

The Web-Application Development for Analysis of Social Graph of the User in Network

Yuliia Leshchenko¹[0000-0001-9232-697X], Alina Yelizeva²[0000-0002-8228-9383],
Yuliia Rudchenko³[0000-0002-4085-2254]

¹⁻³ National Aerospace University "Kharkiv Aviation Institute",
Chkalov str.,17, Kharkiv, Ukraine
j.leshchenko@khai.edu¹,
a.elizeva@khai.edu²,
rudchenko96@gmail.com³

Abstract. The paper is devoted to the research of the social graph analysis of the network user. The review of existing methods of analysis and visualization of the social graph is carried out represented. Comparative analysis of existing systems of processing social graphs is carried out represented. Tools for development a web-based application and interacting with the Facebook API are selected. The facilities of using social networking data are analyzed. Features of social networking information analysis is discussed in the paper. The basic algorithm of the program's work is developed, which will ensure the receipt of publications from the social network and their creation in the WEB-application. A consolidated resource for the analysis of social networking information in the content of publications is developed, taking into account its algorithm of work. It will be performed the assignment of a certain interest to the user by analyzing the text of its publication, and displaying interest statistics in the graph, as well as mapping the location of interests on the map of Ukraine.

Keywords: web application, social networks, social graph, visualization, analysis, data processing.

1 Introduction

Nowadays Internet takes important part in human life. Internet of "early" period was represented simply by sites containing some information and links to other resources or file directories. This is not just a collection of services such as online translators, electronic dictionaries or WEB interfaces for e-mail accounts. In the modern Internet, WEB services integrate with desktop applications, taking advantage of their typical interface, behavior and capabilities [1, 2].

Nowadays social networks become more popular. The main idea of social networks is very simple. A social network is a collection of social elements (people, groups, organizations, communities) that exchange different data among themselves [3]. The data can be text messages, images, video and audio files. Formal-

ly such networks are convenient to represent in the form of graphs [4, 5] and apply developed mathematical models for their analysis [6].

Social networks analysis [7] turn into the main research method in contemporary sociology [8, 9], anthropology [10, 11], geography [12-14], social psychology [15], computer science [16, 17], business and organizations [18-20], and also it's a wide-spread theme for investigation and discussion [21].

Today's social network is the perfect place to get information about people [22]. By various data and criteria you can get a lot of information that can be useful in different fields.

The novelty of the research is in the fact that the developed resource analyzes [23] the user's posts in the social network Facebook that can be useful for law-enforcement agencies [24, 25], politics [26, 27] and business [28]. The resource makes it possible to obtain users interest [29] and visualization of research results.

2 Main Material Presenting

Social networks on the Internet continue to be at the height of popularity fame and have become the object of researcher's persistent attention. Moreover, there are new opportunities for their use: they are tightly integrated into many spheres of business and become its integral part.

All popular social networks provide their Application Programming Interfaces (API) which is a set of definitions for the interaction of software various types [30]. The API is usually (but not surely) an abstraction method between low-level and high-level software.

The main features of the Facebook API [31] for the right information getting are conceded.

Linux was chosen as the operating system of Facebook, the bulk of the code is written in PHP (which compiled in C ++ by HipHop). The MySQL database was used as a repository. Memcached was used for caching.

Graph API is the basis of the Facebook platform, which allows you to read and write information to the social network and represents objects (users, photos, events, pages) and links between them (relationships, content and photo tags). Each object in the graph has a unique identifier, using which you can it access as follows: *https://graph.facebook.com/v2.5/<ID>*.

The answer will be given in the form of json.

All objects in the graph can be linked by a relationship. You can check the presence of connection with objects in the following way:

https://graph.facebook.com/<ідентифікатор_об'єкта>/<тип_зв'язку>.

It should be noted if the information is not complete, then the json-response will has the data object with our data and the paging object containing the fields previous and next, which contain the link for the previous page and the next one, respectively. This is very convenient if you need to remove all information from a social network from a particular keyword and its volume is not known in advance.

When review the Graph API [32], you should pay attention to Real-Time updates as the most interesting feature. The presence of this item shows the seriousness of Facebook treatment to integration with other applications. Real-time updates allow outside developers to retrieve information about some of their users' changes. Facebook itself reports that the user wrote a message in the news feed, added a friend, or clicked the Like button. This point is required for periodic interrogating by APIs resources.

The review and analysis of existing systems of social networks processing and visualization is performed. Today, there are several systems for analyzing social data such as Gephi [32], Igraph [33], NetworkX [34] and others. Their comparative analysis is presented in Table 1.

Gephi is a system intended for visualization and basic analysis of medium size graphs (up to 1,000,000 nodes). It has a multifunctional user interface and a rich set of network visualization methods (including real-time visualization of dynamic graphs). The system architecture and open source code provide the addition of new features through the development of plug-ins.

Unlike Gephi, Igraph is not a finished system, but a library of computational functions. Igraph [33] is implemented on C, but has software shells for Python, Ruby and R. The library contains a lot of graph theory methods, both classical and recent, for example, methods for ranking nodes, community search, etc. The Igraph library is designed to analyze large networks in research and training projects.

NetworkX [34] is a Python library for studying graphs and networks. NetworkX is suitable for working on large real-world graphs: for example, for graphs containing more than 10 million nodes and 100 million lines. Depending on the Python Dictionaries data structure, NetworkX is a very efficient, highly scalable and easily portable framework for working with networks and analyzing social networks (see Table 1).

Table 1. Comparative analysis of existing systems for processing and visualization of social networks.

Characteristics	Gephi	Igraph	NetworkX
Site	gephi.org	igraph.sourceforge.net	networkx.lanl.gov
Users	Scientific, educational organizations	Scientific, educational organizations	Scientific, educational organizations
Data level	Up to 1 million nodes and lines	Up to several million nodes and lines	Up to several million nodes and lines
Data acquisition	Missing	Missin	Missing
Data sources	Missing	Missing	Missing
Methods	Visual analysis. Basic statistical methods. Graph theory basic	Multiple methods of graph theory	Multiple methods of graph theory

		methods.		
Analysis mode		Retrospective analysis	Retrospective analysis	Retrospective analysis
Methods		Visual analysis. Basic statistical methods. Graph theory basic methods	Multiple methods of graph theory	Multiple methods of graph theory
Considered objects	ob-	Network structure (nodes, directed and non-directed links)	Network structure (nodes, directed and non-directed links)	Network structure (nodes, directed and non-directed links)
Distribution terms		OpenSource (CDDL 1.0, GPL 3.0)	OpenSource (GPL 2.0+)	OpenSource (BSD License)
Language support		English	English	English

How you can see, the presented systems are primarily focused on the retrospective analysis of graphs, which consist of a relatively little number of nodes (up to one million nodes). Also, such systems work only with one level of representation – the graph, which requires a lot of preliminary preparation and data structuring. There is no trouble of the level of texts analysis and, as a rule, a very scanty arsenal of data mining methods [35].

3 Features of social networking information analysis

Whole various social networks can be described by a tuple of two sets: multiple sets of users and plurality of content. The social graph of users is a multi-graph, which means users form a multi-faceted relationship and may contains additional structural data

$$G = \{V, E_1 \dots E_k, p, d_1 \dots d_k\}, \quad (1)$$

where V – actors; $E_1 \dots E_k$ – types of relationships in the network; p – profile of the actor $V \rightarrow p$; $d_1 \dots d_k$ – corresponding relation $E_i \rightarrow d_i$.

We can describe the content graph of the network, similarly

$$C = \{T, R_1 \dots R_m, O, y_1 \dots y_m\}, \quad (2)$$

where T – set of heterogeneous content of the network; $R_1 \dots R_m$ – relationship between the content; O – content parameters $T \rightarrow O$; $y_1 \dots y_m$ – parameters of the corresponding relation in the middle of the content $R_j \rightarrow y_j$.

Finally, $A \in V \times T$ – the ratio of actors and content

$$A = \{L_1 \dots L_m, e_1 \dots e_n\}. \quad (3)$$

Relation between the actors of network and the content can be considered through the author's message (post) and comments (comments). Messages can be formed like text,

image, audio or video. They notify network participants about some event or occurrence. Comments – is a fixed attitude of actors to messages in a certain form. The simplest analysis of social networks is based on an egocentric graph.

The peaks of this graph correspond to the central actor and his closest neighbors. This graph does not reflect the characteristics of the network, but it can be applied by studying the social roles in the group. Some templates can be detected by analyzing the topology of the network. But in most cases you need to know the characteristics of the network. Such as centrality, density, average number of passed paths through the actor and the level of structural equivalence with a certain pattern.

Sometimes statistical characteristics are required: dispersion, centralization, topography distribution histogram, and others. Relationship patterns in the egocentric network are reflected in the content of the position, social or professional activity of the actor.

Each social role forms network structure and hence can be defined from it. Thus, the nature and structure of the links, actors can be classified. For example, we can identify potentially useful users and include them to the context-search system, detect "provocateurs", "spammers", "bots", "flooders" and protect them from other actors.

More complex studies are conducted on a full-scale graph of the network. The study of popular social networks: Flickr, YouTube, LiveJournal Orkut and Facebook shows their scalable nature. This means that the network has a power-sharing distribution between actors. So, most actors have the same number of connections, and the part of actors with an excessively large number of connections is a rarity. Therefore, the online network $P(k) \sim k^{-\gamma}$ can be represented by a fairly branched structure with hierarchically related vertices, such as computer networks.

4 Development of algorithms for the operation of the program interface

Thus, on the basis of research of existing software tools and existing mechanisms for obtaining information for a software complex that will implement the process of obtaining data from a particular user from Facebook social network and will perform the assignment of a certain interest to the user by analyzing the text of its publication, and displaying interest statistics in the graph, as well as mapping the location of interests on the map of Ukraine. Separately, a development algorithm that will independently attribute the text of a publication to a certain interest, and also need to develop an algorithm for processing data and displaying them to the user.

A concept has been developed that reflects user interaction and consolidated resource. To do this, we break the process of obtaining and processing data in separate steps:

1. Authentication of the social network user.
2. Obtaining access to user data, and recording them to a local database.
3. Conducting data analysis.
4. Creation of interest standards with the help of the administrator.

5. Automatic launch of the system of self-determination of interest on the basis of preserved standards.
6. Output of statistics in the form of a graph, and mapping of interests on the map of Ukraine.

The general concept of interaction between the user and the consolidated resource to solve the problem of determining the user's interest in publishing it is possible to depict schematically as shown in Fig. 1 [36]. The features of the analysis of the information of the social network Facebook to develop a universal approach for implementing an effective consolidated resource is considered below.

We need to consider the scheme of the algorithm (Fig. 2) of the web-based application that will be able to analyze the information of social networks based on the developed concepts [36].

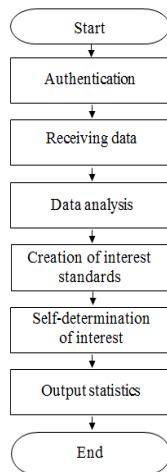


Fig. 1. Scheme of the interaction interface algorithm.

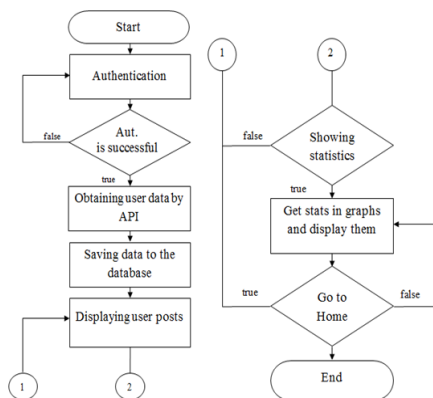


Fig. 2. Scheme of the web application algorithm.

5 Software Implementation

The web application is designed for users of the Facebook social network. It allows registered user to provide analysis and visualization of the selected data. The application uses integration methods with the Facebook API.

Nowadays there are quite a number of tools permitting to develop applications. It was decided to use the Tomcat7 [37] server for the application development. The server is entirely implemented in Java programming language. Tomcat is a very popular and widespread server for web development. Maven [38] is used to build the web application.

In order to integrate the web application with Facebook, you must perform "Log on through Facebook", after which the application receives an access token that provides

temporary secure access to various Facebook APIs [39]. Fig. 3-4 show the integration of the web application with Facebook.

The screenshot shows a form titled "Создайте идентификатор нового приложения" (Create New Application Identifier). The subtitle is "Начните интегрировать Facebook в свое приложение или на веб-сайт" (Start integrating Facebook into your application or website). The form includes the following fields and options:

- Отображаемое название** (Display Name): A text input field with the placeholder "название вашего приложения или веб-сайта" (name of your application or website).
- Это тестовая версия другого приложения?** (Is this a test version of another application?): A toggle switch set to "Нет" (No), with a link "Подробнее" (More details).
- Эл. адрес для связи** (Email address for contact): A text input field with the placeholder "Используется для важных сообщений о вашем приложении" (Used for important messages about your application).
- Категория** (Category): A dropdown menu with the option "Выберите категорию" (Select a category).
- Buttons:** "Отмена" (Cancel) and "Создайте идентификатор приложения" (Create Application Identifier).
- Footer:** "Продолжая, вы соглашаетесь с правилами Платформы Facebook" (By continuing, you agree to the Facebook Platform Terms).

Fig. 3. Development of Application identifier.

The screenshot shows the "Клиентские настройки OAuth" (Client OAuth Settings) page. It contains several settings:

- Клиентская авторизация OAuth** (Client OAuth authorization): A toggle switch set to "Да" (Yes). Description: "Позволяет использовать стандартную форму OAuth с кодом клиента. Обеспечьте безопасность своего приложения и предотвратите его взлом — ограничьте URL-адреса перенаправления в настройках ниже. Отключите эту функцию полностью, если они не используются. [?]"
- Веб-авторизация OAuth** (Web OAuth authorization): A toggle switch set to "Да" (Yes). Description: "Позволяет использовать клиентскую веб-авторизацию OAuth для построения пользовательских форм входа. [?]"
- Запустить принудительную повторную веб-аутентификацию OAuth** (Run forced OAuth web re-authentication): A toggle switch set to "Нет" (No). Description: "При включенной настройке люди будут перенаправляться на страницу ввода пароля Facebook для выполнения входа. [?]"
- Авторизация OAuth, встроенная в браузер** (Browser OAuth authorization): A toggle switch set to "Да" (Yes). Description: "Позволяет использовать URL-адреса перенаправления, контролируемые через браузер, для клиентской авторизации OAuth. [?]"
- Действительные URL-адреса для перенаправления OAuth** (Valid OAuth redirect URIs): A text input field with the placeholder "Действительные URL-адреса перенаправления OAuth".

Fig. 4. User Settings.

In the web application development the page access tokens were used. These tokens are used in API Graph [40-42] and called for managing Facebook pages. To create such a token, the Page Administrator must provide additional access permission to manage pages. With this permission, you can get the Page access token using the user access token with the required permissions.

The D3.js [43, 44] library was selected to visualize the Facebook user's social graph. D3.js is a JavaScript [45] library for document management based on data. D3 helps you translate data into life using HTML [46], SVG [47] and CSS [48]. D3 allows you to bind arbitrary data to the DOM [49, 50], and then apply the results of manipulations with them to the document.

The start page opens after the program starting, which is shown in Fig. 5.

Data Analysis Of Social Networks | Welcome Page

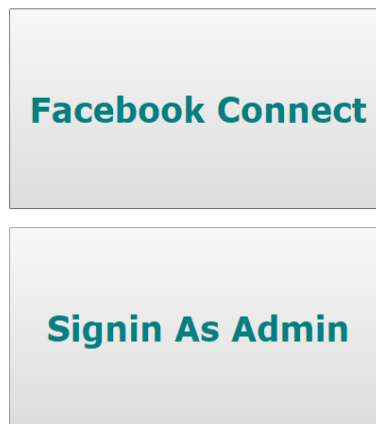


Fig. 5. Web-application Start Page.

The user has a choice between two buttons, called "Facebook Connect" and "Signin As Admin". The "Facebook Connect" button requires the Internet connection, because processes starts after clicking this button are accessed by the Facebook social network for user authentication. For the "Signin As Admin" button, the Internet connection is not required.

Clicking the "Facebook Connect" button is considered. When you click the button, a Facebook authentication page appears (Fig. 6).

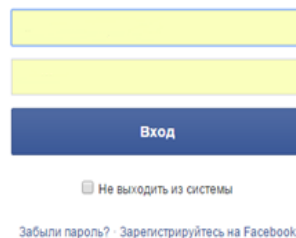


Fig. 6. Authentication on Facebook Page.

If you have already passed the authentication, you will immediately find yourself on the main page (Fig. 7).

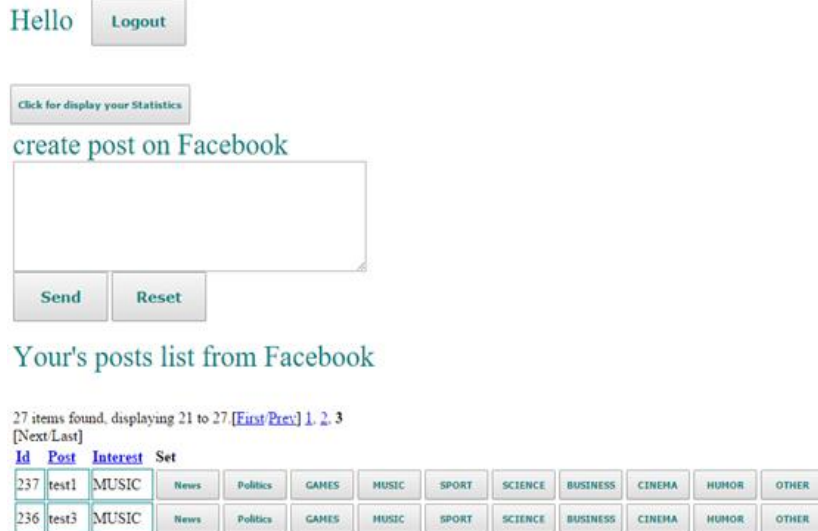


Fig. 7. The Home Page.

The web-application home page appears after successful authenticating and agreeing to personal data access (Fig. 7).

A welcome message with the logged-in username and all his posts from the Facebook page will appear on the home page. There are buttons with the interest title by the post. The user can click the same title button that best describes the content of the text.

The user can resign a post on his Facebook page in a form called "create post on Facebook". There are two buttons "Send" and "Reset" under the form (Fig. 7). When you click the "Send" button, the post appears immediately on the Facebook page. If you press the "Reset" button after entering the text, the text will be deleted. When user clicks the "Click for display your Statistics" button (Fig. 7), the page with graphs and the map appears as shown in Fig. 8-9.

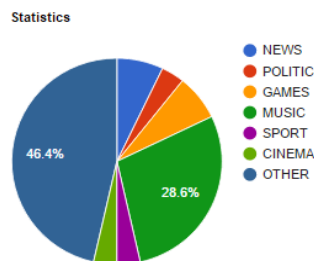


Fig. 8. The Statistics Page View.

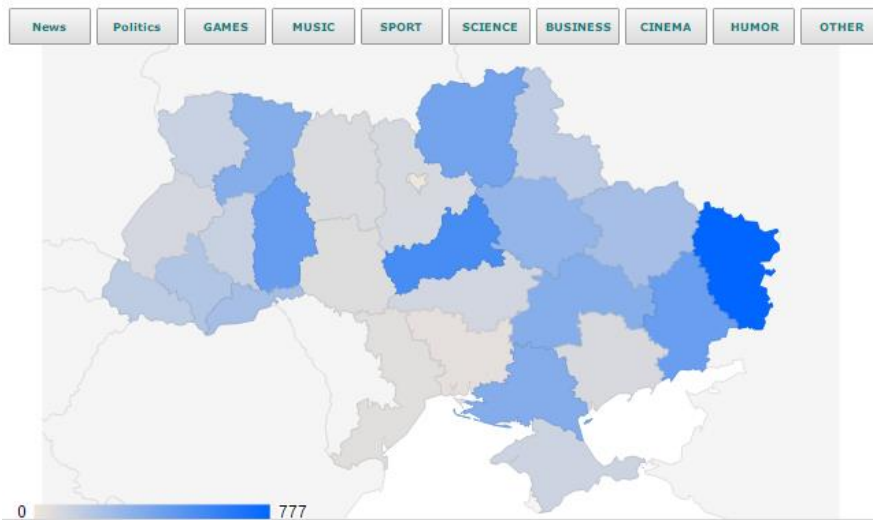


Fig. 9. The Statistics Page View.

6 Conclusion

The web application for processing and visualization of the Facebook user social graph is developed. The analysis of existing methods of analysis and visualization of the social graph is carried out. A comparative analysis of existing systems of processing social graphs is carried out. Tools for the web application development and interacting with the Facebook API are chosen. The web application is implemented using the Java language; the social graph visualization is implemented using the D3.js library.

The basic algorithm of the program's work, which ensures the publications receipt from the social network and their creation in the web-application, is developed.

After testing the program, it can be concluded that the program works correctly. The algorithm for analyzing information, based on existing interactions, is effective, which suggests that the system has a learning function.

The recognition percent is 98-99. Since analogues that analyze information from social networks don't conduct a publications analysis, it can be concluded that the developed consolidated resource is an ideal solution for analyzing the information of social networks in the content of publications, taking into consideration its algorithm of work.

References

1. Terekhov, D.: Social networks – a new communication reality, p.459 (2011).
2. Fedotchenko, Y.: Social networks: Yesterday comes today, p.303 (2010).
3. Newman, M.: Networks: an introduction. Oxford University Press, Oxford, p. 720 (2010).

4. Sheriff, S.: Processing large volumes of graph data: a guide to modern technology, pp. 130–137, <https://www.ibm.com/developerworks/ru/library/os-giraph/index.html> (2013).
5. The theory of graphs and social networks. <https://medium.com/egghead-science/778c92d20cea> (2014).
6. Mazurenko, V. V., Shtovba, S. D.: Overview of models for social network analysis, vol. 2, pp. 62–74 (2015).
7. Demkiv, O.: Development and main areas of network analysis, p.166 (2003).
8. Morgan, V.: Social structure of several networks, p. 880 (2006).
9. Ritzer, G.: Modern sociological theories, p. 688 (2002).
10. Iakymchuk, O. Social networks and their analysis. *Religion and socium*, №1(7) (2012).
11. Wellman, B.: Structural analysis: from method and metaphor to theory and substance. In: *Social Structures: a network approach*. New York: Cambridge, Cambridge university press, pp. 19–61 (1988).
12. Puhach, S., Mytchyk, Yu.: Spatial analysis of social networking services in the Volyn region. In: *Ekonomichna ta Sotsialna Geografiya*, vol. 79, pp. 14–21 (2018). doi: 10.17721/2413-7154/2018.79.14-21
13. Glückler J., Doreian, P.: Editorial: social network analysis and economic geography – positional, evolutionary and multi-level approaches. In: *Journal of Economic Geography*, vol. 16 (6), pp. 1123–11345 (2016)
14. Laniado D., Volkovich, Y., Scellato, S., Mascolo, C., Kaltenbrunner A.: The impact of geographic distance on online social interactions. In: *Information Systems Frontiers* (2017). doi: 10.1007/s10796-017-9784-9
15. Bondarchuk, O.: Psychological features of interaction in virtual social networks. In: *Bulletin of the Chernihiv National Pedagogical University named after Taras Shevchenko*, Bulletin 114, pp. 7–11 (2013).
16. Shelepaeva, A.: Educational opportunities of social networks, №2 (42), pp. 64–70 (2011).
17. Melnyk, V., Cheb, S.: Characteristics of some laws of influence of social network internet on personal development students. In: *Youth and market*, №12 (83), pp. 98–102 (2011).
18. Batura, T. V.: Methods of social networks analysis. In: *Comp. science*, №34, pp. 55–72, <https://cyberleninka.ru/article/n/metody-analiza-kompyuternyh-sotsialnyh-setey> (2012).
19. Batura, T. V., Kopylova, N. S., Murzin, F. A., Proskuryakov, A. V.: Methods for analysis of data from social networks, pp.1–22 (2013). doi: 10.31144/bncc.cs.2542-1972.2013.n35.p1-22
20. Gubanov, D. A., Chkhartishvili, A.: Conceptual approach to the analysis of online social networks. In: *Big Management Systems*, No. 45, pp. 222–236. (2013).
21. Chepik, P. I., Khachatryan, M. G.: Analysis of the social networking sites and the machine learning. *Politechnical student journal*, №2, pp. 1-6 (2018). doi: 10.18698/2541-8009-2018-02-253
22. Cogan, K.: Social networks as elements of a new social environment. In: *International Science Forum: sociology, psychology, pedagogy, management*, № 16, pp. 61–71 (2014).
23. Kulikova, A. A.: Approach to classification of users of social networks. In: *East – European Journal of advanced technologies*, pp. 14-18 (2011).
24. Nelyubin, K.: Analysis of the social networks as one of the key algorithms for disclosure of murders. In: *Actual problems of Russian law*, № 8 (57) (2015).
25. Voitovych, O., Buda, A., Holovchenko, V : Analyses methods research of social networks as a environment of warfake, pp.76–80 (2016).
26. Bysyuk, V., Novozhilova, O.: Methods of identifying spam in social networks. In: *Materials of Ukrainian Scientific Conference, Kropivnitsky*, pp. 84–85 (2016).

27. Voitovych, O., Holovchenko, V.: Research of social networks as a environment of warfake, №1 (2016).
28. Social Networking for Business – We Use It Right! <https://webprofit.com.ua/sotsialni-merezhi-dlya-biznesu/> (2018).
29. I calculate for you on networks: we use the API of the largest social networks for the self-serving purposes. <https://habr.com/ru/company/xakep/blog/254129/>, last accessed 2019/02/17.
30. Freeman, J.: What is an API? Application programming interfaces explained. <https://www.infoworld.com/article/3269878/apis/what-is-an-api-application-programming-interfaces-explained.html> (2018).
31. Dobrovolsky, A.: Integration of applications: methods of interaction, topology, tools, <https://www.osp.ru/os/2006/09/3776464/> (2006).
32. Gephi – The Open Graph Viz Platform. <https://gephi.org/>, last accessed 2019/02/17.
33. Ognyanova, K.: Network Analysis and Visualization with R and igraph. In: NetSciX 2016, School of Code Workshop, Wroclaw, Poland. http://www.kateto.net/wp-content/uploads/2016/01/NetSciX_2016_Workshop.pdf (2016)
34. NetworkX for convenient work with network structures. <https://habr.com/post/125898/>
35. Data Mining. <https://sites.google.com/site/upravlenieznaniami/tehnologii-upravleniaznaniami/data-mining>, last accessed 2019/02/18.
36. Rudchenko, Yu., Leshchenko, Yu.: Creating algorithms for analyzing and visualizing the user's social graph. In: Problems of informatization: theses of additional. VI Intern. sci. pract. conf., Cherkasy – Baku – Belsko-Byala – Kharkiv, p. 104 (2018).
37. Apache Tomcat. <https://tomcat.apache.org/download-70.cgi>, last accessed 2019/02/16.
38. Maven. <http://www.apache-maven.ru/>, last accessed 2019/02/15.
39. Permissions Reference – Facebook Login. <https://developers.facebook.com/docs/facebook-login/permissions/v2.5>, last accessed 2019/02/12.
40. Facebook for Developers. https://developers.facebook.com/docs/graph-api/overview?locale=eng_EN, last accessed 2019/02/18.
41. Singh, A., Srivastava, S.: Development of Facebook applications with Graph API Cookbook, p.366 (2011).
42. Analysis of information – the transformation of data into analytical findings <https://developers.facebook.com/docs/graph-api/reference/v2.5/post>, last accessed 2019/02/17.
43. Meeks, E: D3 is not a Data Visualization Library, https://medium.com/@Elijah_Meeks/d3-is-not-a-data-visualization-library-67ba549e8520 (2018).
44. Introduction to D3 <https://habr.com/ru/company/datalaboratory/blog/217905/> (2014).
45. HTML and HTML5. Description of tags by main sections – HTML5Book, last accessed 2019/02/18.
46. Introduction to SVG graphics. <https://habr.com/ru/post/157087/>(2012).
47. CSS and CSS3. Properties for formatting html elements. <https://html5book.ru/css-css3/>, last accessed 2019/02/16.
48. Object document model: what is DOM and what is not? <https://proglib.io/p/what-is-dom/> (2018)
49. Manipulates with DOM on pure JavaScript. <https://tproger.ru/translations/dom-javascript/>, last accessed 2019/02/19.
50. Javascript. What is javascript really? Client-side web APIs. https://developer.mozilla.org/ru/docs/Learn/Getting_started_with_the_web/JavaScript_basics, last accessed 2019/02/20.