An Ontological Approach to Information Identification in Tasks of Document Retrieval: A Practical Application

O. L. Golitsina, N. V. Maksimov, O. V. Okropishina, and V. I. Strogonov

INTRODUCTION

The concepts of information and information retrieval are always connected in some way with processes with undefined results and if it is a controlled process with choices that, in their turn, make use of data outside an information retrieval system (IRS)—material knowledge. This indefinite choice is due to sequential transformations in the links that an IRS brings together: “knowledge—information—document—document search image (DSI)” and “problem situation—task—query—query search image (QSI).”

The process of information retrieval lies in the search of set of documents formally in compliance with a QSI by means of the procedures that implement some retrieval model. It is to be taken into account that each transformation in the link “knowledge → DSI,” as well as in the link “problem situation → QSI” is a reflection into spaces with less variety. Furthermore, spaces in both links for each transformation are similar (are of the same nature) but are not equal. And at the same a machine search as a process that comes down to selection via the comparison of a generally hypothetical object with objects that are stored in a massive, to selection via the comparison of a generally hypothetical object with objects stored in massive, to selection via the comparison of a generally hypothetical object with objects that are stored in a massive.

our mind, images are formed mostly due to actions and do not in fact exist outside 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 (links), normally by the purposeful (assumed) use of a value.

In this sense, search images that are created using the suggested ontology approaches enable one to work with both features that define properties and with features that determine their mutual connection (behavior).

1. OPERATIONS ON ONTOLOGIES DURING AN INFORMATION RETRIEVAL

The formal determination of an ontology for information retrieval that was suggested in [2] enables one to use it as an operation object, i.e., a means of selecting documents based on their semantics.

1 Let us note that as variable as search models and similarity criteria (in the computing environment) are, the corresponding algorithms are reduced to binary logic.

2 An ontology of a domain of interest is formally defined as $O = (S_f, S_c, S_t, =)$, where $S_f$ is a functional system (the "workspace interface" of the ontology in the activity of the subject); $S_c$ is a conceptual system (the logical and semantic base of an ontology); $S_t$ is a term system (signs that are used to fix ontology on the carrier); and $=$ is the operation of comparing the elements in different systems at the level of signs, which provides their identity in the functional, conceptual, and term systems. From the point of view of structure (for the implementation of operations on ontologies) a functional system can be a marked, weighted, and directed multigraph and a conceptual system with a marked, weighted, and directed graph. A term system is described by an $n$-linked graph, where each connection component is a complete graph (equivalency), a tree (inclusion), or the result of union of complete graphs and trees (if there are common nodes).
An ontology that is created using a certain unit of a document flow can be regarded as a semantic document search image. A QSI, in its turn, can be presented both traditionally (by a list of descriptors) and in an extended form by functional links between concepts (descriptors). Implementation of search engines in this case is therefore based on the use of operations on ontologies.

Union and intersections, as well as projections and scaling, were defined as the major operations on ontologies in [2].

Traditional graph-theoretical operations of union and intersection in ontologies are supplemented by possible comparison of the objects of initial ontologies using the operation of identity, not only in the functional but also in the conceptual and term systems. The union operation for ontologies, for instance, can be used for the design (and/OR sequential expansion) of ontologies of a subject area (SI) based on ontologies of certain scientific studies, or to design the general ontology of a group of researchers based on combining different points of view, etc. The result of the operation of intersection of ontologies enables one to establish the common and particular features in scientific works and can also be a sign of borrowing.

The operation of projection and scaling (expansion and specification) requires a prior design for the initial ontology of an aspect ontology or scaling ontology and is reduced to the operations of intersection or union of the designed and initial ontologies.

According to this definition, the operation of aspect presentation (examination and description) is specified by the functional system $S_f = \langle M_f, A_f, R_f, Z_f \rangle$, and the result of the operation is the intersection of the initial $O = \langle S_f, S_r, S_n, \equiv \rangle$ and the aspect $O = \langle S_f, S_n, \equiv \rangle$ ontologies: $O_{proj} = O \cap O_r$.

The operation of projection is aimed at designing a subgraph of the functional system that reflects a "view" on the initial ontology according to some specified aspect. The aspect of reviewing (presenting) a publication (in this case, it is a scientific work) is seen as some ontology that fixes objects of interest and the links (relationships) between them.  

The result of the intersection of the aspect and initial ontology enables one to define the context of a specified aspect in a certain document and further consider this context from the point of view of implementing search engines, algorithms of automated classification/clustering, etc.

However, in designing an aspect ontology it should be borne in mind that the correct design is possible only if a functional relation can be presented by an arc that joins the nodes of the multigraph of the functional system or if a functional relation are not used at all (the multigraph of the functional system is empty and the aspect is specified by a set of sign descriptions).

The operation of scaling (expansion or specification) enables one to change the level of abstraction of the presentation of a scientific study based on subsumption relation of the links of the conceptual ontology system.

In order to describe the operation of scaling (expansion or specification) of an ontology for the initial ontology $O = \langle S_f, S_r, S_n, \equiv \rangle$ the ontology of scaling $O_m = \langle S_f, S_r, S_n \rangle$ is defined and the operation of scaling is reduced to designing the ontology $O \cup O_m$. The ontology of scaling should contain sign descriptions of objects that are chosen to be scaled and their conceptual trees. The conceptual tree of an object is a fragment of the graph of the conceptual system with the root that is specified by the object. Generic links (like "bottom" ones) are replaced by the functional relationship "to be a part/particular case."

After union of the initial ontology and scaling ontology the links "to be a part/particular case" are reduced to the final (in the case of expansion) or initial (in the case of specification) object.

The operation of scaling can be applied to the initial ontology if there is no scaling ontology (e.g., if scaling objects were not matched with the conceptual system). In this case the operation of extension/specification comes down to reducing the relations "to be a part/particular case" in the initial ontology.

Scaling of ontologies can be used to reduce them to a single conceptual context before making use of the operations of union, intersection and projection.

2. EXAMPLES OF USING PROJECTION AND SCALING

Let us look at the operation of aspect presentation where a QSI is used in order to design the aspect. The following cases are possible.

1. A QSI is specified traditionally as a list of descriptors, which is conformed to when the aspect is specified by a set of sign descriptions.


\[ \text{AUTOMATIC DOCUMENTATION AND MATHEMATICAL LINGUISTICS} \quad \text{Vol. 47} \quad \text{No. 2} \quad 2013 \]