International Review of Management and Marketing

ISSN: 2146-4405

available at http: www.econjournals.com

International Review of Management and Marketing, 2016, 6(S6) 214-220.

Special Issue on "Management of Systems of Socio-Economic and Legal Relations in Modern Conditions of Development of Education and Society”

The Formation of User Model in Scientific Recommender Systems

A.I. Guseva[1]*, V.S. Kireev[2,3], P.V. Bochkarev[4,5], D.S. Smirnov[6,7], S.A. Filippov[8]

[1]National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, Russia, [2]National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, Russia, [3]Event Service Provider Ltd., (ESP, OOO SKT), Yegor’ye Village, Medynskiy Area, Kaluga Region, Russia, [4]National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, Russia, [5]Event Service Provider Ltd., (ESP, OOO SKT), Yegor’ye Village, Medynskiy Area, Kaluga Region, Russia, [6]National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, Russia, [7]Event Service Provider Ltd., (ESP, OOO SKT), Yegor’ye Village, Medynskiy Area, Kaluga Region, Russia, [8]National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, Russia. *Email: aiguseva@mephi.ru

ABSTRACT

In this paper are considered questions which connected with building of the hybrid user model in scientific recommender systems. Offered model is consistent with conceptual model of the presentation scientific Common European Research Information Format data. The authors propose an approach using semantic networks. The model considers the main entities of a physical data model, received from generally available sources. During creation of model approaches from area of marketing were used, in particular psychographic approach. Quantitative experiment for profiles of the users of the “International congress conference ‘Information technologies in education’” portal. Results of experiment have confirmed adequacy of the offered models. The offered detail level of model and methods of its representation are directed to further improving of a pertinent of all personalisation algorithms. This study was financially supported according to Federal Target Program “Research and development on priority directions of scientific-technological complex of Russia for 2014-2020” (grant No. RFMEFI57614X0068).

Keywords: Recommender System, Formal Conceptual Common European Research Information Format Model, Hybrid User Profile
JEL Classifications: M15, M12, O32

1. INTRODUCTION

Recommender systems (RS) is a software and methods for forecasting of the user behaviour relative of the information searching object and recommendation forming for objects which not seen before (Ricci et al., 2011). Formed recommendations help users in different decision making process such as choice of the scientific article, choice of discipline for study, etc. Similar recommendations are built on the grounds of features these object and (or) of the user profile.

Development of the RSs possible on base discipline approach only, when is used together intellectual analysis data (Data Mining), decision theory, statistical methods of data processing, marketing, theory of consumer behavior, expert systems, etc. Formation of recommendations made on the basis of the data. Data used by RS belong to three types of objects: Elements, users and transactions (ratio between the users and items).

In scientific RSs the scientific article, the book and the patent can be elements (or objects of information search). The element can consist of primary terms, i.e., information units such as word, author, name of style, etc. Elements can be described by its complexity, value or usefulness. The value of the element can be positive, if the item is useful for the user and negative if the element is not necessary, and the user has taken a negative decision in choosing it.

The system user can have their own tastes and preferences. Information about users can be collected in different ways. There is always a user profile, explicit or implicit at the user model. The explicit profile is created by means of filling of questionnaires, inquiries, etc. In this case the user is personalized. The implicit user profile is created by means of the analysis of its actions on the website.

Transaction is understood as the recorded interaction between the user and RS. For example, the transaction log may contain the link to the element which selected by the user and the description of a context (for example, user query) for formation of the recommendation. Such transaction can also reflect existence of explicit back communication which the user provided, in the form of an assessment of the selected element.

Actually, estimates which are collected by RS are the most popular form of transaction data. These estimates can be collected explicitly or implicitly. In an explicit set of estimates of the user ask to provide her judgement on an element on a rating scale. In the transactions which collect implicit estimates the system aims to output judgement of the user on the basis of its actions. The dialogue system support interactive process. In such systems more advanced transaction models. In this case the user can request the recommendation and the system can make the list of a sentence. Thus RS reports to the user the best results on the basis of additional user preferences.

One of the relevancy problems is creation of adequate model of the user of system for increase of accuracy in case of demand personalization. This situation arises by development of RSs. This model must take into account the advisory system specifics, i.e., content and behavior of the real human interaction with that content.

In this paper it's offered to construct model of a user profile of scientific systems for pertinent increase of recommendations of responses in the field of scientific and educational information.

2. THE PRINCIPLES OF RS CREATION

RSs are classified as content, collaborative and hybrid. In case of content filtering user profiles and objects are created. User profiles can include demographic information or responses to a certain set of questions. depending on an object type profiles of objects can include names of monographs and articles, names of authors, etc.

Content filtering is focused on a clear classification of users and objects that appear in the information offer. In this case direct compliance between users and objects on the basis of their characteristics is set direct compliance between users and objects on the basis of their characteristics is set. In general strategy well works in areas with finite and rather small amount of the evaluation criteria following from the nature of things in case of big information streams. And also allows a large number of criteria in case of a small information flow. The main problems are a classification and creation of new information sentences.

In case of collaborative filtering information on behavior of users in the past is used. For example, information on its orders or estimates. In this case doesn't matter with what object types operation is carried, but at the same time implicit characteristics which would be difficult to be considered during creation of a profile can be considered. Main problem of this type of recommendatory systems is "cold start." It's mean absence of data on the users who recently appeared in system or objects.

Collaborative filtering is guided by the analysis of a path of the user to a large extent, i.e., it is considered according to what links he transferred as evaluated on what he delivered a tab where used social buttons that looked for, and also implicit actions-where the analysis of cumulative actions and behavior, the analysis of habits was delayed longer where it is less. The considered strategy is considered the most perspective for today, including competitions among scientific groups for search of the best algorithms are regularly held.

Hybrid RSs use both types of filtering.

The assessment of result of operation of recommendatory systems can be carried out from the point of view of relevant and pertinent (Kireev et al. 2015). As a rule, relevant is understood as compliance of the acquired information to a query. Pertinent is understood as compliance of the documents found the information retrieval system to information needs of the user irrespective of the fact how fully and precisely this need is expressed in the form of request. Pertinent is defined by subjective perception of the person. Because of subjectivity of pertinent it is impossible to achieve exact coincidence. Any search engine is customized for information needs of the average, but not specific user. Insufficient pertinent of searches is caused by the following basic reasons: Excessive decomposition of requests and service on excessively broad requests, without personal properties of the user.

Excessive decomposition of requests is that information need of the user revealed in the form of a series from 7 to 10 very specific requests. In case of service on excessively broad requests for one request the subscriber receives from hundreds of thousands to hundreds of millions of documents, web pages. Though directly there corresponds to request only the small part of information.

Scientific RSs have the features. Object of information search in them is the scientific result (SR): Theses of reports, articles, monographs and textbooks, preprints, the defended dissertations and results of intellectual activities. Scientific networks, electronic libraries and databases can be a source of such information. The behavior of users of such systems is almost not studied. Thus, question of improvement of quality of information search (pertinent) at the expense of the accounting of features of behavior of users of scientific systems is especially actual (relevancy).

3. THE QUESTION STATUS

Scientific operations are carried out by the organizations, research teams and certain scientists. One of approaches which is used in scientific recommendatory systems relies on properties of scientific documents. One of the most known is formal conceptual model of scientific data Common European Research Information Format (CERIF), which is constructed on entities such as Publication, Project, Person, Organization, Event ("CERIF 2008-1.2 Full Data Model. Introduction and Specification," 2008; "CERIF 2008-1.2 Semantics," 2008; "CERIF 2008-1.2 XML Data Exchange Format Specification," 2008). The basis of the CERIF model is the publication and its attributes, type of users in it aren't considered.

In other researches it is offered to use the graph of the concepts constructed on keywords for the recommendation scientific documents (De Nart and Tasso, 2014). This RS is based on separation of keywords for generation of meta data in documents which are used for provision of the useful recommendations. This semantic approach allows to create the user's model. This approach is planned to be used for recommendations in different areas, such as news, patents and to legal documentary archives. The recommendatory Rec4LRW system is focused on three research tasks. The first is creation of the list for reading scientific articles. The second is search of similar documents on the basis of a set of documents. And the third is abbreviation of the list for switching on in the manuscript based on article type. The system is intended for assistance, to researchers in search of the purposes of writing of research and development operations and the review of literature. Methods of result of recommendations in Rec4LRW are based on the intermediate set of criteria which capture characteristics of the scientific article (Raamkumar et al., 2015).

Opposite approach suggests to build model of the user of recommendatory system, transferring its characteristics from other systems. For the solution of the task of cold start in certain cases use cross-models of transfer of model of the user from scientific networks (Wongchokprasitti et al., 2015). For the solution of the task of cold start in certain cases use cross-models of transfer of model of the user from scientific networks. It generally is based on keywords on the basis of researches of relevance and on a social network between researchers.

Also more difficult approaches to establishment of communications are offered. For example, in (Yang et al. 2015) new measure of similarity at the institutional level which measures communication force between associated institutions of researchers is proposed. In other researches the hybrid model of the user which includes a combination of joint filtering and filtering on the basis of content is offered (Hao et al., 2016). This hybrid model uses estimates, subjects or classes of elements. In this regard, creation of hybrid models of the user is are very perspective.

On the other hand, there is a row of researches of behavior of users on social networks which still weren't carried out for users of scientific networks.

These researches prove that all users of social networks it is possible to divide into several classes, i.e. patterns of behavior of users are selected. The quantity of classes into which divide Internet users is in an interval from 6 to 10 depending on approach to their formation. There are operations where types of users are selected, considering a way of life of the person in general (psychographic approach) or on the basis of communications with the professional sphere and the sphere of personal interests (social-demographic approach). In our case, a field of activity and the professional sphere of our users one. It is science. And both approaches are the insufficiently informative for our researches.

But approaches when patterns of behavior are selected based on actions of users which are carried out on a network are known. Treat them: Social technographic approach, socio-political approach, empirical approach.

Certain actions of the user on the Internet are the cornerstone of social technographic approach. On its basis five active are selected and one inactive class of users, and a share of inactive users who didn't manage to be classified makes 23% (Nechaev et al., 2014; Al Zamal et al. 2012). The following types belong to the classes of the active users.

- Creators. Their principal characteristic are the activities directed to creation and the publication of content: Have blogs, write and spread in article network, reviews
- Critics of which the activities directed to expression of the relation to already created and published on a network are characteristic. They use forums, express the opinion on web pages, in web blogs, on pages of a social network of other users; on the websites, forums of the companies, shops, firms of vendors, etc
- Collectors which classify and will organize Internet content: Participate in compilation of ratings of the websites, use RSS flows, etc
- Joiners, i.e., users of the websites and blogs of social networks
- Spectators, which activities are directed to consuming of content: Download, listen to audio and video, read forums, visit the sites of recommendations, etc.

In paper (Ryabchenko and Gnedash, 2014) user classification based on socio-political approach and roles which they play in the formation of a specific semantic concept in online-social networks and online-social communities. In this case "Opinion leaders," "Sensors," "Retailers," "Readerships," "Reputational players" were allocated. "Opinion leaders" change an information and news field. The active authors treat them. "Sensors" collect significant information within this or that semantic concept they leave a context of that information which is entered by leaders of judgements. "Retailers" launch a blogwave as it is conscious, and isn't present. "Readerships" permanently are in an online-social network and practically don't generate public messages. However, under certain conditions these users can change the role on "Retailer." "Reputational players" use a social network for solidifying of the image. Most often they are included into online-network community while popularity of community is sufficient is high. As "reputational" players representatives of business and the power appear.

In paper (Chijik, 2014) the results of studies of the behavior of the Russian-speaking Twitter users. It is shown that the active Twitter users can be divided into three types: Informers which disseminate information and have at least formal contact with users; aggregators which publish information but don't have contact with other users; and users which publish news only for friends. In a percentage ratio these types correspond with each other as 15%, 5% и 80% respectively. As informers in Twitter there are politicians and their assistants, representatives of firms which are engaged in advance and are interested in feedback with the information readers, etc. Aggregators members of the media most often appear and mass-media persons. Their only purpose is to alert the audience. The remaining 80% are regular users who, after registering on Twitter, and install the application on a smartphone, can quickly read news from their interest accounts.

Thus, summarizing the above results, it can be concluded that the types of users scientific networks in a certain way must be correlated with the internet users types. Naturally, at the same time it is necessary to consider specifics of scientific networks.

4. THE OFFERED APPROACH

Considering the scientist as the user of scientific system, the main actions of the user with system were selected: Search, read, download, discuss (reviewing), publication or establishment of authorship, quote. Generalizing the considered earlier different approaches it is possible to select categories (types) of users of scientific networks and systems:

- C1 - The active users who are publishing SRs and actively using scientific systems for involvement in life of community
- C2 - The critics who are actively discussing foreign SRs tracking the communications popularizing SRs, etc
- C3 - Pragmatists which study others materials, correctly publish the materials, without allowing any discussions
- C4 - The collectors of content tracking the communications using quotes, etc
- C5 - Readers, the beginning scientists (undergraduates, post-graduate students) who collect and download the necessary materials.

The model of interaction of the user with scientific system is provided in Figure 1.

It's possible to provide (to set) the user's types considered earlier through a set of actions taking into account specifics of scientific systems (Table 1).

Introduction of similar types allows to construct model of a user profile of scientific systems more precisely. It leads to increase of the information sentence pertinent.

Table 1: Types of the user profiles

Process	C1	C2	C3	C4	C5
Search	1	1	1	1	1
Read	1	1	1	1	1
Download	1	1	1	1	1
Discuss	1	1			
Publish	1		1		
Quote	1	1	1	1	

5. DISCUSSION OF THE RESULTS

The RS work with two data sources-database of user preferences и database of SRs. The main sources of data for the user can be considered his personal data which he indicates in the questionnaire for registration on the website. Also the activity log which is automatically recorded by the system, based on what actions the user performed in the scientific system.

The process of the reception recommendation by final user consists of several steps. Personal data from a user profile, logs of actions of the user which he makes by operation with system will be basic data. Recommendations are given based on user based on the user belongs to some group of users with similar interests. The definition of such classes can help to create user demand more accurately, i.e., to increase it's pertinent. This operation is called clustering. Reference of the new user to the most suitable class is executed by means of classification operation. As a result the recommendation is issued to the user in the form of information sentence if it's known what class the user treats and what other participants with similar interests are interested in.

For scientific RSs as main features of the user follows to select-age (date of birth), received education, degree and rank (if present). Also here possible refer an affiliation with the organization (name) and its location (country + city/town).

The SR can have a different type (for example, article, monograph, patent, thesis, etc.), can contain different keywords, pertain to concrete UDK, have a ISDN-number, ISBN-number, can be published to certain date in concrete journal (journal, number, publishing). Except this SR has such features as name, abstract and reference.

For example, key importance are brought for different action of the authorized user in system-a publications and reading (Table 2).

As quantitative restrictions for published material was taken 20 works that at the average satisfies majority scientific employee. Aside from this, names and keywords of the published material must be processed and kept in separate fields that will allow hereinafter more exactly to execute categorization. Also keywords on which cites the author of the work will be taken into account. For user which reads scientific papers was taken 30 papers monthly.

The user interacts with SR or its properties, making different actions which are fixed in a system log (create, read, assign evaluation, download, etc.). Each action also has the spatial and temporal characteristics when and where the action is made). Except actions which are fixed in system the user can also be the author or the coauthor of SR. It must be recorded by means of separate type of the relations. In general, it makes sense to present such model in the form of a semantic network (graph) which provides a possibility of existence of heterogeneous

Figure 1: The model of user interaction with scientific system

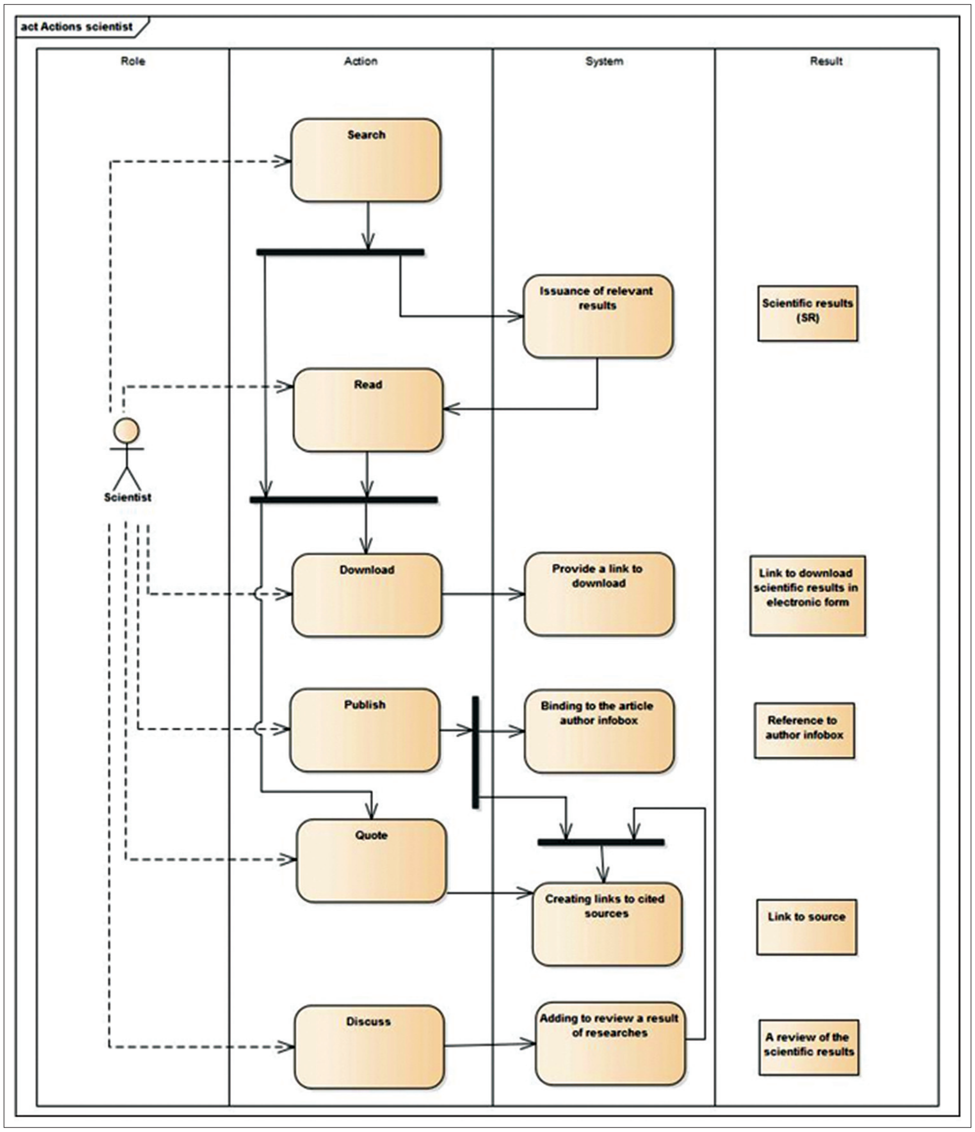

act Actions scientist
Role
Action
System
Result
Search
Issuance of relevant results
Scientific results (SR)
Read
Scientist
Download
Provide a link to download
Link to download scientific results in electronic form
Publish
Binding to the article author infobox
Reference to author infobox
Quote
Creating links to cited sources
Link to source
Discuss
Adding to review a result of researches
A review of the scientific results

Table 2: Information about the man's scientific activity

What man has done in the system				
Field	**Comment**	**Field**	**Comment**	**Type**
CB_author=publishes	Publishes 20 articles per an num	CB_act=read	Reads 30 articles at month	Number
ID_Scient Res		ID_Scient Res		Number
Name	5 words	Name	5 words	Text
Abctract		Abctract		Text
Keywords	10 keywords	Keywords	10 keywords	Text
Type		Type		Number
UDC		UDC		Number
Date		Date		Number
Refers to	5 additional key words	Refers to	5 additional key words	Number
HP_organization				Text

communications. Then it is possible to select entities (vertices) people, Science result, actions and organizations which are connected among themselves, form the full-meshed graph.

Such full-meshed graph is offered to present in the form of a semantic network, i.e., information model of the data domain having an appearance of the oriented graph which vertices correspond to objects of data domain and arcs (edges) set the relations in between (Figure 2).

As a result of data preparation for creation of recommendations the relational database is used. It represents a projection of a semantic network, taking into account restrictions for the volume of the communications characterizing quantity of information units (SRs)

Figure 2: The user profile of the model scientific system in the semantic network form

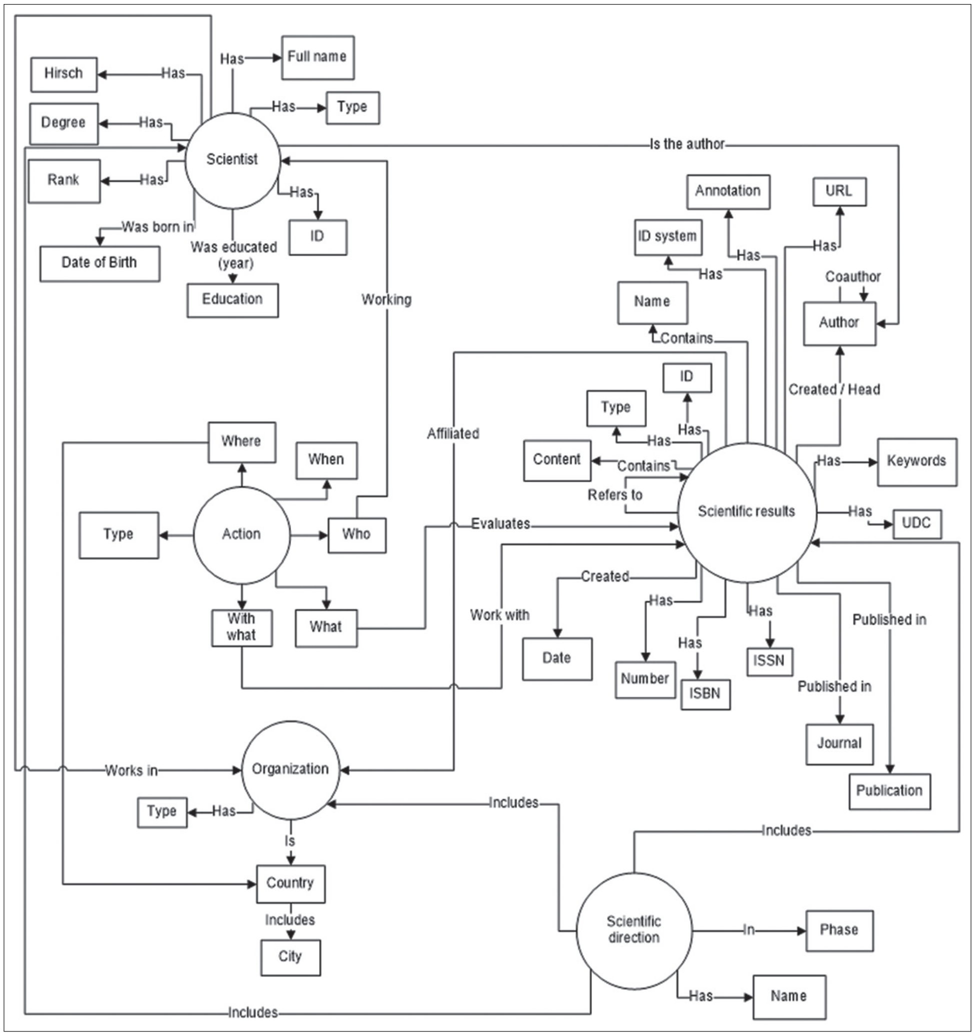
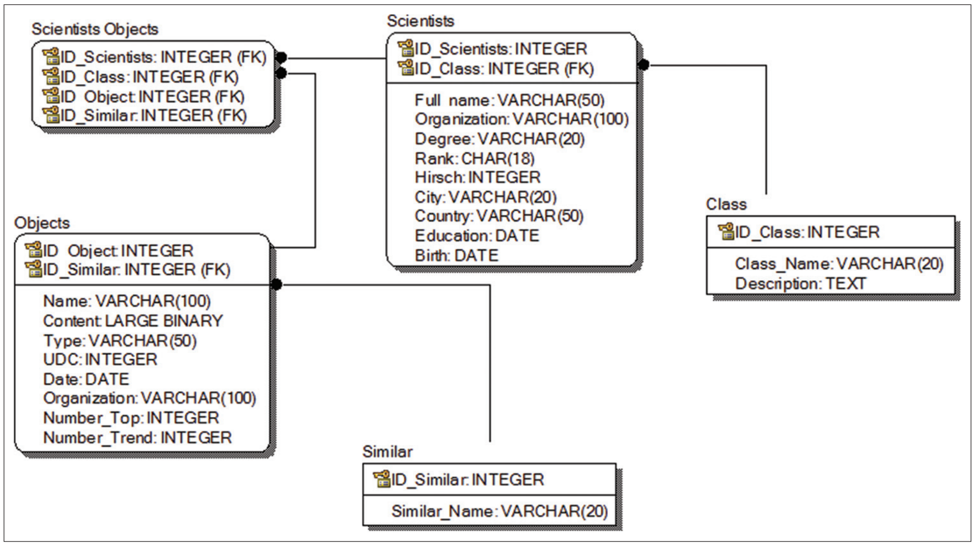


Figure 3: The relational profile user model of scientific system


Scientists Objects

ID_Scientists: INTEGER (FK)
ID_Class: INTEGER (FK)
ID_Object: INTEGER (FK)
ID_Similar: INTEGER (FK)

Scientists

ID_Scientists: INTEGER
ID_Class: INTEGER (FK)

Full name: VARCHAR(50)
Organization: VARCHAR(100)
Degree: VARCHAR(20)
Rank: CHAR(18)
Hirsch: INTEGER
City: VARCHAR(20)
Country: VARCHAR(50)
Education: DATE
Birth: DATE

Objects

ID_Object: INTEGER
ID_Similar: INTEGER (FK)

Name: VARCHAR(100)
Content: LARGE BINARY
Type: VARCHAR(50)
UDC: INTEGER
Date: DATE
Organization: VARCHAR(100)
Number_Top: INTEGER
Number_Trend: INTEGER

Class

ID_Class: INTEGER

Class_Name: VARCHAR(20)
Description: TEXT

Similar

ID_Similar: INTEGER

Similar_Name: VARCHAR(20)

with which the user of system interacts (Figure 3). Quantitative experiment for profiles of the users of the "International congress conference 'Information technologies in education'" portal was carried out (http://ito.evnts.pw). The total quantity of attributes in the database of users preferences in the made experiment was 1279 where 9 attributes describe a person, 520 attributes describe scientific activity of this person и 750 attributes describe activity in the system of this person. Results of clustering by the users actions of the portal the hypothesis proposed allocation has led to the allocation of the five classes.

6. CONCLUSION

Offered in this paper hybrid user model in scientific RS will with conceptual model of the presentation scientific CERIF data. When making the models were used not only characteristic scientific publication and main essence to physical model data, which got from available (for all) sources, but approaches from area of marketing, in particular psychographic approach. Presentation to models is realized with use the semantic networks.

Quantitative experiment for profiles of the users of the “International congress conference ‘Information technologies in education’” portal have confirmed adequacy of the offered models. The offered detail level of model and methods of its representation are directed to further improving of a pertinent of all personalisation algorithms.

Thus, the received results of a clustering can be used for framing of personal recommendations to the user on the most often made actions. Further development of approach is presented in the form of extension of a feature set at the expense of the accounting of properties of scientific publications.

REFERENCES

Al Zamal, F., Liu, W., Ruths, D. (2012), Homophily and Latent Attribute Inference: Inferring Latent Attributes of Twitter Users from Neighbors. In: Sixth International AAAI Conference on Weblogs and Social Media. Homophily and Latent Attribute Inference: Inferring Latent Attributes of Twitter Users from Neighbors: Association for the Advancement of Artificial Intelligence.

CERIF 2008 - 1.2 Full Data Model (FDM). Introduction and Specification. (2008), Eurocris. Available from: http://www.eurocris.org/Uploads/Web%20pages/CERIF2008/Release_1.2/CERIF2008_1.2_FDM.pdf. [Last retrieved on 2016 Apr 27].

CERIF 2008 - 1.2 Semantics. (2008), euroCRIS. Available from: http://www.eurocris.org/Uploads/Web%20pages/CERIF2008/Release_1.2/CERIF2008_1.2_Semant ics.pdf. [Last retrieved on 2016 Apr 27].

CERIF 2008 - 1.2 XML Data Exchange Format Specification. (2008), euroCRIS. Available from: http://www.eurocris.org/Uploads/Web%20pages/CERIF2008/Release_1.2/CERIF2008_1.2_Semantics.pdf. [Last retrieved on 2016 Apr 27].

Chijik, A. (2014), Social and linguistic research of some tendencies of the publication of posts on Russian-speaking Twitter. New information technologies in automated systems. Collection of Scientific Articles, 17, 337-347.

De Nart, D., Tasso, C. (2014), Personalized concept-driven recommender system for scientific libraries. Procedia Computer Science, 38, 84-91.

Hao, J., Yan, Y., Wang, G. (2016), Probability-based hybrid user model for recommendation system. Mathematical Problems in Engineering. Nasr: Hindawi Publishing Corporation.

Kireev, V., Kuznetsov, I., Bochkarev, P., Guseva, A., Filippov, S. (2015), Development of model of the user of scientific networks on the basis of the concept. Open science. Fundamental Research, 12-5, 907-913. Available from: http://www.fundamental-research.ru/ru/article/view?id=39649.

Nechaev, V.D., Brodovskaya, E.V., Kair, Y.V., Dombrovskaya, A.Y. (2014), Classification of Russian Internet users: Preliminary results of cluster analysis. Life Science Journal, 11(12), 330-335.

Raamkumar, A., Foo, S., Pang, N. (2015), Rec4LRW - Scientific Paper Recommender System for Literature Review and Writing. Advancesin Digital Technologies: Frontiers in Artificial Intelligence and Applications, 275, 106-119.

Ricci, F., Rokach, L., Shapira, D., Kantor, P. (2011), Recommender Systems Handbook. Berlin: Springer Science.

Ryabchenko, N., Gnedash, A. (2014), Types of Users of Online-social Networks: The Teoretiko-Methodological Bases for Classification. Technologies of Information Society in Science, Education and Culture: Collection of Scientific Articles. XVII of the All-Russian Integrated Conference “Internet and the Modern Society” (IMS-2014): Saint Petersburg National Research University of Information Technologies, Mechanics and Optics. p143-148.

Wongchokprasitti, C., Peltonen, J., Ruotsalo, T. (2015), User Model in a Box: Cross-System User Model Transfer for Resolving Cold Start Problems. 23rd International Conference on User Modeling, Adaptation, and Personalization (UMAP). Dublin: Springer Verlag.

Yang, C., Ma, J., Silva, T. (2015), Multilevel information mining approach for expert recommendation. Online Scientific Communities, Computer Journal, 58(9), 1921-1936.