

REQUIREMENTS FOR DISTRIBUTED INFORMATION SYSTEMS TO SUPPORT SCIENTIFIC AND EDUCATIONAL ACTIVITIES

S.K. Serikbayeva¹, J.A. Tussupov¹, M.A. Sambetbayeva^{1, 2}, J.K. Serikbayeva³

¹ L.N. Gumilyov Eurasian National University, Nur-Sultan, Republic of Kazakhstan,

² Institute of Information and Computational Technologies CS MES RK, Almaty, Republic of Kazakhstan,

³ Educational and Methodological Center for the Development of Education of the Karaganda Region, Karaganda, Republic of Kazakhstan

To increase the effectiveness of research, it is necessary to have access to systematic information resources of scientific work. Therefore, in any field of science, it begins with research, the search for scientific information, but with the growing number of scientific articles, books, monographs, patents, the search for information becomes more and more difficult. Creating a unified information system that allows scientists to quickly get acquainted with the results of other scientific research and prevent their duplication. The main tasks of creating a model of a distributed information system that supports scientific and educational activities, the functional capabilities of the model, the concept of metadata, and the requirements for the metadata profile are described. The task, subject area, subjects, objects, the main functionality of the information system are defined, a list of the main types of information resources is provided. The paper analyzes the functional requirements for such systems. **Purpose.** In the process of creating distributed information systems that support scientific and educational activities, it is necessary to study the requirements imposed on the information system. **Methods.** The article discusses in detail technological methods for constructing models of information systems for supporting scientific and educational activities. **Results.** Using the requirements, models, metadata of a distributed information system supporting scientific and educational activities, the architecture of the information system was developed and a scheme for exchanging information in the information system through information protocols was presented. **Conclusion.** The article discusses technological methods of a distributed information system that supports scientific and educational activities. The main tasks for building a model of a distributed information system supporting scientific and educational activities, ensuring the functional functioning of the model, the concept of metadata for this system, and the requirements for the metadata profile are described. Based on the proposed requirements, the architecture of a distributed information system supporting scientific and educational activities has been developed and the structure has been clearly described.

Keywords: distributed information system, electronic library, scientific and educational activities, metadata, functional requirements, Z39.50, LDAP, PostgreSQL, Solr.

Introduction

In recent years, high technologies in the field of information transmission and processing have been rapidly developing, in particular, the creation of modern telecommunications systems has led to the emergence of new opportunities for organizing all levels of scientific and educational activities, which has led to a qualitative increase in the needs of information systems supporting scientific and educational activities [1].

Distributed information systems that support scientific and educational activities can work with various information systems. The main goal of creating a distributed information system supporting scientific and educational activities is to accelerate the pace and improve the quality of information exchange in the scientific environment. One of the most pressing issues is the division of the unified compatibility of the information system and the work on the systematization of information resources into professional areas. These can be scientific articles, scientific documents, electronic collections, ontological descriptions, data sets, logical descriptions, and so on. Semantic connections between information resources increase their value and provide additional opportunities for searching and identifying information [2].

The paper considers technological methods for constructing models of information systems designed to support scientific and educational activities. The model under consideration is that the developed model of an information system for working with scientific materials should solve the problems of long-term storage of information, organizing data search by attributes, accumulating and replacing metadata.

Model of a distributed information system that supports scientific and educational activities

The rapid development of global information and computing systems leads to a change in the fundamental paradigms of data processing, which is characterized as the transition to the use and development of distributed information resources [3]. Therefore, the most important goal associated with the technology of working with information is to study approaches to the compatibility of distributed data sources.

Compatibility of information resources refers to the presentation of them depending on the purpose of their use, the storage of various information in their composition, and the provision of user-friendly interfaces. At the same time, resource aggregation should not be carried out physically. The main thing is that it provides the user with the information available in a single way. In particular, regardless of the specifics of access to information systems in electronic libraries, it is necessary to ensure the operation of heterogeneous databases or data, ensuring the effectiveness of the user in searching for Information [4].

An urgent problem is the creation of a model of a distributed information system that supports scientific and educational activities:

- Unification of efforts to exchange the results of scientific work;
- Work with data and documents integrated into an open semantic space;
- Description, presentation, and conversion of resources in accordance with the user's needs.

The model should provide the following functions:

- It is necessary to describe the article resources, including registration procedures, annotation procedures and steps for determining the format;
- Analytical analysis of resources;
- Access to published resources;
- Monitoring the automated system of resources and updating their metasispatas;
- Notify the user of the appearance and update of new resources.

Select metadata profile

Due to the large amount of scientific data in the internet, the problem of document search is at the forefront, and metadata can be used to systematize it.

Metadata is an effective tool for describing information objects. Meteors have a special character in relation to the field of resource use. Metadata can describe objects in the information space and in relation to the real world. Metadata of information resources can be stored separately (Fig. 1) [5].

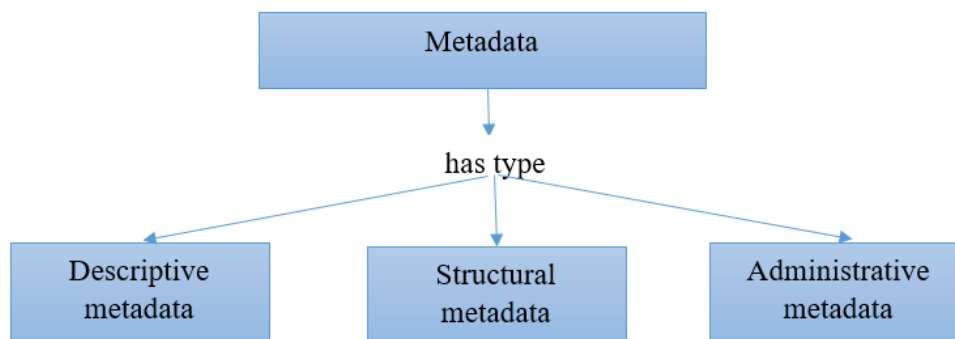


Fig. 1. Metadata structure

Metadata can provide a description of additional information about the document. For example, you can see the author, title, short abstract, and so on.

Metadata – data about data – data describing entities presented in information systems are the characteristics of the entities described for the purposes of their identification, search, evaluation and management.

Metadata is structured information that describes, explains, and indicates the location of an information resource.

Different approaches to metadata classification are possible. Metadata classifications have the right to exist according to their functions, according to the levels of semantic abstraction, according to their properties, according to the levels of information architecture to which the described resources belong, as well as according to a number of other criteria [6].

The main types of metadata:

Descriptive metadata-metadata describes the content of a resource (for example, it is a set of values of metadata elements of the Dublin Core), its bibliographic data (if it is a publication), annotation, resource identifiers (for example, URI or DOI, etc.);

Structural metadata – metadata characterizes the overall structure of the resource and its components, the volume and other similar properties of the described resource;

Administrative metadata – serves for the management and administration of electronic collections and other information resources.

A special type of data is an identifier, whose task is to unambiguously represent a digital object for the outside world and various applications.

The main catalog of information resources of the information system metadata server is built in accordance with the Dublin Core metadata scheme. The developed scheme takes into account the main requirements of this standard, and the metadata scheme is also expanded according to GOST 7.19 (MEKOF).

The term “metadata schema” is widely used in the literature and, in fact, is synonymous with the term “set of metadata elements”. A metadata schema is a set of metadata elements, each of which has a certain name and semantics, takes values with established semantics, sometimes values from a managed dictionary. According to the recommendations of Dublin Core, an information object should have a basic set of attributes. The set of attributes of an object is expanded depending on its type.

Metadata performs many functions in the systems that use it. Their specific functions and composition significantly depend on the information technologies on which the system is based, on its functionality, the properties of the information resources supported in it, the ways of their organization in the system, the specifics of their processing tasks, and on many other factors [7].

Metadata is necessary for solving the following tasks:

- providing information about an object to obtain its content, structure, methods of use, etc.;
- collecting and systematizing information about objects of description;
- selection from a set of objects of a certain subset by formal characteristics and comparison of objects by formal characteristics;
- internal technological tasks related to the preparation of objects, the placement of objects in the information fund, etc.;
- external technological tasks related primarily to the exchange of data with external information systems.

The implementation of subsystems of information systems should be based on open specifications related to international standards. Distributed information should be data synchronization in the system environment, for example, it should be distributed in the form of replication (Fig. 2). In addition, standard protocols such as OAI-RMN, OAI-ORE, SRW/SRU, Z39.50, and LDAP should provide cross-network interaction.

Metadata is needed to solve the following problems:

- Providing information about the content, structure, application methods, etc. of the document;
- Systematization and classification of needs;
- Stage of organization of subsystems;
- Support the exchange of external information systems.

Metadata is divided into several classes: descriptive, structural, and administrative.

Descriptive metadata-describes the content of information resources, for example, a set of values of metadata elements of the Dublin core Dublin Core, if it contains articles bibliographic Data, Annotations, resource IDs, etc.

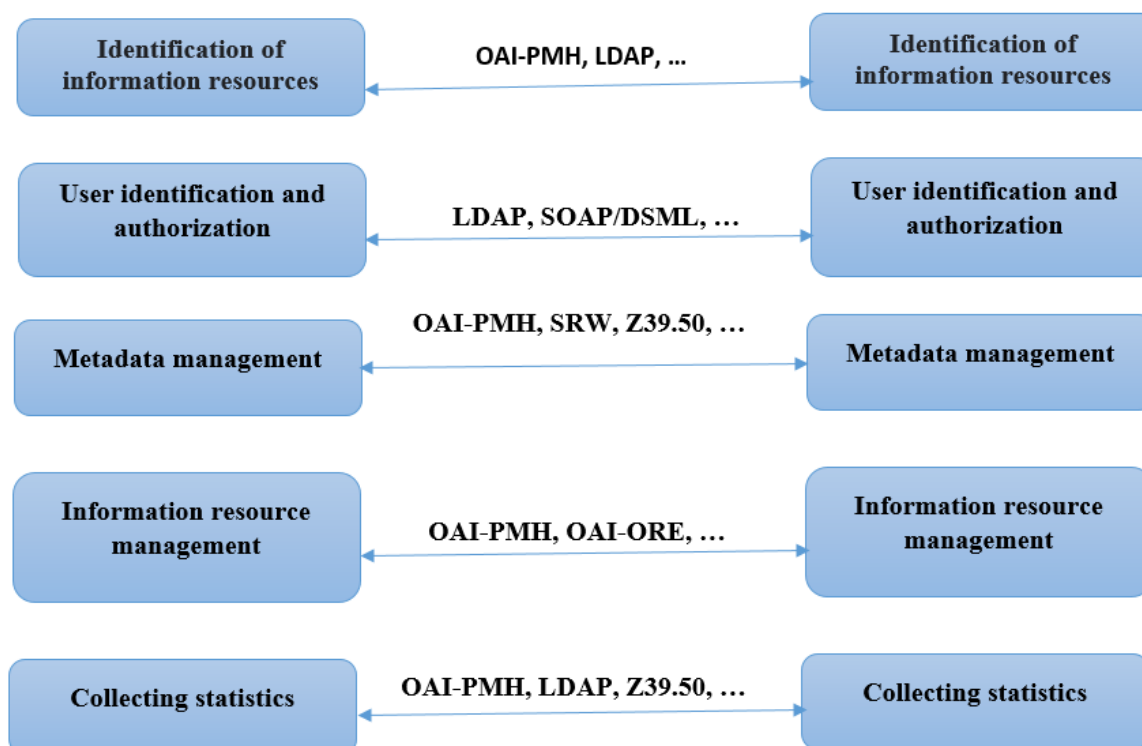


Fig. 2. Protocol of mutual network communication of subsystems of distributed information systems

Structured metadata-describes the overall structure of a resource and its structure, resource size, and other similar properties.

Administrative metadata-describes the date of creation of the resource, who created or changed it, who is the rightholder of resus, user access Authority, data on storage and copies of resources, and resource management data.

For complex search and classification functions for information by attributes, full-text search, it is necessary to provide the ability to view resources by category and dictionary classifiers. Due to insufficient interaction issues at the stage of developing their characteristics: agreements and recommendations for standardization of submitted documents, tools for harmonization of various information resources are poorly used. Interaction of Information Systems is understood as the degree of its ability to interact with other information systems, including people. But if the main burden of achieving mutual understanding in human communication (as in Information Systems) lies with the latter, which can handle poorly organized information, then it is necessary to ensure effective interaction between the correct, special technological methods of Information Systems, and general agreements are required. All this data creates the need to maintain compliance of schematic interfaces and protocols with international standards and recommendations.

The standard application profile is developed on the basis of specific groups or functional tasks of users. This means that it optimizes the creation of metadata processing systems. Metadata profile the label can define the selected standard classes, subsets, additional features, and parameters required to perform a single function [8].

In the field of metadata for articles, it is necessary to have a list of specialized elements of special resources, to set dictionaries to describe the values of elements that complement or expand the set of values allowed in accordance with the standard. In addition, additional characteristics of the elements can be presented.

The basis for the development of a scientific system are standards and international recommendations that form the profile of the scientific and educational system, which are aimed at solving certain requirements, the classes of standards necessary for performing specific functions are selected, represent a set of one or more basic regulatory and technical documents indicating their subsets and options. The metadata profile is the most important in the information system turnover [9].

The metadata profile must meet the following requirements:

- Providing a description of the main types of necessary information supporting scientific and educational activities;
- Ensure that access is open in accordance with the metadata description;
- Provide the ability to specify extended characteristics;
- Integration of information and ensuring information compatibility;
- Ensuring sorting, systematization and classification of information;
- Providing opportunities for placement and search of information in a distributed environment, interaction with other systems;
- Description of information and emphasis on the use of modern technologies [10].

Requirements for distributed information systems that support scientific and educational activities

High requirements are placed on many Information Systems: ease of Use and ease of learning for the user; on the technical side – compatibility with various information systems and ensuring interaction with standard protocols.

In the process of working with scientific documents, several requirements are formulated. It is possible to identify several sets of functional requirements that support scientific and educational activities [11].

1. Accumulation of Information Resources. Collecting information it is necessary to use different types of data input:

- enter user data;
- data collection through special internet programs;
- exchange of data from other information resources.

2. Relevance of documents. For a distributed information system that supports scientific and educational activities, automatic accumulation of information on the internet can be both important and irrelevant. Therefore, the problem can be solved in the following ways:

- creation of basic systematized formats for providing metadata about resources and structured definitions for classifying topics. It is necessary to describe the metadata of distributed information systems supporting scientific and educational activities and provide users with interactive web pages to create metadata in a certain format when placing resources;
- distribution of information resources by users and experts depending on the version of access to the Web page system, as well as increasing the degree of reliability of information;
- specify the search and classification space for information retrieval tools, as well as criteria for evaluating the quality of information entered;
- use the resource classification scheme according to the needs of users and classify information resources according to the scheme [12].

3. Relevance, completeness, and authenticity of documents. The issues of relevance and completeness of Information documents are similar to the approaches to solving resource issues. And ways to determine the correct origin of information:

- only verified (authenticated) users can enter interactive information for input;
- setting restrictions on the scope of application of an agent performing automated information collection systems;
- conducting sorting in the process of exchanging information with resources of other imported information systems;
- it is necessary to conduct methods of verification and classification of all entered information.

4. Use of intelligent services in processing user requests. By user query processing services-must provide attribute search, Full-Text Search, Resource search by category, and semantic search (optional).

5. Knowledge output. Partial automation of knowledge extraction can be used. The main approach to the proposed text is based on the type of semantic network, and the principle of its construction is based on the use of the frequency of conjugation of concepts in the text. Network user thematic terms are presented in the form of a tree, which means that it allows the user to navigate and greatly simplifies the processes of text search and research. In addition, this approach can be used to solve such problems

as classification and clustering of abstracts, text topics, semantic search, and so on. It is necessary to consider the following requirements for a distributed information system that supports scientific and educational activities that work with various types of information.

6. Support for the architecture of decentralized information systems. This requirement is a necessary condition for the completeness, authenticity and relevance of the information. The experience of using a distributed information system that supports scientific and educational activities has shown the complexity of creating centralized scientific systems that contain scientific information in one area of science or in any other.

7. Structure of the information space. In order to use complex functions of information search and classification, in addition to storing full-text information, it is necessary to search for information by attributes, search for full-text information, view resources by categories and classification dictionaries. This situation is characterized by the degree of specialization of the system in the selection of the classifier.

8. Adaptive presentation of information. According to the requests of users of the distributed information system supporting scientific and educational activities, it is necessary to increase the speed and accuracy of information search and selection without losing the quality of information search, take into account competence, time constraints when working in the system. It should allow users of a distributed information system that supports scientific and educational activities to obtain different levels of abstraction when presenting information from a brief description to a specific description of information objects for a quick search.

9. Historicity of information. A special feature of scientific information is its rapid obsolescence and loss of relevance. For many types of information resources, it is important to keep all changes in the information in the database and the ability to restore it is necessary. For example, when changing a person's last name, place of work, information about authors may change over time. Therefore, it is necessary to use up-to-date information for subjects related to time intervals, taking into account the time factor.

10. Archive. As noted above, the relevance of most scientific information becomes obsolete. But it is necessary to provide access to information resources for a long time. For example, they may require long-term legal documents, patents, or multimedia information at any time. At the same time, scientific reports of scientists in the field of science, speeches of scientists can acquire great historical value, gaining significance over time. Therefore, it is necessary to support the possibility of long-term restoration of information resources.

11. Application of Information Classification in the search for information. In order to use complex functions for searching and classifying information, in addition to preserving full-text characteristics, it is necessary to perform attribute search, full-text search, category and word-classifier search. The degree of selection of classifiers is determined by a specialized system. For the implementation of these functions, there must be dictionary-classifiers that ensure the identification and classification of resources that support scientific and educational activities.

12. Distribution support. With the rapid growth of the development of distributed information systems supporting scientific and educational activities in the world, the following requirements are imposed:

- support for metadata standards for exporting and importing accepted data;
- support for information exchange protocols between other information systems;
- support for both user interfaces for communicating with internal resources and features at the system level [13].

In accordance with the above requirements, we will consider the technology of creating a prototype of a distributed information system that supports scientific and educational activities. The main tasks of information systems are to collect, store, process information resources, register, update and ensure the processing of user requests.

1. Accumulation and registration of Information Resources. Services for collecting and registering information resources can be carried out simultaneously. When performing the functions of collecting and registering information resources, it is necessary to solve the following tasks:

- The task of sorting data. The information system receives information from any data source, so

many of them may be redundant. As a result of sorting information, the necessary data is selected from the information set included in the system;

– The task of verifying data. Provides verification of the reliability and logical integrity of data. It is carried out on the basis of expert analysis and verification of compliance with the information entered in the database. It is carried out at the stage of preliminary information processing with the help of tools for monitoring the logical integrity of data;

– Data compression task. It is carried out in order to minimize the storage of information resources in memory, as well as reduce the cost of transmitting data through communication channels;

– The task of converting data from one format to another. Transfer of data from one information system to another is carried out when it is necessary and when data is transferred between different types of information in the system.

The implementation of these functions is based on the use of the z39.50 and LDAP protocols in solving the problems of cleaning, verifying, compressing data, and converting data from one format to another. Common databases on different systems can be found using the same local client or interface. This does not solve the problem of what the interface should look like or how it should behave, it depends on the user to choose the interface. The connection of library systems with the internet and the development of the Z39.50 protocol open the way for access to an ever-growing array of bibliographic databases and full-text databases through a local automated system. The ability to directly connect users to resources offered by various computing platforms has increased the attractiveness of the Z39.50 protocol for libraries that connect institutional systems. As a result of using this protocol, it is possible to create distributed information systems that include databases of various organizations [14].

2. Storage of Information Resources. The system provides management and data storage of distributed information systems, the structure in which various types of data are stored, their integrity and access to them.

3. Processing of Information Resources. Processing information resources stored in the database allows you to present the processed information to the user. When performing information processing functions in the system, it performs data collection, search for dictionaries and indexes.

4. Relevance of Information Resources. Allows you to describe models and support the subject area of information resources in Information Systems. It is necessary to increase the relevance of Dynamic models of information systems by subject area. The relevance of Information Resources is achieved by including, deleting, and changing links to links between documents. When the structure of the subject area of information systems changes, the database scheme changes when updating information.

5. Providing information resources to users. The main purpose of Creating Information Systems is to provide the necessary information resources at the user's request and to meet their information needs. To provide resources, you can use pull and push technologies:

– Pull-technology is implemented on the basis of user initiative, providing mechanisms for searching and navigating information resources through user interfaces.

– Push technology is used to distribute various types of information between users in accordance with certain rules and based on a specific user framework. Users registered in the system are notified of the receipt of new documents in scientific and educational activities [15].

To meet these requirements, it is necessary to create an Information Service or environment for the provision and exchange of metadata – structured information about information resources and principles of access to them. Currently, many information centers engaged in the collection and distribution of metadata are actively interested in organizing interaction in order to exchange resources in them. As a rule, the basis for such a combination of funds is the development of a standard for the presentation of metadata, as well as the integration of arrays of normative reference information [16].

Within the framework of the tasks set, an Information System Architecture has been developed (Fig. 3), a multi-level DL architecture consisting of a data warehouse, a repository, a metadata server, an application server, a reference dictionary, as well as a software implementation of the developed architecture will be used to systematize the resources of the Digital Library.

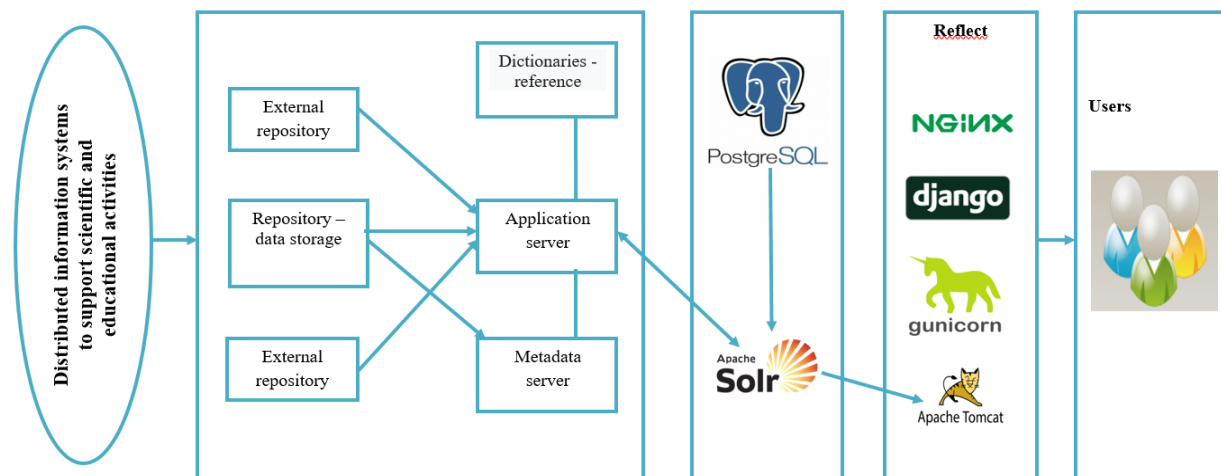


Fig. 3. Architecture of a distributed information system for supporting scientific and educational activities

It was developed taking into account the need for scalability and scalability of the architecture due to the need to store and process large amounts of data, as well as the use of resources of cost-effective algorithms for machine (including deep) learning. Let's look at the main components of this system, which are shown in Fig. 2.

The data warehouse is designed to store additional metadata about collections and their structure. They ensure the safety of electronic versions of articles, books, etc. and allow them to be accessed by external systems and users [16].

The use of electronic library materials depends on the availability of metadata to ensure effective and accurate viewing of content. Metadata must be generated when content is added to the digital library. Metadata and data should be logically linked to each other and, over time, there should be a reliable basic technology for managing logical communication between platforms and excessive geographical separation, everything will be implemented in a network distributed system.

For effective operation of the application server, it is necessary to use a set of classifier dictionaries containing both classification signs and a set of basic terms (with order relations), in which the systematization and classification of the material is carried out.

A dictionary reference is a set of terms that make up a dictionary for describing the content of a document. It is supported by standardization bodies to develop a standard method for categorizing materials included in the archive.

A clear categorization of material using a reference dictionary increases the likelihood that documents related to the search expression will be found when organizing a search in one or more electronic libraries as a result of the search.

1) PostgreSQL-serves as a permanent storage for structured data. The main types of data stored in this database are: a) news and metadata; b) processed data at the level of various basic units of analysis (lexema / word / phrase / sentence / text), including vectorization, lemmatization, cleaning results, etc.; c) thematic modeling results; results of classification of news by various criteria (tonality, political science, social significance, etc.)

2) ApacheSolr is a popular, rapidly developing open source search platform built on ApacheLucene. Solr is highly reliable, scalable, and fault tolerant, providing distributed indexing, replication, and load balancing, automatic disaster recovery, centralized configuration, and more. Solr supports the search and navigation functions of many of the largest Internet sites in the world. Since Solr has distributed search and replication capabilities, Solr is highly scalable. Here are some of the main features that solr provides:

- a) advanced full-text search capabilities;
- b) optimized for high volume traffic;
- c) open interfaces based on standards-XML, JSON and HTTP;
- d) comprehensive administration interfaces;
- e) easy monitoring;

- f) high scalability and fault tolerance;
- g) flexible and adaptable with simple configuration;
- h) next to the real-time index;
- i) extensible Plugin architecture;

Distributed Search in Solr has the following limitations. Each indexed document must have a unique key.

If Solr detects duplicate document IDs, Solr selects the first document and deletes subsequent documents.

Solr offers simple keyword search support for complex queries across multiple fields and multi-faceted search results. Search has more information about search and queries.

If Solr's capabilities aren't impressive enough, its ability to handle applications with very large volumes should help.

The figure will contain the following subsystems:

– Subsystem – a repository of digital objects that provides user and administrative WEB interfaces for accessing digital objects and collections, as well as compatibility interfaces with other subsystems based on open international standards.

– Subsystem for managing current research information, which includes articles by employees, information about their participation in conferences and the implementation of research projects.

– The subsystem will include user and administrative interfaces, as well as compatibility interfaces with other subsystems based on open international standards.

– Subsystem for integrating distributed information resources based on Apache Solr technology.

– Subsystem for accessing distributed information resources the basis of the technology is Nginx, Django, Apache Tomcat.

Conclusion

The main functional requirements for working with documents of a distributed information system supporting scientific and educational activities are formulated. To date, all the necessary components for creating a qualitatively new scientific information system have been developed and clearly systematized. A large part of science-centered distributed systems allow us to create a single environment for the exchange of scientific information.

This research has been funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. AP09057872).

References

1. Fedotov A.M. [Methodologies for building distributed systems]. *Selected reports of the X Russian Conference. Distributed information and computing resources. (DICR-2005)*, 2005, no. 11, pp. 3–16. (in Russ.)

2. Fedotov A.M., Zhizhimov O.L., Fedotova O.A., Baraxnin V.B. [A model of an information system to support scientific and pedagogical activities]. *Bulletin of the Novosibirsk State University Ser. Inform. Technologies*, 2014, vol. 12, no. 1, pp. 89–101. (in Russ.)

3. Zhizhimov O.L., Fedotov A.M., Fedotova O.A. [Building a standard model of an information system for working with documents on scientific heritage]. *Bulletin of the Novosibirsk State University Ser. Inform. Technologies*, 2012, vol. 10, no. 3, pp. 5–14. (in Russ.)

4. Fedotov A.M. Leonova Yu.V. [Requirements for the prototype of an information resource management system in distributed information systems for supporting scientific research]. *Computational Technologies*, 2018, no. 5, vol. 23, pp. 82–109. (in Russ.)

5. Kogalovsky M.R. [Metadata, its properties, functions, classification and presentation tools]. *Proceedings of the 14th All-Russian Scientific Conference "Electronic Libraries: promising methods and technologies, electronic collections" – RCDL 2012, Pereslavl-Zalessky, Russia, 2012*, pp. 4–14. (in Russ.)

6. Bezdushnyy A.N., Bezdushnyy A.A., Serebryakov V.A., Filippov V.I. [Integration of metadata of the Unified Scientific Information Space of the Russian Academy of Sciences]. Moscow, Computing Center of RAS named after A.A. Dorodnitsyn, 2006. 237 p. (in Russ.)

7. Shokin Y.I., Fedotov A.M., Zhizhimov O.L., Fedotova O.A. [The evolution of information systems: from Web sites to information resource management systems]. *Bulletin of the Novosibirsk State University. Series: Information Technologies*, 2015, vol. 13, no. 1, pp. 117–134. (in Russ.)
8. Shokin Yu.I., Fedotov A.M. [Support and development of distributed information and computing resources of SB RAS]. *Bulletin of the Al-Farabi Kazakh National University. Series: Mathematics, mechanics, computer Science*, 2004, vol. 4, no. 3, pp. 324–334. (in Russ.)
9. Zhizhimov O.L., Mazov N.M., Bolvanov A.Yu. [Experience in building a distributed information system based on the Z39.50 protocol]. *Libraries and associations in a changing world: new technologies and new forms of cooperation: 6th International Conference "Crimea 99" (June 5–13, 1999, Sudak): Materials of the conference*. Simferopol, Tavrida Publ., 1999, iss. 1, pp. 249–252. (in Russ.)
10. Serikbayeva S.K., Tussupov D.A., Sambetbayeva M.A., Yerimbetova A.S., Taszhurekova Zh.K., Borankulova G.S. [EduDIS construction technology based on Z39.50 protocol]. *Journal of Theoretical and Applied Information Technology*, 2021, vol. 99, no. 10, pp. 2244–2255.
11. Kutsenogy K.P., Kutsenogy P.K., Molorodov Yu.I., Fedotov A.M. [Development of the metadata structure for atmospheric aerosols based on an information model]. *Special issue: Proceedings of the International conference "Computational and Information Technologies for Environmental Sciences" (CITES 2003)*. Tomsk, 2003, vol. 9, pp. 25–33. (in Russ.)
12. Kogalovsky M.R. [Metadata in computer systems]. *Programming*, 2013, no. 4, pp. 28–46. (in Russ.)
13. Soviets B.Ya., Tsekhanovsky V.V. *Informatsionnyye tekhnologii: uchebnik dlya vuzov* [Information technologies: a textbook for universities]. Moscow, Higher School Publ., 2005. 327 p.
14. Kogalovsky M.R. *Perspektivnyye tekhnologii informatsionnykh sistem* [Promising technologies of information systems]. Moscow, DMK Press; IT Company Publ., 2003. 288 p.
15. Kogalovsky M.R., Parinov S.I. [Information resources, scientometric indicators and indicators of the quality of metadata of the Socionet system]. *Proceedings of the Ninth All-Russian Conference "Electronic Libraries: promising methods and technologies, electronic collections" – RCDDL'2007. Pereslavl-Zalessky, Russia*, 2007, pp. 45–54. (in Russ.)
16. Kogalovsky M.R., Novikov B.A. [Electronic libraries – a new class of information systems]. *Programming*, 2000, no. 3, pp. 3–8. (in Russ.)

Received 12 June 2021

УДК 004.7:004.75

DOI: 10.14529/ctcr210315

ТРЕБОВАНИЯ К РАСПРЕДЕЛЕННЫМ ИНФОРМАЦИОННЫМ СИСТЕМАМ ДЛЯ ПОДДЕРЖКИ НАУЧНО-ОБРАЗОВАТЕЛЬНОЙ ДЕЯТЕЛЬНОСТИ

С.К. Серикбаева¹, Дж.А. Тусупов¹, М.А. Самбетбаева^{1, 2}, Ж.К. Серикбаева³

¹ Евразийский национальный университет им. Л.Н. Гумилева,
г. Нур-Султан, Республика Казахстан,

² Институт информационных и вычислительных технологий КН МОН РК,
г. Алматы, Республика Казахстан,

³ Учебно-методический центр развития образования Карагандинской области,
г. Караганда, Республика Казахстан

Для повышения эффективности научных исследований необходимо иметь доступ к систематизированным информационным ресурсам научной работы. Поэтому в любой области науки она начинается с исследований, поиска научной информации, но с ростом числа научных статей, книг, монографий, патентов поиск информации становится все более и более сложным. Создание единой информационной системы позволяет ученым быстро знакомиться

с результатами других научных исследований и предотвращать их дублирование. Описаны основные задачи создания модели распределенной информационной системы, поддерживающей научную и образовательную деятельность, функциональные возможности модели, концепция метаданных и требования к профилю метаданных. Определены задача, предметная область, субъекты, объекты, основные функциональные возможности информационной системы, приведен перечень основных видов информационных ресурсов. В статье анализируются функциональные требования к таким системам. **Цель.** В процессе создания распределенных информационных систем, поддерживающих научную и образовательную деятельность, необходимо изучить требования, предъявляемые к информационной системе. **Методы.** В статье подробно рассматриваются технологические методы построения моделей информационных систем поддержки научной и образовательной деятельности. **Результаты.** С использованием требований, моделей, метаданных распределенной информационной системы поддержки научной и образовательной деятельности разработана архитектура информационной системы и представлена схема обмена информацией в информационной системе посредством информационных протоколов. **Заключение.** В статье рассматриваются технологические методы создания распределенной информационной системы, поддерживающей научную и образовательную деятельность. Описаны основные задачи построения модели распределенной информационной системы, поддерживающей научную и образовательную деятельность, обеспечивающую работоспособность функциональной модели, концепции метаданных для этой системы и требования к профилю метаданных. На основе предложенных требований была разработана архитектура распределенной информационной системы поддержки научной и образовательной деятельности и четко описана структура.

Ключевые слова: распределенная информационная система, электронная библиотека, научная и образовательная деятельность, метаданные, функциональные требования, Z39.50, LDAP, PostgreSQL, Solr.

Данное исследование финансируется Комитетом науки Министерства образования и науки Республики Казахстан (Грант № AP09057872).

Литература

1. Федотов, А.М. *Методологии построения распределенных систем* / А.М. Федотов // *Избранные доклады X Российской конференции. Распределенные информационно-вычислительные ресурсы. (DICR-2005).* – 2005. – № 11. – С. 3–16.
2. *Модель информационной системы для поддержки научно-педагогической деятельности* / А.М. Федотов, О.Л. Жижимов, О.А. Федотова, В.Б. Баракшин // *Вестник Новосибирского государственного университета. Сер. Информ. технологии.* – 2014. – № 1 (12). – С. 89–101.
3. *Жижимов, О.Л. Построение типовой модели информационной системы для работы с документами по научному наследию* / О.Л. Жижимов, А.М. Федотов, О.А. Федотова // *Вестник Новосибирского государственного университета Сер. Информ. технологии.* – 2012. – № 3 (10). – С. 5–14.
4. *Федотов, А.М. Требования к прототипу системы управления информационными ресурсами в распределенных информационных системах поддержки научных исследований* / А.М. Федотов, Ю.В. Леонова // *Вычислительные технологии.* – 2018. – № 5 (23). – С. 82–109.
5. *Козаловский, М.Р. Метаданные, их свойства, функции, классификация и средства представления* / М.Р. Козаловский // *Труды 14-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL2012, Переславль-Залесский, Россия.* – 2012. – С. 4–14.
6. *Интеграция метаданных Единого научного информационного пространства РАН* / А.Н. Бездушный, А.А. Бездушный, В.А. Серебряков, В.И. Филиппов. – М.: *Вычисл. центр РАН им. А.А. Дородницына.* – 2006. – 237 с.
7. *Эволюция информационных систем: от Web-сайтов до систем управления информационными ресурсами* / Ю.И. Шокин, А.М. Федотов, О.Л. Жижимов, О.А. Федотова // *Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии.* – 2015. – № 1 (13). – С. 117–134.
8. *Шокин, Ю.И. Поддержка и развитие распределенных информационно-вычислительных ресурсов СО РАН* / Ю.И. Шокин, А.М. Федотов // *Вестник КазНУ им. Аль-Фараби. Серия: Математика, механика, информатика.* – 2004. – № 3 (4). – С. 324–334.

Краткие сообщения

9. Жижимов, О.Л. Опыт построения распределенной информационной системы на базе протокола Z39.50 / О.Л. Жижимов, Н.М. Мазов, А.Ю. Болванов // Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества: 6-я междунар. конф. «Крым 99» (5–13 июня 1999 г., г. Судак): материалы конф. – Симферополь: Таврида, 1999. – Вып. 1. – С. 249–252.

10. EduDIS construction technology based on Z39.50 protocol / S.K. Serikbayeva, D.A. Tussupov, M.A. Sambetbayeva и др. // Journal of Theoretical and Applied Information Technology. – 2021. – No. 10 (99). – P. 2244–2255.

11. Разработка структуры метаданных по атмосферным аэрозолям на основе информационной модели / К.П. Куценогий, П.К. Куценогий, Ю.И. Молородов, А.М. Федотов // Специальный выпуск: Труды международной конференции «Вычислительно-информационные технологии для наук об окружающей среде» (CITES 2003). – Томск, 2003. – Т. 9. – С. 25–33.

12. Когаловский, М.Р. Метаданные в компьютерных системах / М.Р. Когаловский // Программирование. – 2013. – № 4. – С. 28–46.

13. Советов, Б.Я. Информационные технологии: учеб. для вузов / Б.Я. Советов, В.В. Цехановский. – М.: Высшая школа, 2005. – 327 с.

14. Когаловский, М.Р. Перспективные технологии информационных систем / М.Р. Когаловский. – М.: ДМК Пресс: Компания АйТи, 2003. – 288 с.

15. Когаловский, М.Р. Информационные ресурсы, наукометрические показатели и показатели качества метаданных системы Соционет / М.Р. Когаловский, С.И. Паринов // Труды Девятой Всероссийской конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2007, г. Переславль-Залесский, Россия. – 2007. – С. 45–54.

16. Когаловский, М.Р. Электронные библиотеки – новый класс информационных систем / М.Р. Когаловский, Б.А. Новиков // Программирование. – 2000. – № 3. – С. 3–8.

Серикбаева Сандугаш Курманбековна, PhD докторант, Евразийский национальный университет им. Л.Н. Гумилева, г. Нур-Султан, Республика Казахстан; inf_8585@mail.ru.

Тусупов Джамалбек Алиаскарович, д-р физ.-мат. наук, профессор, Евразийский национальный университет им. Л.Н. Гумилева, г. Нур-Султан, Республика Казахстан; tussupov@mail.ru.

Самбетбаева Мадина Аралбаевна, д-р философии (PhD) по специальности 6D070300 – Информационные системы, Евразийский национальный университет им. Л.Н. Гумилева, г. Нур-Султан; Институт информационных и вычислительных технологий КН МОН РК, г. Алматы, Республика Казахстан; madina_jgtu@mail.ru.

Серикбаева Жаркынай Курманбековна, магистр информатики, методист, Учебно-методический центр развития образования Карагандинской области, г. Караганда, Республика Казахстан; jarkinai_ser@mail.ru.

Поступила в редакцию 12 июня 2021 г.

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Requirements for Distributed Information Systems to Support Scientific and Educational Activities / S.K. Serikbayeva, J.A. Tussupov, M.A. Sambetbayeva, J.K. Serikbayeva // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». – 2021. – Т. 21, № 3. – С. 149–160. DOI: 10.14529/ctcr210315

FOR CITATION

Serikbayeva S.K., Tussupov J.A., Sambetbayeva M.A., Serikbayeva J.K. Requirements for Distributed Information Systems to Support Scientific and Educational Activities. *Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics*, 2021, vol. 21, no. 3, pp. 149–160. DOI: 10.14529/ctcr210315