

Fedushko Solomia

Ph.D. in Technical Sciences, Professor Assistant, Department of Social Communication and Information Activities, orcid.org/0000-0001-7548-58
Lviv Polytechnic National University, Lviv

**MODELLING SOFTWARE COMPLEX FOR WEB
USER PERSONAL DATA VERIFICATION**

Abstract. This article has resolved an important scientific task of modelling software complex of computer and linguistic verification of web community members' accounts personal data. Methods for verification of Ukrainian web community members' socio-demographic characteristics on the basis of computer-linguistic analysis of web community members' information tracks were suggested. Software complex for verification of web members' socio-demographic characteristics as based on results of the computer-linguistic analysis of web communities' content was developed. Effectiveness and efficiency of use of the developed methods and means for solving tasks in web communities administration is proved by approbation of software complex of computer and linguistic verification of web community members' accounts personal data.

Keywords: validation; software complex; web users; personal data; socio-demographic characteristic; web-community

Introduction

Virtual community stored a huge database of information and profiles, which contain a lot of data about the web user [1]. The auditorium of web-communities is a large number of individuals regardless of age, gender, occupation, education, ethnicity, social status etc., who in order to register need to fill in a form with their personal information. Today, the necessity of verifying of web community users' personal data [2-7] is the up-to-date problem. The importance of this research lies in verifying basic socio-demographic characteristics of communities' user based on comprehensive analysis of information track of web users which gives an opportunity to the identifying socio-demographic profile of web user. Each socio-demographic characteristic of web user is defining by analysis of linguistic features in communication between virtual community users. The difference in style of writing posts by web community's members is the basis for developing effective methods of verifying of accounts personal data.

Background study

For virtual community user data validation, in order to improve virtual communities' management and to improve target techniques in online advertising is enough to analyze such basic socio-demographic characteristics [8-10]: Age; Education; Sphere of activity; Gender; Geographic location.

Obtaining data from real users of social networking pages ("Facebook", "Twitter", "Habrahabr", "Google+", etc.), the popular online newspapers and magazines,

dating sites ("LoveUA", "FindLove.in.UA", "12 Kisses", "Dating foreigners website") and web forums (Ukrainian forum of programmers "Replace.Org.Ua", "Domivka. Net", "The first web site of the Ukrainian community in Italy") is urgent task for virtual community administrators [11], police [12, 13], private detectives (services that have been increasingly needed by users of global environment Internet) and individual user of any information resource. Verification and setting of web users' personal data who performs illegal actions that are offensive to opponents from psychological [14], financial, legal angle, in particular, who levies blackmail [15], sends hate mails, spreads false and aspersion information and conduct other electronic bullying.

Socio-demographic characteristic: Age

Age category is chosen for specified age validation in virtual community member due to a number of important factors: the presence of real online threats to Internet users with the age of 6 to 17 years (these include disclosure of personal confidential information, access to content that do not meet age peculiarities and adversely affect the physical and mental health of the child, online abuse, internet marketing crimes, etc.) and the need for screening age group of children who have applied or already are virtual community member, that is intended only for adult web space users.

The characteristic features of a virtual identity are self-expression and experimentation with the aim to make a definite impression on the users of the virtual community and this is peculiar especially to growing age.

The problem of age differentiation is that the penetration rate of children is increasing in virtual

communities, which are intended for adult users' communication, and this destroys communicative atmosphere of community and vice versa, children are increasingly claiming to be adult users. This situation can lead community administrator to criminal charges. Examples of web communities for teenagers' online communication are: "Student Forum | UNIVER-SITY", "Chat for teens", chat "Teenager", "Chat for teens", "Just chat" and for adults "Our jokes with Pepper" and others.

Socio-demographic characteristic: Education

Many community administrators require virtual community users to follow certain conventions of web communication. Web communication convention depends on the objective goal and projected scenarios of the virtual community owners. An important factor in the possibility of participating in a virtual community is a high literacy level of virtual community member, level of higher education and skills. The method of screening illiteracy and with low education level virtual community member will significantly reduce the time and financial costs, and managers' efforts to moderate virtual community. In order to screen illiterate virtual community member it required to classify all users by the level of literacy, which will help to determine the likely level of each virtual community member education. In the task automation research of detecting errors in the text are only considered character errors. Error analysis and data consolidation that were conducted by scientists, allows us to offer the errors typology and to classify users in the order of education level.

Socio-demographic characteristic: Sphere of activity

A participant in a greater or lesser degree can belong to several areas of interest, as the impact of the content creation in web communities serve a direction of education, professional activity [16] and range of interests in spare time. However, the result of the analysis is to determine the scope of activity of virtual community member. In the general scope of activities are classified as science fields in the following areas: natural sciences; social sciences; formal science. This division is based on analysis of web community members Internet communication with different scopes of activities.

In web-community moderation that is intended for communication between members of certain areas, there many questions appear about level of professionalism, ethical principles, so filtering of web users by scope of activities is a necessary and important task for the administration of this community (government, policy portals) to improve the virtual community position.

Socio-demographic characteristic: Gender

In order to avoid gender conflicts in virtual communities [17]. That are intended for women (*Lviv Women's Forum "Cult of Beauty"*, "Girls chat",

Forum "Lvivmama" et al.) and for men (*"Man Forum"*, *Forum "Antybabskyi site" etc.*) administrators of these communities need to enter a strict gender division.

"Virtual gender change" [17], that is a web user creation identity of the opposite sex, it is common on the Internet and is caused by the following factors as the impact of cultural gender stereotypes, expression of homosexual tendencies or transsexual or diffuse gender identity. So-called virtual sex change is more peculiar to men due to several reasons: desire to control and manipulate others, for women it is easier to get help and attract attention, desire to power over other men, the study of the relationship between sexes and get new experience of Internet communication.

Socio-demographic characteristic: Geographic location

Detect locating of web user with no verification data on computer IP-address of the virtual community member is possible using special software tools (*CNGeoip*, *GeoLite Country*, *GeoLite City*, *IpGeoBase*, *GeoIp etc.*) that can determine the belonging of IP-address to the level of countries and cities.

Software complex modeling for web users' personal data verification

Socio-demographic characteristics [18-21] are determined by markers contained in a specialized dictionary. The structure of the dictionary developed according to scheme of lingvo-communicative indicators. For modelling information scheme was chosen the Barker's notation.

The information scheme of vocabulary and software complex for web users personal data verification is shown in Fig. 1, and comment on their content and their key attributes.

The essence of *entity "Web-community member"* contains all the necessary information about web-community member.

Attribute "Status" contains information about the presence or absence of validation of socio-demographic characteristics of web member. Attribute can take the following values: verified or not verified.

The essence of *entity "Socio-demographic profile"* contains generated socio-demographic portrait of web-community member on the basis of verified SDH. For example, this attribute can take the following values: reliable results and equivocal result and false result.

The essence of the *entity "Information content"* contains contributed by the web community members information which is well structured [22-25].

The essence of *entity "Information track"* is the set of all consolidated data by virtual community members and the results of their communication activities.

The essence of the *entity "Account"* contains data that participant pointed out in his account.

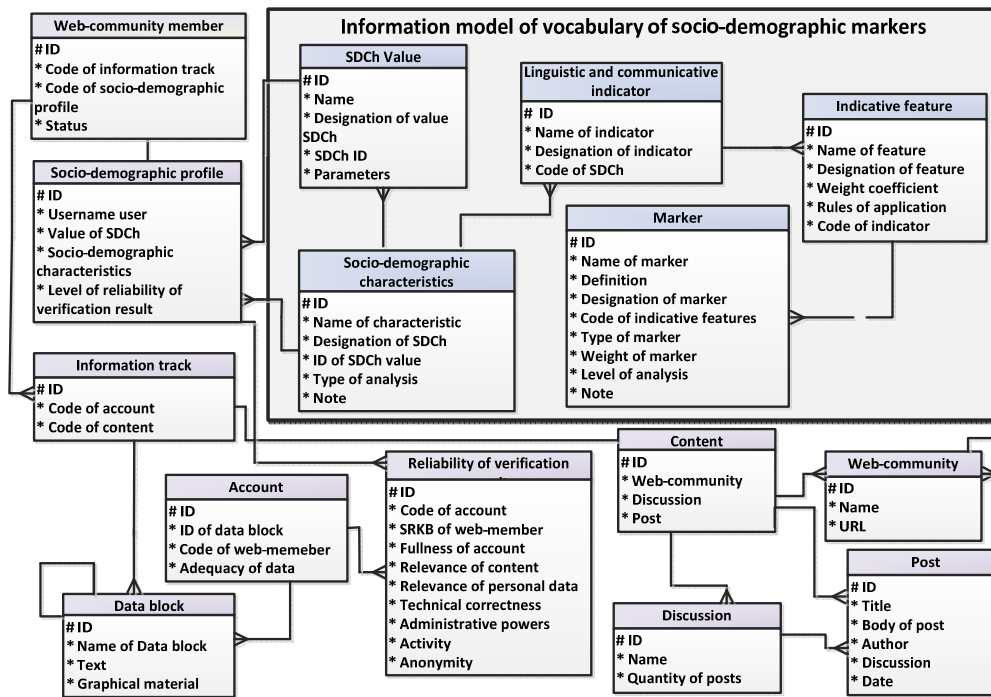


Figure 1 – Entity-relationship model of software for socio-demographic characteristics verification (Barker's notation)

The main feature of this essence is an attribute "Adequacy of data" that contains information about the level of probability analysis account of virtual community member to reference account of web user based on real and relevant information about the virtual community member.

Moreover, the level of adequacy depends on the reliability of the results of computer-linguistic analysis of information track of virtual community member.

The essence of *entity "Data block"* containing personal information of web-community users. These data are structured according to belonging the blocks to the account.

Attribute "Name of block" who can take the following values: *BasicInfo*, *EduInfo*, *WorkInfo*, *ContactInfo*, *FotoInfo* and *InterestsInfo*.

Attribute "Text" and attributes "Graphic material" depending on the type of information containing all personal information web users that needs to check its reliability. For example, in contact information block of web user – Contact Info (Igor Paslavskyi) may contain the email addresses of the personal social networking pages and websites, in which the web user creates content, so, attribute "Text" can take the following values: ICQ: 78316866, Website: http://www.linux.org.ua.

The essence of *entity "Web community"* provides data on the virtual community, social and demographic characteristics of users which require verification.

The essence of *entity "Discussion"* include themed online chat of web user, members are all registered web forum audience.

Entity "Post" contains data contributed by the web forum posts, which are part of surveys and discussions.

The essence of *entity "Reliability verification results"* contains data on the reliability of the verification result socio-demographic characteristics.

Attribute "Rules of web community" includes information about the level of compliance system of rules of communicative validation of virtual community members and this attribute, depending on compliance with the rules of the system.

Attribute "Fullness account" includes information about the importance of weighting coefficients filling level account of web user, providing the amount of personal data in the account.

Attribute "The relevance of information content" contains information about the degree of relevance of content web community [26].

Attribute "Relevance of personal data" includes information about the level of relevance of personal data in an account.

Attribute "Technical correctness account" includes information about the importance of weighting coefficients of technical correctness.

Attribute "Admin Powers" contains data on administrative powers [28] of web community.

Attribute "Activity web user" contains information about the activity level of web user. The possible values for this attribute: active author, a frequent author, passive and episodic author.

Attribute "Anonymity of web user" includes information about the level of anonymity [29] of web-community members.

Conclusion

In this article an analysis of actual development tendencies of web communities as a part of a global system of social communications is carried out.

Conceptual determination of the socio-demographic characteristics of a web community member and substantiate of their selection are suggested.

The use of socio-demographic profile of a web personality in applied tasks of WWW, its components and demand for it building for effective administration tasks in web communities are analyzed.

Model of a specialized socio-demographic markers vocabulary, which is integrated into software complex of personal data validation, is created. Model of software for verification of a web community member

socio-demographic profile by forming sets of indicators based on result of computer-linguistic analysis of web members' information tracks is developed.

The software complex for verifying socio-demographic characteristics of web community members, which includes specialized socio-demographic markers vocabulary, is developed.

The software complex allows implementing decision support system for professionals in web community management in the field of processing personal data.

Effectiveness and efficiency of use of the developed methods and means for solving tasks in web communities administration is proved by their approbation.

References

1. Bedrii, D.I. (2015). *Human resources management in Scientific projects. Management of Development of Complex Systems*, 24, 16-22 [in Ukrainian].
2. Fedushko, S., Peleschyshyn, O., Peleschyshyn, A., Syerov, Yu. (2013.) *The verification of virtual community member's socio-demographic characteristics profile. Advanced Computing: An International Journal*: 4, 3, 29-38, DOI: 10.5121/acij.2013.4303.
3. Syerov, Yu., Peleschyshyn, A., Fedushko, S. (2013). *The computer-linguistic analysis of socio-demographic profile of virtual community membe. International Journal of Computer Science and Business Informatics*: 4, 1, 1-13.
4. Iegorchenkov, O.V., Lisitsyn, O.B. & Kataev D.S. *Optimization of information management in product design management system. Management of Development of complex systems*, Kyiv, 13, 28-31.
5. Fedushko, S. (2014). *Development of a software for computer-linguistic verification of socio-demographic profile of web-community member [Electronic source]. Webology*. – 11(2). – Article 126.–: <http://www.webology.org/2014/v11n2/a126.pdf>, DOI: 10.6084/m9.figshare.205664.
6. Fedushko, S. (2011). *Peculiarities of definition and description of the socio-demographic characteristics in social communication. Bulletin of Lviv Polytechnic National University: Computer Science and Information Technology*, 694, 75-85.
7. Reynolds, D., Quatieri, T., Dunn, R. (2000). *Speaker Verification Using Adapted Gaussian Mixture Models. Digital Signal Processing*: 10, Issues 1–3, 19-41, DOI:10.1006/dspr.1999.0361
8. Fedushko, S.S., Biluschak, H.I. (2014). *Formation of linguistic and communicative indicators of socio-demographic characteristics of the web-community member// Management of development of complex systems*: 18, 112-122.
9. Fedushko, S., Syerov, Yu. (2013). *Design of registration and validation algorithm of member's personal data. International Journal of Informatics and Communication Technology*: .2, 2, 93-98, DOI: 10.11591/ij-ict.v2i2.3960.
10. Korzh, R., Peleschyshyn, A., Syerov, Yu., and Fedushko, S. (2014). *The cataloging of virtual communities of educational thematic.[Electronic source]. // Webology*, 11(1), pp. 1-16. – <http://www.webology.org/2014/v11n1/a117.pdf>, DOI: 10.6084/m9.figshare.2056650.
11. Andrukhiv, A., Sokil, M., Fedushko, S. (2014). *Integrating new library services into the University Information System. Library management*: 1(6), 79-87.
12. Fedushko, S., Bardyn, N. (2013). *Algorithm of the cyber criminals identification. Global Journal of Engineering, Design & Technology*: 2, 4, 56-62.
13. Bushuev, S.D., Lazareva, M.V. (2014). *Organization trust as integration factor of success of projects and programs. Project management and development of manufacture: Collection of scientific papers*, (20), 11-16.
14. Fedushko, S., Peleschyshyn, A., Korzh, R., Syerov, Yu. (2014). *Verification of socio-demographic characteristics of virtual community members. Proceedings of the XIIIth International Conference "Modern Problems of Radio Engineering, Telecommunications and Computer Science"*, 632. DOI: 10.13140/RG.2.1.2176.7761.
15. Biloshchytskyi, A., Dikhtiarenko, O. (2013). *The effectiveness of methods for finding matches in texts. Management of development of complex systems*, 14, 144-147.
16. Fedushko, S., Melnyk, D., and Syerov, Yu. (2011). *Analysis of the function and organization of virtual communication environments of Ukrainian scientists. Bulletin of Lviv Polytechnic National University: Computer Science and Information Technology*: 732, 293-305 [in Ukrainian].
17. Peleschyshyn, A., Fedushko, S. (2010). *Gender similarities and differences in online identity and Internet communication // International Conference of Computer Science and Information Technologies. Lviv*, 195-198.

18. Fedushko, S. (2011). Analysis architecture and modern trends of virtual communities' development. *Bulletin of Lviv Polytechnic National University: Information Systems and Networks*: 699, 362-375.
19. Fedushko, S., Peleschychyn, A., Syerov, Yu. (2011). Definition methods of socio-demographic characteristics of the users' social communications. *7th International Academic Conference of Young Scientists "Computer Science and Engineering"*, 358-361 [in Ukrainian].
20. Fedushko, S., Biluschak, H. and Syerov, Yu. (2015). Statistical methods of virtual community users age verification. *International Journal of Mathematics and Computational Science*: 1, 3, 174-182. DOI: 10.6084/m9.figshare.2056653
21. Fedushko, S., Biluschak, H. and Syerov, Yu. (2014). Investigations of web-Community members age verification by statistical methods // *International Journal of Social Networking and Virtual Communities*: 3, 1. Available at: <http://www.iaesjournal.com/online/index.php/VirCom/article/view/5909/0>
22. Bekesh, Yu., Fedushko, S. (2014). The approaches to attract and inform customers of travel agencies in social networks. *International Academic Conference "Information, Communication, Society"*. 216-217. [in Ukrainian].
23. Chop, N., Fedushko, S. (2013). The basic requirements to information resources of company of construction materials. *International Academic Conference "Information, Communication, Society"*, 234-235. [in Ukrainian].
24. Averkova, O., Fedushko, S. (2013). Informational support of the Reference and Information sector of Migration Service. *International Academic Conference "Information, Communication, Society"*, 206-207. [in Ukrainian].
25. Fedushko, S.S. (2015). Methods and means of computer-linguistic verification of web communities members socio-demographic characteristics. Thesis for a Ph. D degree in specialty 10.02.21 – structural, applied and mathematical linguistics. – Lviv Polytechnic National University, Lviv, 2015. – 20 P. DOI: 10.13140/RG.2.1.2183.9121 [in Ukrainian].
26. Peleshchychyn, A., Syerov, Yu., Berezko, O., Peleshchychyn, O., Tymovchak-Maksymets, O., Markovets, O. (2012). *Processes of Regulating Interactive Social Communications in the Developing Information Society*. Lviv Polytechnic National University Publishing House. 368. [in Ukrainian].
27. Peleshchychyn, A. N., Trach, O.R. (2014). Determination of the stages of life cycle of virtual association. *Management of Development of Complex Systems*, 20, 131-137 [in Ukrainian].
28. Syerov Yu., Kravets, R. (2009). Software for determination of behavior patterns of web-forum members. *International Journal of Computing*: 8 (2), 119-125.
29. Fedushko, S. (2010). Disclosure of Web-members Personal Information in Internet. *Conference of Modern information technologies in economics, management and education*. Lviv, 163-165.

Стаття надійшла до редколегії 25.03.2016

Рецензент: д-р техн. наук, проф. А. М. Пелещин, Національний університет "Львівська політехніка", Львів.

Федушко Соломія Степанівна

Кандидат технічних наук, асистент кафедри соціальних комунікацій та інформаційної діяльності,
orcid.org/0000-0001-7548-58
Національний університет "Львівська політехніка", Львів

**МОДЕЛЮВАННЯ ПРОГРАМНОГО КОМПЛЕКСУ ВЕРИФІКАЦІЇ ПЕРСОНАЛЬНИХ ДАНИХ
КОРИСТУВАЧІВ ІНТЕРНЕТУ**

Анотація. У статті розв'язано важливе наукове завдання – моделювання програмного комплексу комп'ютерно-лінгвістичної перевірки достовірності соціально-демографічних характеристик учасника веб-спільноти. Запропоновано метод верифікації соціально-демографічних характеристик користувачів віртуальних спільнот українського сегмента Інтернету за допомогою комп'ютерно-лінгвістичного аналізу інформаційних слідів користувачів Інтернету. Розроблено програмний комплекс верифікації соціально-демографічних характеристик учасника веб-спільноти на основі комп'ютерно-лінгвістичних методів перевірки достовірності масиву користувацького інформаційного наповнення Інтернету. Результативність та ефективність використання розроблених методів та засобів для вирішення задач адміністрування віртуальних спільнот доведені апробацією програмного комплексу комп'ютерно-лінгвістичної перевірки достовірності соціально-демографічних характеристик користувачів Інтернету.

Ключові слова: верифікація; програмний комплекс; користувачі Інтернету; персональні дані; соціально-демографічні характеристики; веб-спільнота

Федушко Соломия Степановна

Кандидат технических наук, ассистент кафедры социальных коммуникаций и информационной деятельности,
orcid.org/0000-0001-7548-58

Национальный университет "Львовская политехника", Львов

МОДЕЛИРОВАНИЕ ПРОГРАММНОГО КОМПЛЕКСА ВЕРИФИКАЦИИ ПЕРСОНАЛЬНЫХ ДАННЫХ ПОЛЬЗОВАТЕЛЕЙ ИНТЕРНЕТА

Аннотация. В статье решена важнейшая научная задача – моделирование программного комплекса компьютерно-лингвистической проверки достоверности социально-демографических характеристик участника веб-сообщества. Предложен метод верификации социально-демографических характеристик участников виртуальных сообществ украинского сегмента Интернета с помощью компьютерно-лингвистического анализа информационных следов пользователей веб-сообществ. Разработан программный комплекс верификации социально-демографических характеристик участника веб-сообщества на основе компьютерно-лингвистических методов проверки подлинности массива пользовательского информационного наполнения веб-сообществ. Результативность и эффективность использования разработанных методов и средств для решения задач администрирования виртуальных сообществ доказаны апробацией программного комплекса компьютерно-лингвистической проверки достоверности социально-демографических характеристик участника веб-сообщества.

Ключевые слова: верификация; программный комплекс; пользователи Интернета; персональные данные; социально-демографические характеристики; веб-сообщество

Link to publication

- APA Fedushko, Solomia. (2016). *Modeling software complex for web user personal data verification. Management of development of difficult systems*, 26, 105 – 110.
- ГОСТ Федушко, С.С. Моделювання програмного комплексу верифікації персональних даних користувачів Інтернету [Текст] / С.С. Федушко // *Управління розвитком складних систем*. – 2016. – № 26. – С. 105 – 110.