Scientific and methodological approaches to the generation of the Internet information space of scientific and educational resources

A.A. Zatsarinny[0000-0002-8872-2774], F.I. Ereshko[0000-0002-1732-2204], V.I. Medennikov[0000-0002-4485-7132]

Federal Research Center "Informatics and Control" of the Russian Academy of Sciences, Vavilova 44-2, 119333,Moscow, Russia
frccsc@frccsc.ru

Abstract. The article considers scientific approaches to the formation of the information space of scientific and educational resources. A methodology for assessing the effectiveness of their use is proposed.

Keywords: scientific and educational resources, performance evaluation, model sites, university rankings.

1 Introduction

Currently the world is transiting to the digital economics, which requires the integration of both information resources and information systems. This is achieved through the implementation of standards for representing information resources, management functions, as well as through the complex approach to the design, development and implementation of information systems, training of relevant specialists.

In connection with the exponential growth of information volumes in specific areas of activity, there were problems of creating new information technologies that provide the ability of necessary knowledge acquisition. These problems are of specific importance in the field of information accumulation in education and science field.

In this connection, the task is to create the unified Internet information space of scientific and educational resources, which will be in demand both in the process of scientific research support and education efficiency improvement at all levels. In our opinion, such the space should become a tool for resolution of the conflict between the volumes of accumulated knowledge and the possibility of its effective use.

The article describes the methodological approaches to generating the Internet information space of scientific and educational resources, as well as the methodology for evaluating the effectiveness of its use in the transition to the digital economics.

ICID-2019 Conference

2 Conceptual approaches to generating the unified Internet information space of the scientific and educational resources of Russia

The lack of integration technologies at the development of databases of various information scientific and educational resources leads to significant overexpenditure of resources and confuses the increasing number of potential users, when searching for the necessary information. For example, the state spends significant resources on the development and maintenance of information scientific and educational resources database: "The unified state information system for accounting results of commercial research-and-development activities as well as technological works, the database of the Federal Institute of Industrial Property (patent documents, trademarks, industrial patterns, computer programs, databases and topologies of integrated microcircuits) and Elibrary.ru, which have heterogeneous representation and rather narrow designated purpose, as well as very specific audience respectively. The purpose of the first database can be easily defined from its name. The commercial research-and-development activities as well as technological works are presented here in the form of scientific reports, drawn up by the budget funds. The second database purpose is to register the above mentioned developments without the possibility of studying them. The purpose of the third database is to create the national science citation index with the prospect of using it for evaluating the results of scientific work of scientists or research groups. However, the narrow focus of these databases, mainly on the accounting function, leads to the heterogeneity of their structures, making them unclaimed for wide range of users who want to have convenient system for getting knowledge. Unfortunately, valuable and up-to-date information from Russian scientific foundations and federal targeted programs is also practically unavailable for use in the innovation sphere.

The roots of this process lie in the lack of the integration approach in the technologies of design and development of information systems in our country, resulting in appearance of "hodgepodge" of tens and then hundreds and thousands of isolated and functionally incompatible local control systems at plants, factories and agricultural-and-industrial enterprises. The failure of the national network project, based on typification and integration of information systems, proposed by academician Glushkov V.M. [1, 2] on the creation of the National Automated System in the 60th of the XX century, also played a negative role. Although significantly increased opportunities and the level of Internet software and hardware development at present allow implementing the ideas of Glushkov V.M. in full.

The transition to the integrated information systems in the Russian Federation is complicated by some circumstances - the lack of the Internet intellectualization, i.e. the providers have no online tools for the development of mathematical models, statistical processing of information, expert systems, etc.

In the meantime, only very few providers have database management systems, which are used by few people. For example, the studies have shown that no agricultural higher educational institution uses database management systems, when creating websites, which could make it impossible to automatically receive information from

the website for its use in other information systems, including the information system of the Russian Ministry of Education and Science [3].

The analysis of websites of research institutions, higher educational institutions, information and consulting services allowed identifying seven types of information resources that are present in some form on these websites: developments, publications, consulting activities, regulatory information, distance learning, application program packages, databases. These are the types of scientific knowledge representation that are most in demand in the economics [4].

At that the improvement of Internet technologies allows integrating them (on the basis of ontological modeling) into the unified Internet information space of scientific and educational resources from unified scientific and methodological positions with simple (understandable to any user) navigation system with placing information resources in the cloud under control of the powerful database management system based on the uniform classifiers, such as the State Rubricator of Scientific and Technical Information and the All-Russian Classifier of Products [5, 6, 7].

For example, the manufacturer, having chosen some development as some disease control means, can immediately receive all publications, all consultations, regulatory information, and distance learning on this subject. Then he can find the necessary drug supplier in the appropriate database. At the same time, the knowledge required for the production is splitted to various databases that are not related to each other. At that a huge amounts of money are spent on their implementation and maintenance. However, only a small part of the databases is accessible for the scientific community, and it is almost inaccessible for manufacturers.

The possibility of creating the unified Internet information space of scientific and educational resources has been checked on the basis of economic and mathematical modeling, as well as practical implementation at developing the portal of the Russian Academy of Agricultural Sciences in 2007-2008. At that the following was registered there: 12321 publications, 2541 developments, 444 consultants for conducting consulting activities on the subject. At that time, Elibrary database contained significantly less number of publications, and there were no other types of scientific and educational resources (and at the moment they are still not available there) [5].

The requirements, which are imposed for the websites of higher educational institutions by the Ministry of Education and Science, the Federal Education and Science Supervision Agency of the Russian Federation and branch ministries to evaluate the activities of educational institutions, force these websites to become more similar to each other. In the nearest future higher educational institutions should perform the transition to standard websites. And this is the first step towards the generation of the unified Internet information space of scientific and educational resources. When implementing standard websites in higher educational institutions and research institutions and their integrating with this space, the information scientific and educational resources would automatically get there.

RePec (Research Papers in Economics) is the most interesting project in the West in the field of integrating information and scientific resources, which is characterized by the systematic approach that allows free global network access to the information resources of scientific papers on economics published in the world.

RePec is a project based on the joint work of hundreds of volunteers from 93 countries to expand the researches in the field of economics and related sciences [8, 9]. The center of the project is the decentralized bibliographic database of scientific papers, reports, documents, journal articles, books, chapters of books and software products released all over the world. RePec can be called the unified information space for economic researches on global scale.

Currently RePec is the world's largest online collection of working documents, journal articles and software products for economics. The databases for organizations and authors in the field of economics is also collected here. The high demand in such the project is evidenced by the fact that the users downloaded 441,497 document files and viewed 1,657,039 annotations in September 2017. In total 97,933,110 files were downloaded and 393,736,069 annotations were viewed since January 1998.

Thus, we can conclude that in large Western universities, the authors do not provide complete lists of their publications on their department website pages, but prefer to represent links in places, where their papers can be viewed completely.

3 Methods for evaluating the effectiveness of using the information scientific and educational resources

The requirements imposed to the website information content of educational institutions by the Ministry of Education and Science, the Federal Education and Science Supervision Agency of the Russian Federation have aroused the interest of various researchers for evaluating their activities based on these data. The approach to the selection of activities and their indicators in the West differs from the approach adopted in Russia. The difference of approaches can be explained by the fact that the scientific research is one of the main activities in foreign educational institutions and in Russia the educational activities are at the first place.

To eliminate this lack and to increase capabilities of Internet technologies based on the above mentioned, it is very important to develop the methods for evaluating the effectiveness of using information scientific and educational resources of higher educational institutions in the Internet space, taking into account both the requirements for the information content of educational institutions websites, imposed by the Ministry of Education and Science, the Federal Education and Science Supervision Agency, and their demand in the economics field, the impact on the training quality of qualified specialists and scientists in educational institutions, the evaluation of websites by webometrics methods, reflecting image and reputation of the higher educational institution, as well as current trends of providing information services by higher educational institutions in the form of electronic labor exchanges and trading platforms in the Internet space.

The special studies were conducted by the example of agricultural higher educational institutions. For that purposes their websites were monitored and analyzed. At that the specially developed original questionnaire was used at the websites, including the indicators from the set of requirements of the Ministry of Education and Science, the Federal Education and Science Supervision Agency of the Russian Federation, as well

as those indicators, reflecting information about developments, publications, consulting activities, regulatory information, distance learning, application program packages, databases, electronic labor exchanges and trading platforms that were not included in this list. Also the websites were evaluated using webometrics methods.

The questionnaire reflects 214 indicators of the activities of higher educational institutions (122 indicators for evaluating the representation of the higher educational institution itself, 40 indicators for evaluating the department, 46 indicators for evaluating the chairs and 6 indicators for overall website evaluation).

When selecting indicators from the set of requirements of the Ministry of Education and Science of the Russian Federation, the Federal Education and Science Supervision Agency of the Russian Federation, the focus in studies was on selecting the most significant indicators of the activities of higher educational institutions, affecting the achievement of their goals: the training of qualified specialists and scientists, the production of scientific products. The indicators, representing the information on developments, publications, consulting activities, regulatory information, distance learning, application program packages, databases, are related to primary information scientific and educational resources.

The information resources, reflecting the requirements of the Ministry of Education and Science of the Russian Federation, the Federal Education and Science Supervision Agency of the Russian Federation, selected on the basis of expert examination, most influencing the achievement of the goals of higher educational institutions, are called secondary information scientific and educational resources. The indicators, reflecting information on consulting activities, will be presented in terms of the number of consultants.

According to modern trends in the field of Internet technologies, when providers begin providing services of website content storage in database management systems, the information scientific and educational resources can be stored either in the form of catalog or in the form of full-format representation (information resource storage form). On the other hand, it can be stored either in the form of disordered list, or in the form of ordered representation. The ordered representation provides the ability of navigation, for example, on the basis of database management systems on thematic rubrication by the State Rubricator of Scientific and Technical Information and the All-Russian Classifier of Products, authors, organizations, keywords. The disordered and ordered representation is called the information resource integration level.

The integral criterion for evaluating the effectiveness of using the information scientific and educational resources of particular educational institution is defined as the sum of weighted groups (the total sum of weights is equal to 1) of the following particular criteria: criterion for evaluating the types of information scientific and educational resources representation, criterion for evaluating the efficiency of using information resources by webometrics methods, criterion for evaluating the effectiveness of using information resources in terms of the state of the electronic trading platform, criterion for evaluating the effectiveness of using the information resources in terms of the state of the electronic labor exchange.

The values of weights for indicators of the criteria for evaluating the effectiveness of using the information resources are determined on the basis of expert evaluations

obtained by analyzing various articles of specialists in the educational field [5, 10], methods for calculating various ratings of educational institutions, questioning teachers of the Russian State Agrarian University - Moscow Timiryazev Agricultural Academy, as well as the use of appropriate statistical methods.

At that the efficiency is considered as the productivity in achieving the goal in terms of operation research methods. In our case, the goals of generating information scientific and educational resources are:

- availability of information scientific and educational resources for wide range of users (applicants, students, teachers, state authority officials, manufacturers, researchers, managers, population, etc.);
- variety of forms and qualities of information scientific and educational resources;
- completeness, efficiency and accuracy of the received information;
- comfort and ease of receiving information;
- minimizing costs of design, development and maintenance of information systems.

Mathematical description of the methodology

i - code of integration level for primary information scientific and educational resources, $i \in I$;

l - code of storage form for primary information scientific and educational resources, $l \in L$;

n - code of representation type for primary information scientific and educational resources, $n \in N$;

j – code of particular criterion for effectiveness evaluation, $j \in J$;

m – number of higher educational institution, $m \in M$;

s – code of criteria indicator for website evaluation in terms of electronic trading platform, $s \in S$;

g - code of criteria indicator for website evaluation in terms of electronic labor exchange, $g \in G$;

h - code of indicator for secondary information scientific and educational resources, $h \in H$;

P_j^m – particular criterion for evaluating the effectiveness of using information scientific and educational resources of the m-th higher educational institution in terms of the j-th indicator;

P^m – integral criterion for evaluating the effectiveness of using information scientific and educational resources of the m-th higher educational institution;

α_i^1 – weight of indicator value for level of integrating primary information scientific and educational resources;

α_l^2 – weight of indicator value for form of storing primary information scientific and educational resources;

α_n^3 – weight of indicator value for the n–th type of representing primary information scientific and educational resources;

β_j – weight of criterion value for evaluating the effectiveness of using information scientific and educational resources in terms of the j-th indicator;

$v_{i\ln 0}^m$ – volume of information scientific and educational resources of the i-th integration level, l-th storage form, n–th representation type at the level of the m-th higher educational institution;

$vf_{i\ln f}^m$ – volume of information scientific and educational resources of the i-th integration level, l-th storage form, n–th representation type at the level of the f-th department of the m-th higher educational institution;

$vk_{i\ln k}^m$ – volume of information scientific and educational resources of the i-th integration level, l-th storage form, n–th representation type at the level of the k-th chair of the m-th higher educational institution;

$\lambda_{i\ln}^m$ – value of criteria for evaluating information scientific and educational resources of the i-th integration level, l-th storage form, n–th representation type of the m-th higher educational institution;

$$\lambda_{i\ln}^m = (v_{i\ln 0}^m + \sum_f vf_{i\ln f}^m + \sum_k vk_{i\ln k}^m) / \max_m (v_{i\ln 0}^m + \sum_f vf_{i\ln f}^m + \sum_k vk_{i\ln k}^m)$$

d_{rm}^2 - volume of r-th indicator for website evaluation by webometrics methods in the m-th higher educational institution, $r \in R$;

q_{rm}^2 - value of the r-th indicator of criteria for website evaluation by webometrics methods in the m-th higher educational institution;

$q_{rm}^2 = d_{rm}^2 / \max_m d_{rm}^2$;

ω_r^2 - weight of value for the r-the indicator of criteria for website evaluation by webometrics methods;

d_{sm}^3 - value for the s-th indicator of criteria for website evaluation in terms of the state of the electronic trading platform in the m-th higher educational institution;

ω_s^3 - weight of value for the s-th indicator of criteria for website evaluation in terms of the electronic trading platform;

d_{gm}^4 - value of the g-th indicator of criteria for website evaluation in terms of the state of the electronic labor exchange in the m-th higher educational institution;

ω_g^4 - weight of value of the g-th indicator of criteria for website evaluation in terms of the state of the electronic labor exchange;

d_{hm}^5 - volume of the h-th indicator for evaluating the efficiency of using secondary information scientific and educational resources in the m-th educational institution;

q_{hm}^5 - value of the h-th indicator for evaluating the efficiency of using secondary information scientific and educational resources in the m-th educational institution;

$q_{hm}^5 = d_{hm}^5 / \max_m d_{hm}^5$;

ω_{hm}^5 - weight of value for the h-th indicator of criteria for evaluating the efficiency of using secondary information scientific and educational resources in the m-th educational institution.

Then:

$P^m = \sum_j \beta_j P_j^m$, where $P_1^m = \sum_{i,l,n} \lambda_{i\ln}^m \alpha_i^1 \alpha_l^2 \alpha_n^3$, $P_2^m = \sum_k \omega_k^2 q_{km}^2$,

$P_3^m = \sum_s \omega_s^3 d_{gm}^3$, $P_4^m = \sum_g \omega_g^4 d_{gm}^4$, $P_5^m = \sum_h \omega_h^5 q_{hm}^5$.

The detailed description of all the parameters of the methodology can be found in [5].

The studies have shown that the completeness of websites is still very far from optimal in general. On average the websites contain just over half (55.4%) of all the necessary information. The completeness of indicators, reflecting research activities, is only 18.3%, which confirms the assumption that the requirements for the websites of higher educational institutions, imposed by the Ministry of Education and Science, the Federal Education and Science Supervision Agency, have some underestimation of the scientific activities of higher educational institutions.

The integral evaluations of the efficiency of using information scientific and educational resources (maximum possible evaluation according to the method is equal to one) do not exceed 40% even for the best higher educational institutions (Kuban State Agrarian University - 39.15%, Orel State Agrarian University - 38.23%, Russian State Agrarian University - Moscow Agricultural Academy - 32.58%, Krasnoyarsk State Agrarian University - 30.89).

4 Comparison of ratings of higher educational institutions based on evaluating the effectiveness of using information scientific and educational resources with regional ratings

The following regional ratings are used for comparison with ratings of higher educational institutions: rating of social and economic development, rating of regional subsidy assistance by the Ministry of Agriculture, rating of agricultural production efficiency in the regions, rating of regional governors efficiency, rating of regional science development. To establish the relationship between the ratings given above, we use two the most well-known methods.

The first relates to the calculation of pair-wise relationships between Spearman's correlation ranks or coefficients [11]. The calculations showed that only for the rating of subsidy assistance and rating of agricultural production efficiency in regions the values of the Spearman's correlation coefficient indicate the moderate relationship and they are statistically significant with probability of <1% with ratings of higher educational institutions.

The second method is the calculation of the Kendall's concordance coefficient. This coefficient characterizes the degree of rankings proximity (in this case, the regional

ratings). The value of Kendall's concordance coefficient $W = 0.32$ indicates low degree of consistency between the ratings presented in the list.

5 Conclusions

5.1 The transition to the platform of the unified Internet information space of scientific and educational resources of the country is one of high priority tasks in terms of the transition to the digital economics. It will significantly (by tens of times) reduce the costs of developing, implementing and maintaining information systems in science and education field.
5.2 The proposed platform will be the powerful tool for bringing the most effective innovative solutions to the economics. When recording publications, developments and other types of knowledge representation on the websites of research institutions and higher educational institutions, they will be automatically placed in other databases. This work would be greatly simplified at transition of all these organizations websites to the standard form.

References

1. Benjamin Peters: How Not to Network a Nation: The Uneasy History of the Soviet Internet. MIT Press, США (2016).
2. Glushkov V.M.: Macroeconomic models and principles of creation of OGAS. Statistics, Moscow (1975).
3. Medennikov V.I., Muratova L.G., Salnikov S. G.: Technique of assessment of efficiency of use of information scientific and educational resources. Analyst, Moscow (2017).
4. Medennikov V.I., Muratova L.G., Salnikov S. G.: Analysis of efficiency of the information and consulting service (ICS). VIAPI, Moscow (2002).
5. Medennikov V.I., Muratova L.G., Salnikov S. G.: Models and methods of formation of uniform information Internet space of agrarian knowledge. Publishing house State University of Land Use Planning, Moscow (2014).
6. Ereshko F.I., Medennikov V.I., Salnikov S. G.: Design of uniform information Internet space of the country. Business in the law. Economical and legal magazine, № 6, 184-187 (2016).
7. Zatsarinny A.A. The digital platform for scientific research. Materials of the International scientific conference "Mathematical Modelling and Information Technologies in Engineering and Business Applications", p. 104-113. Publishing house VSU, Voronezh (2018).
8. LNCS Homepage, http://repec.inecon.org/, last accessed 2018/10/10.
9. LNCS Homepagehttp://logec.RePec.org/scripts, last accessed 2017/10/09.
10. Sirotkin G. V.: The system analysis of factors of quality of education in higher education institution. Caspian magazine: management and high technologies. 2 (22), 109—118 (2013).
11. Ayvazyan S.A.: Probability theory and applied statistics. UNITY-DANA, Moscow (2001).