

Development of an Information Model for the Personality's Social Portrait Formation Using OSINT Technology

Mykhailo Mozhaiev, Pavlo Buslov

Abstract – The results of the development of an information model for the personality's social portrait formation are presented. The modelling has been carried out using OSINT technology that is the technology of legal obtaining and using open source information.

In the result of the analysis, it has been found out that the social portrait is a heterogeneous semantic network consisting of personalized data. It has been defined that people organize formal and official communities of various orientations and the number of such communities associated with a particular person is practically unlimited.

When formalizing the decision-making process, the concept of a group social portrait (GSP) has been introduced, which takes into account the community's social tendencies united by certain common properties, group members' interpersonal interactions and their behavioural patterns.

The obtained information models of personal and group social portraits let to take into account all the main properties of the objects under study, their tonality and significance, as well as to conduct an analysis of the implicit dependencies determination. The next step is to move on to considering the diversity of the digital social environment elements.

Index Terms – decision-making support systems, information model, information technology, open source intelligence, social portrait.

I. INTRODUCTION

Accounting for the rapid development of information technologies and social processes, there is an acute problem nowadays of effective implementation of the managerial tasks and decision-making when working with large unstructured heterogeneous data based on the personality's social portrait. This is connected to the significant complication of both the initial data and the structure of the information management systems.

To model and research these systems, OSINT (Open Source Intelligence) is currently widely used. This is a concept, methodology and technology of the legal obtaining and using information from open sources. At present, the OSINT content is best described in "NATO Open Source Intelligence Handbook" (2006-2017), "NATO Open Source Intelligence Reader" (2006-2017) and "NATO Intelligence Exploitation of the Internet" (2002).

M. Mozhariev is with the National Scientific Center «Hon. Prof. M. S. Bokarius Forensic Science Institute», Kharkiv, Mashynistiv Str., 6A – 26, Ukraine (e-mail: mikhail.mozhayev@hniise.gov.ua).

P. Buslov is with the National University of Radio Electronics, 14 Nauky Ave str., Kharkiv, Ukraine (e-mail: p.buslov@ukr.net).

Open source intelligence (OSINT) is a process during which the detection, selection, collection and analysis of information in the public domain can significantly increase the efficiency of management decision-making systems [1-2].

The method of working according to the OSINT principles has been actively used for a long time in the business circles of the world leading countries to legally search and obtain information about partners or competitors. One of the basic "golden rules" of such a concept is that about 90% of the information required for analysis and appropriate decision-making can be found in open sources. Fig. 1 demonstrates the possible various sources that are used to consolidate information in the OSINT technology.



Fig. 1. Different sources of OSINT information

Against the background of the rapid modern information technologies development, greater emphasis is placed on this type of search activity. The difference between a newbie looking for information on the Internet and OSINT sales is quite clear: where a beginner sees photos, reposts, groups and pages that a person or organization is subscribed to on a social network, a specialist analyses activity, the photos background, social circles, opportunities for publications, possible reasons for subscribing to certain groups. Most often a person uses one or more nicknames in an online environment, which means that after a nickname is found through a request, one can find a person's other Internet activities, for example, on social networks or on forums that a person visits. And these are just a few examples. At the

same time, the collected information serves as the basis for further processing, cleaning (by assessing the sources and information reliability), analytical generalization and interpretation of the final results [3].

Thus, the basic idea of OSINT is the targeted data collection (Harvesting) according to the object of interest with the purpose of further processing and multi-vector content analysis of the data obtained (creating a person's "portrait", identifying the non-obvious facts or connections, predicting a person's behaviour, etc.).

The paper objective is to develop an information model for the personality's social portrait formation using OSINT technology. To achieve this objective, the following specific tasks have been solved in the paper:

- analysis of the social portrait;
- building an information model of a personal social portrait;
- building an information model of a group social portrait.

II. THE MAIN PART

Social portrait is an information structure that describes social characteristics of an individual or community, moreover, this information exhibits a property of clarity for human-machine perception, which makes it possible to automatically process it in various applied tasks.

Based on the definition, a social portrait is a heterogeneous semantic network consisting of personalized data [4]. The social environment model and the methodology for solving the problem of social portraits construction are based on the information model of the social portrait.

Descriptions of social phenomena are important for a holistic representation of the social environment and individual social portraits: they can often be found in many central objects of social studies. Social phenomena have a number of properties: time and place of origin; list of participants in the definition of the characteristics of the phenomenon; links to the information sources confirming the fact of the phenomenon occurrence; other related social objects and phenomena. Social phenomena are represented in managerial tasks by public events and activities, innovative implementations, legislative initiatives, and so on. In accordance with the theory of actor-network analysis, the elements of a social portrait, expressed by characteristics, concepts, events, phenomena, are represented as actants, social connections between elements are presented as predicates [5, 6].

The results of the social portrait development are structured data extracted from dynamic content and associated with an information map during the analysis. They represent a large-scale network, which must be stored in databases (DB) capable of storing graphs and semi-structured information and further be used by analysts.

Accordingly, the information model of the community social portrait is presented as follows (1):

$$PSP = \{P(X, v), R(Y, v), Q\} \quad (1)$$

Where $P = \{P_1, \dots, P_m\}$ – a set of the social objects characteristics associated with the person in question (themes, events, other people, etc.);

$R = \{R_1, \dots, R_n\}$ – a set of social connections between the social objects of the person under consideration (implemented in the form of their predicates);

$Q = \{Q_1, \dots, Q_{m+n}\}$ – a set of assessments of the social characteristics intonations, on the basis of which a complex assessment of the person's mood Q_{avg} is constructed, depending on the number of social characteristics, represents a fuzzy value, and $Q \in [-1; 1]$, m, n is the number of characteristics of the social objects and connections respectively, moreover, $m-1 \leq n; v$ are weighted coefficients of characteristics of the social objects and connections, represented as fuzzy values; X, Y are sets of characteristics of the social objects and connections, respectively (2):

$$\begin{cases} X = AO \times SO; \\ Y = AE \times SE \end{cases} \quad (2)$$

where, $AO = \{AO_1, \dots, AO_f\}$ – a set of attributes (characteristics) associated with a social object;

$AO_i = \{AO_{i1}, \dots, AO_{ix}\}$ – a set of characteristics for a single definition of a social object;

SO – a set of synonyms and repetitions of the social object definition in the source data;

$AE_i = \{AE_{i1}, \dots, AE_{ig}\}$ – a set of property values for a single definition of a social relation;

SE – a set of synonyms and repetitions of the social relations definition in the source data;

f, g – the number of the characteristics of the social objects and connections of the person under consideration, respectively;

fx, gy – the number of the characteristics occurrences (including synonyms) for social objects and connections, respectively.

Traditionally, individuals cannot be considered in isolation from society. People organize formal and official communities of various orientations, and the number of such communities associated with a particular person is practically unlimited. There is a problem of studying group activities and characteristics, because the personality's characteristics can change under the influence of a particular social group, which affects the results of sociological research. When analyzing changes in characteristics, a number of regularities are revealed that must be formalized for information support of the decision-making and while solving management problems. To solve this problem, it is necessary to introduce the concept of a group social portrait (GSP), which will take into account the social tendencies of the community united by certain common characteristics, interpersonal interactions of group members and their behavioural patterns.

The group social portrait includes personal portraits united by a number of patterns in their personal characteristics, as well as characteristics inherent in the community under consideration. The characteristics of the group's activity can be static or dynamic. The difference lies in the possible number and type of key elements: they can be personalities' identifiers – community leaders and events or some real world objects identifiers acting as symbols – in this case, people are linked by a similar attitude to the symbol. In view

of these features, the number of internal connections in GSP is greater than the total number of connections of all social portraits of group members. Accordingly, the information model of the group social portrait is presented as follows (3):

$$GSP = \{PSP, GSP', GP(X', v), GR(Y', v), GQ\}, \quad (3)$$

where – a set of social portraits of the GSP members;

t – the number of PSP in the GSP group;

$\{GSP' = \{GSP'_1, \dots, PSP'_{tt}\}$ – a set of subgroups included in the social group GSP;

tt – the number of subgroups in the social group GSP;

$GP = \{GP_1, \dots, GP_m\}$ – a set of characteristics of social objects associated with the GSP group;

$GR = \{GR_1, \dots, GR_n\}$ – a set of direct social connections of the GSP;

$GQ = \{GQ_1, \dots, GQ_{m+n}\} \cup \{Q_{PSP_1}, \dots, Q_{PSP_t}\} \cup \{GQ_{GSP'_1}, \dots, GQ_{GSP'_{tt}}\}$ – a set of assessments of the group's social characteristics intonations and its subgroups and people, on the basis of which a comprehensive assessment of the community's mood is constructed (4):

$$GQ_{avg} = \frac{\sum_{i=1}^t Q_{avg_i}}{t} + \frac{\sum_{i=1}^{tt} Q_{avg_{ii}}}{tt}, \quad (4)$$

represented as a fuzzy value, and $GQ \in [-1; 1]$;

m, n – the number of the group's social characteristics and connections, respectively, and $m' - 1 \leq n'; v$ – the weighted coefficients of the GSP's social objects and connections, represented as fuzzy values;

X', Y' – sets of the group's social characteristics and connections, respectively, and (5)

$$\begin{cases} X' = AO' \times SO'; \\ Y' = AE' \times SE', \end{cases} \quad (5)$$

where $AO' = \{AO'_1, \dots, AO'_f\}$ – a set of attributes (characteristics) associated with a social object belonging to the group in question;

$AO'_i = \{O'_1, \dots, O'_{fx}\}$ – a set of property values for a single definition of a social object belonging to the group under consideration;

SO' – a set of synonyms and repetitions of the definition of a social object belonging to the group in question in the source data;

$AE' = \{AE'_1, \dots, AE'_g\}$ – a set of attributes (characteristics) associated with a social relation belonging to the group in question;

$AE'_i = \{E'_1, \dots, E'_{gy}\}$ – a set of property values for a single social relation definition belonging to the group under consideration;

SE' – a set of synonyms and repetitions of the social relations definition belonging to the group in question in the source data;

f', g' – the number of characteristics of the GSP group's social objects and connections, respectively;

fx', gy' – the number of the characteristics occurrences (including synonyms) for social objects and connections of the GSP group, respectively.

When analyzing group social portraits, the definition of the personality's implicit characteristics is simplified. For example, "if one person from the group does not have a certain characteristic, and the rest of the group have approximately the same characteristic, then with a high probability we can conclude that the first person has the same characteristic". The developed information models of personal and group social portraits allow accounting for all the main objects' characteristics, their tonality and significance, as well as conducting an analysis to determine implicit dependencies. The next step is to move on to considering the diversity of elements of the digital social environment.

III. CONCLUSIONS

The paper analyses the social portrait, which allows more fully characterize the behaviour of both the social individual and the social environment as a whole.

In the result of the research, information models of personal and group social portraits have been created, which allow accounting for all the main characteristics of the objects under study, their tonality and significance, as well as conducting an analysis to determine implicit dependencies using OSINT technology.

REFERENCES

- [1] Buslov P. Analysis of using the features of communication models in social groups and virtual communities. Ukrainian Scientific Journal of Information Security. 2017. vol. 23, issue 1, P. 39-44. <https://doi.org/10.18372/2225-5036.23.11581>
- [2] Marinov M. Four-Dimensional Encoding of Character Sequences and Evaluation of their Similarities and Differences. Proceedings of the Technical University of Sofia. 2020. Vol. 70, Issue 2, P. 1-20. <https://doi.org/10.47978/TUS.2020.70.02.008>
- [3] Cherneva G. Fractal Models for Approximation of Random Processes. Proceedings of the Technical University of Sofia. 2017. Vol. 67, Issue 2, P.171-176.
- [4] Slavov D. Programming Languages for Artificial Intelligence. Proceedings of the Technical University of Sofia. 2019/ Vol. 69, Issue 2, 2019, P.43-52.
- [5] Буслів 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. Graph-based intelligence analysis. URL: <https://linkurio.us/blog/graph-based-intelligence-analysis>
- [6] Палагин А. В., Кривой С. Л., Петренко Н. Г. Концептуальные графы и семантические сети в системах обработки естественно-языковой информации. Математические машины и системы. 2009. Т. 1. №. 3. URL: <https://cyberleninka.ru/article/n/kontseptualnye-grafy-i-semanticheskie-setiv-sistemah-obrabotki-estestvenno-yazykovoy-informatsii>
- [7] Целых А. А., Дедюлина М. А. Теоретико-графовые подходы к моделированию актор-сетей в исследованиях науки и технологий. Моделирование, оптимизация и информационные технологии. 2018. Т. 6. №. 4. С. 244-259.
- [8] Целых Ю. А. Теоретико-графовые методы анализа нечетких социальных сетей. Программные продукты и системы. 2008. №. 2. С.48-50. URL : <https://cyberleninka.ru/article/v/teoretiko-grafovyemetodyanaliza-nechetkih-sotsialnyh-setey>

Mykhailo Mozhaiev¹ Pavlo Buslov²
mikhail.mozhayev@hniise.gov.ua, p.buslov@ukr.net

¹ National Scientific Center «Hon. Prof. M. S. Bokarius Forensic ScienceInstitute», Kharkiv, Mashynistiv Str., 6A –26, UKRAINE

² Kharkiv National University of Radio Electronics, Kharkiv, Nauky Ave. 14, UKRAINE