Вісник Тернопільського національного технічного університету
https://doi.org/10.33108/visnyk_tntu
Scientific Journal of the Ternopil National Technical University
2022, № 1 (105) https://doi.org/10.33108/visnyk_tntu2022.01
ISSN 2522-4433. Web: ***visnyk.tntu.edu.ua***

UDC 004.021

THE METHOD OF TRUST LEVEL OF PUBLICATIONS HOSTED IN VIRTUAL COMMUNITIES

Anna Synko

Lviv Polytechnic National University, Lviv, Ukraine

Summary*. The proposed model of data collection and analysis from thematic virtual communities using known information analysis techniques: scoring and parsing. Open communities were selected for the study, namely their architecture and main components: information content (title, description, posts, topics of the event) and audience (community members). To select relevant, informative, reliable publications, the scoring method is used which reflects the level of trust of the authors of the publication in the form of weighted indicators of a set of certain characteristics. Data collection is a combined approach, as virtual communities are dynamic in the content of the data and their content depends on the actions of the participants. To parse posts from virtual communities, it was decided to use ImportXML function in Microsoft Excel, which allows you to collect data from different sources, and then sample, analyze, and select the presentation of results using other built-in tools of this program.*

Key words*: data analysis, parsing, data processing, virtual community, score methodology, ImportXML function.*

https://doi.org/10.33108/visnyk_tntu2022.01.068 *Received 26.01.2022*

Problem statement. The tendencies of the development of modern information-communication technologies (ICT) aim at communication support in all spheres of human activity. The main use of the Internet network is to provide society with the functional result in the communication for opinions exchange. That is why numerous websites contain virtual communities of different types and sizes. These are the communities providing the environment of communication, exchange, and propagation according to the request of their creation.

Virtual communities (VC) differ from other virtual forms of social integration as, firstly, they use common for all groups communication channel, and, secondly, have distinct boundaries (membership), and thirdly, they are characterized by common goals/interests of their participants.

Special attention should be paid to the structure of virtual communities. A virtual community can be divided into open or closed groups, clubs according to the interests which have their own moderators. But not all virtual communities have a mechanism of group formation according to the interests.

Nowadays, the number of open virtual communities is constantly growing as they are a very good source of information. For example, in the field of IT the following problems occur while working with information: the information about the developed software is not complete; data about the software are outdated or absent. But we should not forget about some disadvantages of virtual communities: a big volume of available data; enormous gain of information every minute; some extra wrong data and experience (common knowledge) of the authors which place their publications. It is almost impossible to obtain quality, valid information in a very short period of time. So, it is necessary to investigate and develop methods and instruments of search, identification, analysis, classification, and selection of all available information for its further structuring and representation. The development of a method for posts trust level formation based on the latest techniques of data analysis is described in the

Corresponding author: Anna Synko; e-mail: anna.i.synko@lpnu.ua

paper under consideration. This evaluation may contribute to choosing actual and reliable publications.

There are the following techniques of data analysis: graph-based modeling, scoring, crawling, parsing, Big Data. VC can be presented in the form of a graph having interconnected nodes whose vertices are the users and the edges are the relations between them [1]. This approach provides some terminology for further mathematical description but it is not used in the paper under discussion.

The scoring method makes an evaluation as a weighted sum of the selection of certain characteristics for further decisions making and the system behavior forecasting. This technique is very helpful for our study, namely not only for calculation of the trust level to the publications, but for determination of the materials publication dynamics as well. A scoring model can be presented by the following methods of classification: classification tree, neuron networks, statistical methods which are based on discriminant analysis, genetic algorithms, linear programming etc. [2].

The crawling method is used for content detection and collecting on the Internet network. This technique will be useful for the available posts collection in the specified VC by a search agent (crawler) which consists of a set of computers. A crawler can find and choose posts on the specified search request more quickly than a user by means of a web browser. In fact, crawler can ask thousands of different VCs simultaneously. Moreover, it is very important that the crawler checks the information.

The parsing method allows us to collect data regularly from the web pages and withdraw some specific information from them. The technique has the following advantages:

- quick collection of data in any mode;
- error prevention;
- regular check by the specified time interval;
- data presentation in any form;
- providing uniform load on the website where the parsing is taking place (to prevent DOS-attack effect).

The very process of the platform content searching and analysis by means of their page parsing has been described in the papers [3-5].

The main stages of consideration (parsing) of the VC posts are the following:

- receiving the page HTML code (or API).
- information analysis and processing. After receiving the page code one can choose the necessary data by removing extra attributes and unnecessary information. Then the data must be structured.
- Information presentation in a certain structured form – report formation. The data are usually stored in databases, text files, etc.

The parsing method works only for one VC. For data analysis and collection from other VCs one should carry out the same actions with their HTML pages.

We must pay attention to the fact, that virtual communities where useful, topical information can be found are not numerous, so it was proposed to invite an expert who will select proper virtual communities.

While developing software decisions, it is possible to work both with the page code and with API use as well, which is usually provided by platform owners where VC is located. API, as the method of abstracting between two software of different levels, provides the developers with the complete set for quick development of programs. API can be used for web-based systems, databases, operating systems, software libraries, hardware, etc.

Ready API use has simplified the procedure of data parsing from virtual communities, as the developer doesn't need to find HTML page code by himself and extract the required information. API use describes the principle of the system in the best way [6]. But not all platforms provide open API for general use. For example, the platform Facebook does not reveal complete information about users. Thus, virtual communities on this site won't have these data.

Analysis of the known results of the research. The study of virtual communities is a long experience in different sciences (including the science of communication and society, mathematical analysis and the theory of graphs, marketing, artificial intelligence, and software). These branches reflect the behavior of different generations of the population during a certain period of time in the Internet network. The matter of communicative process analysis in the communities and their information impact on society was considered in numerous scientific papers of national and foreign scientists, namely Peleshchyshyn A. M., Peleshchyshyn O. P., Huminskyi R. V., Fedushko S. S. [12-14], H. Rheingold, W. Ebner and others. Scientist O. Trach conducted investigations of virtual communities: analysis of types, construction of directions and stages of life cycle, development of formal models, etc. [11]. All these studies are the basis for further research.

Many papers were devoted to the data analysis use, namely the parsing technique, starting from the term definition, the sphere of usage of this method, and ending with the certain research. The following papers were quite useful for this article writing: A. Britvin [7], David M. W. Philips [8], Beznosyk O. and Kulyk O. [9].

An insignificant number of scientific investigations on the analysis of posts in virtual communities based on their trust level assessment have caused the study under consideration to carry out.

The aim of the paper. To develop the method defining the trust level to the posts which are in the open, preliminary chosen by the expert virtual communities.

To achieve the goal set it is necessary to do the following tasks:

- study the structure of VC – information source;
- develop the indices of publications evaluation.

To present the results of the developed method one should conduct the research using the above-mentioned techniques of information analysis.

Problem setting. Presentation of the virtual community structure.

As virtual communities are a very good source of information, it is necessary to study them.

There are the following types of VC:

- open (anybody can join VC who is registered on a certain site);
- closed (it is necessary to have permission to join the community).

Further, we will study open VCs, which are regulated by the Law of Ukraine «about information»[10] – information access modes – open information (there is no restriction on information receiving, processing, storing, and spreading by all the people involved).

The main components of VC are information content and audience [11–14]. The VC page is structured and divided into sections where a user (owner, member of the community) can post information about the preliminarily determined character. The typical structure of VC involves the following information blocks (Fig. 1.): name of the community, brief description, posts (a set of messages), topics; events, and members of the group.

Components of VC can be presented as a tuple:

$$VC_i =< Name_i, Description_i, Posts_i, Topics_i, Events_i, Audience_i >; \qquad (1)$$

where $Name_i$ – name of i- community, $Description_i$ – brief description of the community, $Posts_i$ – all posts of i- community, $Topics_i$ – topics of i- community, $Events_i$ – events of i- community, $Audience_i$ – participants of i- communit

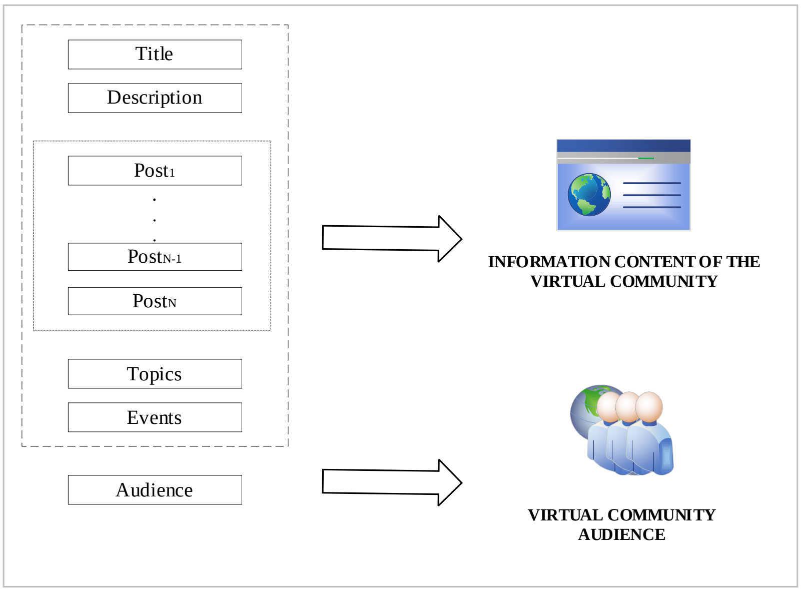

Title
Description
Post1
Post$_{N-1}$
Post$_N$
Topics
Events
Audience
INFORMATION CONTENT OF THE VIRTUAL COMMUNITY
VIRTUAL COMMUNITY AUDIENCE

Figure 1. Virtual community architecture

During their communication the participants of VC create information content presented as a set of posts (messages, publications), which form 90% of all the available content. The number of posts in i-virtual community is represented by the set:

$$Post_i = \{Post_{ij}\}_{j=1}^{N^{(Post_j)}}; \qquad (2)$$

where $N^{(Post_j)}$– the total number of posts in the community.

A post is an atomic unit and involves the following components (fig. 2):

- information about the author;
- title of the message;
- text;
- date of creation;
- reactions to the post (comments, likes, spread).

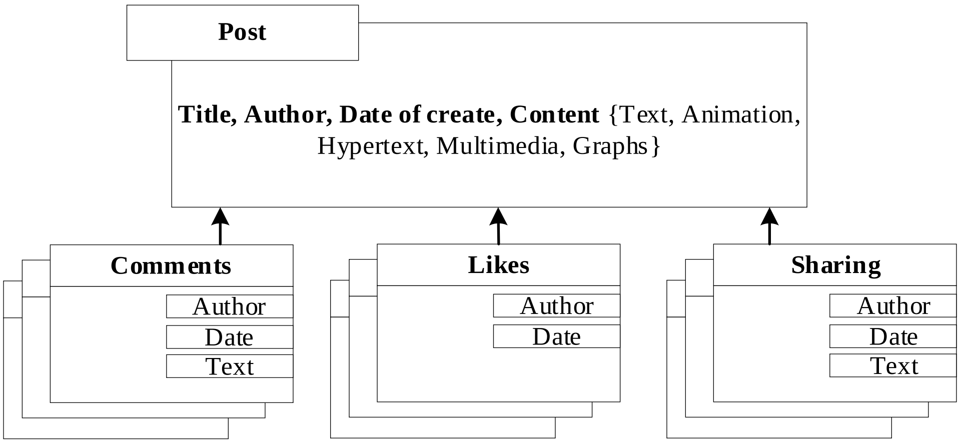

Post
Title, Author, Date of create, Content {Text, Animation, Hypertext, Multimedia, Graphs}
Comments
Author
Date
Text
Likes
Author
Date
Sharing
Author
Date
Text

Figure 2. Components to form a post

A post structure can be written as a tuple:

$$Post_i = < PostTitle_i, PostAuthor_i, PostDate_i, Content_i, PostReaction_i >; \qquad (3)$$

where $PostTitle_i$ – name of the post, $PostAuthor_i$ – the author of the publication, $PostDate_i$ – date of the publication, $Content_i$ – information filling (content) of the post, $PostReaction_i$ – reactions on the publication (comments, likes, spreads).

The posts can be created, edited, deleted, shared, commented, evaluated (to form their rating).

As a rule, section «Information about the author» of the post contains some personal data which were left by the user when he or she was registered (name and surname, sex, age, telephone number, country, place of work/study, interests and hobbies, etc). Some other important data dealing with the software development include: work experience and interests of the user.

Moreover, the evaluation of the author's posts by other users is an important factor (how the specified publication was useful, urgent and meaningful), due to which the user rating can be formed. Having obtained these data, one can calculate «trust assessment to the user».

Development of the assessment method of the level of trust to the post based on the scoring method. To select actual, informative, reliable publications from the virtual communities one should develop assessment describing the level of trust. Such assessment may include (fig. 3): level of trust to the user, post rating and actuality.

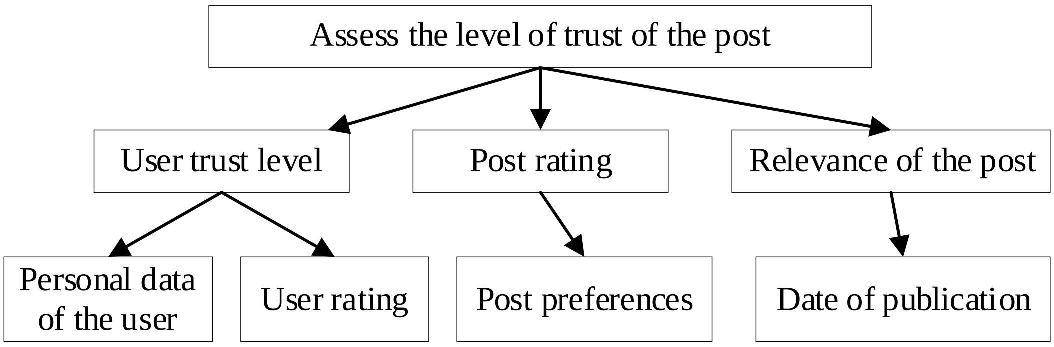


Figure 3. Component of assessment of the level of trust to the post

To calculate the assessment of the level of trust to the post by the natural restrictions, we have obtained $0 \leq LevOfTrustPostEst_j \leq 1$ [15]. In the end, the level of trust to the author of a certain post j is written as follows:

$$LevOfTrustPostEst_j = \frac{LevOfTrustUserEst_j + PostRate_j + RelevPost_j}{n}; \qquad (4)$$

where n is the number of marks which will be used to calculate the level of trust to the post (here $n = 3$); $LevOfTrustUserEst_j$ is the level of trust to the user, who published j post, $0 \leq LevOfTrustUserEst_j \leq 1$; $PostRate_j$ – is the rating of *j* post, formed by the users of the community, $0 \leq PostRate_j \leq 1$; $RelevPost_j$ is the actuality of *j*-post, determined by the date of, $0 \leq RelevPost_j \leq 1$.

$$LevOfTrustPostEst_j = \begin{cases} high\ level, & when\ 1 \le LevOfTrustPostEst_j \le 0{,}75 \\ middle\ level, when\ 0{,}75 < LevOfTrustPostEst_j \le 0{,}5 \\ low\ level, when\ 0{,}5 < LevOfTrustPostEst_j \le 0{,}3 \\ very\ low\ level, when\ 0{,}3 < LevOfTrustPostEst_j \le 0 \end{cases}; \quad (5)$$

The author proposes not to take into account the posts of the very low level of trust while you are searching some information.

The assessment «Trust level of users» consists of the following components:

- personal data of a user;
- user rating.

Personal data of a user are interesting for the estimation from the following points of view: available work experience in the certain branch, as well as likes and interests. Work experience is more important than user's interests, as it is a complex of knowledge, abilities and skills obtained by people (both theoretical and practical). Interests are not always user's experience. So, it is quite reasonable to make work experience more valuable.

The assessment «Trust level of users» is presented by the following formula:

$$LevOfTrustUserEst_j = \frac{PersUserData_j + UserRate_j}{k}; \quad (6)$$

where k is the number of estimations used for calculation of the trust level to the user (here $k = 2$); $PersUserData_j$ – assessment of personal data of the user of *j*-post (formula 7), $UserRate_j$ – rating of the user of *j*-post (in most cases it is an arithmetical average of all estimations made by other users of the author). But not all communities receive the assessment of users. So, it should be taken into account while using formula 6.

The assessment «Personal data of a user» of *j*-post is presented by the following formula:

$$PersUserData_j = \frac{(UserExpEst_j \times 3) + UserInterEst_j}{m}; \quad (7)$$

where m – the number of estimations used to calculate the user's personal data (here $m = 4$); $UserExpEst_j$ is the experience of a user who published the post in a certain branch (as it is more important than the user's interests, it has the coefficient 3); $UserInterEst_j$ are interests and likes of the user.

The rating of a publication of a certain user can be calculated (an arithmetical average of all published posts of the community member) by the following formula:

$$PostRate_j = (\sum_{j=1}^{k} EstimatePost_j)/k; \quad (8)$$

where $EstimatePost_j$ is an assessment of a post of a certain user, k is the general number of posts of a certain user.

Actuality of posts is the degree of adequacy of the specified piece of information at present period of time. Actuality can be determined by the following aspects:

- objectively – according to the date of publication;
- subjectively – user's experience (how actual are the data he/she has).

The section «Trust level to the user» partially covers the subjective estimation. It is the objective assessment that should be applied. It is known, that for each branch the actuality of information is determined by different terms. For example, data in financial sphere are changing more quickly than in any other. So, for each branch it is necessary to adjust these indices of actuality.

As each platform has its own set of characteristics, which do not necessarily correspond to the total set of components required to calculate the trust level (fig. 3), the comparative characteristics of the platforms was given.

Table 1

Use of trust level indices to the post according to platforms

Platforms	Trust level		Post rating	Post actuality
	User personal data	User rating		
Reddit	-	+	+	+
Facebook	+	-	+	+
Habr	+	+	+	+
itProger	-	-	-	+
CyberForum	+	+	-	+
StackOverflow	+	-	-	+
Tproger	+	-	+	+
Proglib	+	-	+	+
DOU (forums)	+	-	+	+

The parsing method use for posts from virtual communities.

Nowadays, the following methods are used to find and collect data:

- manual;
- automated;
- combined – combination of manual and automated in detecting and collecting information in a VC.

Manual collection and analysis of information is universal and individual, but it is rather slow when compared with other methods (for example, the automated method does the same tasks hundreds of times quicker).

It should be mentioned, that virtual communities are dynamic from the point of view of data filling, as their content depends on the actions of participants. It should be taken into account at the automated consideration of the data, as the page code is not complete. So, a good decision is to forecast and imitate the following actions of the user.

The most efficient decision was the use of the combined method of information collection when an expert chooses the sources (virtual communities) and determines the correspondent attributes.

To carry out the parsing of posts from virtual communities, we have decided to use the program Microsoft Excel function ImportXML(), enabling to collect data from different sources, and after that select, analyze and choose the presentation of results due to the in-built instruments of the program.

Syntax of the function ImportXML() is the following: = IMPORTXML(url, xpath_query).

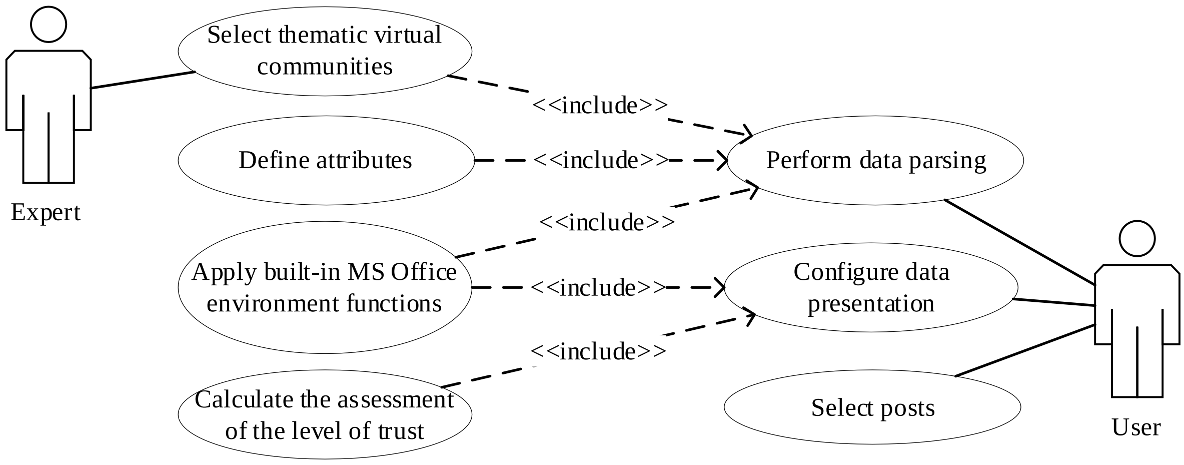

Expert
Select thematic virtual communities
Define attributes
Apply built-in MS Office environment functions
Calculate the assessment of the level of trust
<<include>>
<<include>>
<<include>>
<<include>>
<<include>>
Perform data parsing
Configure data presentation
Select posts
User

Figure 4. Diagram of collecting and analyzing posts from virtual communities

The algorithm of our study conducting:

1. Thematical VCs selection – list formation.
2. Data parsing.

2.1. Determination and adjustment of the attributes which may contain the required information to carry out the investigation:

- for the estimation of trust level to the publication;
- for the content selection.

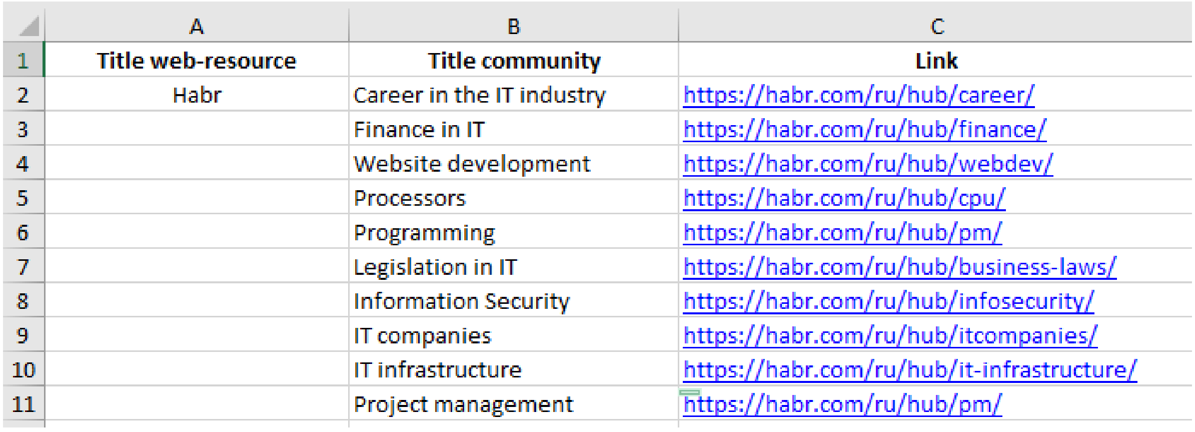
3. Analysis and selection of the posts by the method of estimation of trust level to the posts' authors.

- putting into single calculation system of data received from different platforms.

4. Formation of data presentation to make conclusions.

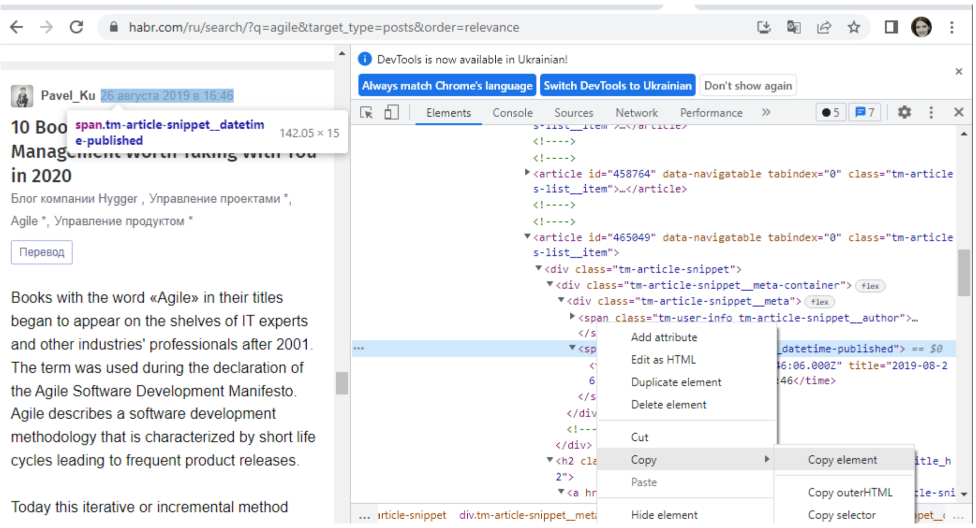
The sources of the data for the analysis may be: one community; many communities from one platform; many communities from different platforms (it is necessary to adjust the indices for different platforms separately).

To carry out the study ten communities were selected on the platform Habr (fig. 5).



	A	B	C
1	**Title web-resource**	**Title community**	**Link**
2	Habr	Career in the IT industry	https://habr.com/ru/hub/career/
3		Finance in IT	https://habr.com/ru/hub/finance/
4		Website development	https://habr.com/ru/hub/webdev/
5		Processors	https://habr.com/ru/hub/cpu/
6		Programming	https://habr.com/ru/hub/pm/
7		Legislation in IT	https://habr.com/ru/hub/business-laws/
8		Information Security	https://habr.com/ru/hub/infosecurity/
9		IT companies	https://habr.com/ru/hub/itcompanies/
10		IT infrastructure	https://habr.com/ru/hub/it-infrastructure/
11		Project management	https://habr.com/ru/hub/pm/

Figure 5. List of selected communities



habr.com/ru/search/?q=agile&target_type=posts&order=relevance

Pavel_Ku 26 августа 2019 в 16:46

span.tm-article-snippet__datetime-published 142.05 × 15

10 Boo... Manag... in 2020

Блог компании Hygger , Управление проектами *, Agile *, Управление продуктом *

Перевод

Books with the word «Agile» in their titles began to appear on the shelves of IT experts and other industries' professionals after 2001. The term was used during the declaration of the Agile Software Development Manifesto. Agile describes a software development methodology that is characterized by short life cycles leading to frequent product releases.

Today this iterative or incremental method

DevTools is now available in Ukrainian!

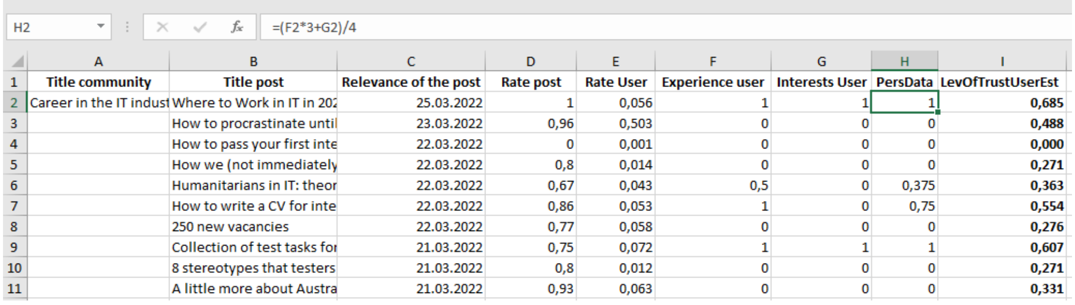
Always match Chrome's language | Switch DevTools to Ukrainian | Don't show again

Elements | Console | Sources | Network | Performance | » | ● 5 | 7

```
<!---->
<!---->
▸<article id="458764" data-navigatable tabindex="0" class="tm-articles-list__item">…</article>
<!---->
<!---->
▾<article id="465049" data-navigatable tabindex="0" class="tm-articles-list__item">
  ▾<div class="tm-article-snippet">
    ▾<div class="tm-article-snippet__meta-container"> flex
      ▾<div class="tm-article-snippet__meta"> flex
        ▸<span class="tm-user-info tm-article-snippet__author">…
         </s
        ▾<sp                                     _datetime-published"> == $0
           <                                     6:06.000Z" title="2019-08-2
           6                                     46</time>
         </s
       </div
       <!---
     </div>
    ▾<h2 cla                                     itle_h
     2">
      ▾<a hr                                     le-sni
```

Figure 6. Selection of attributes in HTML code

Furthermore, when all the necessary information is uploaded, it is necessary to structure it, i.e. the characteristics are within the boundaries [0, 1]. As a result, the assessment of the level of trust to the post can be made (figure 7).



H2 | =(F2*3+G2)/4

	A	B	C	D	E	F	G	H	I
1	Title community	Title post	Relevance of the post	Rate post	Rate User	Experience user	Interests User	PersData	LevOfTrustUserEst
2	Career in the IT indust	Where to Work in IT in 202	25.03.2022	1	0,056	1	1	1	**0,685**
3		How to procrastinate until	23.03.2022	0,96	0,503	0	0	0	**0,488**
4		How to pass your first inte	22.03.2022	0	0,001	0	0	0	**0,000**
5		How we (not immediately	22.03.2022	0,8	0,014	0	0	0	**0,271**
6		Humanitarians in IT: theor	22.03.2022	0,67	0,043	0,5	0	0,375	**0,363**
7		How to write a CV for inte	22.03.2022	0,86	0,053	1	0	0,75	**0,554**
8		250 new vacancies	22.03.2022	0,77	0,058	0	0	0	**0,276**
9		Collection of test tasks for	21.03.2022	0,75	0,072	1	1	1	**0,607**
10		8 stereotypes that testers	21.03.2022	0,8	0,012	0	0	0	**0,271**
11		A little more about Austra	21.03.2022	0,93	0,063	0	0	0	**0,331**

Figure 7. Display of the received data in uniform numerical system

Results of the study. Receiving the evaluation of the trust level to the posts, they were divided into four groups (high, middle, low and very low levels of trust), which were presented in the formula 5. According to the figure 8, a circle diagram was built describing the percentage of publications relative to the trust level.

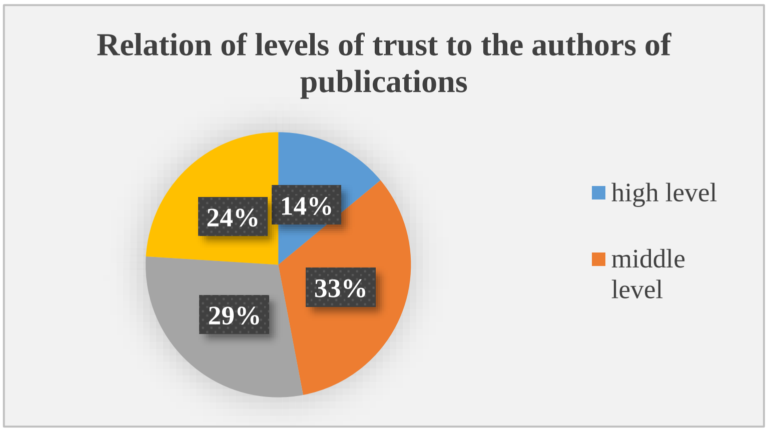


Figure 8. Presentation of results – diagram

As we can see from the picture: 14% of users have a high level of trust to the publications as they give the complete information about them; 33% of the authors of publications have middle level of trust because they give incomplete personal data or these data are not valuable; 29% of messages ore of low level; 24% of posts have very low level. Taking into account all the above-mentioned, the posts of very low trust level should not be taken into consideration, but only 76% of all publications, which, in its turn, have their ranking.

Conclusions. The method of analysis of trust level to the posts based on the modern techniques of data selection has been developed in the article under discussion. According to this method, each post receives its assessment of the trust level as a sum of indices that form it, and it belongs to a certain group presented in formula 5. Due to the method used the obtained estimations of the posts will help to choose actual, reliable publications thanks to the strictly determined trust levels.

Since the virtual communities are dynamic by their data content, the combined method of data parsing was chosen (an expert selects some thematic communities and manually adjusts the attributes containing the required data).

To conduct the study a web resource involving all indices presented in Table 1 was chosen. The results of the study are given in a circle diagram. As we can see, only 14% of all authors have given the complete information about them, and consequently, they obtained the highest level of trust. As, according to the data on Figure 7, it is these attributes that were not filled in most users.

The publications with very low level of trust are not recommended to take into account, they even should be ignored, when the information is analyzed. Since the authors of such publications do not present their personal data and have low user rating in the community.

References

1. Jiang W., Wang G., Bhuiyan Z. A., Wu J. Understanding graph–based trust evaluation in online social networks: Methodologies and challenges. ACM Computing Surveys (CSUR). Vol. 49. No. 1. 2016. P. 1–35. DOI: https://doi.org/10.1145/2906151
2. Lunkina T. I., Velkhovatska K. O. Metody upravlinnia ryzykamy spozhyvchoho kredytuvannia, “Young Scientist”. Vol. 2 (17). 2015. P. 157–160. [In Ukrainian].
3. Zhou B., Zhao H., Puig X., Fidler S., Barriuso A., Torralba A. Scene parsing through ade20k dataset. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. P. 633–641. DOI: https://doi.org/10.1109/CVPR.2017.544
4. Reddy S., Tackstrom O., Petrov S., Steedman M., Lapata M. Universal semantic parsing. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, Association for Computational Linguistics. 2017. P. 89–101. DOI: https://doi.org/10.18653/v1/D17-1009
5. Zhao H., Shi J., Qi X., Wang X., Jia J. Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. P. 2881–2890. DOI: https://doi.org/10.1109/CVPR.2017.660
6. Prylutskyi P. V., Servis z ahrehuvannia platizhnykh system. Kyiv: Igor Sikorsky Kyiv Polytechnic Institute, 2020. 98 p. [In Ukrainian].
7. Britvin A., Tkachuk R. Parsynh danykh z veb storinok. V All-Ukrainian scientific-practical conference of young scientists, students and cadets “Cybersecurity and Information Technology”. CIT 2021. Lviv. November 26. 2021. P. 13–14. [In Ukrainian].
8. Phillips David M.W., Web Scraping with Excel. CreateSpace Independent Publishing Platform. March 6. 2016. P. 62.
9. Beznosyk O., Kulyk O. Automatized parsing of bibliographic references. The 8th International Scientific and Practical Conference “Computer Modeling in Chemistry, Technologies and Systems of Sustainable Development – ChTCTST-2020”. Kyiv: Igor Sikorsky Kyiv Polytechnic Institute, 2020. May 19–22. P. 364–372. [In Ukrainian].
10. Zakon Ukrainy “Pro informatsiiu”, Information of the Verkhovna Rada of Ukraine. № 2658-XII from 02.10.92 edited by 02.06.2016. URL: http://zakon3.rada.gov.ua/laws/show/2657-12. [In Ukrainian].

11. Trach O. R. Mathematical support and software for organization of the life cycle of virtual communities: Thesis for a Ph.D degree, Lviv Polytechnic National University, Ministry of Education and Science of Ukraine. Lviv. 2018. P. 172. [In Ukrainian].
12. Fedushko S. Development of a software for computer-linguistic verification of socio-demographic profile of web-community member. Webology. Vol. 11. No. 2. 2014. Article 126.
13. Fedushko S., Mastykash O., Syerov Y., Shilinh A. Model of Search and Analysis of Heterogeneous User Data to Improve the Web Projects Functioning. Advances in Computer Science for Engineering and Education IV. ICCSEEA 2021. Lecture Notes on Data Engineering and Communications Technologies, Springer, Cham. Vol 83. 2021. P. 56–74. DOI: https://doi.org/10.1007/978-3-030-80472-5_6
14. Fedushko S., Syerov Yu., Skybinskyi O., Shakhovska N., Kunch Z. Efficiency of Using Utility for Username Verification in Online Community Management. Proceedings of the International Workshop on Conflict Management in Global Information Networks (CMiGIN 2019), Lviv, Ukraine, November 29, 2019. CEUR-WS.org, Vol-2588. P. 265–275.
15. Markovets O. V. Mathematical and software of interaction of citizens with authorities in heterogeneous web environments: Thesis for a Ph.D degree, Lviv Polytechnic National University, Ministry of Education and Science of Ukraine. Lviv. 2015. P. 144. [In Ukrainian].

Список використаної літератури

1. Jiang W., Wang G., Bhuiyan Z. A., Wu J. Understanding graph–based trust evaluation in online social networks: Methodologies and challenges. ACM Computing Surveys (CSUR). Vol. 49. No. 1. 2016. P. 1–35. DOI: https://doi.org/10.1145/2906151
2. Лункіна Т. І., Вельховацька К. О. Методи управління ризиками споживчого кредитування. Молодий вчений. 2015. № 2 (17). С. 157–160.
3. Zhou B., Zhao H., Puig X., Fidler S., Barriuso A., Torralba A. Scene parsing through ade20k dataset. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. P. 633–641. DOI: https://doi.org/10.1109/CVPR.2017.544
4. Reddy S., Tackstrom O., Petrov S., Steedman M., Lapata M. Universal semantic parsing. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, Association for Computational Linguistics. 2017. P. 89–101. DOI: https://doi.org/10.18653/v1/D17-1009
5. Zhao H., Shi J., Qi X., Wang X., Jia Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. P. 2881–2890. DOI: https://doi.org/10.1109/CVPR.2017.660
6. Прилуцький П. В. Сервіс з агрегування платіжних систем. К.: КПІ ім. Ігоря. Сікорського, 2020. 98 с.
7. Брітвін А., Ткачук Р. Парсинг даних з веб сторінок: V Всеукраїн. наук.-практ. конф. молодих учених, студентів та курсантів «Інформаційна безпека та інформаційні технології», ІБІТ-2021 (м. Львів, 26 Листопада.). Львів, 2021. С. 13–14.
8. Phillips, David M. W. Web Scraping with Excel. CreateSpace Independent Publishing Platform. March 6. 2016. P. 62.
9. Безносик О. Ю., Кулик А. В. Автоматизований парсинг бібліографічних посилань: VIII міжнар. наук.-практ. конф. «Комп’ютерне моделювання в хімії та технологіях і системах сталого розвитку», КМХТ-2020. (м. Київ, Травень 19–22.). Київ: КПІ ім. Ігоря. Сікорського, 2020. С. 364–372.
10. Закон України «Про інформацію», Відомості Верховної Ради України. – №2658-XII від 02.10.92 відредагований 02.06.2016. URL: http://zakon3.rada.gov.ua/laws/show/2657-12.
11. Трач О. Р. Математичне та програмне забезпечення організації життєвого циклу віртуальних спільнот: див. … канд. техніч. наук / Національний університет «Львівська політехніка». Львів, 2018. С. 172.
12. Fedushko S. Development of a software for computer-linguistic verification of socio-demographic profile of web-community member. Webology. Vol. 11. No. 2. 2014, Article 126.
13. Fedushko S., Mastykash O., Syerov Y., Shilinh A. Model of Search and Analysis of Heterogeneous User Data to Improve the Web Projects Functioning. Advances in Computer Science for Engineering and Education IV. ICCSEEA 2021. Lecture Notes on Data Engineering and Communications Technologies, Springer, Cham, vol 83., 2021, pp. 56-74. DOI: https://doi.org/10.1007/978-3-030-80472-5_6
14. Fedushko S., Syerov Yu., Skybinskyi O., Shakhovska N., Kunch Z. Efficiency of Using Utility for Username Verification in Online Community Management. Proceedings of the International Workshop on Conflict Management in Global Information Networks (CMiGIN 2019). Lviv. Ukraine. November 29. 2019. CEUR-WS.org, Vol-2588. P. 265–275.
15. Марковець О. В. Математичне та програмне забезпечення організації взаємодії громадян з органами влади в гетерогенних веб-середовищах: див. … канд. техніч. наук / Національний університет «Львівська політехніка». Львів, 2015. С. 144.

УДК 004.021

МЕТОД РІВНЯ ДОВІРИ ДО ПУБЛІКАЦІЙ, РОЗМІЩЕНИХ У ВІРТУАЛЬНИХ СПІЛЬНОТАХ

Анна Синько

Національний університет «Львівська політехніка», Львів, Україна

Резюме. *Наведено структуру віртуальних спільнот. Визначено, що спільноти постають гарним джерелом інформації для будь-якої сфери діяльності, але потребують ретельного відбору інформації серед великого об'єму даних для подальшого їх аналізу, класифікації та представлення, адже не все інформаційне наповнення є достовірним та актуальним.*

Для проведення дослідження відповідно до Закону України «Про інформацію» обрано відкриті спільноти, які містять загальнодоступну інформацію.

Маючи таку проблему, запропоновано розробити метод, що надає оцінювання рівня довіри до дописів, які користувачі публікують. Дослідження побудовано базуючись на сучасні техніки аналізу даних. Для формування оцінювання рівня довіри в основу покладено модель скорингу, що надає оцінку у вигляді зваженої суми набору певних характеристик для подальшого прийняття рішення. Оцінка рівня довіри до допису утворюється сумою таких показників: аналіз автора публікації (на основі його особистих даних, інтересів та вподобань), актуальність інформації (відповідно до дати публікування матеріалів), рейтинг повідомлення, який формують інші користувачі спільноти, залишаючи свої реакції. Складність дослідження полягає в тому, що кожне джерело є різнорідним та може містити не весь набір характеристик, що необхідні для проведення розрахунків. Тому наведено таблицю, що відображає наявність показників до спільнот.

Для проведення дослідження здійснено парсинг даних згідно з визначеними атрибутами, адже більшість спільнот не мають відкритих готових API. Надалі отриману інформацію приведено в єдину систему підрахунку даних відповідно до природних обмежень від 0 до 1. Після цього обраховано оцінки рівнів довіри та обрано їх представлення у вигляді кругової діаграми, що містить чотири групи оцінок (високий, середній, низький та дуже низький рівні довіри). Публікації з дуже низьким рівнем довіри рекомендовано не враховувати при аналізі інформації. Адже автори таких дописів не надають персональних даних і мають низький користувацький рейтинг у спільноті. Тому заповнення особистого профілю у спільнотах постає важливим для аналізу дописів.

Ключові слова: *аналіз даних, парсинг, опрацювання даних, віртуальна спільнота, мето-скоринг, функція ImportXML.*

https://doi.org/10.33108/visnyk_tntu2022.01.068 *Отримано 26.01.2022*