

PAPER • OPEN ACCESS

## Methodological support for automating risk analysis of engaging users in the destructive content of the network for sharing content

To cite this article: V Filatov *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **537** 052017

View the [article online](#) for updates and enhancements.



**IOP | ebooks™**

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the **collection** - download the first chapter of every title for free.

# Methodological support for automating risk analysis of engaging users in the destructive content of the network for sharing content

V Filatov, A Ostapenko, N Barannikov and V Yurasov

Voronezh State Technical University, Moscow ave., 14, Voronezh, 394026, Russia

E-mail: vitaliyfilatov3110@yandex.ru

**Abstract.** One of the possible ways of hazard assessment of video hosting resources on the basis of a combination of a way of identification of destructive content and risk of the involvement of users at content distribution is given in the work. The way of identification of destructive content is based on a preliminary calculation of patterns. The method of calculation of patterns is also given in the article. Calculation risk analysis is based on parameters of the channel and content: comments, likes, dislikes, number of viewings, size of audience and others. Hazard assessment consists of obtaining a rated value of integrated risk for the involvement of users into the maintenance of destructive content of the resource. The proposed method is applicable in systems with a high level of automation to identify the most harmful sources of information on video hosting sites.

## 1. Introduction

The Internet is a key element of the modern information sphere. The emergence of social networks became a new round in the development of the Internet. Over time social networks increased the functionality and by the present moment became the powerful media tool capable to influence effectively and multipurpose the users [1-4]. It is natural that such a powerful tool can use also for distribution of destructive influence. For this reason, the question of research of popular social networks, regarding identification and also influence of the destructive content (DC) extending on them on different Internet users becomes the most relevant in the field of safety of the personality and society [1].

On social networks, it is possible to carry content to disruptive content text, audio, video of contents, the image, the files containing promotion of intake of drugs, gamblings, the appeals to racial hostility, the harm of life and health having pornographic focus, etc. Also, special attention should be paid to the illegal content containing appeals to the activity of extremist orientation [5]. Through such popular network for exchange of content as YouTube recruited the considerable number of people in the terrorist organizations. Social media became the tool of "color revolutions" which developed in many respects depended on moods of users of social networks.

The role of the distributor of destructive content leads to an understanding of the importance of hazard assessment of such information [6-8]. One of the main problems of this sphere is the development of methodologies of assessment of the threats arising at the diffusion of disruptive content with the help risk analysis.



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

## 2. Pattern definition

The object of the study is a social network, which is a video hosting. Distributors of content on social network YouTube are channels which hazard assessment includes 2 aspects: assessment of the influence of the channel, width of coverage of audience and entity of content, its orientation (neutral and positive and destructive). For assessment of orientation of content, it is possible to select two approaches: the analytical method assuming exarticulation and the indication of types of the disruptive content fixed legislatively and the intrinsic method assuming the existence of reasonable criteria of injuriousness of content.

At the creation of automated systems, it is more appropriate to apply the second way as it allows to construct formal analysis systems of content on disruptiveness. However, the first approach, because of labor input and lack of adequate and conventional scientific tools for its development at the moment is often used.

For the analysis of the orientation of the content of social network, we will give the method described in work [6]. Originally, it is necessary for the work of a method:

1) define categories of illegal content. The more in details splitting content into categories, the bigger accuracy can be reached when carrying out the analysis;

2) for each selected category to make a pattern, degree of compliance to which of a possibility of information identified as illegal. The concept of a pattern of work is given as follows [9].

Let there be a set of words represented by a set  $W = \{w_1, \dots, w_n\}$  dimension of  $n$  and a set of the numbers corresponding to this mere verbiage presented by a set  $P = \{p_1, \dots, p_n\}$ . At the same time  $p_i$  – a share  $w_i$  words in a pattern. It follows from this that  $p_i$  has the following properties:

$$\begin{aligned} 0 < p_i < 1 \\ \sum_{i=1}^n p_i = 1. \end{aligned} \quad (1)$$

We will also call a set of mere verbiage and a set of the shares corresponding to these words a pattern.

## 3. Algorithm for obtaining a pattern of destructive orientation

Let's describe an algorithm of receiving a pattern of the text of any orientation.

1) it is necessary to create a collection of materials which are referred to a certain category of the forbidden information;

2) to make the content analysis of each material from the created collection; the result of this content analysis will be a set of words  $W$  (only significant words) and a set of shares of these words in the analyzed text of  $P$  corresponding to it will be the result of such content analysis (descriptions of the  $W$  and  $P$  sets are given above at the description of a concept a pattern);

3) to make compression of the data obtained at the previous stage. Compression is made as follows:

a) the set of words of a pattern of  $W = \{w_1, \dots, w_n\}$ , Formation happens by addition of sets of words of each material:

$$W = \bigcup_{i=1}^m W_i, \quad (2)$$

where  $W_i$  – is a set of the words  $i$  of material,  $m$  is the amount of materials in 1 collection created at a stage.

b) the set of shares of a pattern of  $P = \{p_1, \dots, p_n\}$ . Each member of a set of  $p_i$  is found as follows:

4) let's calculate a measure of similarity of the sim as:

$$p_i = \frac{\sum_{j=1}^m \text{map}(w_i, \text{Pat}_j)}{m}, \quad (3)$$

where  $\text{map}(w_i, \text{Pat}_j)$  – the function returning word  $w$  share in  $\text{Pat}$  pattern.

#### 4. Identification of destructive content based on a measure of similarity

Further for carrying out the analysis it is necessary to make an operation of imposing of any content. Content is presented by the structure similar to a pattern (set of words  $W$  and the shares corresponding to them set  $P$ ). Such operation will give as a result a measure of similarity of this material to a pattern. Operation of imposing should satisfy to the following reasons:

1) the more terms of a pattern is present at the text, the similarity of material and a pattern should be bigger;

2) influence of each term of a pattern on the value of similarity is unequal, such influence depends on 2 factors:

a) from term frequency in a pattern (than frequency is higher, influence on a similarity measure is higher there)

b) from a term frequency difference in a pattern and common rate of this term.

According to these reasons we will offer the following scheme of the procedure of imposing:

1) we will deliver to each term from a pattern in compliance with the frequency of the use (common-language). If the term from a pattern is unique (the word is a jargon, a proper name), then such frequency is equated by 0.

2) let's calculate the module of a difference of the frequency of each term of a pattern and the frequency calculated at the previous stage. A set of the received differences we will designate  $D = \{d_1, \dots, d_n\}$ ;

3) let's make the analysis of the occurrence of each word of a pattern in the text. During this procedure we will receive a set of  $W = \{w_1, \dots, w_n\}$ , equal to a set of words of a pattern and a set of  $P_{imp} = \{p_1, \dots, p_n\}$ , which  $p_i$  elements are equal:

$$p_i = \begin{cases} \text{map}(w_i, W_k), & w_i \in W_k, \\ 0, & w_i \notin W_k; \end{cases} \quad (4)$$

where  $W_k$  – is a set of words of the analyzed content.

4) let's calculate a measure of similarity of the  $\text{sim}$ , as:

$$\text{sim} = 1 - \frac{\sum_{i=1}^n (\text{map}(w_i, \text{Pat}) - p_i)^2 * \text{map}(w_i, \text{Pat}) * d_i}{\sum_{i=1}^n \text{map}(w_i, \text{Pat})^3 * d_i}. \quad (5)$$

Such normalized ( $0 \leq \text{sim} \leq 1$ ) the measure of similarity can be interpreted as the probability that the studied content has destructive focus.

#### 5. Calculation of the integral risk of resource user involvement

The second aspect of hazard assessment of sources of destructive contents is assessment of the involvement of the audience. It can be expressed by such parameter as a risk of the involvement of users at a distribution of destructive content during its activity [1]:

$$\text{Risk} = \frac{K_v + K_l + K_{sh} + K_p}{\text{number of users of a resource}}, \quad (6)$$

where:

1) coefficient of likes,  $K_l$  :

$$K_l = \frac{\text{Quantity of likes}}{\text{Quantity of views}}; \quad (7)$$

2) coefficient of dislikes,  $K_d$  :

$$K_d = \frac{\text{Quantity of dislikes}}{\text{Quantity of views}}; \quad (8)$$

3) coefficient of positive comments,  $K_p$  :

$$K_p = \frac{\text{Number of positive comments}}{\text{Number of viewings}}; \quad (9)$$

4) coefficient of negative comments,  $K_n$  :

$$K_n = \frac{\text{Number of negative comments}}{\text{Number of viewings}}. \quad (10)$$

As for many real video hosting sites like the fact, the user of the video leads to a repost of this material, the formula (5) will take a form (10):

$$\text{Risk} = \frac{K_v + 2 * K_l + K_p}{\text{number of users of a resource}}, \quad (11)$$

For integrated hazard assessment of the channel, combining above-mentioned aspects of hazard assessment of a source of distribution of disruptive content it is possible to apply the following formula:

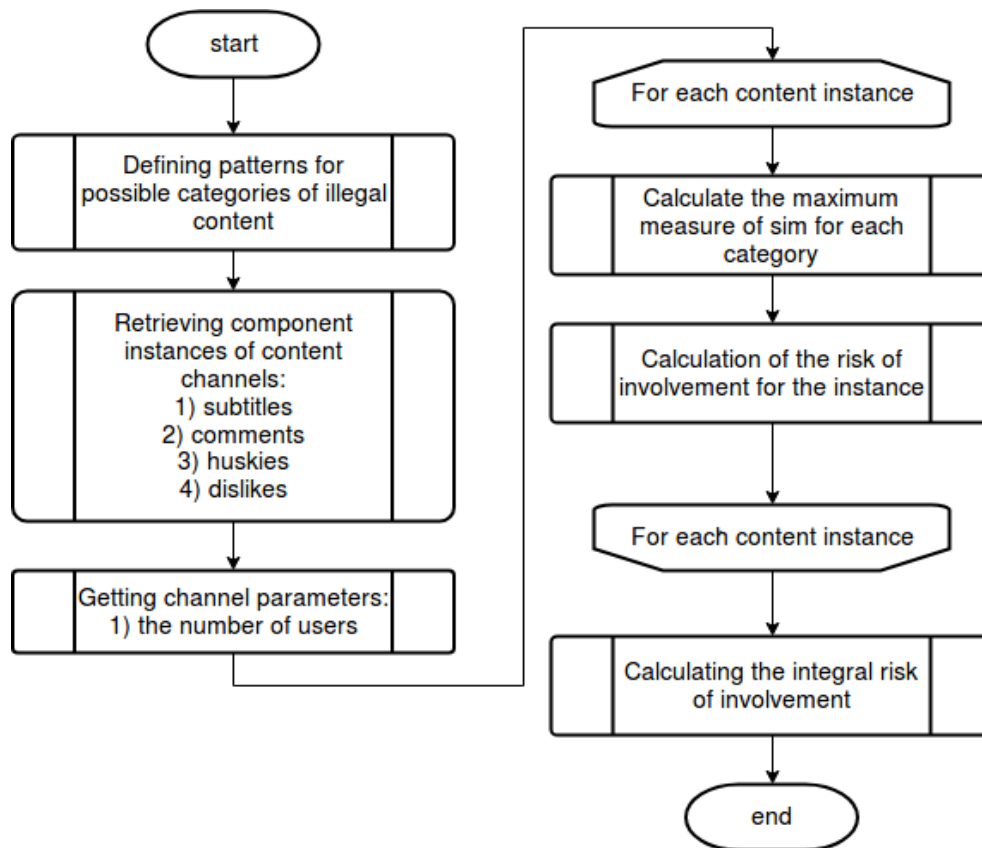
$$\text{Risk}_{\text{channel}} = \frac{\sum_{i=1}^N \text{Risk}_i * \text{sim}_i}{N}, \quad (12)$$

where:  $N$  – quantity of content of the channel,  $\text{Risk}_i$  – risk of the involvement for  $i$  of a copy of content of the channel,  $\text{sim}_i$  – similarity of  $i$  of a copy of content with one of patterns of destructive orientation (as concrete category that at which sim is maximum is used). Generalizing the above-stated stages, it is possible to show an algorithm, for the description of the process of automation risk analysis of the involvement of users in the maintenance of destructive content of channels of social network (figure 1).

## 6. Conclusion

Thus, during the research the methodology of hazard assessment of sources of distribution of content on the basis of social network YouTube which includes studying and the analysis of 2 aspects was offered: orientations of content and involvement of audience of a source. As a result of accounting of these two aspects received a way of determination of a risk of the involvement for the channel in general. The described stages are presented in the form of an algorithm that allows you to automate the process of obtaining a final risk measure for the involvement of users in the content of the destructive content of the video hosting channel. It should be noted that the proposed method of identifying destructive content is more informative, accurate and formalized than the previously proposed, based

on the analysis of emotions caused by content, positive comments and content, the presence of mark words in the title [10-12]. The rather high degree of formalization of the proposed method for assessing the risk of destructive content makes it applicable in systems for automated monitoring of a social network to identify the most dangerous sources of distribution of harmful information.



**Figure 1.** Algorithm of the description of the process of automation risk analysis of the involvement of users in the maintenance of destructive content of video hosting channels.

## References

- [1] Ostapenko A G, Parinov D G, Kalashnikov A O *et al.* 2018 *Social networks and destructive Content* (Moscow)
- [2] Karampelas P 2015 *Techniques and Tools for Designing an Online Social Network Platform* (Springer)
- [3] Golback J 2015 *Introduction to Social Media Investigation: A Hands-on Approach* (Syngress)
- [4] Cha M, Perez J A N and Haddadi H 2009 Flash Floods and Ripples: The Spread of Media Content through the Blogosphere *Data Challenge Workshop* **12** 92–6
- [5] Gubanov D A, Novikov D A and Chkhartishvili A G 2010 *Social network YouTube: models information influence, management and confrontation* (Moscow)
- [6] Vasilyev V V and Shamsutdinov R R 2019 Intelligent network intrusion detection system based on artificial immune system mechanisms *Modeling, optimization and information technologies* (7) **1**
- [7] Minaev V A, Sychev M P, Kulikov L S and Vaitz E V 2019 Modeling manipulative influences in social networks *Modeling, optimization and information technologies* (7) **1**
- [8] 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

- [9] Parinov A V, Sokolova E S, Urasov V G, Tolstykh N N and Filatov V V 2018 Destructive content in multinetwork socio-informative space: formalization of the procedure of detection *International Journal of Pure and Applied Mathematics* **119(15)** 22651–5
- [10] Ostapenko G A, Parinova L V, Belonozhkin V I, Bataronov I L and Simonov K V 2013 Analytical models of information-psychological impact of social information networks on users *World Applied Sciences Journal* **251** 410–5
- [11] Islamgulova V V, Ostapenko A G, Radko N M, Babadzhanov R K and Ostapenko O A 2016 Descreet risk-models of the process of the development of virus epidemics in non-uniform networks *Journal of Theoretical and Applied Information Technology* **86** 306-15
- [12] Shvartskopf E A, Zaryayev A V, Parinova L V and Popova L G 2016 Modeling of layering growth virus epidemic and spread of harmful content on Poisson networks *Research Journal of Pharmaceutival* **7** 2321–31
- [13] Romansky R 2017 A Survey of Digital World Opportunities and Challenges for User’s Privacy *International Journal on Information Technologies and Security* **4 (9)** 97-112