Formation of the Knowledge Base of Digital Libraries Based on Semantic Models of Authority Files

Odiljon Naziraliyevich Urinkulov
Department of Training Scientific and Pedagogical Personnel
Tashkent University of Information Technologies Named After Muhammad Al-Khwarizmi
Tashkent, Uzbekistan
ourinkulov@gmail.com

Marat Rakhmatullaev
Department of Information Library Systems
Tashkent University of Information Technologies Named After Muhammad Al-Khwarizmi
Tashkent, Uzbekistan
marat56@mail.ru

Gulchiroy Erkinovna Ziyodullayeva
Department of Information Library Systems
Tashkent University of Information Technologies Named After Muhammad Al-Khwarizmi
Tashkent, Uzbekistan
gulchiroy.ziyadullayeva.81@bk.ru

Abstract. **This article is devoted to the analysis of methods for creating a knowledge base of a digital library based on the semantic model of authority files. The article also describes the practical aspects of creating an electronic library legal encyclopedia based on the semantic model of authority files. It is worth noting that the creation of a semantic network of authority files significantly optimizes the quality and speed of searching for resources in electronic libraries and creates the basis for meaningful literature search by readers. The article examines in detail the existing approaches to creating a knowledge base of authority files, including files based on semantic models, and forming a knowledge base structure. The main purpose of the study is to create an improved model and develop software for creating a knowledge base of an electronic library of semantic authority files for various subject areas. 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. Used object-oriented programming, intelligent search, reverse indexing capabilities. The field of jurisprudence was chosen for the application and approbation of the created model and software. A knowledge base has been created that studies frequently used specialized terms and semantic relationships between them.**

Keywords: authority files, knowledge base, reverse indexing, semantic network, thesaurus, term.

I. Introduction

In this article, we consider the methods of forming a knowledge base for information systems designed to quickly and qualitatively search for literature in specialized digital libraries. The article uses theoretical research methods (comparison and critical thinking method) and empirical research methods (data collection and document analysis).

In recent years, information and information services have become an extremely important and, at the same time, changing part of the mission objectives of the library institution. While ever-expanding electronic resources and ever-increasing access to digitized materials are changing the nature of information, the content of service remains central. Every day, librarians and their readers are confronted with a large number of information sources that must be constantly evaluated for reliability. Even when a person is in front of a computer and facing a world of documents, he is sure to feel like he is drowning in a sea of information. After all, new media and technologies, like tributaries, contribute to this new sea of knowledge and serve to increase its size and scale. And the information services are naturally both a lifeboat and a map and compass for those who are overboard and lost. By providing timely and special services to such readers, libraries are practically confirming that they are one of the central institutions of the 21st century. Broadly speaking, referral service is the

process of assisting library users in identifying information sources that may answer their specific question, interest, task, or problem. In the modern world, the introduction of literature search and delivery services that can fully satisfy the needs of readers has become a rather complicated process. Especially scientific and educational questions are somewhat more complex to answer and may require more time and information resources to provide a complete answer.

At the same time, it is important how complete and reliable the information or information source provided by the librarian is to the reader's request. Accuracy, reliability, coverage and relevance determine the value of the information provided.

Accuracy is the most important criterion for the information presented. Given that this information will later be used to make decisions for another situation, the accuracy of the information is of the utmost importance. Therefore, any information provided must be accurate.

Reliability. In most cases, the credibility of the publisher or author who published the literature containing the information determines the level of credibility of the information presented.

Coverage. Information on a narrow topic is important to be effective. Readers are interested in a wide coverage of a narrow topic of literature and getting additional information about it. Adequate coverage determines the completeness of the information being conveyed. If the librarian is satisfied with the answer, he will not spend too much time to send a new request for additional information.

The purpose of the article. The main goal of this study is to improve the semantic models of creating a knowledge repository that serves to improve the efficiency of semantic search of authority files.

II. MATERIALS AND METHODS

The methods and technologies used in creating an electronic library system are determined based on the number and scope of users of this system. Nowadays, electronic information-library networks of a smaller size, which provide specific information on a specific specialty, are mainly used in the systems of higher education institutions, scientific research institutes, and some state organizations. From this point of view, in order to increase the effectiveness of their activities, to form professional, spiritual qualities and a broad worldview in students and employees, they are creating information-library systems based on their needs and using them effectively. One of such directions is the direction of jurisprudence, which creates a wide range of opportunities for exchanging information with each other through the mutual integration of unique electronic resources accumulated over many years in higher educational institutions engaged in education in this specialty.

Many legal scholars and statesmen have worked in the Republic of Uzbekistan. During their career, they achieved many scientific achievements, published the results of their scientific activities in one form or another. One of the main problems in this direction is that a single electronic database of the developments of such mature scientists has not been created. In order to solve this problem, using the capabilities of modern information technologies, we have conducted research in the direction of creating a unified system of semantically linked authority files for the specialty of jurisprudence based on intellectual search by conducting new scientific experiments and semantically linking authority files with the help of improved methods.

Searching for the necessary information quickly and qualitatively in the information space, which is increasing day by day, cannot be done by one action. To implement this complex task, it is important to implement several information processes, and most importantly, to limit the human factor. The reason is that the volume of information is increasing so much that it has become impossible to process it with human participation. Therefore, solving these tasks requires the effective use of artificial intelligence technologies. Artificial intelligence technologies, on the other hand, originated from the field of science, and widely use large amounts of data to find effective solutions to given tasks. Several scholars have explored the issue of using big data as a knowledge repository, and classification issues in big data are discussed by Pijush Kanti Dutta Pramanik (https://orcid.org/0000-0001-9438-9309) Moutan Mukhopadhyay & Saurabh Pal, A Machine Learning Platform for NLP in Big Data issues Mauro Mazzei, Semantic Information Retrieval Systems Costing in Big Data Environment Khalid Mahmood, M. Rahmah, Md. Manjur Ahmed & Muhammad Ahsan Raza, Semantic Model Driven Engineering Challenge Milward, D. M. (2020).[1]-[2]

Evaluation of Distributional Semantic Models for the Extraction of Semantic Relations for Named Rivers from a Small Specialized Corpus issues Rojas-Garcia, J., & Faber, P. (2019). A Bi-model based RNN Semantic Frame Parsing Model for Intent Detection and Slot Filling was studied by Wang, Y., Shen, Y., & Jin, H. (2018) in Searching Data: A Review of Observational Data Retrieval Practices in Selected Disciplines Matter Gregory, K., Groth, P., Cousijn, H., Scharnhorst, A., & Wyatt, S. (2020). Machine learning based data retrieval for inverse scattering problems with incomplete data problems Gao, Y., & Zhang, K. (2019). The Search for Thinking Machines issues Ashri, R. (2020).[3]-[8]

The issues considered in their work are very deeply analyzed and rich in content, mainly dedicated to working with large volumes of giant data located on the open Internet. Such methods may not provide sufficient results when used with small and medium-sized data. For this reason, the development of accurate information return methods and the creation of a knowledge repository remain a very urgent issue when working with specialized data.

III. RESULTS AND DISCUSSION

Knowledge warehouse (KW) is a type of information warehouse that contains information on human experience and knowledge in a certain field of science and is created to manage, collect, store, search, and present this knowledge. BO means a set of facts and rules that allow meaningful processing of data and logical conclusions. BO is expressed in the form of rules and specific facts, providing knowledge and details about people, objects, events, facts, making logical conclusions on information processing processes and data storage.

Three main principles should be embodied on the basis of a successful, comprehensive knowledge base:

• Understandability

• Payment

• To be useful

To be understandable to everyone. Regardless of who uses it, information in the knowledge base should be understandable. Information should not be too long or too short. It is desirable that the text should consist of a sequence of words with the same meaning and should not contain complex sentences. In some cases, knowledge repository information may seem convenient to the "learner", but it may be too simple for the "professor". In short, finding the "norm" becomes a difficult task.

There are several ways to solve such problems:

1. Review by a senior expert in the field of science. This process is aimed at increasing the quality of the knowledge repository (KB) materials. The expert checks the correct expression of the material and its logical sequence.
2. By establishing a feedback relationship with the user who used the material, getting his opinions and comments about expanding, shortening, and creating the material. Through feedback, the user can leave any suggestions that are important for him. Also, the user can leave his feedback in the form of various signs and thank you notes.
3. Systematization. This tool can be conditionally divided into two effects:
• administrative effects - bringing information to a unified regulatory view, the procedure for identifying changes, etc.
• technical solution - use additional settings to hide details, display only in necessary situations. It solves the problem of providing information for users of different levels: it meets the requirements of maximum concise and professional language for experts, and the meanings of terms are explained in detail by providing details to novice users.

Always relevant. It is necessary to pay great attention to the fact that the information placed in the repository of knowledge is constantly available. It is necessary to maintain the confidence of all users in the repository of

knowledge and its materials. It is important that the knowledge warehouse manager has a monitoring system that provides a good service that responds to incidents. For example, for the knowledge base of the information system support service, there will be interpretations, changes in legislation and other tasks. Self-starting services determine the people responsible for responding to the incident and distribute the tasks. In order to successfully carry out updates in the system, it is necessary to have an understanding of how to improve the system at the stage of its creation and form it taking into account all situations.

We can use a classifier and a known transition algorithm model as a tool for automatic classification of materials included in the knowledge repository.

List of neighbors. An adjacency list represents a graph as a list of linked arrays. An array index represents a node, and other vertices in the list represent other vertices connected to it by an edge. The neighbor list only stores values that have connections through edges. On large graphs, this saves significant memory.
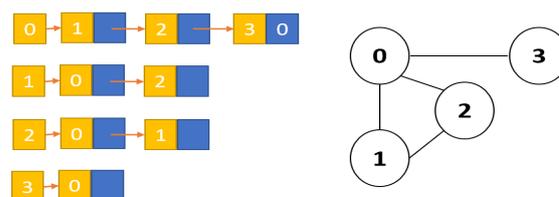

Fig.1. List of neighbors.

A spanning tree is an undirected graph partition that connects all the nodes of the graph with the minimum possible number of edges. The length (value) of the edges can be specified or unspecified.

If we have n=4 the chance of creating spanning trees is $4^{4-2} = 16$.

One of the most important steps in creating a knowledge base is to organize the existing information. In this process, we use the most popular Merge sort algorithm.

Marge sort algorithm. Marge sort is one of the most popular sorting algorithms and works on the principle of "divide and conquer". According to the principle of operation of this algorithm, a large task is divided into several small separate tasks and each small task is performed in a separate order. Finally, sub-tasks are again iteratively combined into a holistic form.

In our project, the researched task is directly related to the hierarchical data structure, and we used the recursive function because we aimed to achieve the most optimal results. The recursive function is mainly used in database design and creation. In this article, we will use the following program code implemented in the database of our research.

```
WITH RECURSIVE terms_tree as (
select
 e1.id,
 e1.parent_id,
 e1.term_username,
 0 as term_level,
 e1.id::VARCHAR as term_id_id
from terms e1
where e1.id = 2
union all
select
 e2.id,
 e2.parent_id,
 e2.term_username,
 term_level+1,
 term_id_id::VARCHAR || ',' || e2.id::VARCHAR
 from terms e2
join terms_tree et on et.id = e2.supervisor_id
) select * from terms_tree;
```

IV. CONCLUSIONS

One of the most important steps in creating a knowledge base is to organize the existing information. In this process, use the most popular Merge sort algorithm.

REFERENCES

[1] P.K.D. Pramanik, M. Mukhopadhyay, and S. Pal, "Big Data Classification: Applications and Challenges." In: Manoharan, K.G., Nehru, J.A., Balasubramanian, S. (eds) Artificial Intelligence and IoT. Studies in Big Data, vol 85. 2021. Springer, Singapore. https://doi.org/10.1007/978-981-33-6400-4_4

[2] K. Mahmood, M. Rahmah, M.M.Ahmed, and M.A.Raza, "Semantic Information Retrieval Systems Costing in Big" Data Environment. In: Ghazali, R., Nawi, N., Deris, M., Abawajy, J. (eds) Recent Advances on Soft Computing and Data Mining. SCDM 2020. Advances in Intelligent Systems and Computing, vol 978, 2020. Springer, Cham. https://doi.org/10.1007/978-3-030-36056-6_19

[3] M. Mazzei, "A Machine Learning Platform for NLP in Big" Data. In: Arai, K., Kapoor, S., Bhatia, R. (eds) Intelligent Systems and Applications. IntelliSys 2020. Advances in Intelligent Systems and Computing, vol 1251. 2021.Springer, Cham. https://doi.org/10.1007/978-3-030-55187-2_21

[4] D.M. Milward, "Semantic model driven engineering (Order No. 30370501)". Available from ProQuest Dissertations & Theses Global. (2773977250). 2020. Retrieved from https://www.proquest.com/dissertations-theses/semantic-model-driven-engineering/docview/2773977250/se-2

[5] J. Rojas-Garcia, and P. Faber." Evaluation of distributional semantic models for the extraction of semantic relations for named rivers from a small specialize corpus". Procesamiento Del Lenguaje Natural, 63, 51-58. 2019. Retrieved from https://www.proquest.com/scholarly-journals/evaluation-distributional-semantic-models/docview/2768727388/se-2

[6] K.Gregory, P. Groth, H. Cousijn, A. Scharnhorst, and S. Wyatt. Searching data: A review of observational data retrieval practices in selected disciplines. Ithaca: Cornell University Library, arXiv.org. 2020. doi:https://doi.org/10.1002/asi.24165

[7] Y. Gao, and K. Zhang. "Machine learning based data retrieval for inverse scattering problems with incomplete data". Ithaca: Cornell University Library, arXiv.org. 2019. Retrieved from https://www.proquest.com/working-papers/machine-learning-based-data-retrieval-inverse/docview/2310531881/se-2

[8] R.Ashri. "The Search for Thinking Machines". In: The AI-Powered Workplace. Apress, Berkeley, CA. 2020.https://doi.org/10.1007/978-1-4842-5476-9_1