Copyright © 2024 by Cherkas Global University

Published in the USA
European Journal of Technology and Design
Issued since 2013.
E-ISSN: 2310-3450
2024. 12(1): 10-15

DOI: 10.13187/ejtd.2024.1.10
https://ejtd.cherkasgu.press

Ontological Models of Information Retrieval

Nikita S. Kurdyukov [a], [*]

[a] Department of IPPI, Institute of Information Technologies, Russian Technological University (RTU MIREA), Moscow, Russian Federation

Abstract

The article explores ontological models. A special type of models is considered. related to information retrieval. The ontological model of information retrieval is a specific model. The information design model is the closest model to an ontological model from a number of information models. The ontological model of information retrieval is generalized and allows for information uncertainty. The connection between the semantic model and the ontological model is shown. The semantic model of information retrieval complements the ontological model. Semantic proximity is a mandatory component of the ontological model. The article describes three methods for forming ontological models in information retrieval: indicator method, probabilistic method, fuzzy method. It is shown that logical weight is only an indicator and a qualitative characteristic, while probabilistic weight is a quantitative indicator. The article introduces several types of weighting coefficients for ontological models in information retrieval. The article introduces the definition of an ontological model of information retrieval. Three key indicators of the ontological model are described. It is shown that the user's information needs in many cases are unclear, uncertain and depend on the individual characteristics of the user. The article introduces the concept of ontological proximity. The article shows the difference between contextual metadata and contextual metamodels. The article introduces the concept of direct and contextual information resource in information retrieval. The difference between these resources is shown. The contextual information resource is associated with the ontological model. The types of relationships for direct and contextual search results are shown. The principles of forming a semantic proximity graph, which is used in ontological models, are described.

Keywords: information set, morphological search, semantic search, ontological search, content.

1. Introduction

The term “ontological models” is widely used in different directions. One of these areas is information retrieval (Kurdukov, 2023; Vallet et al., 2005). Ontological models mainly use formal descriptions. Semantic Web, which uses ontological models, is used to search networks. The Semantic Web (Hitzler, 2021) uses a number of specialized languages: RDF Schema, Ontology Web Language, Resource Description Framework and others. Widespread work is being done to create tools for working with ontological models. All this emphasizes the relevance of the study of ontological models. In many works on the research and application of ontologies, the concept of

[*] Corresponding author
E-mail addresses: nskurdyukov@gmail.com (N.S. Kurdyukov)

ontology is given vaguely. Ontology in computer science is significantly different from ontology in philosophy. Conceptually they are the same. But morphologically they differ. Let us recall that ontology (Guarino et al., 2009) is usually called a formal conceptual description of a field of knowledge. Such a schema typically includes data structures, object classes, relationships and connections, rules and theorems, restrictions) adopted in this area. Ontologies are used in design and modeling. In the information field, ontology is a form of representing knowledge about reality. However, in practice, especially in the field of programming, ontology is reduced to private information or other models. The concept of universality disappears for such models. However, we can talk about a mature ontological approach (Falbo et al., 2002). This information retrieval approach uses the following concepts: semantic environment (Tsvetkov, 2014a), semantic proximity (Gadasin et al., 2022), ontological model, formal logic, semantic frames, semantic classification, intelligent annotation, semantic graph, hyper spatial analogue of language, latent semantic analysis, information-cognitive semantics (Tsvetkov, 2016) and others.

2. Results and discussion

Features of the formation of ontological models

Methodologically, there are three approaches to the formation of ontological models: indicator or logical, probabilistic, fuzzy.

The indicator or logical approach to the formation of ontological models is based on the use of formal logic. He uses the concept of a logical predicate. The "disadvantage" in this approach is that it uses boolean variables. Logical variables have two oppositional values 0 and 1. As an indication of the presence or absence of an ontology, this approach makes sense. It is not suitable as a tool for describing ontology because it simplifies reality. For example, a picture of reality (image) can be described using black and white pixels. She will be rude. This picture corresponds to a binary black and white image. If you use gray halftones, you can get a gray halftone image. This is a more accurate depiction of reality. If you use colors and saturation, you can get a color photo that most accurately conveys the image of reality. A logical description using two values 0 and 1 is a black and white image of reality. This description is quite suitable for programming and computing, but is not suitable for describing the picture of the world (Heidegger, 1977; Tsvetkov, 2014b, Lazier; 2011).

The probabilistic approach to the formation of ontological models is based on the use of probability theory and is analogous to the formation of a halftone and color image of reality

The fuzzy approach to the formation of ontological models is based on taking into account fuzziness in the initial situation and the application of the theory of fuzzy sets. Its analogue is a photograph of a moving object with a blurred image. An example of the logical approach can be demonstrated with an example. The ontological model, denoted O, is given as a – tuple model

$$O= < C, P, I, L, T> \quad (1)$$

the parameters of model (1) are interpreted as follows: C – set of concepts; P – set of properties. Properties are expressed by two-place or one-place predicates. More precisely, a property is a one-place predicate, and a two-place predicate is a relation. However, a relation can be thought of as a property. Parameter I is a set of concept instances. The parameter L is a set of concept values and property values. The parameter T specifies the order on C and P. For the indication case, integer values of the weights are introduced (Tsvetkov, 2014c).

In accordance with parameter I, the semantic weight p is introduced. Semantic weight $p \in [0, 1]$ specifies the semantic proximity for the subject and object of the statement (relationship). This is where indicator properties appear. A boolean variable has the value 0 or 1 and has no values in between. Therefore, this approach suggests that intimacy either exists or does not exist. This model does not evaluate the degree or level of closeness. Analysis of this example shows a typical error in the use of logical values. Logical variables do not give the degree or level of strength of connections, but only state their presence or absence.

In our opinion, this drawback is eliminated by the probabilistic model. The weight in which has many values on the real interval [0-1], including the boundaries of the interval. In this model, two qualitative logical values “yes”, “no” are replaced by a set of quantitative values from the interval [0-1]. The quantitative value characterizes the closeness of belonging as it approaches one and the weakness of the connection as it approaches 0.

Probabilistic ontology model

To eliminate the shortcomings of the logical model, we propose an ontological model (OM), similar in structure to model (1)

OM <T, P (Re, Pr), MMT, SPV, Or >, (2)

In expression (2) T is a set of terms (signs); P – set of predicates (Re – relations, Pr – properties); MMT - set of meanings of terms; SPV – set of property values; Or is a partial order on the set T and P.

Using a set of predicates P, ontologies can describe various relationships between terms, their meanings and properties. Relationships are defined using simple statements

{s, Re, o*} (3).

In expression (3) s is the subject of the statement, o* is the object of the statement,

Re $\in$ P is a predicate of the OM ontology. Let us define a set of characteristic weights.

Any property Pr $\in$ P can be given a probabilistic morphological weight pm $\in$ [0, 1], which specifies the morphological proximity between the subject and object of the statement. For pm = 1 – there is a complete morphological correspondence, for pm =0 there is a complete morphological discrepancy

Let us introduce the concept of a set of information search results RIR. r is a private search result, pat is a search query.

Pat$\to$r (4)

Expression (judgment) (4) means that the request entails the appearance of a particular result of the request. We believe that any result of an information search r $\in$ RIR can be given an interpretative weight rv $\in$ [0, 1], which specifies the interpretive proximity between the request and the interpretation of the search result. For rv =1, the request and the result of the information search are fully interpretable. When rv =0, the result of the information search is completely uninterpretable. When rv = 0.5, the result of the information search is half interpreted.

We believe that any result of an information search r $\in$ RIR can be given an information weight ri $\in$ [0, 1], which specifies the probabilistic proximity between the information need (IN) and the search result. For ri = 1, the result fully satisfies the information need. This is a state of persistence. When ri =0 , the result does not completely satisfy the information need. This is a state of uncertainty. When ri = 0.5, the result of the information search satisfies the information need by half. You can conditionally estimate ri = 0.55 – there is formal relevance, ri = 0.8 – relevance.

The pattern can be formed as a set of terms, as a compound predicate, or as a certain semantic function of the values pat = sf (e). Since patterns are compiled by different people, for the same information need they can differ due to the cognitive and intellectual factors of the individual.

Any information search pattern pat $\in$ T can be given a search weight pati $\in$ [0, 1], which specifies the proximity between the information need and the pattern. For pati = 1, the pattern fully corresponds to the information need. For pati = 0.8, the pattern partially corresponds to the information need. For pati < 0.5, the pattern does not correspond to the information need. Here we can draw a parallel with correlative analysis

Features of the ontological model

Let us define the ontological model of information retrieval as the conceptual correspondence of search results to the semantic information needs of the user. It is advisable to analyze the features of this model.

The key indicator of the ontological model is the parameter "user information needs" (UIN or IN). By its formal name, IN is an information characteristic. The conditions for the formation of IN are associated with three factors:

information uncertainty (IU) (Ferracuti, 2022) in which the user finds himself;

user intelligence (individual's intelligence – II) (Wang et al., 2011);

user cognitive resources (cognitive resource – CR) (Christensen et al., 2020).

In fact, IN is informational and cognitive needs. Sometimes information search, especially in scientific research, is carried out using intuition. Intuition is characterized by information uncertainty and vaguely expressed formalism. It follows that the user's information needs in many cases are unclear, uncertain and depend on the individual characteristics of the user.

The second key indicator of the ontological model is the "conceptual correspondence" (CC) parameter. Conceptual correspondence is a generalized characteristic that allows for multiple interpretations. The basis for checking the truth of conceptual correspondence is comparative analysis. Conceptual correspondence is agreement on the most important parameters and inconsistency on less important parameters. Conceptual correspondence is always not a complete parametric correspondence.

The third key indicator of the ontological model is the "semantic needs" (SS) parameter (Veksler et al., 2007). This parameter is introduced as an alternative to morphological matching. The same form does not mean the same content. The same words can have different meanings. The content of the search is more important. than the form of presentation. The form of representation can be: a constant (yes/no, correct/incorrect), a relation (including modal relations), an analytical formula, a rule. output or table. It is not the form of presentation that is important, but the content of the model in relation to the user's request. Ontological needs are more general than semantic needs.

The search result contains information models and information resources. Let's call a direct information resource a resource described by a direct interpretation. A direct resource is created by the relationship between the parameters of the request property (a) and the result property (b)

$$\{a, Re, b\} \quad (5).$$

For example:

$$a=b \quad (6);$$

$$a \approx b; \quad (7)$$

$$a \neq b \quad (8).$$

Expressions (6), (7), (8) hold for formal parameters and for parameter values. Formal parameters are important for ontology. Parameter values are important for semantics. Expressions (6), (7), (8) are the most important relationships for evaluating search results. A model that consists only of properties and their values is parametric. If most of the search result parameters correspond to relations (6), (7), then such a result is relevant for the parametric model. Relation (7) characterizes the state of uncertainty. If most of the search result parameters correspond to relations (8), then such a result is not relevant.

An alternative to the direct resource and direct model is the context resource and context model. A model that consists only of relationships. is contextual. A model that consists of relationships and parameters is called mixed. Let's call it a contextual information resource, a resource described by statements.

$$\{ra, Re, b\} \quad (9).$$

In expression (9) ra is the subject of the statement, b is the object of the statement, $Re \in P$ is a predicate of the ontology OM. Let's call a context metamodel a set of statements. As a rule, this is a postfix metamodel (Tsvetkov et al., 2020)

$$MMk= \{ti=< ra, Rei, bi>\}, \quad (10)$$

where i=1 n – number of relations

$$b \in T \cup MMT \cup SPV \quad (11)$$

An example of a simple relationship is

$$a=d; a=10; a>c; a<H; a \approx k. \quad (12)$$

RIR contextual metadata is a set of weighted interpretations

$$Md= \{ti=< ra, Rei, bi, wi>\}, \quad (13)$$

In expression (13) w is the weight. The difference between (10) and (13) is that in the first case the formal parameters are studied, in the second the values of the parameters. Ontological proximity is associated with semantic proximity, with interpretive proximity. The ontological similarity of the parameters is assessed by relations (6) (7). Let L1 be the number of parameters corresponding to relations (6) (7), and L2 be the number of parameters corresponding to relations (8), If

$$L1>L2 \quad (14)$$

Then there is ontological similarity in parameters. Semantic proximity (semantic similarity) is determined by attributive characteristics Pr and contextual characteristics Re.

Let Sim(a, в) be the semantic proximity between (elements, resources)

a and b, where $a, b \in T \cup MMT \cup P$.

One method for calculating Sim(a, b) is based on graph theory. This method involves constructing an undirected graph SG from all relations (10), (13). The graph SG is formed in accordance with the principles:

use relations that have weighting coefficients other than zero ($w \in 0$);

the graph has subjects and objects of relations as vertices, and the edges of the graph are relations. The edges have weights w;

the graph admits an inverse relation, which replaces one relation and not two auxiliary ones;

the graph admits a symmetric relation that adds two edges with equal weights to the graph.

the graph has a route or PATH (a, b) as a set of edges connecting vertices a and b, taking into account their direction.

Semantic proximity is calculated as the optimal PATH (a, b).

In this case, the value Sim(a, b) between these vertices is calculated as:

$$Sim\ (a, b) = \min\ (Sim_{PATH}\ (a, b)),(15)$$

The value of semantic proximity is determined by the formula:

$$Sim_{PATH}\ (a, b)= \prod_{i=1}^{n} w\ (16)$$

Thus, the calculation of weights determines semantic proximity. Semantic proximity allows us to assess ontological proximity. However, these concepts are not equivalent. An ontology is a conceptual model and aims to use qualitative features and categories. The semantic model uses quantitative estimates of parameters.

3. Conclusion

The ontological model of information retrieval is a special model, unlike most information parametric models. Of the information models, the model of information design is closest to her. The ontological model includes a double environment – ontological and semantic. The ontological model primarily includes qualitative assessments and secondarily quantitative ones. The semantic model first of all includes quantitative estimates of parameter values and secondly qualitative ones. The semantic model works primarily with meanings. Ontological works with the qualities of features and meanings. The ontological model of information retrieval uses semantic proximity and complements it with ontological proximity. Semantic proximity is determined by parameter values. Ontological proximity is determined by qualitative characteristics. Currently, most ontologies are based on semantics. Therefore, information retrieval is actually about semantic correspondence rather than ontological correspondence. In our opinion, reducing the search for ontologies to semantic proximity is a narrowing of the concept of ontology. In our opinion, a promising direction for constructing ontological models of information search is correspondence theory (Bode et al., 2020).

References

Bode et al., 2020 – *Bode, T., Weißenfels, C., Wriggers, P.* (2020). Peridynamic Petrov–Galerkin method: a generalization of the peridynamic theory of correspondence materials. *Computer Methods in Applied Mechanics and Engineering*. 358: 112636.

Christensen et al., 2020 – *Christensen, J., Aarøe, L., Baekgaard, M., Herd, P., Moynihan, D.P.* (2020). Human capital and administrative burden: The role of cognitive resources in citizen-state interactions. *Public Administration Review*. 80(1): 127-136.

Falbo et al., 2002 – *Falbo, R.A., Guizzardi, G., Duarte, K.C.* (2002). An ontological approach to domain engineering. Proceedings of the 14th international conference on Software engineering and knowledge engineering. Pp. 351-358.

Ferracuti, 2022 – *Ferracuti, E.* (2022). Information uncertainty and organizational design. *Journal of Accounting and Economics*. 74(1): 101493.

Gadasin et al., 2022 – *Gadasin, D.V., Shvedov, A.V., Vakurin, I.S.* (2022). Determination of Semantic Proximity of Natural Language Terms for Subsequent Neural Network Training. *2022 Systems of Signals Generating and Processing in the Field of on Board Communications*. – IEEE, 2022. Pp. 1-5.

Guarinoet al., 2009 – *Guarino, N., Oberle, D., Staab, S.* (2009). What is an ontology? Handbook on ontologies. Pp. 1-17.

Heidegger, 1977 – *Heidegger, M.* (1977). The age of the world picture. In Science and the Quest for Reality (pp. 70-88). London: Palgrave Macmillan UK.

Hitzler, 2021 – *Hitzler, P.* (2021). A review of the semantic web field. *Communications of the ACM*. 64(2): 76-83.

Kurdukov, 2023 – *Kurdukov, N.S.* (2023). Ontologies in Information Retrieval. *European Journal of Technology and Design*. 11(1): 9-14.

Lazier, 2011 – *Lazier, B.* (2011). Earthrise; or, the globalization of the world picture. *The American Historical Review*. 116(3): 602-630.

Tsvetkov et al., 2020 – *Tsvetkov, V.Ya., Shaitura, S.V., Minitaeva, A.M., Feoktistova, V.M., Kozhaev, Yu.P., Belyu, L.P.* (2020). Metamodelling in the information field. *Amazonia Investiga*. 9(25): 395-402.

Tsvetkov, 2014a – *Tsvetkov, V.Ya*. (2014). The Semantic environment of information units. *European researcher*. 6-1 (76): 1059-1065.

Tsvetkov, 2014b – *Tsvetkov, V.Ya*. (2014). Worldview Model as the Result of Education. *World Applied Sciences Journal*. 31 (2): 211-215.

Tsvetkov, 2014c – *Tsvetkov, V.Ya*. (2014). Integer Coordinates as an Nanotechnological Instrument. *Nanotechnology Research and Practice*. 4(4): 230-236.

Tsvetkov, 2016 – *Tsvetkov, V.Ya*. (2016). Information-Cognitive Semantics. *European Journal of Technology and Design*. 4(10): 164-175.

Vallet et al., 2005 – *Vallet, D., Fernández, M., Castells, P*. (2005). An ontology-based information retrieval model. The Semantic Web: Research and Applications: Second European Semantic Web Conference, ESWC 2005, Heraklion, Crete, Greece, May 29–June 1, 2005. Proceedings 2. Springer Berlin Heidelberg. Pp. 455-470.

Veksler et al., 2007 – *Veksler, V.D., Grintsvayg, A., Lindsey, R., Gray, W.D*. (2007). A proxy for all your semantic needs. *Proceedings of the Annual Meeting of the Cognitive Science Society*. 29(29).

Wang et al., 2011 – *Wang, L., Song, M., Jiang, T., Zhang, Y., Yu, C*. (2011). Regional homogeneity of the resting-state brain activity correlates with individual intelligence. *Neuroscience letters*. 488(3): 275-278.