Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 169 (2020) 81–89

Procedia Computer Science

www.elsevier.com/locate/procedia

Postproceedings of the 10th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2019 (Tenth Annual Meeting of the BICA Society)

Knowledge Representation Models and Cognitive Search Support Tools

Nikolay Maksimov*, Olga Golitsina, Kirill Monankov and Anastasia Gavrilkina

National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, Russia

Abstract

The purpose of the work is to develop a cognitive-process-oriented model of data search, as well as tools of human interaction in according to this model. Cognitive search is considered as process of formation of ontology of the subject area (of target object) as a system of three systems (functional, conceptual and terminological). Accordingly, the results and trajectory of information retrieval are a reflection and component of the process of cognition. In order to reduce the dimension of the graph (to perception possibilities) aspect projections operation based on the taxonomy of relationships and entities is used.

Keywords: information retrieval; cognitive search; knowledge extraction; text processing; interactive interface; ontology graph; aspect projection.

1. Introduction

Any new knowledge is based in one way or another on already existing knowledge. Information retrieval systems (IRS) are traditionally intended specifically to provide information support for searching information blocks in external resources that will be useful for synthesizing new knowledge. At the same time, there is a clear trend of the natural merging of technologies of main activity and information activity.

* Corresponding author.
E-mail address: nv-maks@yandex.ru, kmonankov@yandex.ru, asgavrilkina@yandex.ru

10.1016/j.procs.2020.02.118

The use of information retrieval technologies in cognitive processes is based on the presentation of information (published knowledge and information needs) in a well-formalized form, adequate to the computing environment capabilities, and the creation of resources focused on the search tasks, extracting and analyzing knowledge. In turn, the formation of such resources is reduced to the analysis, systematization and formalization of the information collected. That is, the functional sequence of synthesis of new knowledge closes: for synthesis of new knowledge search and the analysis of already existing knowledge is necessary.

In the domain of automated IRS there is no established definition of the concept of cognitive search at the moment. For example, in [1], a cognitive search is presented as a tool focused on solving the following tasks: removal of synonymy and polysemy, analysis of abbreviations, morphological analysis of text, removal of ambiguity, calculation on ontologies. In [2] cognitive search considers as a tools of enriching the search index through the use of artificial intelligence methods such as pattern recognition — extracting text from images, videos, voice recordings, and other unstructured data sources, as well as the use of deep natural language processing. In [3] author discusses the tools of interactive user interaction with information resources containing big data of heterogeneous information, and provides methods for supporting the formulation of a query (hint, error handling), the formation of facets, ranking of results, highlighting query terms and geodata referencing.

Note, however, that the nature of the functions attributed in most works on cognitive search rather corresponds to the concept of conceptual search (search by concepts).

The proposed approach to the cognitive search definition has a fundamental difference. Cognitive search is not only the search by concepts, but also the construction relationships of concepts, which provides the transformation of information from the source data to the practical results of the activity.

The paper presents an approach based on the assumption that the sequence of operations of automated search and information analysis in the computing environment is similar to the sequence of operations performed by a person in the extraction and synthesis of knowledge elements.

Human consciousness operates with interconnected (often associative) images of objects, which can be represented in the form of a certain scheme suitable for solving current problems. Schematic representation of knowledge and the possibility of their generalization is the basis of methods for constructing conclusions, such as induction, deduction, transduction and abduction.

At the same time, the synthetic role of language is important. The language has two functions: (1) communicative – ensuring the exchange of messages between subjects in space and in time, and (2) cognitive – modeling of reality, which establishes a relationships both between sign (linguistic) representations of concepts and relations, and between image spaces of model (entities, concepts and relationships) and originals (elements of reality).

IRS plays the role of a communication environment and serves for the interaction of consciousness with an information resource containing previously accumulated knowledge. The “cognition” of the system in such interaction can be provided by the use of search technologies similar to the operations of consciousness, such as combining, ordering in accordance with the scheme, determining the relevance, significance and usefulness. Accordingly, in terms of information and linguistic support, this will require the availability of deep indexing methods and technologies that allow extracting meanings from the text, representing them in an extended and schematic form, convenient for manipulation.

In this context, the aim of the work is to develop a set of relevant linguistic components and functional tools based on the datacentric model of human cognitive activity and a systematic approach.

Experimental studies were carried out on the arrays of full texts of scientific (articles, dissertations, scientific reports), technical (design, structural documentation), operational (reports and investigations of faults, etc.), regulatory (standards, regulations) and other (for example, court decisions) documents on the platform of information-analytical system xIRBIS[1].

[1] Information and analytical system xIRBIS (©1992-2018) is used by a number of leading information centers and organizations to create industrial documentary databases.

2. Datacentric model of cognitive process

According to [4] the cognitive process and the scientific search process have a general scheme, which including the following stages:

- search and extract information blocks from the environment according to goals and tasks;
- ordered or random combinatorial validation of these blocks importance;
- expanding knowledge due to those combinations of information blocks, which together with the available knowledge form a system of concepts (correspond accepted axiomatic and criteria);
- presentation of the new "personal" knowledge of the subject in form (technical solution, message, document, etc.), which will ensure its "identification" and reuse inside and outside the cognitive process.

At the same time, the relationship of principles, goals and tasks is dialectical and their choice is almost always conditioned by the requirement of "constructability". Goals and tasks are relative and mutually complementary, and the hierarchy between them is established only in the process of forming a holistic view of the task. In addition, in general case, the solution of the problem from the perspective of solvability of problems and achievability of the goal that is, as a solution to the problem that succumbed to the solution [5].

Modeling the cognition process "technology" (deepening knowledge based on structuring and systematization) is based on an iterative correlation of model and practice. This approach proposed in [6], on each turn of the spiral can include:

- the creation of structural schemes that specify the model or the system of models of subject area;
- justification (correlation and verification of adequacy) of these schemes from the point of view of a wider whole;
- knowledge construction that describes the different aspects of the considered reality, that is, knowledge that can be used practically.

Such a procedure of correlation of schemes and objects of reality, the transition from the particular to a wider whole, can be considered as the construction and justification of ontology. Cognition is an iterative (spiral) process of construction a model of reality[2] represented by its individual fragments.

The generalized formalized representation of the new knowledge based synthesis process on the existing is given within the framework of the General theory of Systems [7]: cognition is reduced to search of a set of interrelated elements by their properties, building and applying the law of composition, then decomposing the result into elements by characteristic features, and so on.

These provisions form the basis of the approach considered in this paper.

3. Ontology as a semiotic model of knowledge

In a society, knowledge generated mainly in the human mind and used as a result by man. Therefore, they represented mainly by linguistic signs, which are characterized by ambiguity. In the computing environment, unambiguous mathematical signs should represent knowledge. This predetermines the need to use such a formalization of the representation of knowledge, which will ensure the transformation of the representation of knowledge from systems of linguistic signs into a mathematically computable system. Of course, the purpose of creating such transformations is to build such logic that would ensure not only the identification of knowledge, but also the formalization of logical inference.

From the point of view of semiotics, knowledge is an object of cognitive and production processes that has a complex nature and is closely related to the "concept" and "sign" categories. Knowledge can be represented by a

[2] Reality is that part of the real world that directly or indirectly (e.g., through measurements) interacts with the subject. Model – an image of how reality works.

semiotic triangle with a connotate placed in it (see Fig. 1). This allows to take pragmatics into account, linking the meaning to the circumstances of its use.

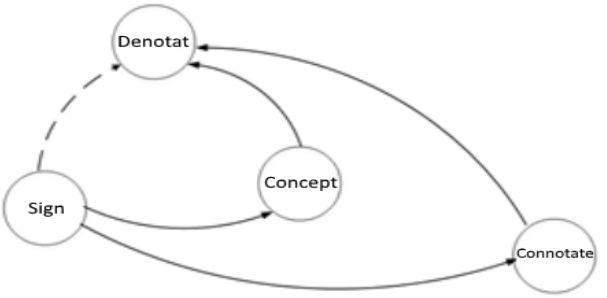

Fig. 1. Semiotic knowledge scheme

Note that the correlation of the concept and the connotate, in essence, ensures that the model (conceptual image) is consistent with practice (compliance with the properties and behavior of a real object).

Cognitive search (formation of systematized knowledge in the subject area) is considered as the process of forming the ontology of the subject area, represented by a system of three systems[3] (functional, conceptual and terminology):

$$S_i = <M_i, A_i, R_i, Z_i>, \text{ where} \qquad (1)$$

- M_i – a set of knowledge elements (fragments), such as the basic concepts of the subject area, separate facts or situations, the goals of the process participants, etc.;
- A_i – the characteristic properties of the knowledge elements: significance, usefulness, relevance, etc.;
- R_i – relations between elements: functional or structural relationships between objects, the order of events, the transformation of an object, etc.;
- Z_i – the composition law, for example, a set of axioms, hypotheses, principles, etc.

Note that the knowledge representation in the form of a system of three interconnected systems corresponds to the dynamism and semiotic nature of the processes of generation, preservation and transfer of knowledge. The objects and relations analysis are performed in a real environment (which is reflected by a functional system). The analysis, generalizations and the construction of conclusions are performed in consciousness (where concepts, entities, relationships, etc. are the operational objects). The fixation and transmission of the results and the developed methods are performed by means of a language (sign system). Each of these systems exists and develops relatively independently. The law of composition in the definition of a knowledge system corresponds to the property of relativity and variation of knowledge: for the same object, process or reality phenomenon, depending on the axiomatic, several alternative or complementary models can be constructed.

---

[3] Ontology is defined as a collection of three interconnected systems [9]: $O = <S_f, S_c, S_t \equiv>$:

S_f – is the functional system (objects and connections of reality). $S_f = <M_f, A_f, R_f, Z_f>$, where M_f is a set of objects (entities), A_f is a set of characteristic properties, R_f is a set of functional relations (represented by typed situational connections characteristic of the subject area), Z_f is the composition law, i.e. rules and ordering schemes for objects (subject domain taxonomy);

S_c – is the system of concepts. $S_c = < M_c, A_c, R_c, Z_c >$, where M_c is a set of subject area concepts, A_c is a set of signs of systematization of concepts (meronomy), R_c are classes / subclasses of paradigmatic relations, Z_c is the composition law (representation scheme that defines which concepts, in which relationship, and in what order will be included in classes);

S_t – is the terminology system. $S_t = < M_t, A_t, R_t, Z_t >$, where M_t is a set of terms, A_t is a set of terms properties, R_t is the relationships of equivalence and inclusion, and linguistic relations, Z_t is the composition law (grammar);

$\equiv$ – the operation of comparing elements of different systems at the level of signs, ensuring their identity in functional, conceptual and terminological systems.

4. Cognitive information retrieval model

Search in information resources, as a substituting part of some part of the knowledge generation process, includes the following steps:

- selection of documents from information resources, each of which represents at least one information component or its image;
- combinatorial construction of information components based on some set of characteristic features of clusters and determination of the degree of "integrity" of these clusters as new information components;
- ordering these clusters according to their "value". The purpose of such ordering is to reduce the size of the sample viewed by the subject. It is assumed that the measure of value corresponds to the probability of the content of the desired new in the cluster.

Thus, the combinatorial combination is a peculiar method of generating knowledge and a common technological basis for relatively independent and, at the same time, interdependent processes of the main and informational activities. At the same time, according to the theory of dynamic systems [10], the IRS, forming unequal combinations (sample of documents), performs the function of a "mixing layer", thus ensuring the acceleration of a non-equilibrium state.

Note that the process is stochastic, while there is implicitly a factor of "directionality", which can be set, for example, the goal or the problem situation.

The trajectory of information retrieval is a reflection of the trajectory of purposeful cognition, and therefore is complex, in particular, it includes referring to a certain technological object, performing a function, as well as analyzing and evaluating the result.

A person usually forms such a trajectory implicitly. However, an automated system should have the means to support the construction of a path in the space of accumulated knowledge, to offer people various ways of searching and forms of presenting information.

A means of systematization of information needs and search results is a Cognitive subject tree (CST) [11]. CST is a hierarchical classification structure of a user representation of a subject area. Tree nodes are assigned by the user according to the elements of the knowledge (documents, terms, fragments of ontologies, etc.) corresponding to the topic. The hierarchical orderliness of the structure reflects the mechanics of the process of cognition: knowledge is improved according to the specialization pattern, usually by dividing the current whole into parts according to the values of the selected division feature.

Methods for analyzing the content of rubrics allow to determine its importance, to identify contradictions, to determine the incompleteness of the existing system of knowledge, to identify dependencies between subject areas.

The ontology graph as a search tool is a convenient form of representing a set of interrelated heterogeneous objects, and allows one to visualize the basic concepts of the subject area and the links between them.

However, an ontology built on the full text, taking into account not only meaningful concepts and relationships, generates a graph, the size of which is practically unacceptable for constructive display. Reducing the dimensionality of operation is implemented by aspectual projections.

An aspect is a person's point of view on a system, depending on which a certain subset of characteristics is of interest for solving the tasks of the main activity. The set of aspects[4] are defined from a system point of view in accordance with the activity model. At the top level, the following two classes are defined: structural and behavioral aspects. The set of structural aspects include aspects that reflect the internal and external morphology of the system. Internal reflects the relationship and composition of the elements that make up the system. External reflects the properties that manifest themselves in the interaction, in particular, with the environment of the system, including

[4] The aspect is formally given by the functional system $S_a = <M_a, A_a, R_a, Z_a>$, where M_a is the set of reference concepts, A_a is the set of characteristic properties, R_a is the set of functional relations characteristic of this aspect, Z_a are the rules that determine the construction of chains (limiting the inclusion of a particular triad in the path on the graph based on the analysis of the characteristic properties of the corresponding arcs and nodes).

qualitative assessments. Examples of structural aspects are: "Composition", "Following", " Accordance", "Form", "Change", "External signs", etc. Behavioral aspects reflect dependencies and ratios, which are determined, among other things, by a generalized functional activity model. Examples of behavioral aspects are "Efficiency", "Manageability", "Security", etc.

Aspects can also reflect the nature of the representation of meaning, allowing to analyze the diversity and intensity of using a particular concept (forming the corresponding hypergraph) or its evolution (forming a directed subgraph in which the concept will be presented in the "historical" sequence of its use in the text contexts, related relationships of different types with different concepts).

Taxonomy of aspects is used as part of the linguistic support of information retrieval. Such taxonomy (being an object open to expansion and modification) defines a set of possible aspects, each of which fixes the connection of an aspect with classes of characteristic relations. Relations classes, in turn, are associated with linguistic constructions[5] that represent the corresponding relations.

Let us give an example of application of the aspect projection for an automatically constructed ontology of the document “General description of the power unit” (Fig. 2[6] shows only a fragment, since the ontology includes several hundred vertices).

Figure 3 shows the “Accordance” aspect of graph for the entity “station”. The aspect was defined by the basic concept of “Station” and the following classes of characteristic relations: “Dependence”, “Be a basis”, “Be a part/whole”, “Be a goal/purpose”, “Locativity in space”, etc.

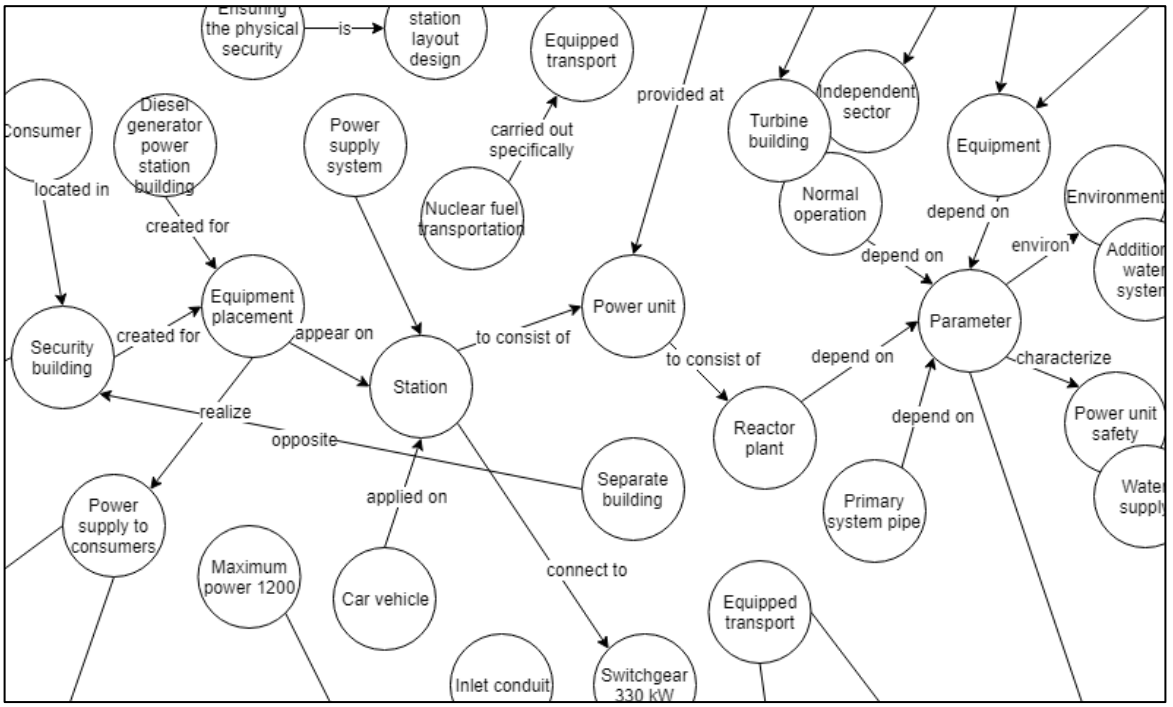

Fig. 2. Fragment of the ontology graph of the document “General description of the power unit”

[5] The correspondence of relations and linguistic constructions is given in [12].
[6] Fig. 2 and Fig. 3 are translated from Russian.

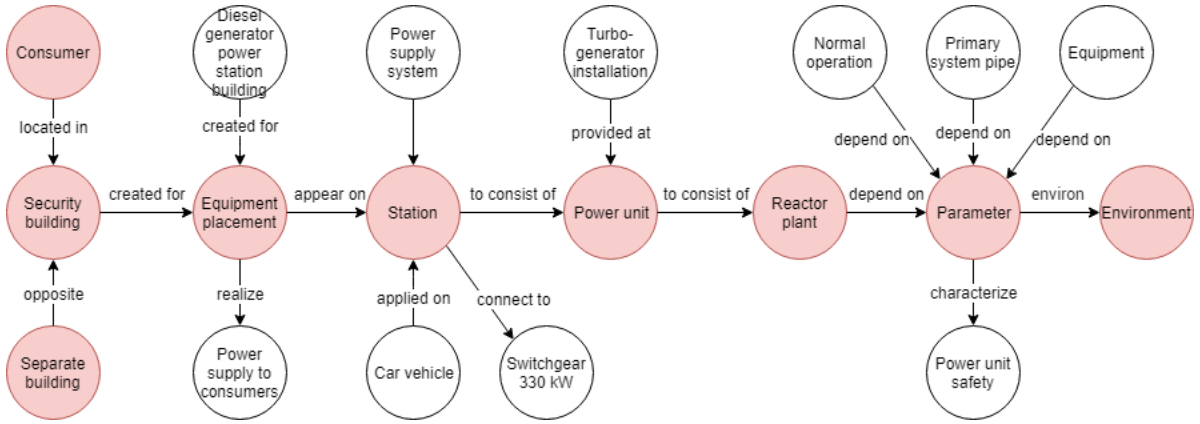

Fig. 3. Aspect projection of the ontology graph of the document “General description of the power unit”, aspect “Accordance” of the entity “Station” is represented by a sequence of vertices highlighted in red

Consider the correspondence of cognitive information retrieval operations to cognitive processes. According to [13], the process of cognition is phased (the stages are shown in Table 1) and iterative, and the stimulus for the transition from iteration to iteration is the presence of contradictions in the existing knowledge. Table 1 shows the stages of cognition, which correspond to the operation of cognitive search.

Table 1. Correspondence of cognitive search operations to the stages of the cognitive process.

Stage of the cognitive process	Operation of the cognitive process	Operations of the cognitive information retrieval
Check and study your own knowledge of the object under study. Search for contradictions in existing knowledge	Identification of the components of the object of knowledge.	Research and addition of the structure and content of the CST.
	Formation of the corresponding set of concepts. Matching analysis.	The use of methods for comparing the states of filling individual rubrics.
Obtaining as facts of the elements of knowledge, proven theory and practice.	Selection of known properties and relationships of the object of cognition. Check for compliance with the laws and regularities of the subject area, as well as practice.	Search for documents (definitions and descriptions of facts), their fragments and interrelationships using dictionaries, classification systems and ontologies. Verification by finding supporting documents.
The resolution of contradictions in the knowledge gained.	Internal and external comparison of the components of the object of cognition and its properties.	Search for contradictions in the structure and content of the rubrics of the CST.
	Selection of classes of objects and classification features.	Comparison of graphs of ontologies of documents (intersection, search for analogues).
Decomposition/consolidation and ordering of the knowledge gained in accordance with the new methodological (classification) scheme.	Generalization and detailing of the object. Detection or modification of the properties of the object of cognition.	Construction of aspect projections. Scaling ontology graphs. Selection of vertices and connections of the graph, reflecting the characteristic and distinctive properties, as search signs.
	Reforming the class system.	

Cognitive search is a directional process in which the global direction is set by the ultimate practical goal. At the same time, objectives are reflected in the CST indirectly: through the names of information blocks (thematic areas) and the knowledge decomposition scheme.

5. Formalized presentation and extraction of knowledge from the text

Extraction of knowledge from the text for a further schematic representation consists in dividing the text into elementary constructs corresponding to the concepts and relations. That is, the objects and links of reality correspond to the concepts and relations, which, in turn, have a symbolic image.

Building a formalized conceptual structure of the ontology of the text includes the following steps:

- tokenization - separation of characters sequences according to the assigned delimiters;
- preliminary identification of tokens based on the analysis of graphematic signs, regular expressions, the use of dictionaries of names and units, in particular, abbreviations, proper nouns, units of measure, numerical values, etc.;
- identifying of semantic units (concepts, properties, nominal groups, etc.) by combining tokens based on an analysis of their morphological characteristics, order, group structures, dictionaries, thesauri, etc.

For extracting and identifying concepts and relations, patterns of typical Russian language structures are used, which takes into account morphological features, compatibility, word order, and the presence of certain morphemes or specific vocabulary. Nouns and noun phrases are considered as concepts. Relation can be represented by a single word or a combination of words of the following parts of speech: verb, participle, adverbial participle, short adjective, short participle, adverb, particle, and preposition. Relationships are typed in accordance with the taxonomy of relationships [12], which allows to correlate semantic links that are similar in meaning, but presented in the text by different verbal constructions.

6. Conclusion

Modeling the cognitive search process from the point of view of the Systems theory is a complex approach to solving knowledge management problems. The data-centric model of the information retrieval process is similar to the model of the human cognitive activity processes and allows the search development trajectory to be fixed in the computing environment, which corresponds to the cognition trajectory.

Representation of cognitive search as a process of explicit formation of the subject area ontology, the use of a CST as a means of accumulating and preserving knowledge and the use of aspect modeling allows to form a subject area decomposition scheme in the computing environment by identifying relatively independent aspect representations, which together form a complex system.

Formation of the ontology graph of the document is one of the approaches for extracting and presenting knowledge from natural language texts. Such an approach allows visualizing the basic concepts and relations of the subject area in a convenient form. And the use of special operations allows reducing the set of objects to a brief form acceptable for display and perception. The approach to the construction of ontology is based on the unification of semantic relations through the use of taxonomy and classification of relations, which allows the use of comparison operations based on meaning, regardless of its verbal expression, and thus identify contradictions and gaps in knowledge.

The purpose of cognitive search is achieved by iterative addition of the existing model of reality to meet the user needs. A decrease in the number of iterations and an increase in the number of relevant results can be achieved by increasing the completeness of knowledge extracted, which requires the accumulation of lexical and semantic patterns and typed linguistic constructions, as well as the creation of inference rules for expanding the capabilities of the aspect representation.

A systematic approach to the knowledge synthesis, the rational organization of the process of working with objects of different nature (goals, documents, terms, evaluations) and the process of human-computer interaction (search, result evaluation) can effectively solve the problems of knowledge management systems.

References

[1] Dahlgren, K. (2010) “Demonstration of Cognition Search.” *Proceedings of the IEEE Fourth International Conference on Semantic Computing*: 454–455. https://doi.org/10.1109/ICSC.2010.100

[2] What is "cognitive search" in Azure Search?, https://docs.microsoft.com/en-us/azure/search/cognitive-search-concept-intro, last accessed 2019/05/25.

[3] Microsoft Azure official YouTube channel, How to build AI applications with Cognitive Search, https://www.youtube.com/watch?v=k5xScEyyI4M, last accessed 2019/05/25.

[4] Druzhinin, V. V., Kontorov, D. S. (1976) “Problems of System Science [Problemy siste-mologii].” *Sovetskoe radio*, Moscow.

[5] Penrose, R. (1994) “Shadows of the Mind.” *Oxford University Press*, Oxford, (Vol. 4).

[6] Shchedrovitsky, G. P. (2004) “Problems of research logic and analysis of the structure of science [Problemy logiki nauchnogo issledovaniya i analiz struktury nauki].” *Put'*, Moscow, (Vol. 7).

[7] Urmantsev, Y. A. (1988) “General Theory of Systems: State. Applications and Development Perspectives [Obshchaya teoriya sistem: Sostoyanie, prilozheniya i perspektivy razvitiya].” *Sb. sistema, simmetriya, garmoniya*, Mysl', Moscow: 38–124.

[8] Golitsina, O. L., Kupriyanov, V. M., Maksimov, N. V. (2015) “Information and technological solutions applied for knowledge-management tasks.” *Scientific and Technical Information Processing*, **42**(**3**): 150–161.

[9] Golitsyna, O. L., Maksimov, N. V., Okropishina, O. V., Strogonov, V. I. (2012) “The onto-logical approach to the identification of information in tasks of document retrieval.” *Automatic Documentation and Mathematical Linguistics*, **46**(**3**): 125–132.

[10] Chernavskiy, D. S. (2004) “Synergetics and information [Sinergetika i informaciya].” URSS, Moscow.

[11] Maksimov, N., Golitsyna, O., Usenko, A. (2014) “The structure and components of the operational visual space for scientific interactive information retrieval.” *Scientific visualization* **4**(**6**): 96–106.

[12] Maksimov, N. V., Gavrilkina, A. S., Andronova, V. V., Tazieva, I. A. (2018) “Systematiza-tion and Identification of Semantic Relations in Ontologies for Scientific and Technical Subject Areas.” *Automatic Documentation and Mathematical Linguistics*, **52**(**6**): 306–317.

[13] Yablonskiy, A. (2001) Models and methods of science [Modeli i metody issledovaniya nauki]. Editorial URSS, Moscow.