averaged risk assessment of the distribution of inappropriate content of popular Internet resources

References

[1] Islamgulova V V, Ostapenko A G, Radko N M, Babadzhanov R K and Ostapenko O A 2016 Discreet risk-models of the process of the development of virus epidemics in non-uniform networks Journal of Theoretical and Applied Information Technology 86(2) 306-15
[2] Eshchenko A, Ostapenko G, Bataronov I and Tolstykh N 2019 The automated networks and regional users: risk analysis of their reactions to the attacks of different destructive orientation *IOP Conf. Series: Materials Science and Engineering* **537** 052020

[3] Surkov I, Ostapenko G, Belonozhkin V and Razinkin K Metric parameters of diffusion in destructive fillers automated networks 2019 *IOP Conf. Series: Materials Science and Engineering* **537** 052018

[4] Filatov V, Ostapenko A, Barannikov N and Yurasov V 2019 Methodological support for automating risk analysis of engaging users in the destructive content of the network for sharing content *IOP Conf. Series: Materials Science and Engineering* **537** 52017

[5] Kravets O Ja, Choporov O N and Bolnokin V E 2018 Mathematical models and algorithmization of monitoring control an affiliated network service distributed organizations *Quality: Access to Success* **19**(167) 68-72

[6] Tsaregorodtsev A V, Kravets O Ja, Choporov O N and Zelenina A N 2018 Information Security Risk Estimation for Cloud Infrastructure *International Journal on Information Technologies and Security* **10**(4) 67-76

[7] Stepanov L V, Koltsov A S, Parinov A V, Parinov D V and Soloviev B A 2019 Practical aspects of application of theory of games to the assessment of system security *Modeling, Optimization and Information Technology* **7**(4) DOI: 10.26102/2310-6018/2019.27.4.028

[8] Kunizheva L A 2019 Mathematical model of the distribution of the global chain epidemic of network viruses on the fracturing graph *Modeling, Optimization and Information Technology* **7**(4) DOI: 10.26102/2310-6018/2019.27.4.012

[9] Minaev V A, Sychev M P, Kulikov L S and Vaitz E V 2019 Modeling of counter action to destructive influence in social networks *Modeling, optimization and information technology* **7**(1) DOI: 10.26102/2310-6018/2019.24.1.004

[10] Minaev V A, Sychev M P, Kulikov L S and Vaitz E V 2019 Modeling manipulative influences in social networks *Modeling, optimization and information technology* **7**(1) DOI: 10.26102/2310-6018/2019.24.1.003

[11] Tsaregorodtsev A V, Zelenina A N and Savelev V A 2017 Vulnerability classification of cloud tools in the problem of quantitative risk assessment *Modeling, optimization and information technology* **5**(4) 4217

[12] Goncharov A A, Sokolova E S, Chapurin E Y, Pozdysheva O V and Kulikov S S 2019 Methodological and algorithmic support for evaluating the destructiveness of content circulating in a social network *Information and security* **22**(3) 373-86

[13] Goncharov A A, Chapurin E Y, Belonozhkin V I, Radko N M and Ruzhitsky E 2019 Development of methods and construction of algorithms for searching and classification of constructive content circulating in a social network *Information and security* **22**(3) 345-60

[14] Moskaleva E A, Nezhelsky E R, Volkova M E, Sibirko K V and Safronova V V 2019 Use and development of methods for assessing the tonality of the contents of social networks in order to ensure information security of users *Information and security* **22**(3) 433-44

[15] Ostapenko A G, Parinov A V, Kalashnikov A O, Shcherbakov V B and Ostapenko A A 2017 *Social networks and destructive content* (Hotline-Telecom, Moscow) 276

[16] Ostapenko A G, Radko N M, Kalashnikov A O, Ostapenko O A and Babajanov R K 2017 Epidemics in telecommunication networks (Hotline-Telecom, Moscow) 284

[17] Ostapenko A G, Zheleznyak V P, Chapurin E Y, Ostapenko A A, Trubitsyn S D, Ostapenko O A and Miroshnichenko T Y 2019 Monitoring and risk management of the socio-information space in order to ensure regional and national security *Information and security* **22**(4) 557-72

[18] Ostapenko G.A, Moskaleva E.A., Chapurin E.Y., Ostapenko A.A., Volkova M.E., Khokhllova D.S., Miroshnichenko T.Y. Possibilities of using artificial intelligence systems and big data in the implementation of risk monitoring of regional socio-information space *Information and security* **22**(4) 587-600
[19] Telegin E N, Chapurin E Y, Razinkin K A, Plotnikov D G and Popov A V 2009 Methods of thematic modeling, their development and application for content circulating in regional online communities *Information and security* **22**(3) 325-44

[20] Schwarzkopf E A, Savinov D A, Pasmurnov S M, Nezhelsky E R, Tikhonova S S, Lantyukhov N M and Bartenev A N 2009 Organization of monitoring the socio-information space and building a model of a regional Internet user in the context of ensuring its security *Information and security* **22**(4) 601-8