World Applied Sciences Journal 27 (12): 1536-1545, 2013
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2013.27.12.13629

Distributed Information Resources: Compatibility Analysis and Retrieval Means

N.V. Maksimov and A.E. Okropishin

Department of System Analysis, National Research Nuclear University "MEPhI", Moscow, Russia

Abstract: An approach to modeling and create integrated information retrieval environment is described. The ways to build up an object model of information resource and languages describing are considered. Solving the problem of compliance setting between data elements of different resources is proposed. Mathematical models giving an assessment of semantic losses when converting search queries, which is used in procedures of resources ranking, are constructed.

Key words: Document information retrieval systems . heterogeneous distributed information resources . interoperability . information interoperability models

INTRODUCTION

A continuous increase in the amount of information presented on electronic media, organized as a set of multiple distributed documentary resources has become a constitutive attribute of modern society in recent decades. The development of information retrieval means within separate resources cannot compensate their processing complexities that are increasing both quantitatively and qualitatively.

With the complex use of information resources (IR) one of the major problems remains unsolved: The inability to provide sufficiently high recall and precision measures, required for information support of scientific research, due to information resources amount and diversity (in structure, content and search tools) that are beyond the capability of human perception.

While solving this problem, the main trend of development is designing the unified access means to heterogeneous distributed information resources, based on the ompatibility models and providing controlled selective use of heterogeneous resources according to its architecture and implementation features.

EXPERIMENTAL BASE

In order to study the dependence of information retrieval efficiency on factors determined by the properties of heterogeneous resources, a series of information searches over one hundred, mainly scientific, topics were made, during which experts selected documents to meet their information needs by fixing the following parameters: document type (article, monograph etc.), document source (resource they were derived from) and search tools used.

The experimental base was as follows: (1) industrial retrospective bibliographic databases (DB) including VINITI RAS' subject-oriented scientific & technical information DB (DB CompSci (The resource symbols that are used in the pictures below are shown in brackets)), INION RAS' social sciences database (DB PublSci), CITIS' dissertation information cards database (IC Dissertations), research and development information cards DB (SRW IC), FIPS RF's patent information DB (Patent DB), IAEA INIS database, (2) electronic catalogs of Russian State Library (RSL), ScienceDirect and SpringerLink Publishers; (3) search engines Yandex, Google, Bing, AltaVista, Yahoo, Nigma, Rambler.

INFORMATION RETRIEVAL IN SCIENTIFIC RESEARCH

1.1 General: Modern conditions determine that a separate domain knowledge (DK) can be represented in different conceptual and terminological sets of different completeness and detail level and from different IR.

In general, this corresponds to the following provisions [1]:

- Typical search queries are not static, but rather evolutive;
- Searchers gather information in bits and pieces, iteratively, instead of collecting all information in response to a single query;
- Searchers use a wide variety of search techniques, including similarity search and relevance feedback.

Corresponding Author: Okropishin, Department of System Analysis, National Research Nuclear University "MEPhI", Kashirskoe Highway d.31, 115409, Moscow, Russia

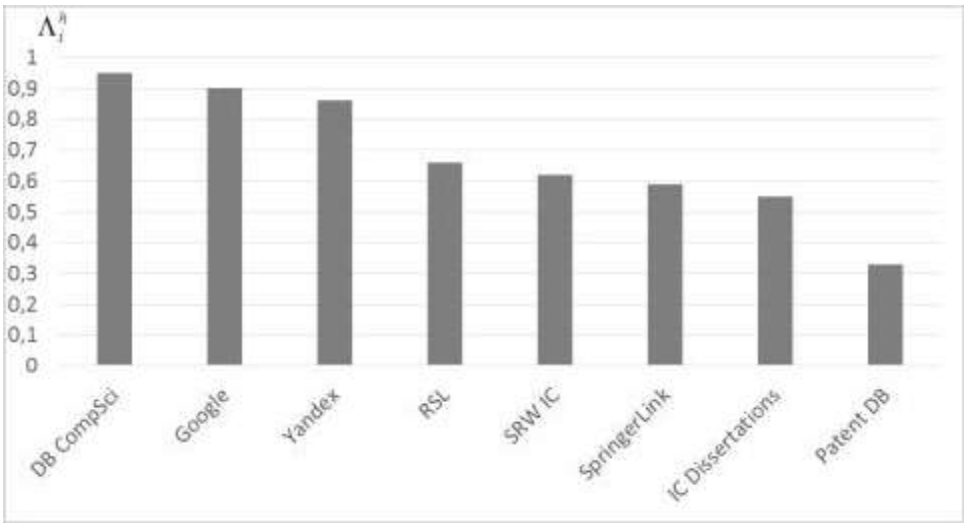


Fig. 1: Completeness and uniformity of presentation of subjects

This implies that the search navigation is a sequence of operations, which includes resource, documents and terminology retrieval, aimed at removing the following types of uncertainties caused by the information transformation at different stages of processing when creating or using a resource:

- Semantic and linguistic uncertainties arising from the analytical description dependence on the resource subject area, publication type and linguistic support used for content identification;
- Meta-information uncertainty due to metadata set of the specific IR.

Due to the information communication semiotic nature and relation between known and unknown, three types of search tasks in the subject search are defined [2].

The first type of search tasks-attributed search-relates to a search of the known properties of the object (e.g. search for a particular author's works).

The second type of search tasks-thematic search (e.g. review of scientific problems)-is a search of descriptions of real or hypothetical objects whose properties can be fully identified by the known set of attributes, but their values may be unknown.

The third type-problem search-relates to a search of descriptions of the objects or their components, potentially existing in DK and in the aggregate, perhaps forming something new, whose properties explicitly do not match their own attributes, but possibly could be defined as a combination of known attributes.

Thus, an iterative interactive search should facilitate the solving the following tasks:

- Resource search that could ensure the completeness of awareness in aggregation;
- Terminology search, that could adequately represent the essence of the information need;
- Document search in information resource;
- Search for retrieval's view point (with due account of the view point that is used in a particular resource).

1.2 "Heterogeneity" property as a factor of efficiency of search: The laws, reflecting properties of openness of comprehensive system of information resources, dictate that, in general, no resource is exhaustive, either in the subject matter or the document type.

To evaluate the fullness of resources on the themes an integrated indicator of the completeness and uniformity of the distribution of the subjects was proposed:

$$\Lambda_i^h = \frac{\alpha_i^h + \beta_i^h}{2}$$

where α_i^w -is the composition characteristic of the i-th resource documents subject spectrum, reflecting the presence of documents on the different subjects. β_i^w -characteristics of the uniformity of the spectrum of the i-th resource documents subjects.

$$\alpha_i^h = \frac{\sqrt{\sum_{j=1}^{t}\left(\lambda_{ij}^{\alpha}\right)^2} - 1}{\sqrt{t} - 1} \in [0,1], t > 1$$

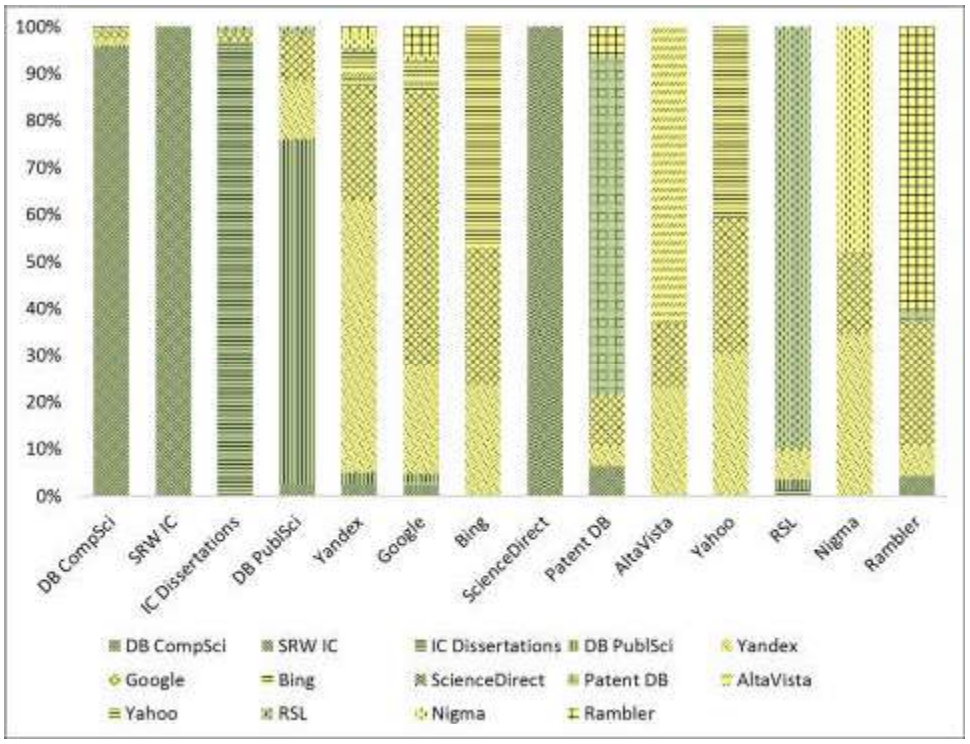


Fig. 2: The intersection of resources by documents

$$\beta_i^h = \frac{\left(1-\sqrt{\sum_{j=1}^{t}(\lambda_{ij}^{\beta})^2}\right)}{\sqrt{\sum_{j=1}^{t}\left(\lambda_{ij}^{\beta}\right)^2}\left(\sqrt{t}-1\right)} \in [0,1], t>1$$

where $\lambda_{ij}^{\alpha} \in \{0,1\}$ -discrete value reflecting the occurrence of documents of the j-th subject in the i-th resource; $\lambda_{ij}^{\beta} \in [0;1]$ -probability of finding the documents of the j-th subject in the i-th resource; t-cardinality of a subjects set.

Experimentally obtained values of the integral characteristic $\Lambda_i^h \in [0,33;0,95]$ shown in Fig. 1 confirm that none of the resource is exhaustive for the document subjects.

For the effective ranging of resources for the relevance of the search subject, the values of the integral characteristics Λ_i^h and Λ_i^w (a similar characteristic for the fullness on the document types) must be considered. In case of a highly specialized search the user will first be offered specialized resources (on type and/or topics). For polythematic search, the ranging will be made in the reverse order. This approach will reduce the number of resources.

1.2.1 Resource intersection on retrieved relevant documents: Figure 2 shows the distribution of the documents relevant to the topic, which are found in various information resources, with mutual intersection of the sets of documents found.

The experiment results show that the largest generality of the search results is observed for search engines (30-40% of identical documents). Documents found in certain specialized resources are almost never met in other specialized resources. There is some intersection between the documents in the specialized resources and in search engines, but it is relatively small. This is because the search engines, with rare exceptions, have no abilities to index resources containing professional information.

1.2.2 Publications types distribution: Figure 3 shows the experimentally obtained distribution of documents by types of publication.

The results show a significant scattering of documents of considered types in a set of resources: for the most localized type of documents-a monograph-about a quarter of the documents are located outside a single resource. For other types, this scattering is even more evident, which once again confirms the need to search in several resources to satisfy information needs.

1.2.3 Dynamics of the iterative search: Information resources differs not only in subject and type, but also in the sets of used metadata (some of them have dozens

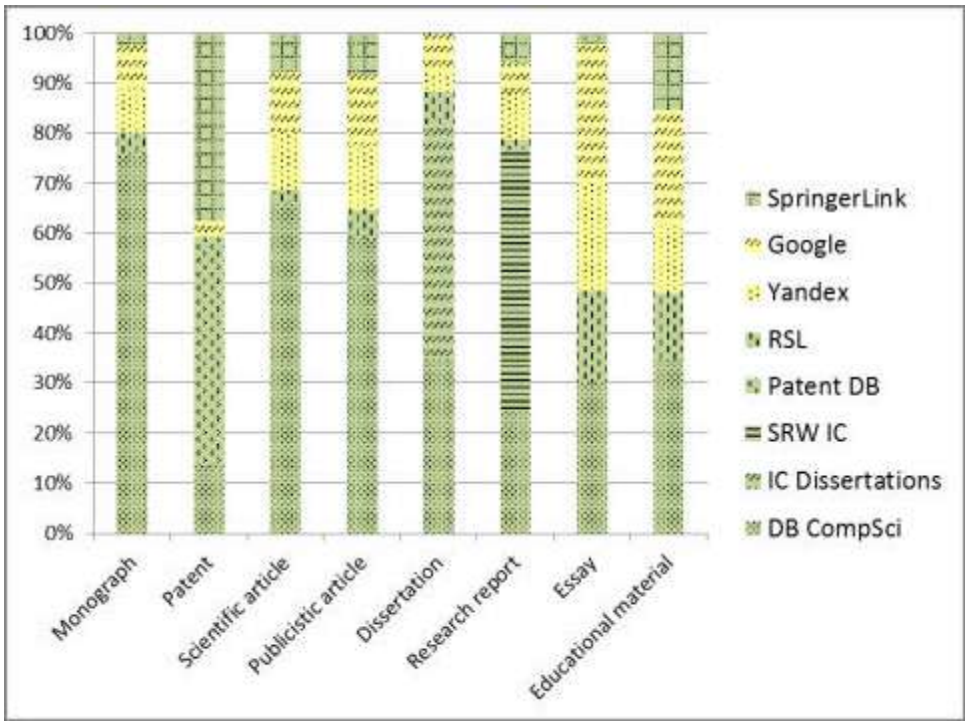


Fig. 3: Distribution of types of publications on resources

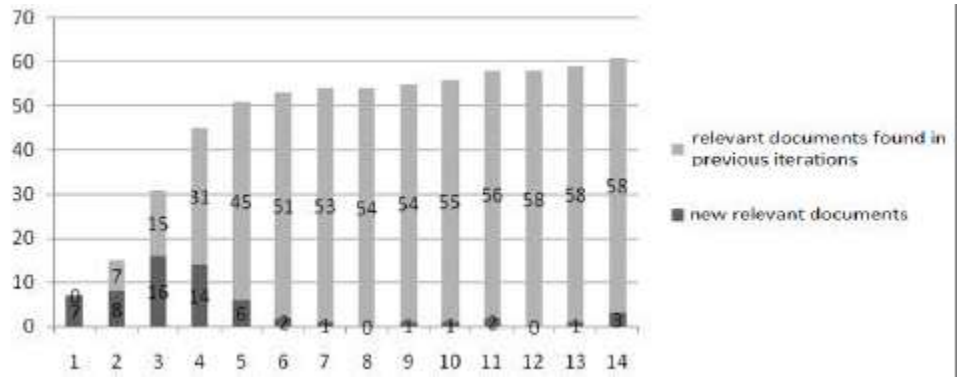


Fig. 4: Distribution of relevant documents by search iterations

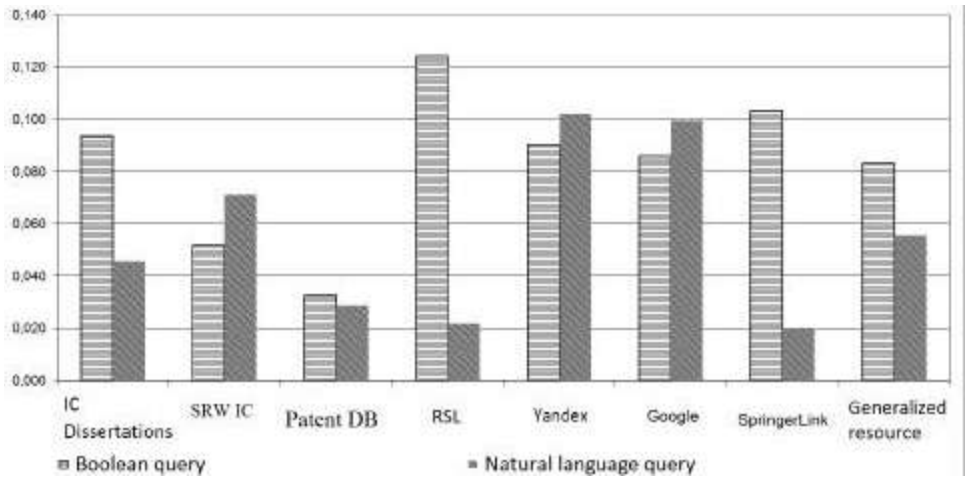
of data items, while others such as title, author and abstract, only a few) that almost always results in differences of using vocabulary and search engines.

Using different data elements can improve the completeness of the search. This is due to the fact that instruments used for indexing and search for different data elements are different: even the same vocabulary in queries addressed to different metadata allows obtaining various results. This means that owing to selective usage of vocabulary, the recall and precision of the search result can be increased. Similarly, nonverbal search tools (analogues search, search by reformulating query based feedback on relevance) allow the user to get outside his "personal" vocabulary and increase the recall of the total search result, especially in case of a problem search.

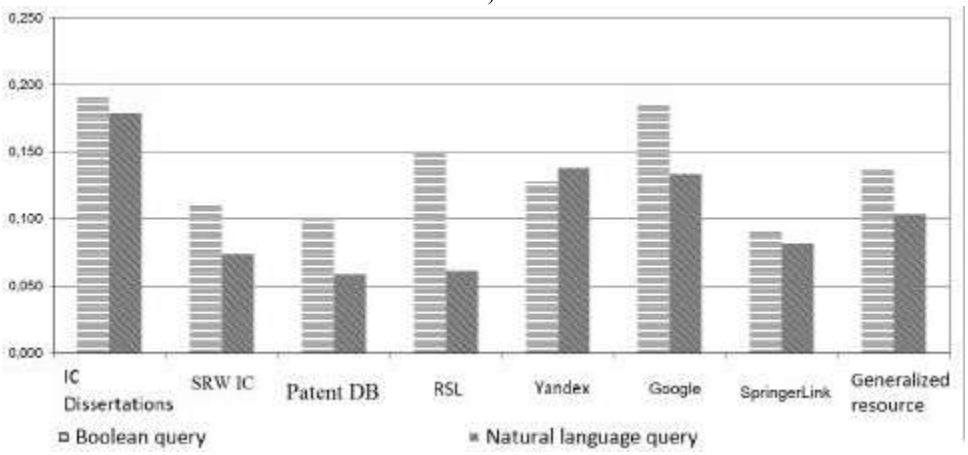
Dynamics of the iterative search effectiveness while using search tools systems is illustrated by cumulative histograms (Fig. 4).

As for queries on "general" subjects (not shown in the Fig.), in contrast to the special subject (e.g. "influence of the defects nature on the critical current value in type 2 superconductor"), an increase of a relevant documents practically terminates after 3-5 search iterations.

1.2.4 Query language type as a search effectiveness factor: To determine the effect of the query



a)



b)

Fig. 5: Retrieval quality indicators recall (a) and precision (b) for different resources

representation method, an experiment was held, in which a group of experts prepared pairs of queries: the natural language query and the query using well formalized boolean-type language.

In general, the search that used queries, prepared with due account for capabilities of the query language syntax (QL), gives, on an average, better document relevance indicators and level of redundancy of a search result. Averaged indicators of recall and precision on a series of searches for a number of resources are shown in Fig. 5.

According to the results obtained the well formalized query languages give higher information retrieval quality indicators. The exceptions were the search engines with well-developed linguistic processors and multilingual dictionaries (Google and Yandex); however, for highly specialized queries, the situation was reversed: the use of formal query language gives a gain of 10-20%.

In addition, the volumes of search results in the natural language by orders of magnitude are much higher than those obtained using the well-formalized (usually boolean-type) query language that significantly complicates the identification of relevant documents.

COMPATIBILITY OF INFORMATION RESOURCES

2.1 Review of approaches and solutions: Despite the effectiveness of approaches as concerns the quality of information search within individual resources, the situation with its aggregation and harmonization of representation forms in the space of distributed resources is not distinctly changed [3].

Existing integration platforms are classified [4] on the basis of relative position of indexed data, search tools and ways of interaction of the uniform resource with primary sources as follows: (1) resources aggregating data and having own search engine; (2) resources-catalogues which interact with sources and return links to search results; (3) unified search shells with possible integration and ranging of results.

For example, search environment introduced in [5] focusing on the interaction with Web services, provides

integration based on resources descriptions in a standardized form and presentation of data in RDF/XML-form.

Metasearch engine ProThes [6] supports a basic set of Boolean operators and allows building a search query in a graphic form using a thesaurus of the selected subject area.

During information monitoring in the Internet (search engines, wiki-like systems, Internet-catalogues), a metasearch system, which provides an automated search interaction with external resources over HTTP using for these purposes the formalized representation of a resource is created [7, 8].

To provide the federated search in the resources that support Z39.50, the Moscow State University specialists jointly with the Library Computer Network Company have developed the SIGLA system [9].

The System [4] provides a search in heterogeneous distributed resources belonging to search engines and local databases and uses Dublic Core set as a data schema.

Under the project MetaQuerier [10-13], a search engine providing user support when searching in distributed heterogeneous resources (related to the consumer market), including mechanisms of automatic matching of data elements and carrying out the translation of the search query and the union of search results is developed.

A search in databases, electronic libraries and other sources with the subsequent retrival of results in the unified format (using a pre-known protocol, for example, OpenURL) is provided by information and bibliography managers such as Bibliographix [14], Reference Manager [15], BiblioScape [16] and some others. These programs implement a request-response search and in this case the linguistic support is not provided.

The majority of the systems considered do not allow working with resources that use non-trivial query languages and advanced sets of metadata, generally used in scientific research.

2.2 Typology of resources compatibility: The effectiveness of search in distributed heterogeneous resources is determined by the following factors:

- Metadata sets used by resource;
- Resources subject and type spectra;
- Resource information retrieval language;
- Characteristics of the search interface and query language;
- Network protocols and clients-means used to interact with the resource.

Proceeding from the generalized level view of scheme of information interactions in a network of distributed resources, the following types of interoperability are singled out:

- Organizational defined by organizational and law aspects of the resource functioning;
- Technical defined by architecture and program-technical solutions relating to resource access;
- Informational compatibility divided into:
- Meta-informational caused by differences in access interfaces at the data element level (entry points);
- Linguistic caused by differences in information retrieval languages;
- Lexical caused by differences in vocabulary used in documents.

Only information interoperability, the basic and variable factor determining the effectiveness of information support of user tasks, will be considered below.

2.3 Models of information compatibility: For the analysis and provision of informational compatibility, we will consider the models of the three above-mentioned kinds of compatibility.

2.3.1 Model of metainformation compatibility: Provision of meta-information compatibility in accordance with [17, 18] is based on two strategies of the formation of the global data schema (schema mediator) which ensures matching of data elements. The strategy Global-As-View (GAV) envisages the formation of a global scheme based on the schemes of local resources and the strategy Local-as-View (LAV) which envisages the introduction of a global scheme independently from the local resources. The second approach allows operation in conditions, when a set of resources is not known in advance. In this paper, we propose a combined strategy, when to build a mediator the GAV strategy is used, followed by the creation of the taxonomy tree with uncertain definition and ambiguous division basis. The use and development of the taxonomy further takes place in accordance with the LAV strategy.

We assume that an individual item of data, as a component of metadata of different resources, can occur repeatedly and named differently within the whole tree. We introduce the notion of class of data elements θ_A, as an abstract element that is not related to any specific data schema. Then, in accordance with the obtained structure, abstract elements detailing some element, we will call the child classes and this class will be the parent class for them.

As an assessment of the meta-information equivalence (from the point of view of the possible replacement of some data elements with others), we introduce the concept of distance ρ-measure of difference between classes of data elements θ_A and θ_B which can be determined by their coordinates in the hierarchy.

Any class can be uniquely identified by its coordinates in the hierarchy:

$$\exists\Theta:\theta_A=\Theta(a_1,a_2,\ldots,a_k,0,\ldots,0)=\Theta(A),k=\overline{0,n}$$

We determine the class $\theta_B = \Theta(B)$ as a parent class for

$$\theta_A=\Theta(A)\Leftrightarrow A=(a_1,\ldots,a_{k-1},a_k,0,\ldots,0)$$

and

$$B=(a_1,\ldots,a_{k-1},0,0,\ldots,0)$$

We introduce the operation of obtaining the coordinates of the parent class (*): $A^* = B$
We denote:

$$A\equiv A^{*(0)},A^*\equiv A^{*(1)},(A^*)^*\equiv A^{*(2)},\ldots$$

The distance function between the classes with the coordinates

$$A=(a_1,\ldots,a_k,0,\ldots,0)\ \text{ and }\ B=(b_1,\ldots,b_l,0,\ldots,0)$$

we define as follows:

$$\rho(A,B)=\begin{cases}0 & ,if\ \ A=B^{*(|l-k|)}\\ \frac{1}{2^{k-1}}(1-\frac{1}{d_A}) & ,if\ A^*=B\\ \frac{1}{2^{k-1}} & ,if\ (A^*=B^{*(|l-k|+1)})\wedge(A\neq B^{*(|l-k|)})\\ \rho(A,A^*)+\rho(A^*,B) & ,\text{otherwise}\end{cases}$$

where d_A the number of child classes for A^*.

Practically, meta-information compatibility during the query translation is provided by using a global table of classes of data elements by establishing correspondences between global table elements and elements of a particular resource.

2.3.2 Model of linguistic compatibility: Based on the generalized model of machine search, we represent an elementary request in the following form: q = {q^F, q^C, q^T}, where q^F is a search area, q^C is a semantic matching criterion and q^T is a query term (simple or compound) and its qualifiers. Then the query language L^Q = {o,q,z} is described by the following attributes:

o-set of available operators-ligaments, q-set of available structures of elementary queries, z-rules of shared use of terms and operators in the query (syntax of search queries).

Determination of a distance between a pair of query languages L_1^Q and L_2^Q is reduced to determining the distances between their respective components.

The distance between two classes of operators-ligaments

$$\beta_1=\mathrm{B}(c_1^u,c_1^d,c_1^s)\ \text{ and }\ \beta_2=\mathrm{B}(c_2^u,c_2^d,c_2^s)$$

we define as follows:

$$\eta(\beta_1,\beta_2)=\frac{c^{u0}\eta^u(c_1^u,c_2^u)+c^{d0}\eta^d(c_1^d,c_2^d)+c^{s0}\eta^s(c_1^s,c_2^s)}{c^{u0}+c^{d0}(1-p^d)+\frac{1}{2}c^{s0}}\in[0,1]$$

where c^{d0}, c^{u0} and c^{s0}-weight coefficients of distances between the individual attributes, p^d-probability of keeping the meaning of the search query during transition to the operator that does not take into account the distance between the query terms. $c^u \in$ {AND, OR, NOT}-variable, reflecting the attribute of boolean operator-ligament; c^d = {0,1}-variable, reflecting the attribute of operator that takes distance into account; c^s = {0,1}-variable, reflecting the attribute of operator that takes sequence into account.

$$\eta^u(c_1^u,c_2^u)=\begin{cases}0,if\ c_1^u=c_2^u\\ 1,\text{otherwise}\end{cases}$$

$$\eta^d(c_1^d,c_2^d)=\begin{cases}0,if\ c_1^d\le c_2^d\\ 1-p^d,\text{otherwise}\end{cases},0<p^d<1$$

$$\eta^s(c_1^s,c_2^s)=\begin{cases}0,if\ c_1^s\le c_2^s\\ 1/2,\text{otherwise}\end{cases}$$

A distance function for two sets of operators of semantic match $\varepsilon_C(\varsigma_1, \varsigma_2)$ and two sets of available qualifiers $\varepsilon_T(\sigma_1, \sigma_2)$:

$$\varepsilon_C(\varsigma_1,\varsigma_2)=\frac{|\varsigma_1\setminus\varsigma_2|}{|\varsigma_1\cup\varsigma_2|}\in[0,1]$$

$$\varepsilon_T(\sigma_1,\sigma_2)=\frac{|\sigma_1\setminus\sigma_2|}{|\sigma_1\cup\sigma_2|}\in[0,1]$$

where ς_i-the set of all available operators of semantic match, σ_i-the set of all available qualifiers of query term.

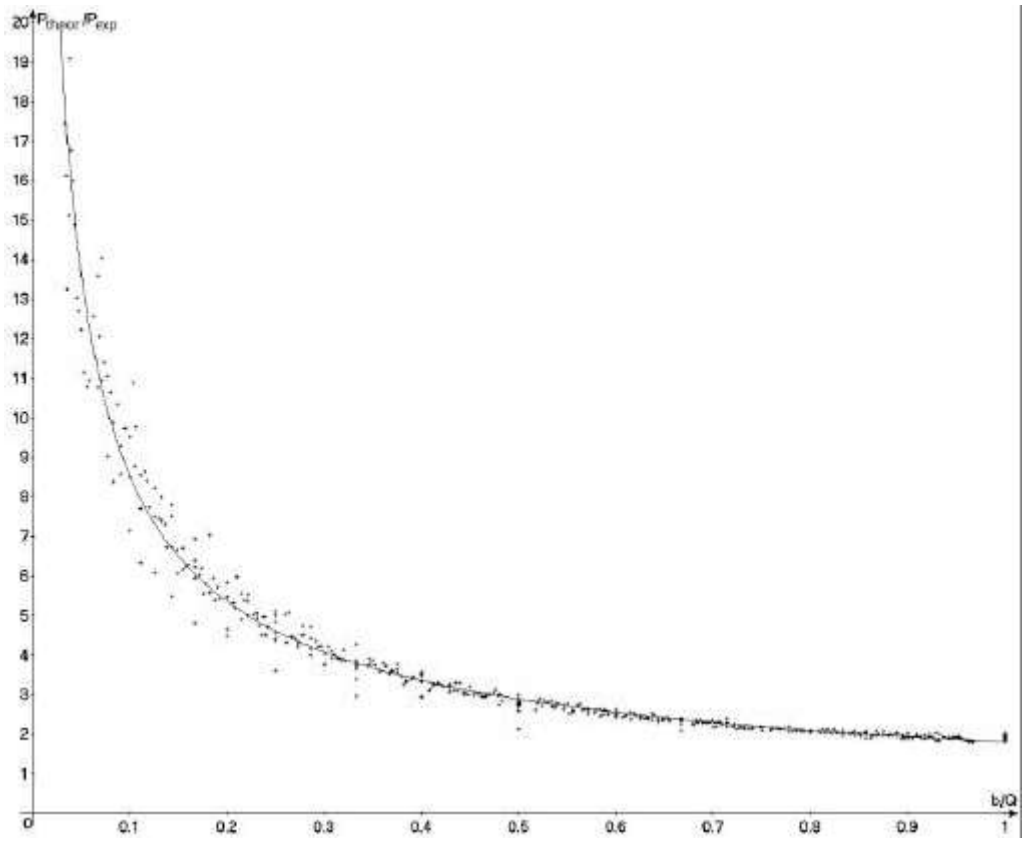


Fig. 6: Experimental and calculated P(q,b)

Thus, for a pair of languages L_1^Q and L_2^Q, we obtain:

$$\mu(L_1^Q, L_2^Q) = \mu^o \eta^\Sigma(o_1, o_2) + \mu^F \rho_1^\Sigma(A_1, A_2) + \mu^C \varepsilon_C(\varsigma_1, \varsigma_2) + \mu^T \varepsilon_T(\sigma_1, \sigma_2)$$

$$\eta^\Sigma(o_1, o_2) = \frac{\sum_{\beta_{1i} \in o_1} \min_{\beta_{2j} \in o_2} \left(\eta(\beta_{1i}, \beta_{2j})\right)}{|o_1|}$$

$$\rho_1^\Sigma(A_1, A_2) = \frac{\sum_{A_{1i} \in A_1} \min_{A_{2j} \in A_2} \left(\rho_1(A_{1i}, A_{2j})\right)}{|A_1|}$$

where ρ_1-normalized function of the distance between two classes of data elements; $o_1 = \{\beta_{1i}\}$ and $o_2 = \{\beta_{2i}\}$-the set of all classes of boolean operators with instances in the first and second languages respectively; the set of all coordinates of the available classes of data elements $A_1 = \{A_{1i}\}$ and $A_2 = \{A_{2i}\}$; μ^o, μ^F, μ^C, μ^T-weight coefficients, which determine the influence of the distance between those or other components of the languages on the distance between L_1^Q and L_2^Q. $\mu^o+\mu^F+\mu^C+\mu^T = 1$.

2.3.3 Model of lexical compatibility: For evaluation of lexical compatibility as a component of informational one, which characterizes the closeness of resources in terms of content, measure l_X is introduced. It is defined by P(q,b) which is the probability that the document D, formally relevant to arbitrarily given query Q by field A of resource R_1, will be relevant to query Q when searching on the field B of resource R_2.

Based on the set-theoretic approach, it has been found that the dependence on the value l_X from the query length and the number of query terms from the dictionary B has the following form:

$$l_X = \frac{1}{\alpha\left(b/q\right)} P(q, b)$$

where b-the number of terms in the query simultaneously belonging to dictionaries fields A and B; q-the number of terms in the query;

$$\alpha\left(b/q\right) = \frac{c}{\left(b/q\right)^\gamma}$$

experimentally obtained coefficient allowing us to take into account the unevenness of distribution of frequencies of terms in the resource.

Figure 6 gives an example of a dependence of values of ratios of theoretical and experimental values of probabilities for different b/q (calculated for one pair of resources), for which values c = 1,82 and $\gamma = 0{,}67$ were defined using the approximation method.

This assessment allows judging about the lexical closeness of resources pairs and it can be used as a basis for selection of the target resource through resources ranging, in accordance with their lexical closeness to the original resource.

OBJECT MODEL OF THE INFORMATION ENVIRONMENT OF THE FEDERATED SEARCH

The object model forms the basis of the integrated information environment (IIE) and search management navigation. The structure of the IIE is determined by the three-part nature of comprehensive information system "user-retrieval system-information resource", where search engine, as an intermediary, shall ensure the coordination of the parties-the user and the information resource, each having its own specifics in the organization and conduct.

The object model of the information environment is defined [2] as interaction between user workspace and user interface, a means of workspace control:

IS = <WS, IUse, Mf>, where WS-workspace; IUse-user interface; Mf-matrix of the functional interaction between objects in the workspace and the interface objects.

The user workspace is presented by three objects:

$$WS = \langle IR_w, IR_U, F_{WS} \rangle$$

where IR_W-associated external resources; IR_U-local user resources; F_{WS}-the structure organizing the workspace and reflecting the user's point of view on subject area integrating both information (documents, queries and links to associated resources) and meta-information components (dictionaries of subject areas, classification, rubricators, thesauri, ontologies), as well as the results of analytical processing.

Information resource as an object of workspace is determined on the basis of three components corresponding to the higher levels of the OSI:

- Identification of the resource, including properties such as a network protocol and type of resource;
- Contents of the resource, reflected by the resource subjects, metadata, presentation forms, methods of indexing;
- Interaction with resources, including search, API, query language, search tools.

The structure organizing the workspace is presented as taxonomy (hierarchy of objects), capturing user's knowledge about the problem at the documentary, lexical and conceptual levels.

The user interface is represented by the following components:

- Model of an operational space that includes document object model (DOM) and browser object model (BOM);
- Functional model that includes operational objects, functions and relationships as well as representations.

IMPLEMENTATION

The presented approaches and models are implemented in AIS xIRBIS [19, 20].

Interoperability at the application level is provided, in accordance with the linguistic compatibility model, through the application of unified query language. All types of query language syntaxes are supported (prefix, infix and postfix) with possible assignment of the mutual arrangement of the main elements of a query (search scope, criteria operator and condition expression). The characters encoding used in a query is specified and there is a possibility to work with operators different from AND.

Interoperability at the level of presentation is based on a model of meta-information compatibility and a table of classes of data elements.

At the session-level interaction with distributed resources is provided through a search gateway by the Internet-protocol (HTTP or Z39.50). Set of interaction parameters allows us to specify the protocol of interaction with the resource, the network address, the name of the database and schema, as well as the output format of the results obtained.

Depending on Internet-protocol, during interaction, the module of query forwarding (after the query translation and setting the necessary compliance between data elements) transmits a generated query to one of the gateways, which later provides a search session with the resource.

A repository of information resources containing their descriptions in a standardized form and providing the use of these descriptions in the process of automated federated search was designed. The repository consists of two components: a catalog of resource description and a table of data elements classes. The first stores records that contain XML-

descriptions of resources in a standardized form specifying resource parameters. The second describes a tree structure, where all elements are combined in groups of different detail levels, which provides an opportunity of exact or fuzzy conformity of data items in different resources.

CONCLUSION

The developed models of compatibility of information resources allow us to determine closeness of resources as concerns their possible application within federated search. The set of information and meta-information components based on the object model and presenting both information resources and user profiles in various aspects, forms the architecture of the distributed heterogeneous information array of subject area, where repository of resources provides an optimization of the process of accessing to resources in conditions of their partial completeness. This technology of federated search includes: (1) debugging (enrichment and reformulation) of query expression through the use of complementary search tools in a local database and (2) targeted translation of query expression and its redirection to associated resources.

ACKNOWLEDGMENTS

This work was partially supported by the Russian Fund of Fundamental Research (RFFR), grant 11_09_13128.

REFERENCES

1. Bates, M., 1989. The design of browsing and berrypicking techniques for the online search interface. Online Rev., 5 (13): 407-424.
2. Golitsina, O.L. and N.V. Maksimov, 2011. Information Retrieval Models in the Context of Retrieval Tasks. Automatic Documentation and Mathematical Linguistics, 1 (45): 20-32.
3. Strogonov, V.I., N.V. Maksimov, O.L. Go litsina, E.I. Bolotin and A.E. Okropishin, 2012. Models and the efficiency of the federated search in documentary information resources. Management systems and Information Technology, 1 (47): 78-83.
4. Lin Fang, 2004. A Developing Search Service: Heterogeneous Resources Integration and Retrieval System. D-Lib Magazine, Vol: 3 (10).
5. Sysoev, T.M., 2007. Integration and search for distributed data based on semantic web technologies, Ph,D Thesis, CCRAS, Moscow.
6. Braslavsky, P.I. and A.S. Shishkin, 2005. Approaches to the construction and implementation of specialized Metasearch Engine ProThes. Computing technologies, Special, 10: 49-57.
7. Averchenkov, V.I. and E.A. Leonov, 2011. Mathematical model of a universal multi-agent-based metasearch subsystem. BGTU Journal, 2 (30): 101-110.
8. Leonov, E.A., 2011. Formalizing of the process of monitoring of information on the Internet when creating object-oriented data warehouse, Ph.D Thesis, BGTU, Volgograd.
9. Khokhlov, A. Yu, 2003. Organization of adaptive federated search across library catalogs using the Z39.50 protocol. Digital Libraries, 2 (6): 28-43.
10. Kevin Chen-Chuan Chang, Bin He and Zhen Zhang, 2005. Toward large scale integration: Building a metaquerier over databases on the web. In CIDR, pp: 44-55.
11. Kevin Chen-Chuan Chang and Hector Garcia-Molina, 2001. Approximate query mapping: Accounting for translation closeness. VLDB, 10 (2-3): 155-181.
12. Shui-Lung Chuang and Kevin Chen-Chuan Chang, 2008. Integrating Web Query Results: Holistic Schema Matching. CIKM, pp: 33-42.
13. Zhen Zhang, Bin He and Kevin Chen-Chuan Chang, 2005. Light-weight domain-based form assistant: Querying web databases on the fly. 31st international conference on Very large data bases, Trondheim, Norway.
14. Bibliographix Official site. Date Views 01.12.2013 home.mybibliographix.com.
15. Reference Manager Official site. Date Views 01.12.2013 www.refman.com.
16. BiblioScape Official site. Date Views 01.12.2013 www.biblioscape.com.
17. Kogalovskii, M.R., 2006. Trend development of management technologies of information resources in digital libraries. RCDL'2006, Yaroslavl State. Univ. Demidov, pp: 46-55.
18. Riabukhin, O.V., D.O. Bryukhov and L.A. Kalinichenko, 2009. Formation of the expressions of views in task of resources registration in the subject mediators. Proceedings of the 11th Scientific Conference RCDL'2009, Petrozavodsk State University, pp: 343-349.
19. Maksimov, N.V., E.N. Vasina and O.L. Golitsyna *et al*., 2008. Document Information_Analytical System xIRBIS, State Registration Certificate No. 2008611511.
20. Maksimov, N.V., O.L. Golitsyna and A.E. Okropishin *et al*., 2011. Subsystem of documental information analytical treatment. State Registration Certificate, No 2011611694.