Methods to identify the destructive information

V B Barakhnin¹,²,³, R I Mukhamedyev³,⁴,⁵, R R Mussabaev³,
O Yu Kozhemyakina¹,³, A Issayeva⁴, Ya I Kuchin³, S B Murzakhmetov³,⁴
and K O Yakunin³,⁴

¹ Institute of Computational Technologies SB RAS, Russia,
² Novosibirsk State University, Russia,
³ Institute of Information and Computational Technologies MES RK, Kazakhstan,
⁴ Satbayev University, Kazakhstan,
⁵ ISMA University, Latvia

E-mail: bar@ict.nsc.ru, ravil.muhamedyev@gmail.com, rmusab@gmail.com,
olgakozhemyakina@mail.ru, ainur.kaznrtu@gmail.com, ykuchin@mail.ru,
sanzharmrz@gmail.com, yakunin.k@mail.ru

Abstract. The article discusses the methods that ensure the detection of destructive messages in the Internet environment. It is shown that to detect them the approaches based on the analysis of the syntactic patterns; analysis of semantic information included in the text, and correlating it with the text corpus; crowd sourcing methods; identification of the patterns of user behavior in social networks; the consideration of the additional information etc., were applied. Good results can be achieved under certain conditions: an access to traffic social networks and other online news resources, opportunities to organize or to get results from crowd sourcing, etc., when there are the restrictions on these conditions the detection of destructive messages can be performed based on indirect signs. The results of the detection of destructive news in the corpus of news reports in Russian, posted on the websites of the Republic of Kazakhstan are shown.

1. Introduction

Over the past decade and a half or two, the volume of constantly generated information materials has increased. These materials are instantly becoming publicly available due to their placement in the Internet, and they can become both a positive and a negative impact on society. In order to ensure the sustainable development of both individuals and society as a whole, it is necessary to minimize the possible destructive impact of media content. As noted in [1], new protective measures obtained in the course of interdisciplinary research aimed at limiting the spread of false information are needed. Of course, the solution to this problem lies not only in the field of information technologies, but the information technologies are the important tool in ensuring the reliability of information in the media space.

Let’s describe the signs of news of interest from the point of view of the organization of the information system of media monitoring in the interests of maintaining social stability and sustainable development of the individual and society.
First, we discard the non-resonant news, which are published on resources with a small media coverage or having for a certain period of time a small number of reposts. Obviously, such news cannot have a serious impact on social stability.

Secondly, the resonant news are tested for social significance, focusing primarily on the subject of news, as the news with insignificant topics also have a weak impact on social stability, regardless of their other characteristics.

Third, if the news is socially significant, we check its reliability. This is the most difficult part of the decision-making process, which is shown in Figure 1. If the news is unreliable (fake), then we classify it as interesting from the point of view of users of the information system of media monitoring in the interests of maintaining social stability.

Finally, fourth, even formally reliable news, but bearing the signs of bias (unjustifiably high degree of generalization, excessive emotionality, politicization, etc.) will also be classified as of interest from the point of view of the potentially dangerous destructiveness.

![Figure 1. Algorithm of decision-making process](image)

2. Overview of problem solving methods

Let’s describe the approaches used to identify the news that satisfy the above criteria.

The methods for determining the resonance of news are quite obvious and depend on whether the creator of the monitoring system is guided by the information provided by the reviewed site or has access to its traffic.

In the first case, as a rule, you can see only the number of comments. At the same time, to count the number of persons who read the news or left comments taking into account the repetition, it is necessary to use separate means of analyzing the web page: in addition, if the site does not provide authorization of users by means of electronic digital signature (EDS) or SMS messages [2], then a reliable count of commentators is impossible: the same person can leave comments using different names or even anonymously.

An external user can see only the number of readings if the site owner has opened this indicator. In this case, it is necessary to check whether repeated visits from the same computer are not taken into account – in this case, the rate of visits is unreliable, because the interested user can easily "catch up" with the attendance.

The number of approvals/disapprovals can be seen if the site is equipped with a voting mechanism. At the same time, if the authorization of users via EDS or SMS-messages is not provided, the indicator is not reliable, because “cheating” is possible.
A researcher at whose disposal is the website traffic, has the very different possibilities. In this case, it can, using information about the URLs of visitors, track the number of unique visitors and unique commentators, the location of visitors – with accuracy not only to the country but also to the city, etc.

Thus, the reliability of the assessment of the resonance of the Internet publication increases with access of an explorer to Internet traffic, if there is no access, more or less reliable judgment on the resonance of a publication can be made only on the basis of the number of comments, and after further analysis of the uniqueness of the commentators.

The topics of social importance are usually determined by interviewing experts who possess the data on the state of public opinion, which, in turn, can be obtained through sociological surveys of the population.

The most difficult task is to identify the unreliable, in other words, “fake” news (according to, for example, [3] – they are “news stories, invented with the intention to deceive”).

Due to the actuality of the topic of identifying false reports, intensive work in this direction is carried out by groups of scientists around the world. For example, in the work [4], the main recommendations for the formation of the body of texts were formulated, among which one can distinguish a homogeneity in the length of the news and the style of presentation. The works [5, 6] describe the practical results of the usage of machine learning to identify unreliable news. A sign of unreliability are, according to the authors, more positive facts and greater confidence in the above. In these works, the authors did not use any ontology for semantic analysis, limiting themselves to morphological and syntactic analysis. Sources of unreliable news were as news, really collected from various sites, as fake messages created by volunteers on the basis of truthful news. As a result, the corpus of texts was formed, consisting of an equal number of truthful and unreliable news. Frequency characteristics of N-grams, syntactic and psycholinguistic features of the text are used as analyzed characteristics of the texts. The usage of the support vector machine (SVM) method allowed to achieve classification accuracy of 75–85 % on different text corpus. In [7] the authors identified the words that correspond to more reliable and unreliable texts.

The regularities inherent in some classes of unreliable news are analyzed in [8]. The authors suggested that unreliable texts contain fewer nouns, numbers and quotations, redundancy of adverbs and repetitions. As a result, an algorithm was implemented, which in 70% of cases correctly determined unreliable news based on these features.

The so-called “hybrid” approach is promising, which implies, along with multi-level linguistic analysis of the text, fact checking via the Internet, for example, by querying the ontology DBpedia or Google Relation Extraction Corpus (GREC) [9].

The formation of text corpus for training classifiers in the case of unreliable news is quite a difficult task. To reduce the complexity of markup in [10], a method of partial markup and establishing a connection between articles in an array of untagged texts is proposed.

This approach allowed to achieve the accuracy of more than 70 % with only 2 % of marked texts. Let’s note that after increasing the number of marked texts to 30 %, the accuracy of the classification reached only 75.43 %. This approach seems to be one of the most promising due to the low requirements for the body of texts and good results obtained by indirect use of semantic information of texts.

According to [11], the hoaxes, which are also a special case of fake news, as they are created with the intention to mislead readers, are possible to be classified with a high degree of reliability using social networks. In this work, the identification of hoaxes is performed by analyzing user messages in Facebook. It is reported that if it is knowing which users prefer unreliable messages, it is possible to achieve classification accuracy of about 99 % even if only 1 % of messages are marked.

The following approach, which is presented in [12], uses “wisdom of crowds” and is also based on the analysis of the behavior of users of social networks. By highlighting and analyzing of the comments about the news, the unreliable news was revealed. The following factors were used: firstly, if the article was devoted to a local event, the comment from a user geographically close to the place of events was considered more reliable, and secondly, if the user was inactive for a long time, his comment was considered less reliable. In addition, it is noted that the speed of false news distribution in
the media space is higher than the true [13], and the nature of news distribution on the network is significantly different for false and true [14].

Let's review the works devoted to the identification of fake news in Russian.

The review of methods of detection of lies in the written text is given in [15], however, it is mainly about “live written” speech, including Internet communications, but not about the media. One reviewed study notes that “liars produce longer texts, use more words associated with channels of perception (see, hear), use fewer pronouns for self – referencing, more for naming other people. In addition, motivated liars avoided causations, and unmotivated liars did a lot of negations”.

In the article [16] the similar signs of fakes are noted: “With confidence the lexical markers of fakes can be called that are common to advertising and the media: “you'll be amazed”, “endangered”, “world discovery”, “something scary”, “disturbing the minds” – these are just some of the titles are teasers from Martketgid”.

Finally, with the work with Russian written texts corpus [17, 18], which was specially made for the study of language features of texts in Russian, and contains deliberately distorted information, the fact that the models of lie detection in the text can be significantly improved by considering the personal characteristics of their authors (gender, age, psychological profile) was revealed.

However, the professionally created fakes that imitate the style of the media, completely devoid of emotional, personal, etc. characteristics. An example of this is the portal “IA “Panorama”” [19], specializing in fake news. Although the portal does not hide such specialization, but there are cases when after 1–2 reprints fake news generated by “Panorama”, were spread by quite serious media.

At the same time, [15] emphasizes: “When we are choosing the characteristics of the text that can be markers of lies, it is necessary first of all to rely on those that can be extracted from the text automatically. At the moment, the most reliable are morphological parsers, a certain accuracy is given by syntactic parsers. Semantic analyzers to date give a lot of errors and have been excluded from the study”.

Finally, the identification of biased news involves the use of techniques similar to those used in the identification of fake news, since both of these classes of news are characterized by the usage of similar manipulative technologies.

3. Method description

The classification of articles described above is possible in two ways.

First, on the basis of expert assessments, it is possible to compile dictionaries characterizing the belonging of the text to each of these classes, after which for each text under study the set of relevant parameters is determined and then they are aggregated in order to assess the possible belonging of the text to one of the classes.

Secondly, if there is a marked body of articles, it is possible to train an automatic classifier, which could calculate the belonging of the text to the corresponding class automatically, based on a particular formal model of the text.

Each approach has its own drawbacks.

The first version of the classification of media texts is distinguished by a certain subjectivity, due to the fact that both the parameters and their mutual importance are determined by experts.

The second variant of the classification requires a significant amount of marked body of texts. Given that the parameters of the texts that affect the conclusion of the expert are not defined, it may be necessary to use a corpus suitable for deep learning methods (hundreds of thousands of texts). The formation of such a corpus is a long and laborious process. Therefore, at the initial stage of implementation of the system, we use the first approach, which can be called parametric.

Generally the process of text and media analysis consists of the following steps:

- A list of parameters that determine the belonging of the text to one of the above classes is formed.
- The comparative importance of the parameters is estimated.
- For each text to be considered, a parameter estimate is calculated.
The parameter estimates and their relative importance are aggregated to obtain an estimate of the text belonging to a particular class.

On the basis of the received estimates of texts the assessment of mass media as a whole is formed.

Evaluation of the above-mentioned parameters of media texts can be based on natural language processing (NLP) methods: significant progress has been made in the development of models and algorithms for text analysis. In turn, the aggregation of heterogeneous parameter values is often performed using multi-criteria decision making (MCDA). Let’s consider the main features of NLP and MCDA in solving the problem of multi-criteria evaluation of the media.

One of the methods which is used productively in the field of NLP is thematic modeling (thematic analysis). This is a method based on statistical characteristics of document collections, which is used in the tasks of automatic abstraction, information retrieval and classification [20]. The idea of this approach is to intuitively understand that the documents in the collection form groups in which the frequency of occurrence of words or combinations of words varies.

The basis of modern thematic models is a statistical model of natural language. The probabilistic thematic models describe documents by a discrete distribution on a set of topics, and topics are described by a discrete distribution on a set of terms [21]. In other words, the topic model determines which topics each document refers to and which words form each topic are formed it. The clusters of documents which are related to a set of topics allow, in particular, to solve the problems of synonymy and polysemy of terms [22].

The probabilistic latent semantic analysis (PLSA), additive regularization (ARTM) [23] as well as a very popular latent Dirichlet allocation (LDA) [24, 25] are used to build a mathematical model of the corpus of documents.

Each topic can be considered as a separate parameter of the text or the media as a whole, the value of which is determined by the mentioned conditional probabilities. Where it is necessary to make a decision on the basis of a variety of heterogeneous parameters and alternatives, a number of methods are used to ensure that the mutual importance of the parameters is taken into account and that their values are aggregated as one or a small number of estimates. By their nature, such tasks belong to the field of multi-criteria decision-making, widely used in decision support system (DSS).

The method of analysis of hierarchies (AHP) [26] has shown its effectiveness in solving many practical problems [27]. Therefore, in this paper, this method is used to assess the weight of parameters in the process of classification of media texts.

In turn, the thematic model built with the use of LDA provides multimodal evaluation of media texts by calculating the conditional probability distributions of documents by features ($p_1$), topics ($p_2$) and classes ($p_3$). The implemented method is notable for the fact that the thematic model created by cluster analysis (teaching without a teacher) is then applied in combination with expert-defined classes and features. Thus, the semantics of the required distribution is set by the user (expert), although the initial thematic analysis depends only on the body of documents. The application of the thematic model provides, in our opinion, the reduction of errors associated with the calculation of compliance of texts and dictionaries parameters.

The initial data and the resulting conditional probability matrices are shown in Figure 2, where MMS (mass media sources) is a set of text sources. Media (MMS) are the source of m documents (papers) that are obtained through the application of data collection systems (process 2). The resulting body of documents M is divided into thematic clusters $T$ (process 1). Experts form classes $C$ (process 4) and define properties or parameters of classes $Q$ (process 3). Properties are described by dictionaries of words, expressions (dictionaries) or procedures of their identification in the text (features).

Using a set of corpus topics, firstly, we obtain a discrete distribution of conditional probabilities of articles and topics – $p_2(k|m)$, where $k \in T$, $m \in M$. 

Second, we obtain a conditional distribution of dictionaries (parameters) and topics $p_1(k|q)$, where $q \in Q$, that is, we determine to what extent the parameter describes a specific topic. Thirdly, using the analytical hierarchical process (AHP) we calculate the importance of the parameters for the classes (for each class) and obtain $p_3(c|q)$, where $c \in C$. Then, using $p_1$ and $p_3$, we expect the conditional distribution of subjects by grade $- p_4(k|c)$. Knowing the probability distribution of the topic by class ($p_4$) and the probability distribution of the article by topic ($p_2$), you can calculate the distribution of the article by class $p_3(m|c)$. In turn, the distribution of the article by features or dictionaries $- p_6(m|q)$ depends on $p_1$ and $p_2$.

The LDA algorithm is used to calculate the set of topics of the corpus $T$. The Jacquard coefficient [28, 29] is used to calculate the probabilities $p_1$ and $p_2$. To calculate the probabilities $p_3$ – AHP.

4. Adaptation of pymorphy library for a work with pre-reform spelling

The algorithm described above (referred to below as an aggregator) is used to calculate the distribution of media texts by parameters ($p_6$) and the distribution of articles by classes ($p_3$). At the same time, the experts have created the dictionaries of signs: Facts (657 lexical units), Negative tonality (2344), Positive tonality (1207), Politicization (35), Call to action (145), Manipulation (1072). Using the expert estimates, the $p_3$ matrix is formed (the importance of parameters for classes):

\[
\begin{bmatrix}
0.23, 0.18, 0.18, 0.2, 0.21, 0.0, \\
0.0, 0.22, 0.22, 0.26, 0.0, 0.3, \\
1.0, 0.0, 0.0, 0.0, 0.0, 0.0
\end{bmatrix}
\]

where the columns (from the left to the right) correspond to the parameters listed above, and the rows (from the top to the bottom) to the classes: “Socially significant”, “Objective”, “Reliable”. Let’s note that this matrix is not complete. Not all parameters for the corresponding classes are defined. Zero means that this parameter is not used to determine the probability $p_3$. For example, it can be seen that the social significance of the text is determined by five criteria, while reliable using only one (facts).

A set of articles by the Tengri News news agency (https://tengrinews.kz/) was used to evaluate the capabilities of the aggregator. The corpus contains 1800 articles marked up by experts, what made it possible to compare the results of the aggregator with expert estimates. Figure 3 shows the tables where 5 news with the largest and 5 with the least probability of matching the parameter or class are displayed (aggregator score). The expert ratings are given in the column Expert Score.
| Text | Title | Aggregator score | Expert score |
|------|-------|------------------|--------------|
| В Астане состоялась встреча Министра оборо Министр обор | 0.12821395 | politicized |
| Лидер КНДР Ким Чен вн, выступая в понедель | 0.127618874 | politicized |
| Лидер партии ЛДПР Владимир Жириновский у Жириновский у | 0.122347869 | politicized |
| Бахытжан Сагинтаев представил коллектив М Сагинтаев о нас | 0.119933087 | politicized |
| Нурсултан Назарбаев направил телеграмму Ге Назарбаев образ | 0.119336992 | difficult to determine |
| Видео о том, как оригинально встретили маму Оригинальную | 0.080228157 | politicized |
| КамАЗ, "Газель" и три авто столкнулись в Астане Массовое ДТП | 0.07948159 | politicized |
| Врачи борются за жизнь беременной женщины 3-летняя девоч | 0.078410558 | politicized |
| Певица Карина Абуллина опубликовала в Инст Карина Абулл | 0.077287639 | politicized |
| Турсенғали Алазов будет ведущим вечера на Турсенғали Ал | 0.077140974 | politicized |

| Text | Title | Aggregator score | Expert score |
|------|-------|------------------|--------------|
| Игрок турецкого "Буджаспора" Абдуллах Бали Турецкий футбол | 0.089326405 | Objective |
| Лидер КНДР Ким Чен вн, выступая в понедель | 0.089229172 | Objective |
| Арестован подозреваемый по делу о провокации Арестован подозр | 0.089027615 | Objective |
| Директора школы осудили в Карагандинской с Директора школы | 0.088805319 | Objective |
| Организаторы зимних Олимпийских игр-2018 Скандал на Клин | 0.088727886 | difficult to determine |
| Житель Великобритании Джеймс Крэйн (James) Реакция соовсе | 0.086347233 | Objective |
| Казахстанские школьники, которые завоевали С главой Google | 0.086273295 | Objective |
| Видео о том, как оригинально встретили маму Оригинальную | 0.086242858 | Objective |
| Известный французский актер Венсан Кассель Одинокий и ви | 0.086116306 | Objective |
| Певица Карина Абуллина опубликовала в Инст Карина Абулл | 0.085758897 | Objective |

| Text | Title | Aggregator score | Expert score |
|------|-------|------------------|--------------|
| Певица Карина Абуллина опубликовала в Инст Карина Абулл | 0.228169926 | reliable |
| Известный французский актер Венсан Кассель Одинокий и ви | 0.22495325 | reliable |
| Видео о том, как оригинально встретили маму Оригинальную | 0.223814277 | reliable |
| Казахстанские школьники, которые завоевали С главой Google | 0.223540345 | reliable |
| Житель Великобритании Джеймс Крэйн (James) Реакция сорв | 0.222874905 | reliable |

| Text | Title | Aggregator score | Expert score |
|------|-------|------------------|--------------|
| Организаторы зимних Олимпийских игр-2018 Скандал на Клин | 0.201449026 | reliable |
| Директора школы осудили в Карагандинской с Директора школ | 0.200752131 | reliable |
| Арестован подозреваемый по делу о провокации Арестован подозр | 0.198751464 | reliable |
| Лидер КНДР Ким Чен вн, выступая в понедель | 0.196937454 | difficult to determine |
| Игрок турецкого "Буджаспора" Абдуллах Бали Турецкий футбол | 0.196062357 | reliable |

Figure 3. Assessment of poh:J.azat:wn, obj:J.вlty and relабlhty of media texts
5. Conclusion
The article provides an overview of methods to detect the destructive messages in the Internet environment. It is shown that for this purpose the approaches based on the analysis of syntactic laws are applied; the analysis of the semantic information contained in the text and its correlation with the text corpus; crowd sourcing methods; identifying patterns of user behavior in social networks; consideration of additional information, etc. Good results can be achieved under certain conditions: having access to the traffic of social networks and other online news resources, the ability to organize or get crowd sourcing results, etc. With restrictions on the fulfillment of these conditions, the identification of destructive messages can be performed based on indirect signs. The results of revealing destructive news inside the news messages in Russian, posted on the websites of the Republic of Kazakhstan using dictionaries containing words indicating the following text properties: the presence of verifiable facts, politicization, a call to action, negative or positive tonality, and manipulativeness are presented.

The results of the application of the algorithm show that in some cases it gives a good agreement with expert estimates. On the other hand, it allows the ranking of the texts in situations where the texts are very close in their characteristics. The above texts are taken from one sufficiently reliable source, therefore the reliability and objectivity of the texts are close. However, this approach requires very careful tuning of dictionaries and the usage of additional algorithms to determine a wider range of text parameters.

Acknowledgments
The work was supported by grant No. BR05236839 of the Ministry of Education and Science of the Republic of Kazakhstan and by the theme of the state task No. AAA-A17-117120670141-7 (No. 0316-2018-0009) of the Russian Federation.

References
[1] Lazer D M J et al 2018 The science of fake news Science 359 (6380) pp 1094–1096
[2] 2017 Anonymous comments on websites are banned in Kazakhstan https://www.zakon.kz/4896289-anonimnye-komentarii-na-saytah.html
[3] Tavernise S 2016 As Fake News Spreads Lies, More Readers Shrug at the Truth The New York Times Dec. 6 https://www.nytimes.com/2016/12/06/us/fake-news-partisan-republican-democrat.html
[4] Rubin V L, Chen Y and Conroy N J 2015 Deception detection for news: three types of fakes Proc. of the Assoc. for Inf. Sci. and Tech. doi:10.1002/pr2.2015.145052010083
[5] Ahmed H, Traore I and Saad S 2018 Detecting opinion spams and fake news using text classification Security and Privacy 1 (1) doi:10.1002/spy.2.9
[6] Pérez-Rosas V et al 2017 Automatic Detection of Fake News arXiv.org/abs/1708.07104
[7] Rubin V L et al 2016 Fake news or truth? Using satirical cues to detect potentially misleading news Proc. of the Sec.Workshop on Comp. Appr. to Deception Detection pp 7–17
[8] Horne B D and Adali S 2017 This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news arXiv:1703.09398
[9] Conroy N J, Rubin V L and Chen Y 2015 Automatic deception detection: Methods for finding fake news Proc. of the 78th ASIS&T Ann. Meet. doi:10.1002/pr2.2015.145052010082
[10] Guacho G B et al 2018 Semi-supervised Content-based Detection of Misinformation via Tensor arXiv preprint arXiv:1804.09088
[11] Tacchini E et al 2017 Some like it hoax: Automated fake news detection in social networks arXiv:1704.07506
[12] Jin Z et al 2016 News Verification by Exploiting Conflicting Social Viewpoints in Microblogs AAAI pp 2972–2978
[13] Chadwick P 2018 Why fake news on social media travels faster than the truth The Guardian https://www.theguardian.com/commentisfree/2018/mar/19/fake-news-social-media-twitter-mit-journalism
[14] Zhao Z et al 2018 Fake news propagate differently from real news even at early stages of spreading arXiv:1803.03443
[15] Litvinova T A and Seredin P V 2013 Search for signs of lies in the written text: modern methods and approaches Proc. of XXIV Int. sci. and pract. Conf. Novosibirsk: SibAK pp 126–133
[16] Issers O S 2014 Media fakes: between truth and hoax Commun. stud. 2 pp 112–123
[17] Litvinova O et al 2017 Deception detection in russian texts Proc. of the Student Research Workshop at the 15th Conf. of the Eur. Ch. of the Ass. for Comp. Ling. pp 43–52
[18] Corpus Russian Deception Bank http://rusprofilinglab.ru/korpusyi-tekstov/korpus-russian-deception-bank/
[19] News agency «Panorama» https://panorama.pub
[20] Mashechkin I V, Petrovsky M I and Tsarev D V 2013 Methods for calculating the relevance of text fragments based on thematic models in the problem of automatic annotation Comp. meth. And programming 14 (1) pp 91–102
[21] Vorontsov K V and Potapenko A A 2012 Regularization, robustness and sparsity of probabilistic thematic models Comp. research and modeling 4 (4) pp 693–706
[22] Parkhomenko P A, Grigoriev A A and Astrakhantsev N A 2017 Review and experimental comparison of text clustering methods Proc. of ISP RAS 29 (2) pp 161–200
[23] Vorontsov K et al 2015 Bigartm: Open source library for regularized multimodal topic modeling of large collections Int. Conf. on Analysis of Images, Soc. Networks and Texts pp 370–381.
[24] Blei D M, Ng A Y and Jordan M I 2003 Latent Dirichlet allocation Jour. of Machine Learning Research 3 (Jan) pp 993–1022
[25] Jelodar H et al 2018 Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey Multimedia Tools and Applications doi:10.1007/s11042-018-6894-4
[26] Saaty T L 1990 Decision making for leaders: the analytic hierarchy process for decisions in a complex world (Univ. of Pittsburgh)
[27] Saaty T L 2015 About measuring of the intangible. An approach to relative measurements based on the main vector of the matrix of paired comparisons Cloud of Sci. 2 (1) pp 5–39
[28] Jaccard P 1901 Distribution de la flore alpine dans le Bassin des Dranses et dans quelques regions voisines Bull. Soc. Vaudoise sci. Natur 37 (140) pp 241–272
[29] Kravtsov H A et al 2018 Trainable model calculations on classifications Electronic simulation 40 (3) pp 63–76