Adequacy of Personal Medical Profiles Data in Medical Information Decision-Making Support System

Solomiia Fedushko [0000-0001-7548-5856]

Lviv Polytechnic National University, Lviv, Ukraine

solomiia.s.fedushko@lpnu.ua

Abstract. The paper describes the verifying methods of medical specialty from user profile of online community for health-related advices. To avoid critical situations with the proliferation of unverified and inaccurate information in medical online community, it is necessary to develop a comprehensive software solution for verifying the user medical specialty of online community for health-related advices. The algorithm for forming the information profile of a medical online community user is designed. The scheme systems of formation of indicators of user specialization in the profession based on a training sample is presented. The method of forming the user information profile of online community for health-related advices by computer-linguistic analysis of the information content is suggested. The system of indicators based on a training sample of users in medical online communities is formed. The matrix of medical specialties indicators and method of determining weight coefficients these indicators is investigated. The proposed method of verifying the medical specialty from user profile is tested in online medical community.

Keywords: Personal Medical Profiles, Decision-Making Support System, Medical Information, Profiles Data, Health-Related Advices.

1 Research Aims

- developing community health workers
- improving quality of e-health systems
- simplifying the patients identification in various medical organizations
- collecting, processing and analyzing Big medical data from various sources
- working-time reduction of medical staff
- detecting non-valid accounts and accounts with incorrect or stale medical data
- consolidating all medical personal data with high level adequacy

2 Problems Solving

To develop an effective methods and means of determining the level of data adequacy of personal medical profiles.

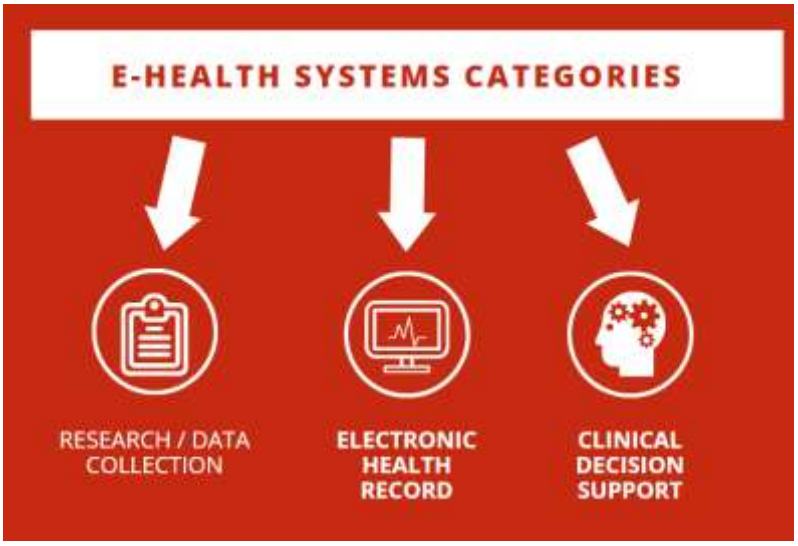

Fig. 1. E-health systems categories

3 Models of personal medical profiles

Proposed method is carried out in order to identify the user personal data in the online community and verify the authenticity of the personal data specified by the user.

The information track of the web user:

$$InfTrack\left(P_i\right)=\left\langle Content\left(P_i\right),PersonalData\left(P_i\right)\right\rangle \tag{1}$$

The components of an information track are: $Content(P_i)$ created by a member of the online community, and personal data – $PersonalData(P_i)$.

$$Content\left(P_i\right)=\left\langle Thread\left(P_i\right),Poll\left(P_i\right),Post\left(P_i\right)\right\rangle \tag{2}$$

$Thread\left(P_i\right)=\left\{Thread_j\left(P_i\right)\right\}_{j=1}^{N_i^{UThead}}$ is set of online community discussions;

$Poll\left(P_i\right)=\left\{Poll_j\left(P_i\right)\right\}_{j=1}^{N_i^{UPoll}}$ is set of online community polls;

$Post(P_i) = \{Post_j(P_i)\}_{j=1}^{N_i^{UPost}}$ is set of online community posts.

Computer-linguistic analysis is made only for the personal data that the user of the online community has specified in user account.

The most prioritized data for forming the data profile of the patients in the online community is the mandatory information about the online community user, less important – important data. Personal data distribution of online community user account to blocks is as follows:

$$PersonalData(P_i) = \left\langle \begin{matrix} BasicInfo(P_i), EduInfo(P_i), InterestsInfo(P_i), \\ WorkInfo(P_i), ContactInfo(P_i), FotoInfo(P_i), \end{matrix} \right\rangle \quad (3)$$

Formal description of the online community member account:

BasicInfo(P_i) is block of basic personal information of the online community user.

$$BasicInfo(P_i) = \left\langle \begin{matrix} Name(P_i), NickName(P_i), Age(P_i), \\ Gender(P_i), Region(P_i), Lang(P_i) \end{matrix} \right\rangle \quad (4)$$

where *Name(P_i)* is full name; *NickName(P_i)* is nick name; *Gender(P_i)*is gender; *Age(P_i)* is age; $Lang(P_i) = \{Lang_j(P_i)\}_{j=1}^{N_i^{Lang}}$ is plural of languages signed by a user; N_i^{Lang} is number of language; $Region(P_i) = \{Region_k(P_i)\}_{k=1}^{N_i^{Region}}$ is set of regions with which a user is associated; N_i^{Region} is set of regions.

EduInfo(P_i) is information block about education.

$$EduInfo(P_i) = \langle EduLevel(P_i), Specialization(P_i) \rangle \quad (5)$$

where *EduLevel(P_i)* is level of education received, *Specialization(P_i)* is specialty.

WorkInfo(P_i) is block of data about the work of online community user.

$$WorkInfo(P_i) = \langle Company(P_i), Position(P_i) \rangle \quad (6)$$

where *Company(P_i)* is institution where works, *Position(P_i)* is a position taken by a member of the online community in that institution.

ContactInfo(P_i) is contact information for the member of the online community.

$$ContacInfo(P_i) = \langle Email(P_i), SocialNets(P_i), Website \rangle \quad (7)$$

where *Email(P_i)* is main email address, $SocialNets(P_i) = \{SocialNets_j(P_i)\}_{j=1}^{N_i^{(Up)}}$ is plurality of pages in social networks, $N_i^{(Up)}$ is the number of pages in social networks; *Website* is website.

FotoInfo(P_i) is graphic information block.

$$FotoInfo(P_i) = \langle Avatar(P_i), Userbar(P_i), Foto(P_i) \rangle \quad (8)$$

where $Avatar(P_i) = \{Avatar_j(P_i)\}_{j=1}^{N_i^{Ava}}$ is set of avatars, N_i^{Ava} is number of avatars, N_i^{Foto} is a number of photos, $Userbar(P_i) = \{Userbar_k(P_i)\}_{k=1}^{N_i^{Userbar}}$ is set of graphic signatures, $N_i^{Userbar}$ is number of signatures, $Foto(P_i) = \{Foto_m(P_i)\}_{m=1}^{N_i^{Foto}}$ is set of photos.

InterestsInfo(P_i) is a block of information about the hobby and interests of the online community user.

$$InerestsInfo(P_i) = \langle Byline(P_i), Activity(P_i), Quot(P_i), Biography(P_i) \rangle \quad (9)$$

where $Byline(P_i) = \{Byline_j(P_i)\}_{j=1}^{N_i^{Byline}}$ is number of signatures, N_i^{Byline} is number of signatures; $Activity(P_i) = \{Activity_k(P_i)\}_{k=1}^{N_i^{Act}}$ is a set of favorite lessons and phrases, N_i^{Act} is number of lessons, phrases, $Quot(P_i) = \{Quot_l(P_i)\}_{l=1}^{N_i^{Quot}}$ is a plurality of quotations, N_i^{Quot} is number of quotations, *Biography(P_i)* is a biography.

Information about contacts and websites where the web user displays communicative activity is placed in the *ContactInfo(P_i)* block. Each account blocks contain information from three groups of personal data of the online user. Preferably, in the *BasicInfo(P_i)* block, compulsory data is placed, without this data registration in the online community is not possible.

4 Method of determining the personal medical profiles data adequacy level

The concept of the personal data adequacy of the patient profile of the medical clinics is presented to compare the personal data of the online communities informational profile with the medical information system data. The information profile in the web communities is engendered by the method of computer-linguistic analysis of the user information track.

Determining the personal data adequacy of the account to the real information of system user consists in the implementation of the main stages of the algorithm of determining the of personal data adequacy of the account.

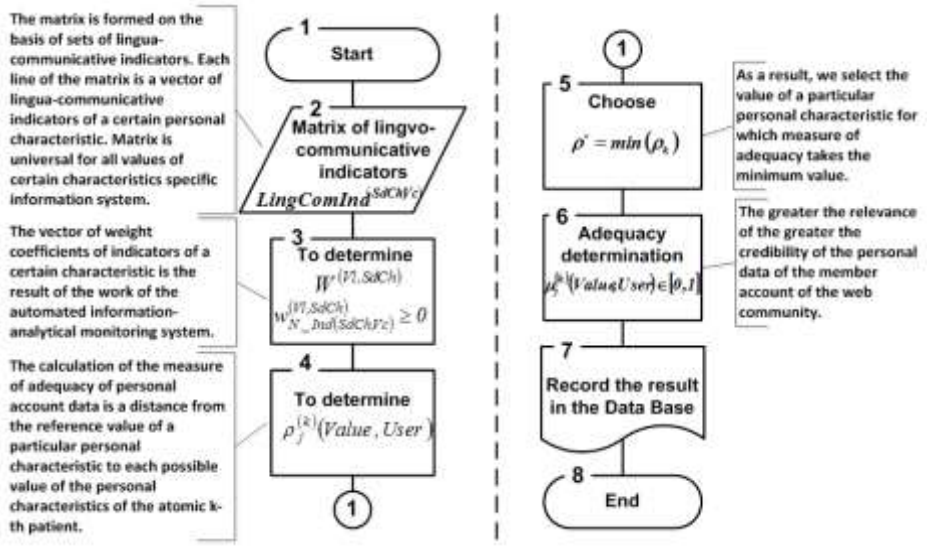

Fig. 2. Block diagram of the algorithm for determining the personal data adequacy of account

The difference between 1 and $\rho_j^{(k)}(Value, P)$ is the distance between the reference value of the personal characteristics and the value of the personal data of the atomic k-th user is determined by the adequacy of the personal data of the k-th user profile.

$$\mu_j^{(k)}(Value, P) = 1 - \rho_j^{(k)}(Value, P) \tag{10}$$

$\rho_j^{(k)}(Value, P)$ is distance to each possible value of the personal data of the atomic k-th user of web community:

$$\mu_j^{(k)}(Value,P) = 1 - \sqrt{\sum_{i=1}^{N\_Ind(PrCh,k)} \left(Ind_{i,j}^{(PrCh,Vc)} - Ind_{i,j}^{(PrCh,P)} \right)^2 * w_i^{(PrCh)}} \tag{11}$$

where $k \in 1 \ldots N\_Vl(PrCh, Vc)$. Moreover, $\mu_j^{(k)}(Value, P) \in [0,1]$.

$\mu_j^{(k)}(Value, P) \to \max$, then the degree of probability of personal characteristics of a particular user in the online community to this user personal data is high.

The proposed method of vectorization consists in transforming the data into a vector form, which will enable to determine the degree of similarity between the values of personal characteristics. The value of a similarity measure between the value of personal characteristics and the control vector indicates the importance of membership by the online community to a certain value characteristics.

5 Approbation of results in the practice

The results of data level adequacy analysis of personal medical profiles of Ukrainian medical centers comparing to patients online information tracks.

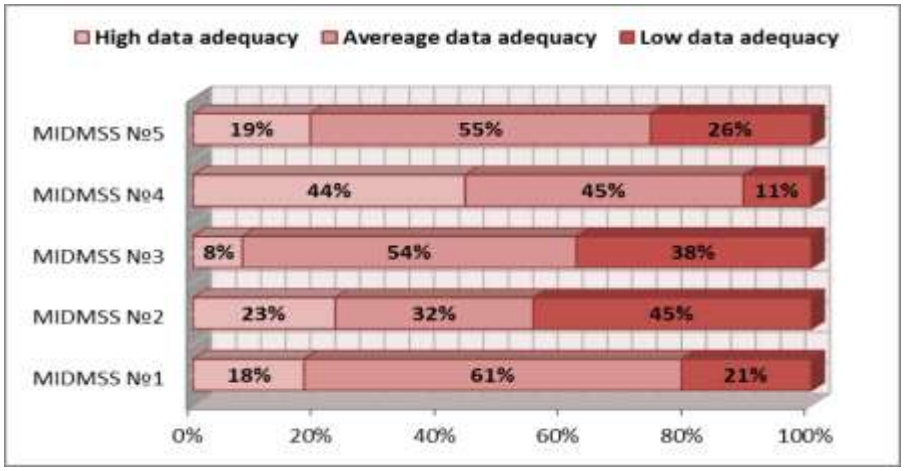

Fig. 3. Results of medical profiles data adequacy analysis of Ukrainian medical information systems

The indicator of the effectiveness of the developed methods of data verification of personal medical profiles is determined in equation (12).

$$\text{Efficiency}=\frac{N^{VerPD}}{N^{VerPD}-N\left(\text{LAdequacy}\right)^{APD}},\ N^{VerPD}{}_1N\left(\text{LAdequacy}\right)^{APD} \qquad (12)$$

N(LAdequacy)APD is number of personal medical profiles with low account data adequacy, *NVerPD* is the total number of verified personal medical profiles.

Based on equation (12) the results of data verification level of personal medical profiles, investigated profiles are classified according of data verification of personal medical profiles (21% of all investigated accounts contained high level of data verification, 41% of all investigated accounts contained average level of data verification, 38% of all investigated accounts contained low level of data verification). These results are presented graphically in Fig. 3.

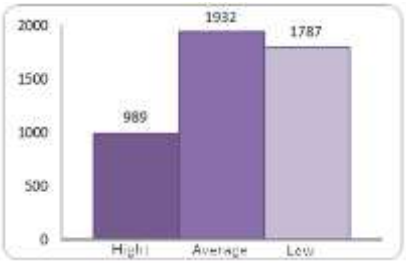

Fig. 4. Results of analysis of level of data verification of personal medical profiles

The results show that 23% of the patients (total of 4708 person) provided reliable information in their accounts. 28% of members updated their credentials in the accounts. 4% of all personal medical accounts are blocked.

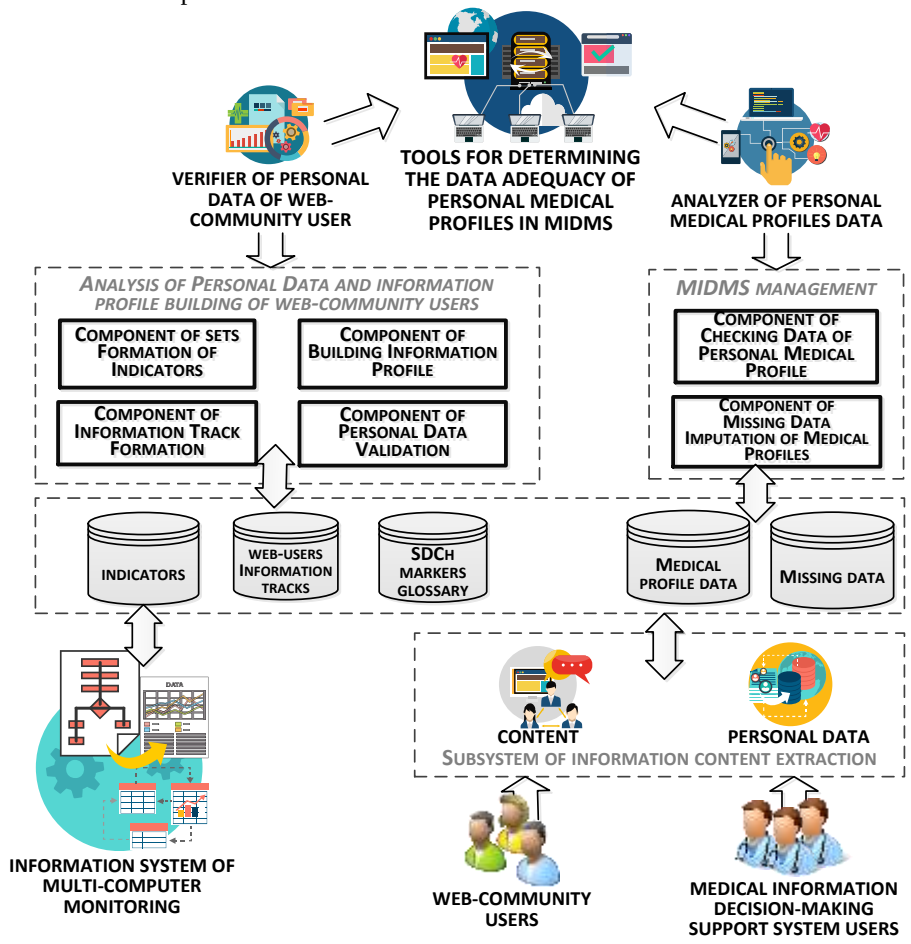

Fig. 5. Software Complex of Determining the Data Adequacy of Personal Medical Profiles Medical Information Decision-Making Support System

Conclusion

- the suggested method of medical information decision-making support system have been tested in 5 medical information systems.
- the efficient in detecting non-valid accounts and accounts with incorrect or fake data.
- the given methods simplifies work of medical staff on analysis of patients' personal data and reduces check times.

References

1. Fedushko S., Shakhovska N., Syerov Yu. Verifying the medical specialty from user profile of online community for health-related advices. CEUR Workshop Proceedings. Vol. 2255: Proceedings of the 1st International workshop on informatics & Data-driven medicine (IDDM 2018) Lviv, Ukraine, November 28–30, 2018. P. 301–310.
2. Babichev, S., Korobchynskyi, M., Mieshkov, S., Korchomnyi, O.: An Effectiveness Evaluation of Information Technology of Gene Expression Profiles Processing for Gene Networks Reconstruction. IJISA. Vol.10, No.7, 1-10 (2018).
3. Boyko, N., Pylypiv, O., Peleshchak, Y., Kryvenchuk, Y., Campos, J. Automated document analysis for quick personal health record creation. Proceedings of the 2nd International Workshop on Informatics & Data-Driven Medicine (IDDM 2019) Lviv, Ukraine, November 11-13, 2019. Vol 2488. http://ceur-ws.org/Vol-2488/paper18.pdf
4. Fedushko S. Adequacy of Personal Medical Profiles Data in Medical Information DecisionMaking Support System. IREHI 2018. http://ieee-rural-elderly-health.com/2018/wp-content/uploads/2018/12/IREHI-Programm-1.pdf
5. Gnatyuk S., Kinzeryavyy V., Sapozhnik T., Sopilko I., Seilova N., Hrytsak A. (2020) Modern Method and Software Tool for Guaranteed Data Deletion in Advanced Big Data Systems. In: Hu Z., Petoukhov S., He M. (eds) Advances in Artificial Systems for Medicine and Education II. AIMEE2018 2018. Advances in Intelligent Systems and Computing, vol 902. Springer, Cham. pp 581-590.