

## Conference Paper

# Information Seek and Retrieval Mechanisms Based on Interactive Dynamics Linking Technology

N.V. Maksimov and O.L. Golitsyna

National Research Nuclear University MEPhI (NRNU MEPhI), Moscow, Russia

## Abstract

The paper discusses the approach to dynamic hypertext technology used for query expansion and development. Two schemes of links creation on the fly are presented. The first one, System controlled scheme, generates (computes) hypertext links on retrieved documents output phase. In the second one, User controlled scheme, terms in viewed document are highlighted (and can be further used as a query in selected resource by one click) in accordance with user profile. The link creation mechanism is specified parametrically through the specification of the resource and selected (i.e. controlled) by the user. Linking mechanism is considered as a component of IR&S process, well controlled by user thought graphic interface.

Corresponding Author:

N.V. Maksimov  
nv-maks@yandex.ru

Received: 22 July 2018

Accepted: 9 September 2018

Published: 8 October 2018

Publishing services provided by  
Knowledge E

© N.V. Maksimov and O.L.  
Golitsyna. This article is  
distributed under the terms of  
the [Creative Commons](#)

[Attribution License](#), which  
permits unrestricted use and  
redistribution provided that the  
original author and source are  
credited.

Selection and Peer-review  
under the responsibility of the  
Breakthrough Directions of  
Scientific Research at MEPhI  
Conference Committee.

**Keywords:** Information Retrieval Interface; Dynamic Linking; Hypertext Technology; System-Human Information Processing; Dynamic Query Reformulation.

## 1. Introduction

It is obvious (see for example [1], that users may not always be successful in using self-representative vocabulary when locating searched objects in a system. This is an important thing to be aware of in case of problem or thematic queries with long-term interests.

This is accounted for by the fact that search process performs transformation of the information needs into a set of documents whose content meets a user's real need (this is a *visceral conscious, formalized, compromised* information needs defined by R.S Taylor [2] and, correspondingly, *Real, Perceived, Expressed, Formalised* information needs, defined by S. Mizzaro [3]. Each transformation (in general corresponds to Stratified model of information retrieval interaction [4] introduces its own uncertainty. The purpose and nature of search technology is to localize and either remove, or fix uncertainties. Thus the query is represented as a definition of the unknown through the known

## OPEN ACCESS

concepts and relations of the object itself, if this supposed object exists (actually or hypothetically). Otherwise (if objects to be retrieved don't exist) query is represented via additions, i.e., concept and object relations that are associated with the object.

Furthermore, a machine retrieval as a process that comes down to selection via comparison of a generally hypothetical object with documents that are stored in a Data Base, is implemented via comparison of well structured descriptions, i.e., images. Machine search is based on the set-theoretic model: i.e., a set of documents defined on the basis of the set of terms that represent the "entry points" to the search array. The selection ("calculation" of the relevant documents set according to the criterion of issue) and ranking or clustering (in accordance with the measures of proximity, for example) are the main types of search operations.

Another feature deals with "working" images that are formed and used in the machine environment and the human mind. Machine images are normally created as a static set of attributes (a stable structure) to reflect the most typical properties. In human mind images are formed mostly due to actions and do not in fact exist without links. Machine selection is thus implemented according to precise criteria that exclusively correlate with the means (values) of properties, but image search in the human mind is always by means of associations, normally by the purposeful (assumed) use of a value.

This evolves into the concept of Information seek and retrieval in Context (IS&R) which concerns the interaction between Documents and IT platform in a Temporal context of (influenced by) socioorg manifestations, domains and tasks [5].

Therefore, IS&R systems should have a mechanism for expanding search process based on user's search interests mining and fixing.

In the simplest case, query adaptation is an attempt to expand the terms or modify expression of the user's information needs. Six techniques are reviewed in [6] for obtaining terms for query expansion, which is classified in terms of whether they are user-focused or not and whether they are implicit or explicit: (1) *processing the user model*, which involves the implicit selection of terms from the user mind; (2) *processing aggregate usage information*, which involves implicitly obtaining terms from the query logs and/or documents; (3) *pseudo-relevance feedback (local analysis)*, which involves performing an initial retrieval round (that takes place behind the scenes) using the source query and then implicitly selecting expansion terms from the documents; (4) *global analysis*, which involves the implicit selection of expansion terms from a thesaurus, a knowledge source; (5) *(explicit) relevance feedback*, which requires the fact that the user explicitly provides relevance feedback about a number of documents

from an initial set of retrieved results where documents marked as relevant are processed to obtain expansion terms; (6) *interactive query expansion*, which involves user interface that allows the user to explicitly select expansion terms from a candidate list of terms suggested by the system.

Summarizing these provisions, we can say that the main ways of development of the searching process are:

- modification and enrichment of a query expression, and the use of complementary search mechanisms;
- using alternative "entry points" in the search space: a search for names of authors, names of projects, etc. related to the topic of the query, for example;
- using other information resources (IR) by forwarding the query and query expression translation to the target resource language syntax.

These technologies are briefly presented below using the example of IS&R system xIRBIS (© 1992-2018) [7].

Reference tools supporting the query development process and created on the fly are the subject of this work. They are not the main search engine, because they are used situationally in query formation and browsing processes. Such links are targeted at extracting of implicit and explicit user knowledge and information needs. In this approach links are inferred by the user-system interactions. We imply that interactive environment of IS&R is focused on formalized highly complicated cognition processes, characterized by non-linearity and unpredictability (it should be noted that due to this there are no means available in practice for measuring the actual performance of such mechanisms). In this search environment become more sophisticated: many functions including analytical moved to the client side. In such conditions a trivial reference mechanism becomes a universal management tool for complementary search functions.

Only reference tools that are targeted at query development will be considered in the paper. We will focus primarily on system-created links. We don't examine a static (predefined by author, for example) hypertext links. A link typology, mechanisms of links generation and use, corresponding control tools that are based on profiled resource specification, will be suggested.

## 2. Related Works

Technology of links creation and use of query-based dynamic linking is not a new emerging trend.

Technology of dynamic linking based on query marking up directly on text while it is being formed for browsing was suggested in [8].

In [9] the topic objects are stored with textual data objects containing references to other topic objects. The textual data objects are string-correlated to the topic objects to determine which topic objects are referenced in each textual data object. Hypertext links are generated for each reference in the textual data objects. In [10] contextual links are created between the selected terms and target documents associated with the identified topics using the URLs for the documents in the contextual links. The target documents are selected by identifying topics that are associated with, or described by, the selected terms.

Some works [11–13] propose concept-based methods to refine and expand queries in order to improve search performance. These models, however, are concentrated on reformulation of given queries but not users' background knowledge.

The diverse and developed query expand technology deals with the conception of User profiles for personalized Web search [14]. A user profile is defined by Li and Zhong in [15] as the topics of interests relating to user information needs. They further categorize user profiles into two diagrams: the data diagram for the discovery of interesting registration data, and the information diagram for the discovery of the topics of interests related to information needs.

Finally, in [16] the effectiveness of generating personalized hypertext is connected with user information media (as integration space).

Advantages and disadvantages of dynamic hypertext analyzed in [17]. Arguments in favor of query-based dynamic hypertext include: a simplified interface to search functionality, reduced authoring effort, and greater opportunity for customization based on the current user's interaction history or specific task context. However, disadvantages include computation of each link at run-time (instead of using stored, predefined links), which can be expensive in a large system with numerous users.

The flexibility of dynamic hypertext deals with respect to how links are created, computed, and selected. But it is clear, that one it creates new opportunities for personalized and customized information retrieval in general.

### 3. System Descriptions: Information Seek and Retrieval Features

Let us briefly characterize the ability of an IS&R system xIRBIS for better understanding of the location and nature of the reference mechanisms.

xIRBIS is an example of IS&R system which operates on domain-specific corpora, where the system allows to search for academic and technical documentary publications for user's long-term interests.

Hybrid Search engines use two kinds of linguistic tool: pre-coordinated structures (taxonomies, dictionaries, etc.) that are adequate for well-defined information needs, and post-coordinated (key words, semantic networks) structures for all other cases.

Search technologies are built on mutual complementarity by explicit (Boolean or vector) and implicit models. Explicit technologies are used for user specified query expression. Implicit «clustered» technologies are based on the dynamic search query re-formulation by relevant feedback. This helps to select statistically significant document subsets, as well as «in-between» documents which contents may not be that significant but include some new information. Technologies of xIRBIS used a several complementary information retrieval Mechanisms are enumerated below:

- verbal – search query in a natural language form, or in the form of expressions using Boolean and context operators;
- cluster – searching for similar documents, a set of relevant documents, or searching by selected terms from relevant documents;
- using taxonomy filters – pre-categorized structures by different kinds of aspects, objects, subjects areas;
- links – static and/or dynamic generated hypertext links of selected terms in browsing document;
- using linguistic and conceptual resources – dictionaries, thesauri, glossaries, typical search queries, etc.;
- using the time series dynamics – searching a set of documents which periodic dynamics is similar for the dynamics of a master set.

Lexical analysis of documents in a resulting set helps in following rating of the possibility of practical use of their contents, also enrich the search query and extend the subject area vocabulary. In conjunction with the visualization tools of statistical

analysis of the documentary flows and subsets (including the search for dependencies in the set of time series of publications) it provides the overall efficiency of the analysis.

Hierarchically organized structures [18] - Cognitive Tree, as called in xIRBIS, are used for personalized presentation of the information in the user profile. These structures are created by the user in a dynamic fashion and reflect his / her subject area personal vision. Moreover, each object is both standard, and individual vision of subject area. Such a view is integral due to the fact that it is implemented by objects as the level of resources (collections of documents, links to associated resources, etc.) and level of terminology (fragments of thesauri or subject heading, dictionaries).

Several complementary indexing technologies are used in the system. Typically, each document is classified in accordance with one or more classification engines and/or categories that are used for indexing in specific IR. For example, UDC, INIS, INION RAS, etc. In addition, the document is indexed by the search images based on key words – main concepts extracted from text. A list of descriptors (concepts relevant for thesaurus of corresponding subject area or specific for this document) stands out. Deep semantic relationships are reflected by the ontology. In the ontology, unlike a linear search image, the specific situation is presented in a connected form. Moreover, a set of standards, so-called functional relations are usually limited by taxonomy of relations that is determined on the basis of the functional domain model. Building of ontologies on the basis of NL-texts, i.e., the formation of connections between the selected text in the concepts is based on the transformation of linguistic relations in functional [19].

Because of this, it will provide a suitable search trajectory and an opportunity to find new dependencies for knowledge.

## 4. Hypertext Typology and Interactive Search Features

It should be noted that due to the nature of Web-interfaces, many functions of search engines are implemented by means of hypertext links.

Different forms of link types are dependent on how and when a query is constructed. In [20] the taxonomy of functional link types can be empirically derived by involving actual hypertext users who were asked to rank the types according to their perceived usefulness. However, it must be noted that such typology is more likely useful for ontology presentation of subject area with functional links as details of a particular object or a process.

To define suitable typology and so on to determine the diversity of links in general we consider the three level information architecture of a resource that are common to almost all professional IS&R. From the point of view of the nature and form of data representation (and therefore, search logic) classic architecture of information system includes three levels: the level of the actual documents (full text), search images level and meta-information level. The nature and logic of interrelations between the tiers' elements are displayed in Figure 1, where in parentheses some examples of the relations, characteristic for the automated or traditional information systems are presented.

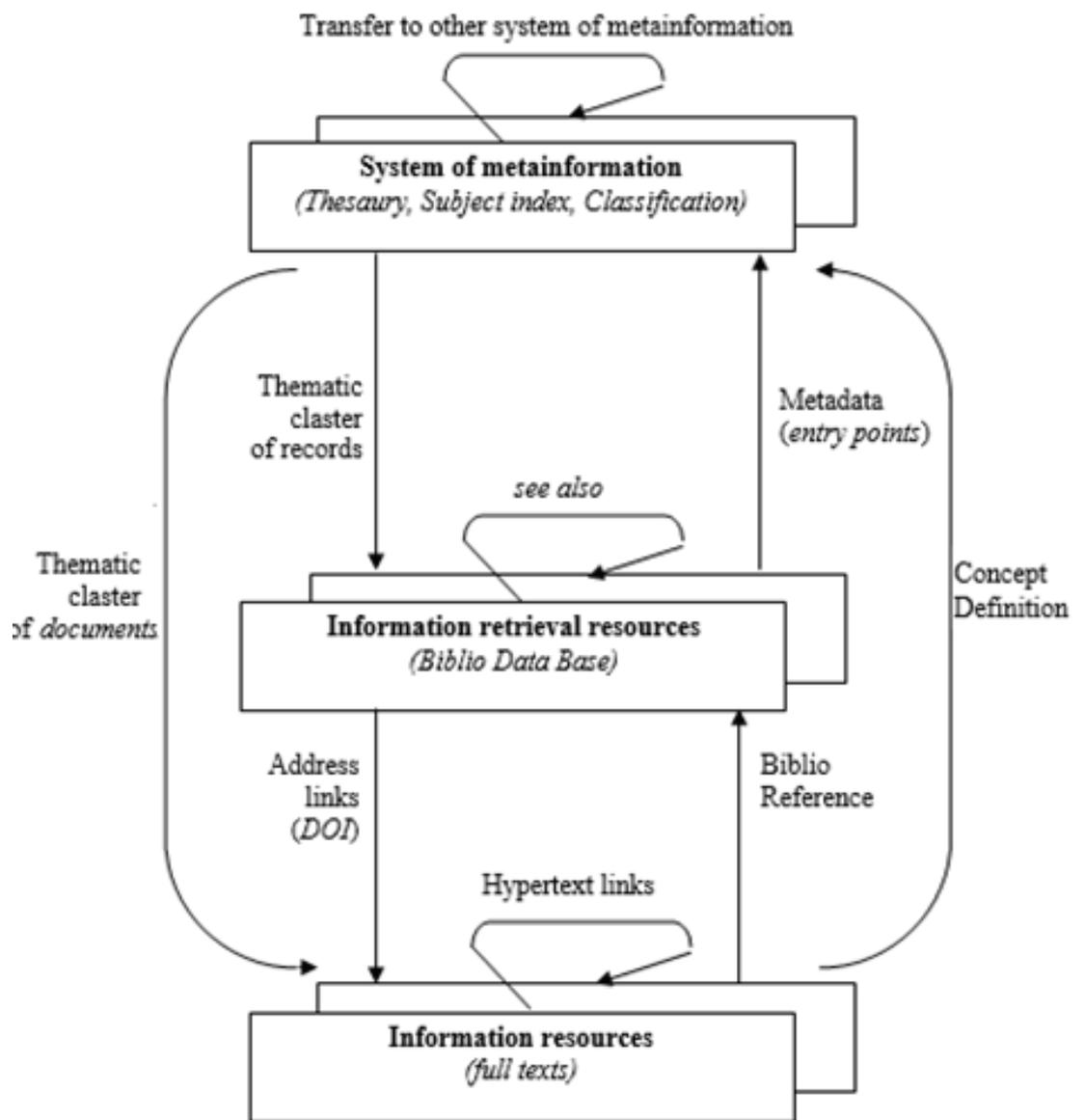


Figure 1: Interrelations between the tiers' elements of information resource.

For example, links to other publications of the author, or the publication where his role was different (editor, supervisor, etc.) are structurally defined. Links to text fragments that contain the parameter name to search for semantically close to the text are context-defined. Another example is the ability to move from data points to documents that match that points. All this gives the opportunity to develop the search process directly as a result of the analysis.

In particular, the documents obtained at the next stage search can be used through the reference mechanisms for the extension of user space terminology. The two following technological solutions for the links implementation are used to identify the entry point to the information space.

1. Performed by the system during the issuance formation, under the corresponding rules specified by the scheme of information resource:
  - dynamically generated links (hypertext deterministic) - search queries for documents containing the selected term;
  - indirect links established dynamically, for example, through some standard dictionary or search index;
  - dynamically resolved references (lookup or substitution) defined by the search function in the context of a specific document field.
2. User-initiated links to document elements allocated by the system in accordance with the dictionaries, controlled by the user:
  - by terms highlighting (color) within a document in accordance with the dictionaries (of the query terms, the operational lexicon, reflecting the interests of the user, for example), which allows to focus the user's attention, not only on the terms of the current query, but also on the terms alternative or corresponding more closely to the interests of the user;
  - by initiating a search for term, e.g., in Thesaurus, Glossary or the index, allowing to refine the understanding of the meaning of the term and other entry points.

Note that the first group of dynamic links represents queries that reflect typical (structurally-defined) communication. The second group (see example on Fig. 2) – quickly performs links and situational focused on the peculiarities of information needs of the user and his cognitive state.

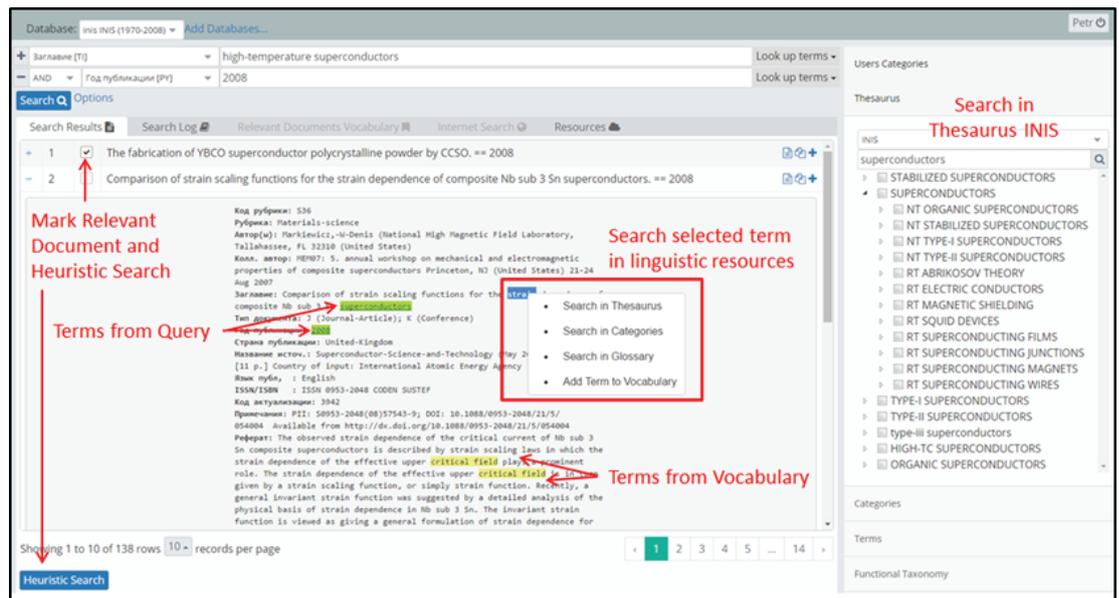


Figure 2: Example of different type links for xIRBIS system.

That is, the development of a query is essentially harmonizing of coordinates (multiple entry points) that are used on the one hand, by the creators of the IR, and on the other - by the user.

## 5. Technology of Links Creation and Use

Technology of dynamic linking is based on query marking up directly on texts while they are being read and browsed [8].

The sentence that the link (i.e., the marked up text that was clicked on) occurs in is sent to a search engine as a query. It is assumed that a user selects a particular link due to his/her interest in the content surrounding the link.

Actually the links that are implemented on specific elements of the current document define the structural metadata of the resource.

The basis of the search navigation management and, in particular, the reference mechanism is the object model of the information environment. The aggregate model is a three-tier information system "user – IS&R – information resource", where IS&R should ensure the agreement between the parties.

The information environment consists of three components:

- associated external IR;

- a search engine of IS&R which supports search and navigation captures the search results, and supports the local information resources as well as gateways to associated resources;
- a user view on the problem systematized in a way reflecting his view About (mentioned above Cognitive Tree).

Metadata (description) of the information resource includes two components important for dynamic links:

- specification of associated resources;
- specification of local IR.

Specification of associated IR determines their occurrence in the interface of the system either in the form of links to go into the operating environment of this resource, or in the form of the function (for example, in the local menu) to call the search function to search the selected fragment of the current document. Specification includes a formal description of the syntax of information retrieval language and names of search fields, call method and the structural analysis of the response.

The functioning of local resource is based on the concept of the extended Document Object Model (DOM). The actual specification of a local resource is defined by the so-called "database scheme". The scheme defines a document as an ordered set of data fields. In addition, it also includes the browsing form, the interface form of the query and the model search entry point. For any database it is possible to have multiple schemas that enable both a variety of the document contents displays and quick shift of the search entry point.

To generate dynamic link on a fly a for a certain field of the browsed document the query template should be specified (in an associated database scheme) for this field. During the formation of output forms of the current document the value of the field for which hypertext reference is specified, will be inserted into the corresponding query template.

If a link is defined as a "lookup" found records in specified IR that are formatted in accordance with the specified schema, are included in the current document. A typical example is including in the browsed form of retrieved document the definition of the subject heading index (as decoding, explanation). A corresponding fragment of scheme presented by next XML-code:

```
<fld fldtype="val">  
  <fldname "SI" />
```

```
<fldtitle "Subject Index" />
</fld>
<!-- lookup link -->
<fld fldtype="refval">
  <refbdname "SIbd" />
  <refquery "SI:#SI" />
  <schemname "SIvwbrief" />
  <fldtitle "Subject Index Headings" />
</fld>
```

This code specified an adding the heading (retrieved in SIbd – Subject Index data base) of Subject Index (assigned to browsing document) formatted according to DB-scheme "SIvwbrief".

If a relationship is defined as a hypertext link, the URL request (which is formed in accordance with the template) is included in the form of the displayed document. A typical examples is represented by generated hypertext links to author web-site or to other author publications.

## 6. Summary

As noted by [4] information searching is a subset of information-seeking, and in the context of information science, it refers to processes used for interrogating different information systems and channels in order to retrieve information. It is the most empirical and pragmatic part of information-seeking studies.

In general, not a single object, or technology can, if used in isolation, provide an comprehensive result. From one side, the structure and behavior of each of the above-mentioned objects are obvious. Whereas from the other side, each of them has its specific construction technology and "element base", and therefore they cannot be integrated analytically.

In this sense, technologically simple tools of dynamic links are a good way of expanding search space. A significant factor is that the link creation mechanism is specified parametrically through the specification of the resource and is selected (i.e., controlled) by the user. Visual interactive interface provides both a documents view and a functionality to analyze group of documents and their relationships. Interface components also enable users to select in the resulting set the subject area he is

interested in by clicking elements values or their attributes. This reduces the dimension of analyzed area without loss of important information for the user.

## References

- [1] Furnas, G.W., et al., 1987. The Vocabulary Problem in Human-System Communication. *Communications of the ACM*, 30, 964-971.
- [2] Tailor, R.S., 1968. Question-negotiation and information seeking in libraries. *College and Research Libraries*, 29, 178-194.
- [3] Mizzaro, S., 1998. How many relevances in information retrieval? *Interacting With Computers*, 10 (3), 305-322.
- [4] Saracevic, T., Spink, A. and Wu, M-M., 1997. Users and intermediaries in information retrieval: What are they talking about? *User modeling. Proceedings of the Sixth International Conference, UM97*. New York: Springer, 43-54.
- [5] Ingwersen, P., Jarvelin, K., 2005. *The Turn: Integration of Information Seeking and Retrieval in Context*, Dordrecht:Springer.
- [6] Ghorab, M.R., et al., 2013. Personalised Information Retrieval: survey and classification. *User Modeling and User-Adapted Interaction*, 23 (4), 381-443.
- [7] Maksimov, N.V., et al., 2014. Documentary information-analytical system xIRBIS: Software. *The state registration certificate of computer programs* No. 2014619640 from 17.09.2014.
- [8] Golovchinsky G., Chignell M.H., 1993. Queries-r-links: Graphical Markup for Text Navigation. In: *Proceedings of INTERCHI'93, Amsterdam*, 454-460.
- [9] Automatic generation of hypertext links to multimedia topic objects, 1998. US Patent #US 5815830 A.
- [10] Dynamic generation of contextual links in hypertext documents, 2000. US Patent # US 6122647 A.
- [11] Fonseca, B.M., et al., 2005. Concept-based interactive query expansion. In: *Proc. of the 14th ACM intl. conf. on Information and knowledge management, New York*, 696-703.
- [12] Grootjen, F.A., van der Weide, Th.P., 2006. Conceptual query expansion. *Data & Knowledge Engineering*, 56(2), 174-193.
- [13] Stojanovic, N., 2005. Conceptual query refinement: The basic model. In: *Proceedings of WISE2005*, 404-417.
- [14] Gauch, S., Chaffee, J. and Pretschner A., 2003. Ontology-Based Personalized Search and Browsing. *Web Intelligence and Agent Systems*, 1 (3/4), 219-234.

- [15] Li, Y., Zhong, N., 2006. Mining Ontology for Automatically Acquiring Web User Information Needs. *IEEE Transactions on Knowledge and Data Engineering*, 18(4), 554-568.
- [16] Steichen, B., Lawless, S. and O'Connor, A., 2009. Dynamic hypertext generation for reusing open corpus content. In: *the 20th ACM conference on Hypertext and hypermedia, Torino, Italy — June 29 - July 01*, New York, 119-128.
- [17] Bodner R., Chignell M., 1999. Dynamic Hypertext: Querying and Linking. *ACM Computing Surveys*, 31 (4), Article No. 15.
- [18] Maksimov, N.V., Golitsyna O.L. and Usenko, A.L., 2014. The structure and components of the operational visual space for scientific interactive information retrieval. *Journal on Scientific Visualization*, 6 (4), 96-106.
- [19] Golitsina O.L., et al., 2014. Semantic Identification of Text in the Class of Tasks of Information Retrieval. In: *Proceedings of ICAI'14, WORLDCOMP'14, July 21-24, Las Vegas, 2014*. Nevada:CRSEA Press, 47-52.
- [20] Kopak, R.W., 1999. Functional link typing in hypertext. *ACM Computing Surveys*, 31 (4), Article No. 16.