MODELS AND ALGORITHMS FOR OPTIMIZING LEGAL INFORMATION RETRIEVAL IN THE CORPORATE NETWORK OF ACADEMIC LIBRARIES

Odil Urinkulov
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Uzbekistan

Murod Abdullayev
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Uzbekistan

Abstract. *With the rapid growth of information in the global network, the challenges of finding information quickly and easily in a narrow range of fields of study and specialization are increasing. People are constantly looking for information in some form throughout their lives. This is the result of the constant striving of human beings for innovation, efforts to improve personal and professional competencies. One of the main objectives of libraries is to meet people's needs for information. In short, this process can be called the type of informational support. The main purpose of this research is to develop models and algorithms to optimize the effective search of information about health information in corporate networks. Electronic libraries in the field of jurisprudence serve not only to train personnel in the field of jurisprudence, but also to increase legal literacy in society, to make citizens aware of their rights and obligations, and to prevent them from becoming victims of various frauds. For organizations, it serves as the most important repository of knowledge for their employees to constantly update their legal knowledge, to draw up normative-legal documents, contracts and agreements within the framework of legal requirements. Despite the fact that the field of jurisprudence is one of the most important areas of activity, the provision of scientific information to this field is not sufficiently systematized. Different organizations and institutions store their existing legal literature in the way they choose, and there is no single mechanism for making it available to users, digitizing, classifying, and searching for it. Most library users rate the efficiency of the library by the availability of the necessary literature. A survey of law students and professors was conducted to examine the interest of library users in legal electronic literature and their use. More than 50% of respondents use the electronic library daily, 93% are looking for legal literature, and 50% of participants said it is difficult to find legal literature. Also, all respondents (100%) approved the need to create a single corporate network by pooling electronic resources of higher education institutions providing legal training.*
Keywords: *algorithmizing, functional modeling, fuzzy set theory, hierarchical analysis, mathematical modeling, mathematical statistics, semantic search.*

Introduction

In this article, we will consider the problems of increasing the efficiency of searching legal literature in the academic electronic libraries of the higher education institution and the methods of their elimination. The main goal of this research is to develop models and algorithms to optimize the effective search for information on corporate networks. In the article the theoretical research methods (method of comparison and critical thinking) and empirical research methods (data collection and document analysis) have been applied (Karimov & Rakhmatullaev, 2008). The provision of informational support is related to the satisfaction of information requests of various categories of consumers: organizations, institutions and individuals. This includes not only statistics, surveys, archival data, but also official institutions, as well as books and journals, scientific lectures, dissertations and other sources of information related to science and education. Libraries are the most common form of this type of information provision, and information analysis services and centres are becoming increasingly important in today's environment.

With the extensive development of information and communication technologies in recent years, modern libraries have been transformed from passive warehouses storing printed documents into active automated information and information resource centres (IRC), electronic libraries, Centres that transmit and receive documents by e-mail, global telecommunications, Internet. Electronic libraries are not created as an alternative to traditional libraries and automated libraries. The EL is more comprehensive than the traditional library concept, it is a collection of annotated and full-text information related to a common system of electronic library catalogues and classification and coding of information (Karimov & Rakhmatullaev, 2008).

Whether libraries are traditional or electronic, finding literature or information that meets user needs has always been a relevant issue. The large amount of data, their relevance to different fields of science, and search requests that are made by different type of users are causing difficulties in the searching process. It can be seen that one of the main problems of searching information in modern libraries is the lack of qualitative and correct classification of the literature stored in them, and it is known that they belong to different disciplines.

From this point of view, in recent years, higher education institutions with the greatest demand for electronic library services need to create electronic libraries in the fields of personnel specialties trained in this educational institution, and integrate them with electronic libraries of other similar institutions, by this way serious attention is being paid to the creation of a specialized corporate information-library network and its joint use.

One of such fields of science is the field of jurisprudence. The field of jurisprudence is particularly important due to its strict rules, management and regulatory nature, normative determination of human, society and state relations. Also, another important aspect of jurisprudential literature is that they are constantly changing and improving according to the demands of the times. This, in turn, creates the need to organize electronic libraries in the field of jurisprudence and to constantly improve them.

The findings of the above research suggest that legal science in corporate networks should pay special attention to searching information based on semantic and unclear rules, improving processing technologies, database and logical knowledge design.

Analysis of foreign experience in the creation of corporate networks and databases

In the world, many scientific research works and practical projects aimed at creating and improving the system of information search in corporate networks, designing of data and semantic knowledge base, intellectual analysis of data, search, storage algorithms and software modules have been implemented. Problems of intelligent data retrieval and processing in computers include mathematical statistics, fuzzy set theory, natural language processing (NLP), machine learning, big data, deep learning, knowledge repository, methods based on expert systems and commercial companies such as IBM, Meta, Yandex, Rambler, Mail.ru and a number of scientists, including keyword and keyphrase extraction techniques (Siddiqi & Sharan, 2015), Keyword Extraction from Documents Using a Neural Network Model (Jo, Lee, & Gatton, 2006), klink search (Soumya, 2021), legal terminology of the Uzbek language (Kuchimov, 2020), analyzing and tuning user queries to search engines (Bufnea, 2012) have conducted reseach work.

Among the scientists who have carried out Understanding Electronic Resources Collection Development Practices Through Selected Theories (Okogwu, 2021), machine-readable cataloging (Avram, 1975), work about creation of corporate information-library systems, as well as creation, the introduction and implementation of models and algorithms of centralized cataloguing and information search systems were studied (Karimov & Rakhmatullaev, 2008).

The Library of Congress of the United States of America (Cassell & Hiremath, 2014) has also begun to work on the creation of a bibliographic database and the creation of an electronic library facility from anywhere in the world via the Internet. The Jackson Digital Library (Cassell & Hiremath, 2014), a project of the Library of Congress (Cassell & Hiremath, 2014), provides its own digital resources in multiple languages via the Internet. The main purpose

of this library is to promote inter-ethnic and intercultural understanding, to enrich the Internet with various cultural resources, to provide resources for teachers, scientists and all interested persons, etc. The global electronic library has more than 11 million electronic resources. Another electronic library is the National Electronic Library of the Russian Federation (The National Electronic Library of the Russian Federation, 2023). This electronic library provides services via the Internet and more than 5 million electronic copies are stored there. In addition, there are large electronic libraries such as Universal Digital Library, Project Gutenberg, Bartleby, Google Books, Internet Archive, Open Library, which have a database of electronic literature around the world and provide online services through the Internet. These electronic libraries mainly provide a wide variety of resources aimed at stimulating various human interests, education, research activities, and are considered public libraries. At the same time, specialized libraries focused on a specific subject in a separate field of science also operate around the world. Specialized libraries can be corporate, legal, medical, military, federal, music, art. Specialized libraries began to appear in the United Kingdom and the United States in the 19th century. The development of various specialized library associations led to the formation of the American Library Association in 1876. In 1898, the Association of Medical Specialized Libraries, and in 1906, the American Law Library Association was founded (Cassell & Hiremath, 2014).

Legal libraries are one of the most common types of specialized libraries. Law libraries are established to provide legal literature to law students, lawyers, judges, other legal researchers and citizens. Today, the largest legal library in the world is the United States Congress Library. It was opened in 1832 and considered as a branch of the main Library of Congress. Initially, the library served only the US Congress and Supreme Court staff. Currently, the library has 2.65 million unit related to legal resources (Cassell & Hiremath, 2014).

Legal information search problems

Searching for information in libraries consisting of electronic copies of legal literature that requires sufficient knowledge in the field of jurisprudence. Depending on the types of law libraries, the requirements for the knowledge of law librarians are determined differently. For example, an academic librarian-lawer who provides legal advice at the Law Library of the US Congress must have at least a master's degree in library science and a doctorate in law.

The imposition of this requirement confirms that finding information related to the field of jurisprudence and evaluating its relevance to the existing problem is a difficult process. Jurisprudence literature differs from other literature in that it is full, does not store redundant information, often uses complex terms that are rarely used in common speech, and at the same time

complicates the task of searching for them. In the legal literature, the information mainly reflects the nature of ordering and regulation. They contain the expression of thoughts based on the rules of jurisprudence, with a certain sequence. In addition, common terms are rarely used in legal literature thus most users do not know whether they exist or enter the terms into the search engine with grammatical errors. As a result, the search engine will not find anything for this keyword. In some cases, users cannot formulate a sufficient query to find exactly the literature they need. This is one of the most common problems faced by search engine. In such situation, the search engine does not know in which section to find the required literature and has to search for the literature in the entire database. The time and resource consumption of the search engine will increase. Another problem of search engine is that presenting the results to the user by sorting them as needed. In this case, the problem of evaluating the degree of relevance of the found results to the given search query can be appeared. In general, it is necessary to process all the processes carried out in it and correctly form the queries in order to make the search engine work well.

Proposed model and algorithms

The goal of information search optimization is to satisfy the needs of users with the necessary information. To achieve this goal, it is necessary to carry out a number of preparatory works, such as analyzing the array of data and making it convenient for storage and processing. One of the main components of information retrieval is the information retrieval model. It describes the criteria and methods of comparison of requests and documents, as well as the form of presentation of comparison results.

There are two main approaches of information retrieval. Based on the first approach, a special thematic catalog is used and a classified information search system is organized. To search for a document, its name, author, subject, and other information are used. According to the second approach, a dictionary information search system is used. In these types of systems, a vocabulary of terms is used to describe the content of the indexed documents. Indexing of documents by terms is performed by a software tool called a search engine. If you use both approaches to optimize search processes, using them in a hybrid form will have a positive effect.

If the lists containing the information available to us are in order, the efficiency of applying search algorithms on them will be several times higher. That's why it's a necessary process to bring the data into a certain order with the help of sorting algorithms and then perform actions on them.

A new hybrid model and algorithm was developed by improving existing models and algorithms in order to improve legal information retrieval.
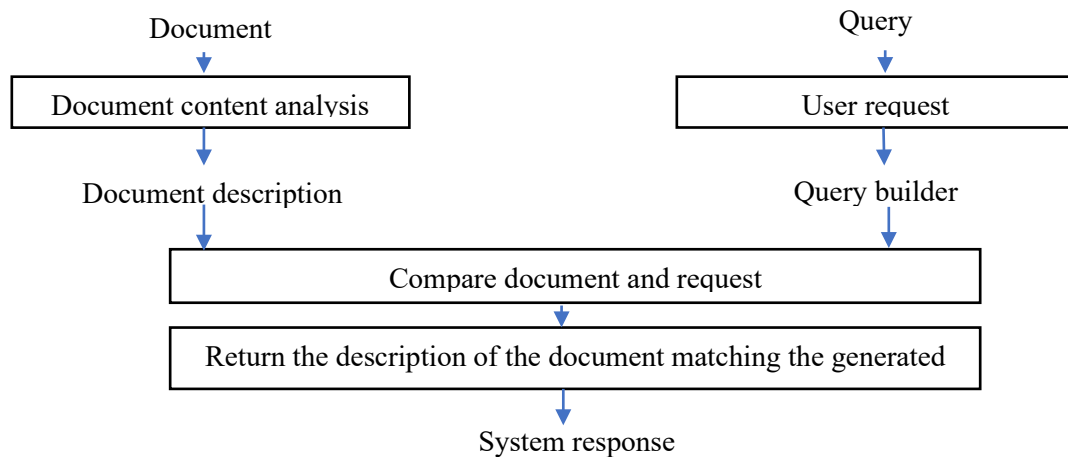
Document   Query

| Document content analysis | | User request |

Document description   Query builder

| Compare document and request |

| Return the description of the document matching the generated |

System response

*Figure 1 **Document search process** (Gunjal, 2016)*

The main idea of this new hybrid algorithm is to create a thesaurus consisting of semantic links of the main terms involved in the search and use this thesaurus as an auxiliary knowledge repository for the search engine. Initially, the terms frequently used in the practice of jurisprudence were distinguished in a hierarchical order depending on the branches of jurisprudence.

*Table 1 **List of speciality** (made by Authors)*

Name of speciality	Identification code
Theory and history of state and law. History of law doctrines	12.00.01
Constitutional law. Administrative law. Finance and customs law	12.00.02
Civil law. Business law. Family law. International private law	12.00.03
Civil procedures law. Economic procedures law. Arbitration process and mediation	12.00.04
…	…
Crime prevention. Ensuring public safety. Probation activities	12.00.14

1-the table lists the names of the fields of science at the first level of the hierarchy. All legal terms are distributed in the form of a tree within these disciplines. As an example, if we take the field of science "Environmental Law", it covers internal science fields that are divided into several levels.

As a result of the hierarchical arrangement of terms, relations between them such as interlevel, generation, ancestor and neighbor appear. (Figure 2)
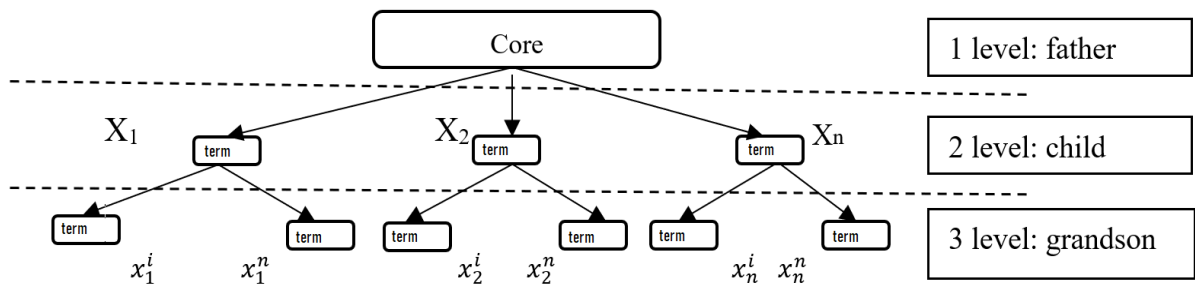
*Figure 2 **Hierarchical arrangement of terms** (made by Authors)*

An interconnected thesaurus of legal terms, a knowledge repository, was created in a very simple way, that is, by running an algorithm that works with the recursive function of the data management system. To create this small but very useful repository of knowledge, a table of the following order is created

For the purpose of the experiment, some legal terms and the identification numbers of their subordinate branches of science were included in the table.

A recursive function consists of three elements:

• Non-recursive term: as a result of the query, the base returns the result.

• Recursive term: combines the result of one or more queries using concatenation operators and refers to its name.

• Termination check: Recursion stops when the iteration does not return a string.

The recursive function that we are creating can work in the following way

1. A non-recursive term is executed to return the base result (x_0).

2. A recursive term is executed as the input value x_i to return the output result x_i+1.

3. The second step is repeated until an empty value is encountered. (Completion check)

4. The final form of the result set x_0, x_1, ... x_n is returned.

In the final result, a result is returned showing the given terms, the level at which the term is located, and the path from the topmost term to the final term.
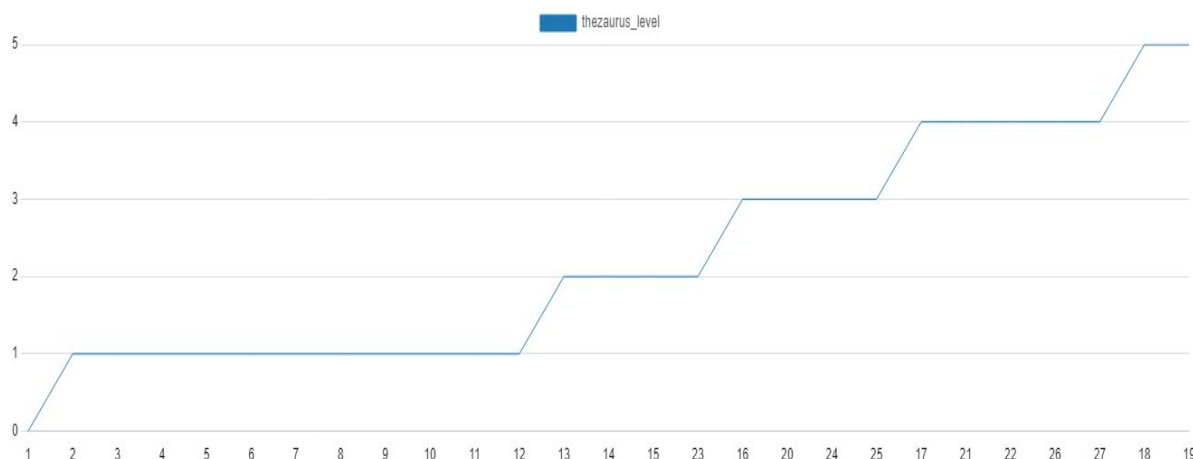
*Figure 3 **A graph of the placement of terms by levels** (made by Authors)*

The thesaurus of scientific terms created in the above processes helps the system by processing the queries entered by the users, to determine which level of the field of science the queries belong to.

Suppose that the set of educational options is expressed in the following form

$$x_{p1}, x_{p2}, \ldots, x_{pm_p} \in X_p, \ p = \overline{1,r}$$

Here

$$x_{pi} = \left(x_{pi}^1, x_{pi}^2, \ldots x_{pi}^N\right), \ i = \overline{1,m_p}$$

An N -dimensional ordered character space $\quad X_p, p = \overline{1,r}$ forms a complex of classes.

A proximity function is introduced to determine the degree of similarity between objects. Let two objects x_p1, x_p2 belonging to the class X_p be given in the ordered character space. Here the proximity function between objects $\rho\_j \,(x\_(p1,)\, x\_p2)$ is introduced as follows:

$$\rho_j\left(x_{p1}, x_{p2}\right) = \begin{cases} 1 & \text{if } \left(x_{p1}^j - x_{p2}^j\right) = 0, \ j = \overline{1,N}. \\[2mm] 0 & \text{else,} \end{cases}$$

In the ordered character space, the size estimation of k-object of arbitrary p-class through other t-object, indicating their degree of similarity, is calculated based on the following formula.

$$\Gamma_t\left(x_{pk}, x_{pt}\right) = \sum_{j=1}^{N} \rho_j\left(x_{pk}, x_{pt}\right), k = \overline{1, m_p}; \ t = \overline{1, m_p}; \ k \neq t$$

261

The proposed method is very compact compared to other methods, easy to use, thus requires less resources. It returns more accurate results when using a smaller knowledge base.

Conclusion

Since the current advanced search engines are mainly specialized for searching information from large datasets in an unordered global network, and use large amounts of BigData to return accurate results in search processes, the application of these models and algorithms in specialized electronic library corporate networks does not provide the desired results. In order to optimize the poultry information search system in corporate networks, it was found out that it is effective to first create a semantically linked thesaurus of terms related to this field of science and use it as a database for the search algorithm.

In order to optimize the bird information retrieval system in corporate networks, it was found that it was effective to pre-create semantically related thesaurus terms related to this field of science and use it as a database search algorithm.

References

Avram, H. D. (1975). MARC: its history and implications. *Library of Congress MARC Development Office*. Retrieved from https://files.eric.ed.gov/fulltext/ED127954.pdf

Bufnea, D. (2012). *Analyzing and Tuning User Queries to Search Engines.* Studia Universitatis Babeş-Bolyai, Seria Informatica. LVII. 41-48.

Cassell, K. A., Hiremath, U. (2014) *Information and information systems in the 21st century*. 2en edition. 131-146.

Davies, D. (2020). *How Search Engines Display Search Results*. Retrieved from https://www.searchenginejournal.com/search-engines/display-search-results/

Gunjal, A. (2016). The Process of Information Retrieval. Retrieved from https://amitgunjal.wordpress.com/2016/11/21/the-process-of-information-retrieval-from-scratch/

Jo, D. T., Lee, M., & Gatton, T. (2006). Keyword Extraction from Documents Using a Neural Network Model. *Hybrid Information Technology,* International Conference on. 2. 194-197. DOI:10.1109/ICHIT.2006.253612

Karimov, U., & Rakhmatullaev, M. (2008). Corporate information-library systems and networks. Monograph. Retrieved from https://scholar.google.com/citations?view_op=view_citation&hl=ru&user=swaB3GoAAAAJ&citation_for_view=swaB3GoAAAAJ:qxL8FJ1GzNcC

Kuchimov, Sh. (2020). Legal terminology of the Uzbek language: problems and solutions. Retrieved from https://cyberleninka.ru/article/n/yuridicheskaya-terminologiya-uzbekskogo-yazyka-problemy-i-resheniya

Okogwu, F. (2021). *Understanding Electronic Resources Collection Development Practices Through Selected Theories*. Retrieved from

https://www.researchgate.net/publication/352835142_Understanding_Electronic_Resou rces_Collection_Development_Practices_Through_Selected_Theories

Siddiqi, S., & Sharan, A.. (2015). Keyword and Keyphrase Extraction Techniques: A Literature Review. *International Journal of Computer Applications. 109.* 18-23. DOI: 10.5120/19161-0607

Soumya, Sh. (2021). *Klink search: enabling exploratory browsing activities in digital libraries.* Retrieved from https://ourspace.uregina.ca/bitstream/handle/10294/14366/Shukla_Soumya_MSC_CS_S pring2021.pdf

Taylor, M. (2020). *Search Results: Predicting Ranking Algorithms With User Ratings and User-Driven Data.* Retrieved from https://scholarworks.waldenu.edu/cgi/viewcontent.cgi?article=9609&context=dissertati ons

The National Electronic Library of the Russian Federation. (2023). *Information for libraries.* Retrieved from https://rusneb.ru/for-libraries/