

FVS-TECHNOLOGY: INTELLECTUAL SEARCH TOOLS

**BAHODIR MUMINOV BOLTAYEVICH* AND
ULUGBEK BEKMURODOV BAKHROM UGLI**

*Fundamentals of Informatics Department
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi,
Tashkent, Uzbekistan*

**Corresponding author: mbbahodir@gmail.com*

(Received: 10th March 2020; Accepted: 29th September 2020; Published on-line: 4th January 2021)

ABSTRACT: It is enough to have 3 basic stages of the modules in the SPD of a diversified corporate network: (F) - the method of submitting the request, i.e. the method of forming the expression of the information needs of the system user (S) - the function of the correspondence of the electronic resource to the request degree of compliance with the request and the found electronic resource; (V) - method of presenting electronic resources. Combining these three stages for models, methods, and software modules of the AML, is referred to as FSV technology (FSV platform, FSV Framework). FSV technology is an instrumental software platform based on a client-server architecture, integration and modification of models, and methods and algorithms of AML in the information environment of corporate networks. The following architecture has been developed for the FSV technology proposed for the search index in data retrieval systems.

ABSTRAK: Tiga peringkat asas modul adalah cukup dalam pelbagai rangkaian korporat SPD iaitu: (F) - kaedah penyerahan permintaan, kaedah membentuk ungkapan keperluan maklumat pengguna sistem (S) - fungsi surat-menyurat sumber elektronik bagi permintaan tahap pematuhan permintaan dan sumber elektronik yang dijumpai; (V) - kaedah penyampaian sumber elektronik. Gabungan tiga peringkat model, kaedah dan modul perisian AML, dipanggil teknologi FSV (platform FSV, rangka FSV). Teknologi FSV adalah platform perisian instrumen berdasarkan seni bina pelanggan-pelayan, integrasi dan pengubah suaian model, kaedah-kaedah dan algoritma AML dalam persekitaran maklumat dalam rangkaian korporat. Seni bina ini telah di bina bagi teknologi FSV yang dicadangkan bagi indeks carian dalam sistem dapatan data.

KEYWORDS: *intellectual search; FVS technology; information resource; search module; program module*

1. INTRODUCTION

In the world to meet the needs of the population in electronic resources, information is given great attention to the creation of intelligent data search systems in corporate networks of information and resource centers. Technologies of using information resources, the number of which is increasing every day, are one of the important issues of today. In this regard, including special attention is paid to improving the technology of searching and processing data based on semantic and fuzzy rules in corporate networks, database design and logical knowledge.

The world conducts research work aimed at creating data retrieval systems, database design and semantic knowledge, creating and improving software modules and algorithms

for mining, search, data storage in corporate networks. In this regard, the important tasks are the design, development of models, methods for searching and processing data, creating software for building data retrieval systems in fuzzy and stochastic information environments of corporate networks.

Modern technologies that do not replace many years of experience and highly qualified specialists, are designed to significantly improve the quality and efficiency of its work. This was discussed at the international forum held and dedicated to modern intellectual systems.

As it was noted, the modern IT industry all over the world is going through a new stage of its development. The catalyst for this process is the latest technology that can solve problems that have traditionally been considered creative. The forum discussed trends in the development of modern intellectual systems, their implementation in various sectors of the economy, the health sector.

With the independence of the republic, special attention is paid to the formation of a database of information resource centers (IRCs), on the creation of corporate information library systems, national content in raising the level of informatization of society based on intelligent management of hardware and software in information technologies. In this direction, the development of databases and software systems for logical search in data retrieval systems is included. The objectives for “... developing national content, improving the mechanisms for creating and promoting modern educational, scientific and educational information resources, multimedia products in the state language, meeting the needs of young people, ... introducing information and communication technologies ...” were defined. The implementation of this task, including the formation of electronic resources, the improvement of data mining systems, the development of a database of semantic data and knowledge in corporate information library systems, is one of the most important issues.

In the world of creating electronic resources, modeling processes and creating high-performance control systems to study the issues of intelligent search and data processing, research is carried out in a number of priority areas, including the creation of intelligent software modules based on the theory of fuzzy sets in solving problems in the fields of computer science and library, creation of methods for analyzing, searching and storing data based on models and methods of Data mining and Text mining, Big Data, development of methods and algorithms for integrating systems data retrieval, creation of algorithms, methods for processing and searching full texts, creating databases of aggregate data of corporate information library systems, creating methods and algorithms for processing requests for artificial and natural languages, creating a knowledge base, a semantic core based on the principles of SMART and SEO, developing software and data retrieval models based on local language.

Studies have shown that issues related to information retrieval, the structure and properties of data, the formation and processing of queries, the creation of a semantic core, the identification of electronic resources rating, the development of models and methods for intelligent search, the presentation and processing of data in fuzzy and stochastic information environments of corporate information biotech systems are not well understood.

The purpose of the research is the development of models, methods for searching and processing data in fuzzy and stochastic information environments, as well as the creation of software for corporate information library systems.

2. MAIN CHAPTER

Methods of forming queries, determining the rating and elements of data search in the corporate network are devoted to the questions of forming queries, determining the rating and elements of searching in fuzzy and stochastic information environments (NSNS).

A “search module” is a subsystem that, upon request, provides the most current ordered set of electronic resources, equipped with a user-friendly interface, including a database and statistical data.

The search module in the corporate network consists of two main elements that are independent of each other: an indexing element and a search element. The user sees only the capabilities of the search item and uses it. The indexing element is used to create the appropriate form for the effective presentation of information when searching for the necessary data in the data retrieval system (DPS).

Suppose that in the database are consistently given and are in the segment $V = [1, n]$, $N = \{0, 1, 2, 3, \dots\}$ with DosID identifiers with the set V , which has n electronic resources:

a) A quote (or reference) is a pair of ordinal electronic resources. $(i, j) \in V^2$. i - incoming link and j outbound link electronic resource.

In the set V between the electronic resources from all the links we form the E set, the graph directed to the links can be called reference graphs $G = (V, E)$.

b) For example, $G = (V, E)$ in this case the graph is the final set of vertices, $E = V \times V$ and $i \in V$, denote $I(i)$ the incoming set of links, and $O(i)$ the outgoing set of links denote as, then:

$$I(i) = \{e \in E \mid e = (i, j), j \in V\}, \quad O(i) = \{e \in E \mid e = (i, j), j \in V\} \quad (1)$$

c) If $O(i)$ and $I(i)$ do not exist and, then they are denoted as in this case $\{\emptyset\}$, the rating of the electronic resource $i \in V$ equals 0.

d) No electronic resource can provide its inbound links $I(i)$ and outbound links $O(i)$ to these links. The rating of incoming links I_r and outgoing links O_r is determined as follows:

$$I_r = \frac{\sum_{i=1}^{|I(i)|} Doc_r^i}{|I(i)| \sum_{count(I(i))} I(i)} \quad O_r = \frac{|O(i)|}{\sum_{count(O(i))} O(i)} \quad (2)$$

Here $count(I(i))$ - is the number of incoming, $count(O(i))$ - the number of all outgoing links.

Using the expression given above (2), we determine the method of preliminary rating of an electronic resource using (3) an expression.

$$D_r = (I_r + O_r) / 2 \quad (3)$$

Expression (2) allows the definition of electronic resources with a high rating and select a set that is recognized as relevant based on the identification of the rating of electronic resources in corporate information library systems (ICBS).

Suppose that $q = \{q_i\}$ a lot of requests and given to q_i the corresponding to a variety of electronic resources $R^i = \{r_j^i\}$, $j = 1 \dots M_d$ $j = 1 \dots N_q$. Here M_d – is the number of electronic resources viewed in the system, N_q – is the number of requests in the system. The method of logical-semantic search for a couple of queries, we will write as follows.

$$f(q_i, q_j) = \langle q_i | q_j \rangle = \frac{|R^j \cap R^i|}{|R^j|} \quad (4)$$

Here, $\langle a | b \rangle$ - denotes the relation and similarity of a with respect to b. $|R^j \cap R^i| \leq |R^j|$ - the condition must be relevant.

Rule 1: If $\langle q_i | q_j \rangle = 0$, then in this case q_i - is not semantically related to q_j , and is also considered not similar to each other.

Rule 2: If $\langle q_i | q_j \rangle = 1$, then in this case it is considered that q_i , q_j is completely semantically related and the similarity 1 is considered (q_i exactly similar).

Rule 3: If the expression (4) is equal $i = j$, then in this case $\langle q_i | q_j \rangle = 1$ it is considered semantically related, the similarity is 1 (it is exactly the same).

Rule 4: If $0 \leq \langle q_i | q_j \rangle \leq 1$, then q_i in this case, it is considered that semantically equal q_j and the similarity is calculated by the expression (4).

The method of logical-semantic search (4) for all queries, we write as follows:

$$f(q_1, q_2, \dots, q_{N_q}) = \sum_{i=1}^{N_q} \sum_{j=1}^{N_q} \langle q_i | q_j \rangle = \sum_{i=1}^{N_q} \sum_{j=1}^{N_q} \frac{|R^j \cap R^i|}{|R^j|} \quad (5)$$

here \sum is the union of the elements of the set. (5) the expression allows you to create a base of semantic knowledge in the data retrieval system (DPS).

Creating models, methods, and software for searching and processing data in fuzzy information environments, using the theory of fuzzy sets and using the linguistic variable model, defines the method for developing fuzzy queries in the data retrieval system (PDS) as follows:

- 1) (β) – defined linguistic variable;
- 2) $T(\beta)$ – we define a linguistically similar variable of a fuzzy set of terms;
- 3) X – for fuzzy sets of terms, we determine the set of values;
- 4) $G(\beta)$ – defined in the set of fuzzy terms, characterized by logical and modifier outputs;
- 5) $\mu_x - X$ defines the membership function of the set.

When designing a parametric membership function, the set $\mu_x - X$ is the linguistic variable of the corresponding object characteristic, the parametric design

β –method of the membership function is an important element in the intellectualization capabilities of the data retrieval system (PDS). μ_x – in the membership function, enter 4 parameters and give the following definition:

Proposition 5: If a, b, c, d parameters are taken from X – the set of values, then in this case the function is considered to be $\mu_x(a,b,c,d)$ the membership function of the set of values - X .

Proposition 6: In the membership function $\mu_x(a,b,c,d)$, the parameters a, b, c, d among themselves must be in ascending order. So one of the conditions $a < b$ or $b < c$ or $c < d$ must be constantly met.

Proposition 7: $T(\beta)$ – the parameters of the membership function for each term of the set of fuzzy terms corresponding to linguistic variables $T(\beta)$, at least one distinguishing parameter must be introduced.

It is recommended to use the choice of classes for the selection of membership functions corresponding to the linguistic variables of a fuzzy set in intelligent search and processing in a fuzzy information environment.

When intelligently searching for data in a fuzzy information environment of corporate information library systems, the main problem is the creation and management of the semantic core. We give the definition of a semantic core for intelligent search and data processing.

In a fuzzy information environment:

The semantic core is a set of relationships recognized as relevant, relevant in content and essence to normalized terms and electronic inquiry. Based on this definition, 3 methods have been developed for creating the semantic core of data retrieval systems in a fuzzy information environment. These include:

- Automatic construction of the semantic core as a result of search and processing.
 - Creating a semantic core by processing the citation of links, data in electronic resources.
 - Creating a semantic core with the help of expert groups. This method is carried out in version 2.
- a) Creation of a semantic core by an expert by creating queries corresponding to an electronic resource.
 - b) Expert processing of the existing semantic core, cancellation, modification, creation by it of the relationship of the request with the electronic resource, if necessary.

A model of the semantic core, based on the implementation of the control of the semantic core and 3 algorithms of methods for creating a semantic core, has been developed.

To determine the relevance of the semantic core model 4 properties have been added to the semantic core. If in property 1 there is a semantic link of the electronic resource and the query is 1, otherwise it takes the value 0, property 2 activates N - the total number of links, property 3 - m - the number of link-references.

In a fuzzy information environment for intelligent search and data processing in the development of models and rules for fuzzy search, the core of the knowledge base will form a fuzzy query q in a variety of fuzzy queries q^* (6) to form the rule condition as follows:

$$f : q \rightarrow q^* = \{q_1^*, q_2^*, \dots, q_n^*\} \quad (6)$$

in this case, $q_i^* - q_{i1}^*$ the term consists of a fuzzy pair of terms q_{i2}^* and can be defined as $\langle q_{i1}^* | q_{i2}^* \rangle$, and, $i = 1 \dots n$, $n \leq 2 \cdot |q_{i2}^*|$, $|q_{i2}^*| - q_{i2}^*$ is the number of fuzzy terms.

Conditions of the rule for the logical action "And" and q_1^* a pair of terms (6) we write as follows:

$$(q_1^* = q) \text{ and } (q_{11}^* = d) \text{ and } (q_{12}^* = t^*) \text{ and } (a_1 = \mu_{1,x_1}(q_{12}^*)) \quad (7)$$

where $(q_1^* = q) - q_1^* - q -$ is the pairing or non-pairing of the terms that make up the query, $(q_{11}^* = d) -$ the semantic relation of the electronic resource d and term, q_{11}^* , $(q_{12}^* = t^*) -$ the presence or absence in the set of fuzzy terms t^* of a database of fuzzy data q_{12}^* , $(a = \mu_x(q_{12}^*)) - (q_{11}^* = d)$, the weight a and the membership or non-membership of the values of membership functions of fuzzy terms q_{12}^* .

Using this condition (7) we write as follows.

Optimally, the fuzzy rules are based on the Mamdani method for intelligent data search in the data retrieval system.

$$\bigcup_{i=1}^n ((q_i^* = q) \text{ and } (q_{i1}^* = d) \text{ and } (q_{i2}^* = t^*) \text{ and } (a_i = \mu_{i,x_i}(q_{i2}^*))) \rightarrow R \quad (8)$$

If, on the basis of fuzzy rules, to write the set D of electronic resources in the knowledge base of rule (8), we obtain a system of fuzzy rules $D = \{d_1, d_2, \dots, d_m\}$, $m -$ the number of electronic resources in the database.

We introduce the concept of the knowledge base kernel for the definition method, a fuzzy term and modifier, parameters corresponding to the membership functions that have been developed and are constantly updated with the help of expert experts. The algorithm for the development of the knowledge base kernel, the IDEF1x model, and relationships in the model have been developed.

For the formation of a variety of electronic resources that meet the requirements in a fuzzy information environment, IDEF0 models have been developed that describe the functional structure and functionality of category 3 related tools to create a knowledge base based on the knowledge base, semantic core, and processing fuzzy requests.

Creation and implementation of software for searching and processing data in the corporate network, is devoted to the development of technology and architecture of FSV, composite modules and models, for the effective operation of search modules in corporate information-library environments it is necessary to have several basic steps.

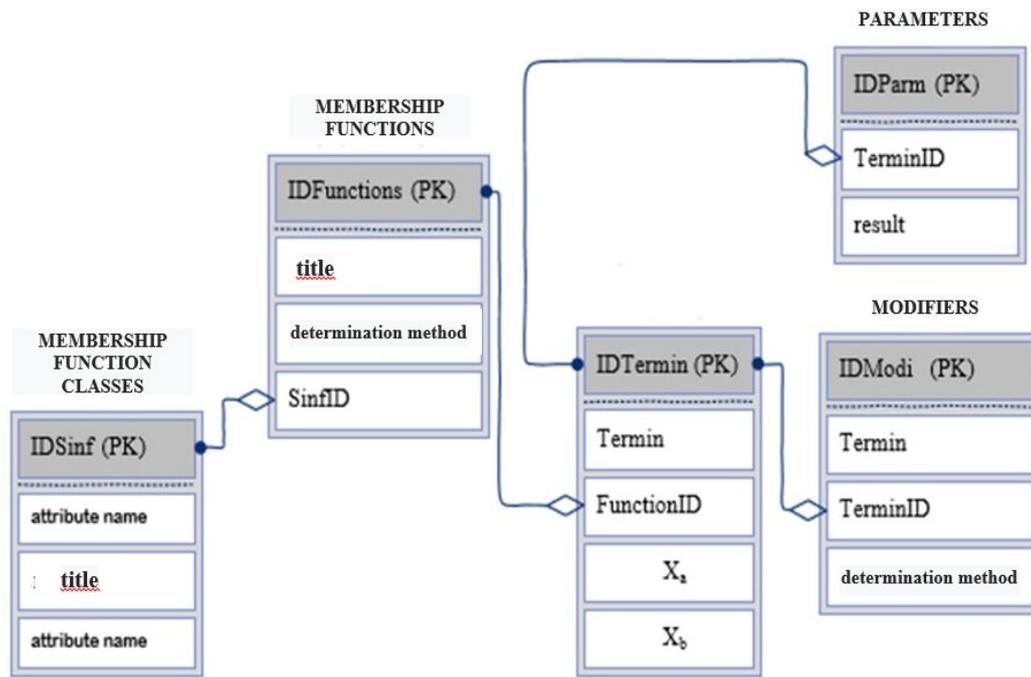


Fig. 1: IDEF1x core model knowledge base.

It is enough to have 3 basic stages of the modules of the search box in the data retrieval system of a multi-profile corporate network. These include:

1. (F) - method of presenting the request, i.e. method of forming the expression of the information needs of the user system.
2. (S) - the function of the correspondence of the electronic resource to the request, i.e. the degree of compliance with the query and the found electronic resource.
3. (V) - method of presenting electronic resources.

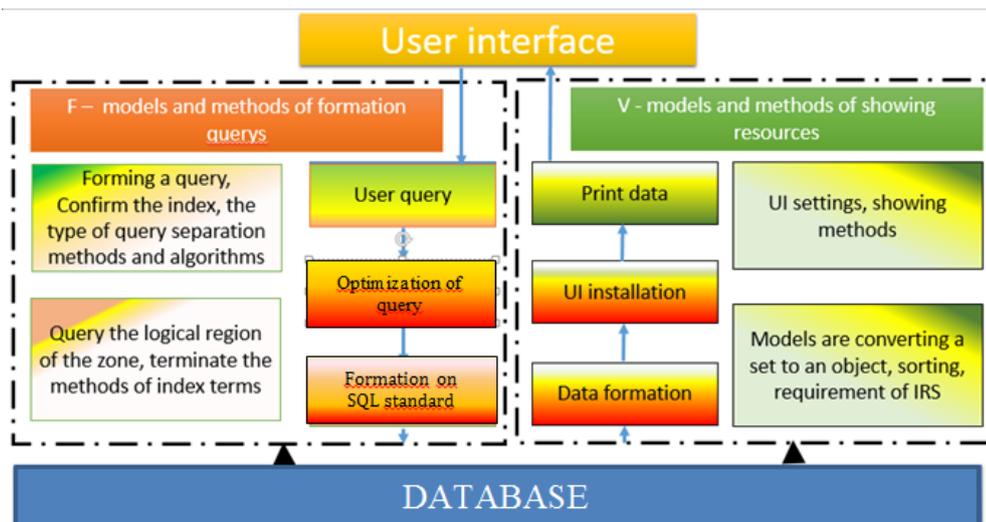


Fig. 2: Architecture of FSV technology for processing data.

Combining these three stages for models, methods and software modules of the AML, we call FSV technology (FSV platform, FSV Framework).

FSV technology is an instrumental software platform based on a client-server architecture, integration and modification of models, methods and algorithms for searching and processing data in the information environment of corporate networks.

The following architecture has been developed for the FSV technology proposed for the search index in data retrieval systems. In the development of the functional structure and IDEF models of the instrumental software module for query processing corporate networks request processing is divided into 2 stages.

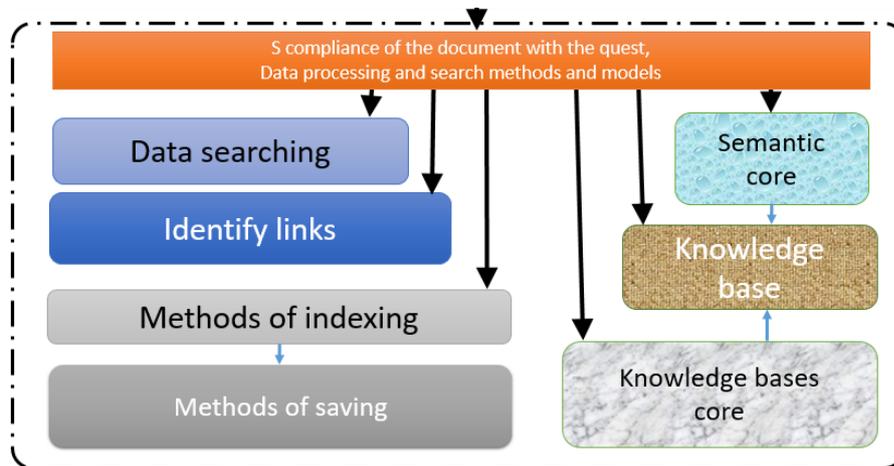


Fig. 3: Architecture of FSV technology for intellectual research.

At stage I, requests are created in natural language, based on the request processing methods, artificially created requests, based on the UI (User interface) methods. The second stage, which is a recurrent sequence, is carried out by clearing additional and auxiliary terms, determining the type of query, creating indices of logical and fuzzy terms, of the zonal field.

When creating IDEF models and functional structure of a software search and presentation software module. In the corporate network, an IDEF0 model was developed based on the methods, data retrieval system models in fuzzy and stochastic environments, traditional search, and S-ER matching process. Presentation V includes the processes of receiving a set of electronic resources, choosing a presentation method, determining the type of UI, providing constraints and filters and representing the set based on them, determining and controlling calls in the user's electronic resources.

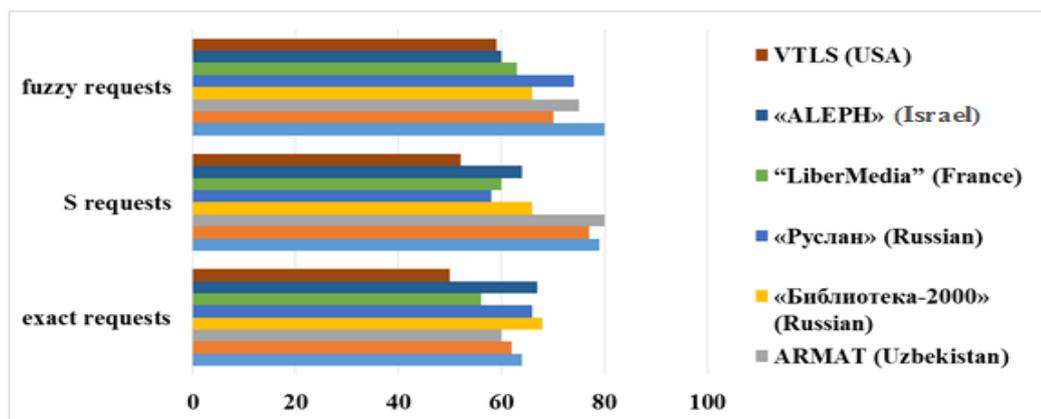


Fig. 4: Search results and data processing for the parameter "Speed".

FSV technology is integrated into such systems as: “ARMAT ++” - an automated system of information resource centers, “Mobile-Library” - mobile software on library resources, “vLibrary” - a virtual library of written cultural heritage of Turkic peoples of Central Asia of the XIX century (Kazakhstan).

For the experiment in MARC21 format, you can use the ISO2709 data exchange standard when exchanging data in corporate information library systems. Therefore, more than 50,000 titles of various electronic resources and full texts were chosen. For comparison and analysis results, such parameters as “Speed”, “Number of found data”, “Average relevance value” were selected and their results are given in the following figures (Fig. 3 and Fig. 4).

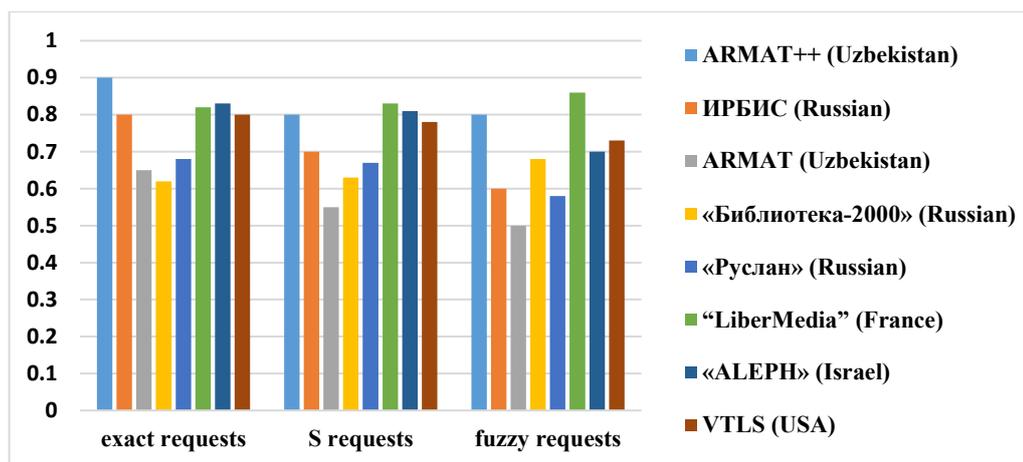


Fig. 5. Search results and data processing for the parameter "Average relevance".

The results of the experiment show that high speed provides the user with, firstly, superfluous information, secondly, it is very difficult to satisfy the information needs, i.e. need to analyze a lot of data. Although the search takes more time, but finding the necessary data indicates the effectiveness of search systems, since the analysis and limitation of redundant information saves 25%, sometimes up to 50% of the time of users.

The implementation of the results. Based on the software "ARMAT ++", created on the basis of FSV technology, models, methods of searching and processing data in fuzzy and stochastic information media:

- a method for determining a preliminary rating of an electronic resource based on references, a software module and an algorithm for designing parametric membership functions based on linguistic variables for fuzzy user requests, a software module, a method and algorithm for creating a knowledge base were introduced into the corporate networks of the main library of the Academy of Sciences (reference dated October 30 2017 No. 33-8 / 7293 of the Ministry for the Development of Information Technologies and Communications). As a result of scientific research, there were opportunities to improve the efficiency of determining electronic resources rating by 30%, using high-rating resources when presenting resources to users, improving the efficiency of processing requests in the corporate network by 15%, increasing the efficiency of collecting, searching, processing and providing data by 12%.
- a mathematical model of logical-semantic search, a software module and an algorithm for designing parametric membership functions based on a linguistic

- variable for fuzzy user requests, a model, methods, algorithms and a software module for creating a semantic core (reference from October 30, 2017 No. 33-8 / 7293 of the Ministry for the Development of Information Technologies and Communications). The results of the research allowed to increase the choice of resources that meet the requirements described in natural language by 30%, increase the efficiency of data search by 8%, reduce the number of corresponding electronic resources found by 2 times and increase accuracy by 30%.
- an information retention algorithm with a recurrent relationship, a model, an algorithm and a software module of the semantic core, a model, an algorithm and a software module for creating a knowledge base were introduced into the information resource center of the Intellectual Property Agency in accordance with Rule m Mamdani (reference dated October 30, 2017 No. 33-8 / 7293 of the Ministry of Information Technology and Communications Development). The results of the research allowed to increase the efficiency of the formation and processing of the original data by 15%, increase the efficiency of data retrieval by 12%, increase the amount of data matching the requests by processing fuzzy thermal queries by 30% and reduce the amount of unnecessary data by 50%.
 - E-Line Press Ltd. introduced a database project and IDEF-architecture architecture of FSV technology, developed to create a method and software module for determining a preliminary rating of an electronic resource based on links, a mathematical method of logical-semantic search, a data retrieval system (reference from October 30, 2017 No. 33-8 / 7293 of the Ministry for the Development of Information Technologies and Communications). The results of the study allowed to increase the efficiency of determining the accuracy of the preliminary rating of scientific resources and periodicals by 1.5 times, increase the efficiency of creating a database by 20%, increase the efficiency of data search by 5%, reduce the number of corresponding resources found by 50% and improve their accuracy in 1.3 times.

3. CONCLUSION REMARKS

As a result of the scientific and practical work “Models, methods of searching and processing data in fuzzy and stochastic information environments” the following results were obtained and the following main conclusions were made.

For the evaluation of data retrieval systems, compliance with the criteria of accuracy and integrity, reducing search time, a logical-semantic search model was developed, which increased the search efficiency by 5%. It can serve to study the issues of indexing sets of predefined ERs and queries, building their semantic links, optimizing search speed based on criteria and mathematical methods for searching and processing data in multi-disciplinary and multi-stage information resources.

The algorithm and software module for determining the relationship between the parameters of parametric membership functions, the choice of fuzzy terms corresponding to the classes of membership functions in the methods of intelligent search and data processing have been developed. Improved processing of fuzzy requests by 22%.

Developed algorithms, software module and 3 ways on creation of a semantic core for data retrieval systems in a fuzzy information environment of corporate information library systems, as well as increased the amount of data matching the query, on 10%.

IDEF0 and IDEF1x models for software modules of data retrieval systems in

corporate information library systems were developed. Serves for use in related research with the design of the functionality of the instrumental software module for analysis, text processing and the development of IDEF models, with questions of indexing, compression, information security, the choice of technical means.

The architecture, technologies FSV-software for creating a system for searching and processing data in corporate information library systems has been created. The results of introducing FSV technology into the ARMAT ++ corporate information library system make it possible to increase the effectiveness of the “Number of found data” by 7%, the Average relevance value by 10%, and the Search by 5%.

REFERENCES

- [1] Anagnostopoulos A, Andrei Z, Kunal P. (2006) Effective search engine model. In Proc. CIKM, pp 208-217.
- [2] Arlitsch K, Obrien P, Rossmann B. (2013) Managing Search Engine Optimization. Journal of Library Administration, 53(2-3): 177-188.
- [3] Azevedo C, Iacob M, Almeida J, Van S, Pires L, Guizzardi G. (2015) ArchiMate. Information Systems, 54: 235-262.
- [4] Bhat NA, Ganaie Sh. A. (2016) E-resources: Y.S. Parmar University of Horticulture and Forestry. Collection Building, 35(1): 16-21.
- [5] Brock EO. (2016) Declarative semantics of transactions in ORM. Information Systems, 60:85-94.
- [6] Cai F, Rijke M. (2016) Learning from homologous queries and semantically query auto completion. Information Processing & Management, 52(4): 628-643.
- [7] Muminov BB. (2014) Methods of information retrieval and cash algorithms. ITPA, 4-5 November, pp.194-196.
- [8] Muminov BB. (2016) The model of logical semantic searching. Proceedings of the Uzbek-Japan Symposium on Ecotechnologies, Tashkent, pp 72-77.
- [9] Newquist H. P. Data Mining: The AI Metamorphosis // Database Programming and Design. - 1996. - № 9. HP. (1996.)
- [10] Muminov BB. (2016.), Malumotlarni izlash usullari. Fan va texnologiya, Toshkent pg 276.
- [11] Muminov BB. (2016.), Data search methods. Science and Technology, Tashkent pg 276.