ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 1, Issue 8, October 2013**

# Method for Searching of «The Ideal Interlocutor» for Users of «Information and Organizational Environment for Pre-University Training and Social Adaptation of Foreigners»

Sergiy Paliy[1], YevgeniyBorodavka[2], MykolaTsiutsiura[3]

Assistant Professor, Chair of Computer Science Basics, Kyiv National University of Construction and Architecture, Kyiv, Ukraine[1]

Associate Professor, Chair of Information Technologies Design and Applied Mathematics, Kyiv National University of Construction and Architecture, Kyiv, Ukraine[2]

Graduate Student, Chair of Project Management, Kyiv National University of Construction and Architecture, Kyiv, Ukraine[3]

***ABSTRACT:*** In this paper we considered preconditions for developing the method for searching of «the ideal interlocutor» within the design of information and organizational environment for pre-university training and social adaptation of foreigners (IOEUTSAF). The database of user's profiles and the ways for it filling were proposed. Also, we defined the list of attributes and state criteria for finding users. As the main result the mathematical method for estimating conformity of users was developed.

**K*eywords*:**pre-university training, training of foreigners, social network, social adaptation, searching of «the ideal interlocutor».

## I. INTRODUCTION

### A. The Context Establishment

References [1]-[3] show the reasons, expediency, possibility and basic principles of making IOEUTSAF with cloud technologies. One of the basic principles of IOEUTSAF is organization of a wide listeners' community for the sake of exchange knowledge and useful experience. For the sake of a fast involving to the IOEUTSAF community numerous of users, who has necessary knowledge and experience, the conceptual model and methods of connecting IOEUTSAF with social networks were proposed [4].

The number of active user accounts in popular social networks (SN) varies from millions to one billion. It is quite hard to find person among plurality of users, who could be useful for particular user of IOEUTSAF. To facilitate this task and make the possibility of it solution in automated/automatic mode we suggest to develop the method for searching of «the ideal interlocutor» based on the recommendation systems and in particular on the methods of filtering by content and collaborative filtering.

### B. The Literature Overview

Based on the results of scientific sources study, we made conclusion that on the development of recommendation systems are working a lot of foreign and national researchers, including Daniel Lemire, Anna Maclachlan, Prem Melville, Raymond J. Mooney, RamadassNagarajan, Hennig-Thurau, Thorsten, André Marchand, Paul Marx, DietmarJannach, Markus Zanker, Alexander Felfernig, Gerhard Friedrich, Pravykov A.A., Glibovets M.M., Gorokhovsky S.S., Pick A.A. and others [5]-[10].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 1, Issue 8, October 2013**

In existing works, the recommendation systems only considered for offering to users certain resources. The recommendation systems need further development for using in information and organizational environment for pre-university training and social adaptation of foreigners as the technology for search of "the ideal interlocutor".

C. THE MAIN PURPOSE

All users in IOEUTSAF are divided into two categories: IOEUTSAF students who have a contract with the university for study and who after the distance conducted initial phase of training and social adaptation will move to some hypothetical country for further study (the listeners); and other users registered in IOEUTSAF and have access to certain resources (mostly foreign students of senior courses, graduate and doctoral students who are studying or studied earlier in universities specified country). This category of users, in comparison with listeners, has a much greater experience of staying and studying in this country, and sees the problem, in contradistinction to teachers and tutors, «by the eyes of foreigners». Very often advice from users may be more important and desirable than the advices of the citizens of the hypothetical country, due to «the fresh look» and the practical experience of being in identical situations.

The purpose of the article is to develop the method for searching «the ideal interlocutor» - the method that will find persons in among multimillion numbers of users (including social networks, aggregated to the environment), who are the most useful and interesting for certain listener.

## II. THEORETICAL BASIS

Preferably, the recommendation systems used in electronic commerce for providing to user some recommendations for some products or services (resources) on the basis of unclear formulated request. In this case, the analysis of the history of ratings resources (which are evaluated by users with buttons "Like" and "Dislike") is take place separately for each users. Therefore it is obtained set of triples in the form:

$$(m, n, r_{m,n}) \quad (1)$$

where$m$ - the user, $n$- resource, $r_{m,n}$ – rating, derived by resource $n$ from user $m$. After that the matrix is composing. Rows of it correspond to users and columns correspond to resources. At the intersection of a row and a column the rating of the resource, given by the user, is placed. Due to the fact that we have numerous resources and users, the obtained matrix is quite sparse. Task is that on the basis of existing ratings find similar users, to forecast yet not set ratings and therefore to predict which of the unrated resources a user will like.

In our case the task is other. For the sake of find «the ideal interlocutor» it is necessary to create a database of user's profiles. This database might be interpretive and in case of searching query should find the most similar profile to the given one.

The database of user's profiles (DBUP) than those referred previously rated resources (or histories ratings) contains users' demographic data, which include: date of birth, place of residence, sex, marital status, education, social status, place of further education and others.

To find the most suitable profile database should contain a maximum number of records, so to fill it makes sense not only with data of users of IOEUTSAF, but also with data received from the social networks.

The structural and functional diagram of database content profiles filling has shown in Figure 1. Listeners and other users of IOEUTSAF are passing questioning at the registration and in process of work. Users, who registered into IOEUTSAF through appropriate services of social networks, additionally transfer their account information through application programming interface (API) of a social network by IOEUTSAF request. IOEUTSAF application, installed in users of a social network transmits to DBUP account information and statistics about visiting resources and ratings (Figure 2). In this way, the data about users, which somehow related to IOEUTSAF, is collected.

To cover wider audience of the social networks users, particularly who is not related to IOEUTSAF, filling of DBUP also carried out by agents of IOEUTSAF (Fig. 2).

As agents we mean some program entity, that operates independently on behalf of another entity (program or user) and perform certain actions [11].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 1, Issue 8, October 2013**

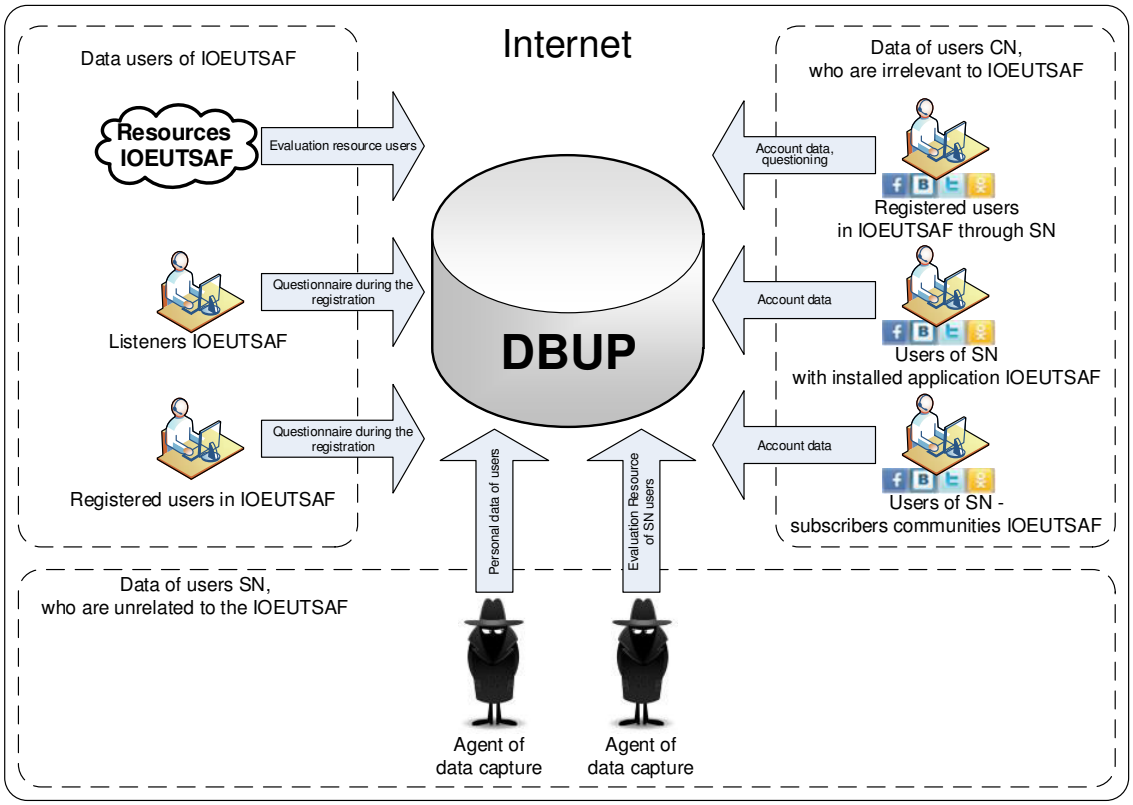


**Figure 1: The structural and functional diagram of data transmission from the social networks to the database of user's profiles of IOEUTSAF**

Agents of IOEUTSAF are consistently traversing users' accounts of biggest social networks using their searching mechanism. For search we define basic criteria such as: native city and place of study. Place of study should correspond to one of universities of the hypothetical country but native city conversely, should be outside of this country. Also DBUP included the data about «friends» of such persons.

There are a number of citizens, who study foreign languages and cultures and have desire to communicate with foreigners. This category of the social networks users may be determined by analyzing «hobby» and «interests» attributes. If they belong to the corresponding communities it is possible to add such accounts «manually» by the moderators of IOEUTSAF.

The process of DBUP filling is permanent and in progress of it the number of records is increasing, that will positively reflect on the search results.

All data of people who are in DBUP belong to two categories: demographic data, which is filled by users (directly into IOEUTSAF or in some social networks), and data of users' ratings on certain resources. To search of «the ideal interlocutor» it is appropriate to apply the hybrid method of filtering that combines filtering by content for some user's data with demographic sense and collaborative filtering for the data of resources rating. In this case, for the total rate, weight coefficients for each of the methods are entered.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 1, Issue 8, October 2013

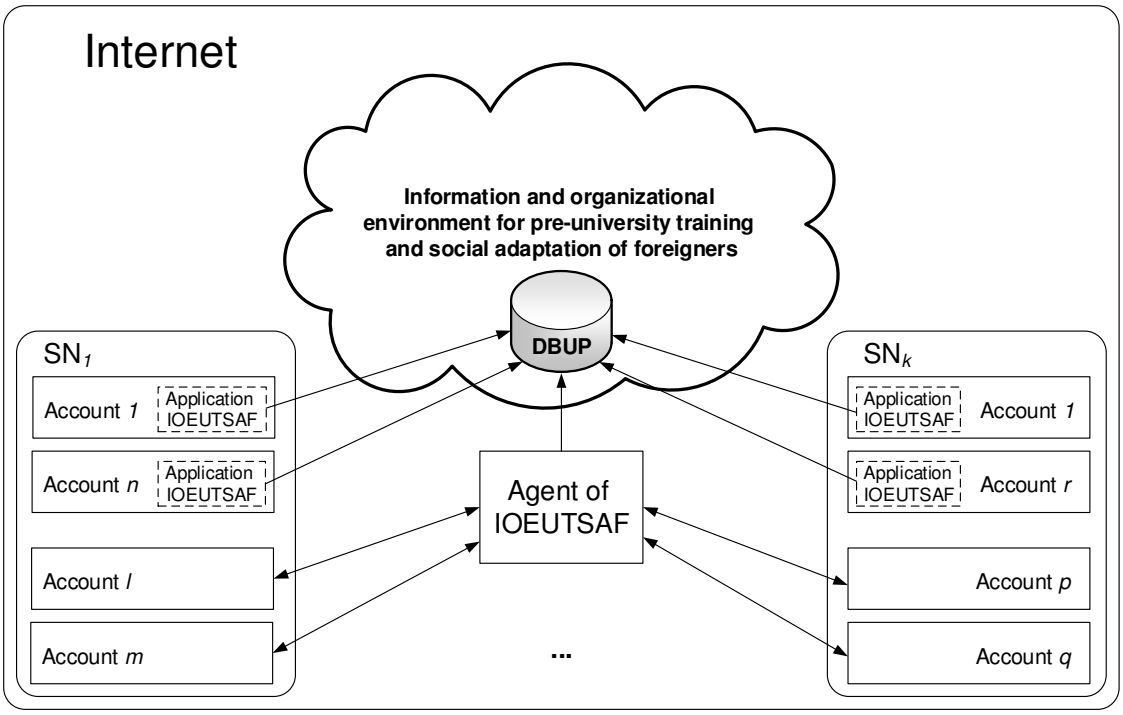


**Figure 2: The structural and functional diagram of filling the database of IOEUTSAF with personal users' data and fixing their ratings stories**

## III. RESULTS

Let us denote a set of students, who has been registered in IOE database, as $S$. In particular case we have database with all information has been received about students from different sources. Different students may give different sets of information – one student has filled some data fields and another one has not. Each student of set $S$is denoted as $s_i$:

$$s_i \epsilon S, \quad \sum_i s_i = S. \tag{2}$$

The personal information received from each student we will call "attributes" and denote as $A$. So we can write an equation:

$$s_i = \sum_k A_k. \tag{3}$$

But, on the other hand, we can describe each student by his preferences ($P$):

$$s_i = \sum_t P_t. \tag{4}$$

Both of embodiments presented above we suggest to use for ideal collocutor searching. Let us combine both embodiments to one:

$$s_i = \sum_k A_k \cup \sum_t P_t. \tag{5}$$

The equation (5) describes each student in IOE database as plural of attributes and preferences. But as we said before, in IOE database we could not have full information about each student, so it may be some attributes are empty. The preferences are not presented in IOE database. These features of student we need to extract from his profile in social networks.

Now we consider how to find ideal collocutor using presented information. Let us denote some student who is searching for collocutors as $s_0$:

$$s_0 = \sum_i s_i. \tag{6}$$

Firstly, we propose to reduce amount of candidates to ideal collocutors. For this purpose we need to select a several most important attributes and use them to separate trustful inappropriate students. As the important attributes we propose to use homeland, language and university.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 1, Issue 8, October 2013

The student $s_i$ is considering as possible collocutor if his attributes have the same values as the student $s_0$ attributes. But in case of absence some values of attributes in student these attributes are considering as matched and the student is still considering as possible collocutor and is put in the list ($L$):

$$L(s_i) \leftarrow (\sum_k A_k(s_0) = \sum_k A_k(s_i)) \vee \forall A_k(s_i) = \emptyset. \quad (7)$$

The list of the selected students we use for additional search based on activity in social networks. We look for any items (photos, movies, groups, celebrities and etc.) liked or not by each student in the list and combine these grades in the matrix M follow the rules (1). This matrix is not as sparse as was said before, because we reduce amount of candidates thanks to personal information from IOE database. But it is still so huge and unsuitable for the processing. To handle this problem we suggest using Singular Value Decomposition (SVD) to matrix $M$:

$$M = U\Sigma V^T, \quad (8)$$

where$U$ – $m$ x $m$ unitary matrix;

$\Sigma$ – $m$ x $n$ diagonal matrix;

$V^T$ – $n$ x $n$ conjugate transpose of the unitary matrix $V$.

The diagonal matrix $\Sigma$ consists of singular values of $M$ denoted as $\sigma$. A common convention is to list the singular values in descending order. In this case the diagonal matrix $\Sigma$ is uniquely determined by $M$. Right now it seems like we just complicated the problem and enlarge amount of data. But because of the matrix $M$ has a low rank $r$ we can reduce size of the result matrix $\Sigma$. We may leave only first $r$ elements in the matrix $\Sigma$ and set to 0 others.

Than size of the matrix $\Sigma$ will $r$ x $r$, the matrix $U$ – $m$ x $r$, the matrix $V$ – $r$ x $n$:

$$M = \begin{bmatrix} u_0^0 & \dots & u_0^r \\ \dots & \dots & \dots \\ u_m^0 & \dots & u_m^r \end{bmatrix} \cdot \begin{bmatrix} \sigma_0 & 0 & 0 \\ \dots & \dots & \dots \\ 0 & 0 & \sigma_r \end{bmatrix} \cdot \begin{bmatrix} v_0^0 & \dots & v_0^n \\ \dots & \dots & \dots \\ v_r^0 & \dots & v_r^n \end{bmatrix}. \quad (9)$$

The maximal value of $r$ in worst case is equal to count of the students in list $L$ and it is far lower than count of the items.

Rows of the matrix $U$ present each student by a set of factors $u_i$ and columns of the matrix $V$ present each item by a set of factors $v_j$. After that we use dot product of these factors $u_i \cdot v_j$ to predict the student $i$ rating for the item $j$. As a result we got the matrix $m$ x $n$ again but it is not sparse and each item has foreseen rating.

Now we can use this foreseen matrix to compute collaborating rank of current student in respect to others. For this purpose we suggest to use Pearson correlation coefficient:

$$w_i^j = \frac{\Sigma_a (r_i^a - r_i) \cdot (r_j^a - r_j)}{\sqrt{\Sigma_a (r_i^a - r_i)^2} \cdot \sqrt{\Sigma_a (r_j^a - r_j)^2}}, \quad (10)$$

where$w_{i,j}$ – correlation coefficient of students $i$ and $j$;

$r_{i,a}$ – rating of the item $a$ by the student $i$;

$r_{j,a}$ – rating of the item $a$ by the student $j$;

$r_i$ – average rating of all items by the student $i$;

$r_j$ – average rating of all items by the student $j$.

The average rating can be founded by a simple equation:

$$r = \frac{1}{n} \sum_{a=0}^{n} r_a. \quad (11)$$

After computation we have for each student $i$ his collaborating rank vector with each student $j$. Each student with rank greater than 0.8 selected as ideal collocutor.

## IV. CONCLUSIONS

During developing of the method for searching of «the ideal interlocutor» within the design of IOEUTSAF we proposed take as a basis and refine existing methods for recommendation systems, namely the method of filtering content and collaborative filtering.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 1, Issue 8, October 2013**

To search of «the ideal interlocutor» we need to create DBUP. A list of attributes, which should be saved in this database, is given.

For the sake of increase the number of users we proposed to use data from the accounts of popular social networks. The criteria of selection of users in social networks were proposed.

The ways of IOEUTSAF and popular social networks users' data collecting for DBUP filling were described.

The mathematical method for assessment conformity of users was developed.

The described method let us to find «the ideal interlocutor» in among of IOEUTSAF and social networks users.

## REFERENCES

[1] SV Paliy ["Formulation of problem of designing of distance education system for foreign students"],VisnykSkhidnoukrayins'kohonatsional'nohouniversytetuim. V. Dalya. Naukovyyzhurnal,Luhans'k, Vol.6 (136) Part 1, pp. 309-312, 2009. http://archive.nbuv.gov.ua/portal/soc_gum/vsunu/2009_6_1/69.pdf [Article in Ukraine].

[2] SV Paliy ["Creating a structural model of information and preparation of the organizational environment and social adaptation of foreign students"],Upravlinnyarozvytkomskladnykh system, Vol.8, pp. 112-116, 2011. http://urss.knuba.edu.ua/files/zbirnyk-8/111-115.pdf [Article in Ukraine].

[3] SV Tsyutsyura, SV Paliy ["Formation of information-organizational environment using calligraphic trainer for pre-university preparatory foreigners"],Systemyobrobkyinformatsiyi,Kharkiv, KhUPS, Vol.09 (107), pp. 264-269,2012. [Article in Ukraine].

[4] SV Paliy ["Social networks as a communication medium e-learning"],Upravlinnyarozvytkomskladnykh system, Vol.13, pp. 152-156,2013. http://urss.knuba.edu.ua/files/zbirnyk-13/152-156.pdf [Article in Ukraine].

[5] D.Lemire, A.Maclachlan. Slope One Predictors for Online Rating-Based Collaborative Filtering. In SIAM Data Mining (SDM'05), Newport Beach, California, April 21-23, 2005.

[6] Hennig-Thurau, Thorsten, A.Marchand, and P. Marx, "Can Automated Group Recommender Systems Help Consumers Make Better Choices?", Journal of Marketing, vol. 76 (5), pp. 89-109,2012.

[7] D.Jannach; M.Zanker; A.Felfernig; G. Friedrich,"Recommender Systems: An Introduction", Cambridge University Press, New York, p 352,2010.

[8] MM Glybovets, SS Gorohovskiy, AA Pika. ["Agents for recommendations in collaborative environments"],NaukovipratsiMykolayivs'kohoderzhavnohohumanitarnohouniversytetuim. Petra Mohylykompleksu "Kyyevo-Mohylyans'kaakademiya", Vol.121, No.134,Komp'yuternitekhnolohiyi, pp. 142-151. [Article in Ukraine],2010.

[9] P. Melville, R.J. Mooney and R.Nagarajan, "Content-Boosted Collaborative Filtering for Improved Recommendations." Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI-2002) Edmonton, Canada, pp. 187-192, July 2002.

[10] F. Ricci, L.Rokach, B.Shapira, P.B. Kantor "Recommender Systems Handbook", Springer, p. 875,2011.

[11] V.Sugumaran. "Application of Agentsand Intelligent Information Technologies", Idea Group Inc, p. 377, 2007.

## BIOGRAPHY

**Mr. SergiyPaliy**

He received B.Sc. and M.Sc. degrees in computer science from Kyiv National University of Construction and Architecture, Ukraine, in 2001 and 2002 respectively.From 2008 to 2012 - Dean of Preparatory Department for Foreign Citizens in Kyiv National University of Construction and Architecture. Since 2007 assistant professor of the Computer Science Basics department in Kyiv National University of Construction and Architecture, Ukraine.Fields of Specialization are E-Learning, Remote Pre-University Preparation of Foreigners, Social Networking, Social Adaptation of Foreigners, Information and Organizational Environment, Calligraphy Trainer, Programming Language etc.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 1, Issue 8, October 2013**

**Mr. YevgeniyBorodavka**
Received B.Sc. and M.Sc. degrees in computer science and Ph.D. degree in computer aided design from Kyiv National University of Construction and Architecture, Ukraine, in 2002, 2003 and 2009, respectively. Since 2005 assistant professor and since 2009 associate professor of the Information Technologies of Design and Applied Mathematics department in Kyiv National University of Construction and Architecture, Ukraine. From 2004 to 2009 - lead engineer and from 2009 to 2012 - senior scientific staff in State Enterprise "State Research and Development Institute of Computer Aided Design in Construction". Since April 2013 - lead engineer in Samsung Research and Development Institute Ukraine (SRK).

**Mr. MykolaTsiutsiura**
In 2012 made off a city council from speciality the "Informative managers of the system and technology". In 2013 made off a city council from speciality of "Managing projects". From November 2012 for present tense is a graduate student of the Kyiv national university of construction and architecture from speciality 05.13.22 –"Managing projects and programs".