Designer of the Private Joint-Stock Company "Institute of Information Technologies", Candidate of Technical Sciences, Professor.

Потій Олександр Володимирович, заступник Голови Державної служби спеціального зв'язку та захисту інформації України доктор технічних наук, професор.
E-mail: potii.oleksandr@gmail.com.
Orcid ID: 0000-0002-2366-0541.
Потий Александр Владимирович, заместитель Председателя Государственной службы специальной связи и защиты информации Украины, доктор технических наук, профессор.
Potiy Oleksandr, Deputy Head of the State Service for Special Communications and Information Protection of Ukraine, Doctor of Technical Sciences, Professor.

Онопрієнко Віктор Васильович, генеральний директор приватного акціонерного товариства «Інститут інформаційних технологій», кандидат технічних наук, старший науковий співробітник.
E-mail: kuznetsov@karazin.ua.
Orcid ID: 0000-0001-5681-3412.

Оноприенко Виктор Васильевич, генеральный директор частного акционерного общества «Институт информационных технологий», кандидат технических наук, старший научный сотрудник.
Onoprienko Viktor, General Director of the Private Joint-Stock Company "Institute of Information Technologies", Candidate of Technical Sciences, Senior Researcher.

Бобух Всеволод Анатолійович, начальник відділу апаратних засобів захисту інформації приватного акціонерного товариства «Інститут інформаційних технологій», кандидат технічних нау
E-mail: bobukhv@iit.kharkov.ua.
Orcid ID: 0000-0002-1175-5092.
Бобух Всеволод Анатольевич, начальник отдела аппаратных средств защиты информации частного акционерного общества «Институт информационных технологий», кандидат технических наук
Bobukh Vsevolod, Head of the Department of Information Protection Hardware of the Private Joint-Stock Company "Institute of Information Technologies", Candidate of Technical Sciences

DOI: 10.18372/2410-7840.22.14977
УДК 004.94

DEVELOPMENT OF UNCLEAR CRITERIA FOR DETERMINING THE SIGNIFICANCE OF A COMPOSITE SOCIAL PROFILE INFORMATION

Mykhailo Mozhaiev, Pavlo Buslov, Viktoriia Shvedun

Summary. The results of development of fuzzy criteria for determination of significance of composite information of social profile are presented. The simulation is carried out using OSINT technology - the technology of legal acquisition and use of information from open sources. As a result of the research, the parameters affecting the significance of individual characteristics of the social profile are selected and their characteristics are determined. The information of varying degrees of structurality is processed in the course of performing of the task of social profiling. At the same time, the following data models are used: network one - for storing the data of the final social profile, and presenting them in the form of graphs; relational one - for storing the information map of the social profile; post-shooting and NoSQL one - for storing unstructured source data and dynamic content, including multimedia. The conceptual model of presentation of these social profiles can be expanded and brought into line with the complex and mathematical models of the digital social environment. In the future, it is necessary to disclose the elements and relationships of the four basic categories of the infological model of presenting the results of social profiling. A new model of fuzzy significance coefficient of social profile parameters is obtained, the difference of which is the possibility of taking into account both formalized objective and difficult-to-interpret subjective indicators for evaluating the initial information of the SP. The advantage of using the significance criterion over the involvement of experts is the lower time spent on analyzing large amounts of data while maintaining the level of objectivity of the assessment. Consideration the diversity of the digital social environment is the next step.
Key words: *information technologies, fuzzy criteria, information from open sources, social portrait, formalization of criteria, information model, infological model.*

INTRODUCTION

Currently, the possession of information on the social profile of both the individual and the social group as a whole is of significant interest in most areas of human and social activity: economics, education, politics, security, comfortable living conditions ensuring, etc. Therefore, formation of a social portrait of an individual, a group, as well as the who-

le society is an important scientific problem for solution of which numerous studies are aimed [1-5].

The starting point of searching for information about the person, group and community of personalities is to get an answer to the question "In the name of what?".

Then the search area, the type of presentation and the place of storage of this information will also become clear.

This means that we fill the information mosaic of a social portrait, where each element represents an independent segment, at the same time closely connected with many other elements connected to each other. OSINT (Open Source INTelligence) -a concept, methodology and technology for the legal acquisition and use of information from open sources [1-3] - can be used as a tool for performing such research.

ANALYSIS OF EXISTING RESEARCHES

One of the most important aspects of the research is an accurate assessment of importance of social objects, which is complicated by verification problems, subjectivity of expert assessments and sensitivity of the data to the effects of the human factor.

Only ratings or estimates presented in numerical or percentage form can be considered as formalized criteria.

Thus, an urgent scientific task arises to select the apparatus for formalizing research objects and determining criteria for the significance of composite information of a social profile.

The purpose of this article is to develop fuzzy criteria for determining the significance of composite social profile information.

In order to achieve this purpose, the following specific tasks should be solved.

1. To define the data type and content requirements for formation of a person's social profile.

2. Analysis of requirements for quality of investigated information.

3. Construction of a model of fuzzy significance coefficient of social profile parameters.

4. Construction of an infological model of representation of static and dynamic parts of a social profile.

THE MAIN PART OF THE RESEARCH

The relevance (no, yes, partially) - the "freshness" of the information depends on the time when the information appears in the sample under study. The period of updating of social profile data is one day. This is justified by the activity of the average modern user of the Internet: the changes taking place over a shorter interval of time will be insignificant, while when setting a larger interval, up-to-date information will be lost. In the future, with the acceleration of the rate of digitalization of society, the period of data updating will decrease. Taking into account the linkage of people's activities to calendar cycles, it is possible to justify the loss of relevance of most of the dynamic social information a month after publication due to the appearance of updated information during this period. Further, the decrease in the relevance of the statement follows a power law. The characteristic function of the relevance indicator in this case takes the following form:

$$\mu_{vol} = \begin{cases} 0, vol = 1 \wedge time = '' \ future \ '' ; \\ 1, x_{vol} \leq 1 \wedge time = '' \ past '' ; \\ \dfrac{1}{1 + a^2}, x_{vol} > 1, \end{cases} \quad (1)$$

where vol – Shows the fact of the availability of more recent information, time – speaking time, x_{vol}– time elapsed since publication, $a = \frac{x_{vol}-1}{30}$ - information obsolescence rate.

The authority (absent, low, high) of the source is a key parameter that affects the assessment of statements by experts when it is difficult to determine its reliability by the evidentiary means available to experts in the disposal. This characteristic is determined by the following empirical arguments:

$$\mu_{auth} = \begin{cases} 0, x_{auth} = 0; \\ 2{x_{auth}}^2, 0 < x_{auth} \leq 0.5; \\ 1 - 2(x_{auth} - 1)^2, 0.5 < x_{auth} \leq 1, \end{cases} \quad (2)$$

where x_{auth} – Shows the proportion of experts who consider the source as reliable one.

The importance (irrelevant, secondary, important, and critical) of information - shows the degree of connection with other characteristics of the social profile. The parameter depends on their weights, so a plurality of high-importance SP elements may not coincide with a plurality of key elements obtained during the evaluation of the centrali-

ty metric. The degree of belonging of the importance of information to the fuzzy value is described by the following formula:

$$\mu_{val} = \begin{cases} 0, x_{val} = 0; \\ \frac{2x_{auth}^2}{nlink^2}, 0 < x_{vak} \le \frac{nlink}{2}; \\ 1 - \frac{2(x_{val} - nlink)^2}{nlink^2}, \frac{nlink}{2} < x_{val} < nlink; \\ 1, x_{val} = nlink, \end{cases} \quad (3)$$

where nlink – The total number of unique links with the research object, x_{val} – the number of links with the social objects whose weight is more than some threshold value.

The arguability (not enough, enough) - takes into account the presence of arguments, evidence and links in order to identify information noise and unsubstantiated statements, depends on the authority of information sources:

$$\mu_{arg} = \begin{cases} 0, x_{arg} = 0; \\ \frac{1}{e^{2na}}, x_{arg} > 0, \end{cases} \quad (4)$$

where x_{arg}– The number of possible arguments, na – a value showing the lack of authority of sources in the average and maximum value, we denote it as follows:

$$na = 2 - \max_{x_{arg}}(\mu_{auth}) - \frac{\sum^{x_{arg}} \mu_{auth}}{x_{arg}} \quad .(5)$$

The uniqueness (previously unknown, known, widely known) of the information - allows separating facts from unsubstantiated statements, determines the degree opposite to the prevalence of statements in various sources of information at the current moment:

$$\mu_{unq} = \begin{cases} 1, x_{unq} = 0; \\ 1 - \frac{2b^2}{ns^2}, 0 < x_{unq} \le \frac{ns}{2}; \\ \left(\frac{ns - 1 - b}{ns}\right)^2, \frac{ns}{2} < x_{unq} < ns; \end{cases} \quad (6)$$

where x_{unq} – the number of sources containing the research statement with the publication date earlier than in the original, ns – the number of sources containing the research statement with the publication date earlier than in the original, ns - the total number of sources in the social profile containing the research statement; $b = \sum^{x_{unq}} \mu_{auth}$ – Total authority of sources containing the investigated statement with the date of publication earlier than in the original.

The validity (falsehood, truth) of a statement in the presence of opposing opinions depends on the most authoritative of them, which is supported by many facts or is innovative. In the absence of contradictory points of view, the characteristic of reliability of the statement is considered equal to the indicator depends on the degree of reasoning and uniqueness:

$$\mu_{ver} = \begin{cases} 1 - \max_{i=1}^{x_{ver}}(ver_i), x_{ver} \ge 1; \\ 1, x_{ver} < 1, \end{cases} \quad (7)$$

where x_{ver}– number of alternative viewpoints, ver – verification indicator, we denote it as follows:

$$ver = \max\left[\mu_{arg}(\mu_{auth}), \mu_{unq}(\mu_{auth}, ns_{min})\right]. \quad (8)$$

Thus, the final weight is described by the following formula:

$$v = \mu_{vol} \cdot \mu_{val}(u_{min}) \cdot \mu_{ver} \cdot ver. \quad (9)$$

The developed coefficient of significance of characteristics of the social profile allows considering the division of the unstructured diversity of the source information by the level of its potential benefit, based on the values of authority of the sources, relevance, arguability, reliability, uniqueness and connectivity of the objects of the digital social environment. The use of the significance criterion in the framework of the developed mathematical model of the digital social environment is designed to influence the consistency of elements of the created social profiles. It can also be used in the analysis of moods, to help determine the semantic content of statements.

In the course of performing, the task of social profiling, information of different degrees of structurality is processed. At the same time, it is possible to use the following data models [4-6]: - network - for storing the data of the final social profile, and presenting them in the form of graphs; - relational - for storing the information map of the social profile; - post-shooting and NoSQL - for storing unstructured source data and dynamic content, including multimedia.

First, data is collected in the general non-relational repository [7], where data is stored "as is," but with the indication of service information (record number, search date, source, and comments) in separate columns to facilitate subsequent processing.

If you have already known key information or find it in the collected data, you need to add it to the information map table, the columns of which indicate the lists of found attributes, the social profile ID and a link to the dynamic content associated with it. The rows of the information map table relate to different social profiles.

Dynamic content data from a non-relational repository, depending on type and degree of structurality, may be transferred to other repositories [8,9] offering specific functionality such as the ability to record content versions, make real-time changes and comments, analyze non-text data, etc.

The information of the final social profile is reproduced in the form of vertices and edges of the graph. Each characteristic must correspond to one of the templates (subject, interest), which have their own sets of attributes.

In addition, for each object of the JV, concepts of weight denoting significance and moods indicating the emotional color of the statement are introduced. Examples of the main categories of objects in the task of building a social profile can be the following:

- The person under consideration - information (name, age...) about the person who is the central object of the social profile under construction and has the largest number of connections with other social objects;

- The mentioned person - brief information about a person who is not considered in the construction of this social profile, but has a direct or indirect connection with the person in question;

- Address - information about the geographical location correlated with some social phenomena and objects;

- Activities - data on social activity of a person or community, such as programming, professional sports or charity;

- Event - information about the social phenomenon that has the exact place and time of holding, as well as allowing to unite many participants into a community and determine their common characteristics;

- Works - describes the various material and intangible achievements of a person or group obtained because of his/their activities;

- Organization - information about various institutions and communities associated with people listed in the social profile;

- Specialization - specialization of a person or a specific area of activity;

- Telephone - phone number, code and related information;

- Hobbies - information about the hobbies and interests of people (it is important not to confuse the concepts of activity and hobbies);

- Account - identifiers of web resource accounts and e-mail addresses belonging to a specific person or community.

It is necessary to take into account the possible types of connections between elements of the social environment.

They make it possible to distinguish the effects of social objects from each other. Possible relationships between objects and social profile link types are shown in Tab. 1.

The sign "±" in the matrix means the possibility of combining homogeneous objects with the selected connection. An important condition when planning links is that there should be only one connectivity component in the PS column, and a vertex of type "Person in question" should be reachable from other vertices.

The links are full-fledged elements of the social profile representation model and can have their own attribute sets. The limitations on the location of links depend only on the convenience of subsequent analysis and presentation of data.

The conceptual model of presentation of these social profiles is proposed to be expanded and brought into line with the complex and mathematical models of the digital social environment.

Further, it is required to disclose the elements and relationships of the four basic categories of the infological model of presentation of social profiling results.

The following category of objects of a person's social profile, taking into account the division of data into static and dynamic parts, is depicted:

$$PSP = \begin{Bmatrix} ppid, PNAME, PDATE, PLOC, PLINK, PGLINK \\ PTLINK, PDLINK, PSOURCE, TESLINK \end{Bmatrix}, \quad (10)$$

where ppid – Unique numeric identifier of the social personality profile;

$PNAME = \{FNAME, LNAME, PATR, ALIAS\}$ – Many names of the person in question (Name, Surname, aliases);

PDATE – Many key dates of life of the person;

PLOC – A plurality of persona-related geolocations including location coordinates pcoord, their names PNAMELOC, important dates PLOCDATE and comments PLOCNOTE, wherein:

$$PLOC = \left\{ \begin{matrix} pcord, PNAMELOC \\ PLOCDATE, PLOCNOTE \end{matrix} \middle| PLOCDATE \subset PDATE \right\}, \quad (11)$$

PLINK – A plurality of references to other JVs with which communication is established, wherein:

$$PLINK = \left\{ ppid \middle| \begin{matrix} \exists x, y \in PSP : x \cap y \neq 0, \\ ppid\langle x \rangle \neq ppid\langle y \rangle \end{matrix} \right\}, \quad (12)$$

PGLINK – Links to multiple connected communities that take into account the social roles of the person role and the nature of relationship with the group matter, wherein:

$$PGLINK = \left\{ \begin{matrix} ppid, pgid, \\ role, matter \end{matrix} \middle| \begin{matrix} \exists x \in PSP, : \exists y \in GSP: \\ ppid \in x, pgid \in y, x \cap y \neq 0 \end{matrix} \right\}; \quad (13)$$

PTLINK – Many references to the most significant social phenomena and characteristics for the person, wherein:

$$PTLINK = \left\{ ppid, idtopic \middle| \begin{matrix} \exists x \in PSP, : \exists y \in FACT: \\ ppid \in x, idtopic \in y, x \cap y \neq 0 \end{matrix} \right\}; \quad (14)$$

PDLINK – A plurality of references to dynamic part objects with the most stable links, wherein:

$$PDLINK = \left\{ ppid, dynid \middle| \begin{matrix} \exists x \in PSP, : \exists y \in PDYN: \\ ppid \in x, dynid \in y, x \cap y \neq 0 \end{matrix} \right\}; \quad (15)$$

PSOURCE – Source information, with weblink sourcelink, names sourcename and source reference dates sourceget, wherein:

$$PSOURCE = \left\{ \begin{matrix} ppid, sourcelink, x \\ sourcename, sourceget \end{matrix} \middle| x \in PSP \cap SOURCE \right\}; \quad (16)$$

where SOURCE – Multiple sources of information in the Internet;

TESLINK – Links to dictionaries.

Dynamic content is seen as the following self-contained set:

$$PDYN = \left\{ \begin{matrix} dynid, DATATYPE, DSERVICE, DNOTE, DTAG, \\ PDLINK, DGLINK, DTLINK, DSOURCE, TESLINK \end{matrix} \right\}, \quad (17)$$

where dynid – Unique numeric identifier of the dynamic social object;

$DATATYPE\{text, audio, video, graphics, geodata, binary\}$ – Type of data;

$DSERVICE = \{servicetype\ servicevalue\}$ – Service information in the form of parameter-value about a dynamic object;

DNOTE – Dynamic object comments;

DTAG – Many thematic tags.

$$DTAG = \left\{ \begin{matrix} Creature, Organization, Activity, Address, Telephon, \\ Email, Achievement, Specialization, \dots \end{matrix} \right\}. \quad (18)$$

PDLINK – many links to CPP information maps (formula 33);

DGLINK – a plurality of links to GPS information maps, wherein:

$$DGLINK = \left\{ dynid, pgid \middle| \begin{matrix} \exists x \in PDYN, : \exists y \in GSP: \\ dynid \in x, pgid \in y, x \cap y \neq 0 \end{matrix} \right\}; \quad (19)$$

DTLINK – many references to social phenomena, wherein:

$$DTLINK = \left\{ dynid, idtopic \middle| \begin{matrix} \exists x \in PDYN, : \exists y \in FACT: \\ dynid \in x, idtopic \in y, x \cap y \neq 0 \end{matrix} \right\}; \quad (20)$$

DSOURCE – source information, wherein:

$$DSOURCE = \left\{ \begin{matrix} dynid, sourcelink, x, \\ sourcename, sourceget \end{matrix} \middle| x \in PDYN \cap SOURCE \right\}; \quad (21)$$

TESLINK – links to dictionaries.

Social phenomena (CP) (topics, facts) constitute a structured set of data, separated from information maps of joint ventures of persons and groups, having its own set of unique attributes:

$$FACT = \left\{ \begin{matrix} idtopic, TNAME, TTYPE, TFEATURE, TNONE, TLINK, \\ PTLINK, GTLINK, DTLINK, TSOURCE, TESLINK \end{matrix} \right\}, \quad (22)$$

where idtopic – Unique identifier of CP;

TNAME – Many names or keywords of a social phenomenon, theme or fact;

$TTYPE = \{event, social\ object, social\ connection\}$ – Kinds of CP;

TFEATURE – Many attributes, characteristics of CP;

TNOTE – Comments to CP;

TLINK – Many references to other CPs, wherein:

$$TLINK = \left\{ idtopic \middle| \begin{matrix} \exists x \in FACT : x \cap y \neq 0, \\ idtopic\langle x \rangle \neq idtopic\langle y \rangle \end{matrix} \right\}; \quad (23)$$

PTLINK – Multiple references to associated CPAs (formula 32);

GTLINK – A plurality of references to linked GPS, wherein:

$$GTLINK = \left\{ pgid, idtopic \middle| \begin{matrix} \exists x GSP, \exists y \in FACT: \\ pgid \in x, idtopic \in y, x \cap y \neq 0 \end{matrix} \right\}; \quad (24)$$

DTLINK – multiple dynamic content links (formula 38);

TSOURCE – information about the source of information about the social phenomenon:

$$TSOURCE = \left\{ \begin{matrix} idtopic, sourcelink, x \\ sourcename, sourceget \end{matrix} \middle| x \in FACT \cap SOURCE \right\}; \quad (25)$$

TESLINK – links to dictionaries. The characteristics of SP groups are not static over time, so they do not include entirely personal social profiles of members. The category of data related to GPS information maps is the following:

$$GSP = \left\{ \begin{matrix} pgid, PGNAME, GTAG, PGDATE, PGLOC, GLINK, \\ PGLINK, DGLINK, GTLINK, GSOURCE, TESLINK \end{matrix} \right\}, \quad (26)$$

where pgid – Unique group ID;

$PGNAME = \{RNLAM, ABBR, TRNM\}$ – Many community names (official, abbreviation, translation) if available;

GTAG – Multiple tags to describe the social status of a group;

PGDATE – Many key group dates;

PGLOC – Multiple community-related locations, including coordinates, locations' name PGNAMELOC, dates PGLOCDATE and comments PGNOTE:

$$PGLOC = \left\{ \begin{matrix} pgcoord, PGNAMELOC, \\ PGLOCDATE, PGNOTE \end{matrix} \middle| PGLOCDATE \subset PDATE \cup PGDATE \right\}, \quad (27)$$

GLINK – a plurality of references to other profiles of the groups with which the links are established, wherein:

$$GLINK = \left\{ pgid \middle| \begin{matrix} \exists x \in FACT: x \cap y \neq 0, \\ pgid\langle x \rangle \neq pgid\langle y \rangle \end{matrix} \right\}; \quad (28)$$

PGLINK – Multiple references to group members (formula 31);

DGLINK – Multiple references to group attributes from dynamic content (formula 37);

GTLINK – References to related SPs (formula 42);

GSOURCE – Information about the source of the data of the group, wherein:

$$GSOURCE = \left\{ \begin{matrix} pgid, sourcelink, x, \\ sourcename, sourceget \end{matrix} \middle| x \in GSP \cap SOURCE \right\}; \quad (29)$$

TESLINK – Links to dictionaries. In the social profile model, for information obtained from various dictionaries and directories, a separate following set is allocated:

$$TESAURUS = \{definition, SEMANTICPROPERTIES, VALUE\}, \quad (30)$$

where definition – a concept, definition or term from a dictionary, SEMANTICPROPERTIES – a plurality of semantic properties of the term, VALUE – meaning of a concept, definition or term.

Table 1

Social Profile Representation Graph Incident Matrix

	Addressing	Communication with organization	Organization Specialization	Participation in the event	Communication with person	Hobbies of a person	Activities of a person	Organization team	Contacts of a person	Specializations of a person	Associated Works
The person in question	1	0	0	-1	-1	1	1	1	±1	1	1
The person mentioned	1	0	0	-1	0	0	1	1	±1	1	1
Address	1	0	0	0	0	0	0	0	0	0	0
Activity	0	0	0	0	0	0	-1	0	0	0	0
Event	1	0	0	1	0	0	0	0	0	0	0
Works	0	0	0	0	0	0	0	0	0	0	-1
Organization	1	-1	1	0	0	0	0	-1	0	0	1
Specialization	0	0	-1	0	0	0	0	0	0	-1	0
Phone	0	1	0	0	1	0	0	0	0	0	0
Hobby	0	0	0	0	0	-1	0	0	0	0	0
Account	0	1	0	0	1	0	0	0	0	0	0

The considered properties of the developed models provide the possibility of subsequent incompatimization of analytical functions to support decision-making persons in the framework of a new social profiling methodology. The given opportunity to take into account the nature and main characteristics of the collected information helps to identify clear and implicit connections between elements of the social environment.

CONCLUSIONS

As a result of the research conducted, fuzzy criteria are developed for determining the significance of the composite information of the social profile, which will significantly increase the adequacy of the social profile model.

When solving the main task of the research the following results are obtained:

1. The requirements for the type and content of data for building a social profile of a person are defined. The original information is publicly available, i.e. the possibility of its processing by third persons should be indicated. It is collected from the social media network of the Internet and can refer to the following types of data: text, multimedia, geodata. The content of the data should correspond to the objects of the social environment: persons, communities and social phenomena. The social profile information can be distributed in space and time.

2. The quality requirements are as follows: data on the social environment should be up-to-date, complete and reliable. Compliance with these requirements makes it possible to create social profiles used as an information basis for the tasks of formalizing the human factor in the social and economic activity.

3. A new model of fuzzy significance coefficient of social profile parameters is proposed, the difference of which is the possibility of taking into account both formalized objective and difficult-to-interpret subjective indicators for evaluating the initial information of the joint venture.

The advantage of using the significance criterion over the involvement of experts is the lower time spent on analyzing large amounts of data while maintaining the level of objectivity of the assessment.

4. An infological model for representing of the static and dynamic parts of a social profile is proposed, which allows designing a scheme for information support of a social profiling system within the framework of a developed methodology for collecting and analyzing heterogeneous social data from open Internet sources. The difference of the model is the use of the principles of the theory of metagraphs and implementation of the confused structure of the social environment in the form of four large categories of social information related to persons, communities and social phenomena, which allows us to overcome the problem of heterogeneity and inequality of the initial data.

REFERENCES

[1] Буслов П.В., Зоренко Д.С., Рябуха Ю.М *Використання технологій OSINT для отримання пошукової інформації: практичний порадник*. Х.: ІПЮК для СБ України, 2021. - 28 с.

[2] *Graph-basedintelligence-analysis* [Електронний ресурс] – Режим доступу до ресурсу: https://linkurio.us/blog/graph-based-intelligence-analysis/.

[3] Tiwari G. Chapter 1: NoSQL: What It Is and Why You Need it > Definition and Introduction, Professional NoSQL, *Packt Publishing*, 2011. - 384 p.

[4] Бодров И. Сильные и слабые стороны NoSQL. *Корпоративный журнал компании "Инфосистемы Джет"* [Електронний ресурс] / И. Бодров – Режим доступу до ресурсу: http://www.jetinfo.ru/stati/silnye-i-slabye-storony-nosql.

[5] Robinson, I., Webber, J. & Eifrem, E. *Graph Databases*, O'Reilly Media, 2013. - 178 p.

[6] *Apache HBase ™ Reference Guide* [Електронний ресурс] – Режим доступу до ресурсу: http://hbase.apache.org/book.html#faq.

[7] Introduction to Views // Apache CouchDB 2.2 Documentation. [Електронний ресурс] – Режим доступу до ресурсу: http://docs.couchdb.org/en/stable/ddocs/views/intro.html.

[8] Hows, D., Membrey, P., Plugge, E. & Hawkins, T., *The Definitive Guide to MongoDB: A complete guide to dealing with Big Data using MongoDB*, Third Edition, Apress, 2015. - 376 c.

РАЗРАБОТКА НЕЯСНЫХ КРИТЕРИЕВ ДЛЯ ОПРЕДЕЛЕНИЯ ЗНАЧИМОСТИ СОЦИАЛЬНОЙ ИНФОРМАЦИИ О СОЦИАЛЬНОМ ПРОФИЛЕ

Приведены результаты разработки нечетких критериев для определения значимости составной информации социального профиля. Моделирование было проведено с использованием технологии OSINT - технологии легального получения и использования информации из открытых источников. В результате исследований были подобраны параметры, влияющие на значимость отдельных характеристик социального профиля, и определены их характеристические функции. В ходе выполнения задачи социального профилирования производилась обработка информации разной степени структурированности. При этом были использованы следующие модели данных: сетевая – для хранения данных конечного социального профиля, и представления их в виде графов; реляционная – для хранения информационной карты социального профиля; постреляционная и NoSQL – для хранения неструктурированных исходных данных и динамического контента, в том числе мультимедийного. Концептуальная модель представления данных социальных профилей возможно расширить и привести в соответствие с комплексной и математической моделями цифровой социальной среды. В дальнейшем, необходимо раскрыть элементы и взаимосвязи четырех базовых категорий инфологической модели представления результатов социального профилирования. Получена новая модель нечеткого коэффициента значимости параметров социального профиля, отличием которой является возможность учета как формализованных объективных, так и трудно интерпретируемых субъективных показателей для оценки исходной информации СП. Преимуществом использования критерия значимости перед привлечением экспертов являются меньшие временные затраты на анализ больших объемов данных при сохранении уровня объективности оценивания. Следующим шагом является переход к рассмотрению многообразия элементов цифровой социальной среды.

Ключевые слова: информационные технологии, нечеткие критерии, информация из открытых источников, социальный портрет, формализация критериев, информационная модель, инфологическая модель.

РОЗРОБКА НЕЯСНИХ КРИТЕРІЇВ ДЛЯ ВИЗНАЧЕННЯ ЗНАЧЕННЯ СКЛАДЕНОЇ ІНФОРМАЦІЇ ПРО СОЦІАЛЬНИЙ ПРОФІЛЬ

Наведено результати розробки нечітких критеріїв для визначення значущості складової інформації соціального профілю. Моделювання було проведено з використанням технології OSINT - технології легального отримання і використання інформації з відкритих джерел. В результаті досліджень були підібрані параметри, що впливають на значимість окремих характеристик соціального профілю, і визначені їх характеристичні функції. В ході виконання завдання соціального профілювання проводилася обробка інформації різного ступеня структурованості. При цьому були використані наступні моделі даних: мережева - для зберігання даних кінцевого соціального профілю, і подання їх у вигляді графів; реляційна - для зберігання інформаційної карти соціального профілю; постреляційна і NoSQL - для зберігання неструктурованих вихідних даних і динамічного контенту, в тому числі мультимедійного. Концептуальну модель представлення даних соціальних профілів можливо розширити і привести у відповідність з комплексною і математичною моделями цифрового соціального середовища. Надалі, необхідно розкрити елементи і взаємозв'язок чотирьох базових категорій інфологічної моделі представлення результатів соціального профілювання. Отримана нова модель нечіткого коефіцієнта значущості параметрів соціального профілю, відмінністю якої є можливість обліку як формалізованих об'єктивних, так і суб'єктивних показників, що важко інтерпретуються, для оцінки вихідної інформації СП. Перевагою використання критерію значущості перед залученням експертів є менші часові витрати на аналіз великих обсягів даних при збереженні рівня об'єктивності оцінювання. Наступним кроком є перехід до розгляду різноманіття елементів цифрового соціального середовища.

Ключові слова: інформаційні технології, нечіткі критерії, інформація з відкритих джерел, соціальний портрет, формалізація критеріїв, інформаційна модель, інфологічна модель.

Можаєв Михайло Олександрович, кандидат технічних наук, завідувач сектора комп'ютерно-технічних та телекомунікаційних досліджень, Національний науковий центр «Інститут судових експертиз ім. Засл. проф. Н.С. Бокаріуса».
E-mail: mikhail.mozhayev@hniise.gov.ua.
ORCID ID: 0000-0003-1566-9260.

Можаев Михаил Александрович, кандидат технических наук, заведующий сектора компьютерно-технических и телекоммуникационных исследований, Национальный научный центр «Институт судебных экспертиз им. Засл. проф. Н.С. Бокариуса».

Mykhailo Mozhaiev, PhD (Engineering Sciences), Head of Department on Computer Engineering and Telecommunication, National Scientific Center «Hon. Prof. M. S. Bokarius Forensic Science Institute».

Буслов Павло Володимирович, аспірант, Харківський національний університет радіоелектроніки.
E-mail: p.buslov@ukr.net.
ORCID ID: 0000-0002-2558-4984.

Буслов Павел Владимирович, аспирант, Харьковский национальный университет радиоэлектроники

Pavlo Buslov, Post Graduate Student, Kharkiv National University of Radio Electronics.

Шведун Вікторія Олександрівна, доктор наук з державного управління, професор, професор кафедри публічного управління та підприємництва, Національний аерокосмічний університет «Харківський авіаційний інститут».
E-mail: applevikvs@gmail.com.
ORCID ID: 0000-0002-5170-4222.

Шведун Виктория Александровна, доктор наук по государственному управлению, профессор, профессор кафедры публичного управления и предпринимательства, Национальный аэрокосмический университет «Харьковский авиационный институт».
Viktoriia Shvedun, Doctor of Sciences (Public Administration), Full Prosessor, Professor of the Department on Public Administration and Business, National Aerospace University "Kharkiv Aviation Institute".