

# Intelligent information retrieval based on Ontology

Yuye Chen

Electronic Information Engineering, Chongqing University of Technology, Chongqing, Chongqing, 401135, China  
cyy\_official\_2000@163.com

**Abstract:** Traditional search engines can no longer meet the increasing amount of information in today's world, and often face problems such as blurred results and long search times for some queries. And the addition of ontology makes information retrieval more intelligent, so that information data between various fields can be shared, to achieve the purpose of getting the answers users want quickly and accurately. This paper firstly introduces the definition of ontology and analyzes the traditional information retrieval system, and finally summarizes some experimental results in recent years in ontology-based intelligent information retrieval and reviews them to get a more efficient and accurate information retrieval system to achieve the effect of improving the accuracy of search engines.

**Keywords:** Semantic; Intelligent information retrieval; Ontology.

## 1. Introduction

In recent years, the Internet has continued to permeate every aspect of human life. Its addition has not only improved people's quality of life, but also enriched their entertainment time. In order to find accurate information in the vast amount of information, search engines are particularly important in human life as an information search tool. However, with more and more information pouring into the Internet, information retrieval is facing a great challenge. Information retrieval techniques based on keyword matching and full-text search techniques face problems such as large amount of information, too much information in query results, confusion in query results, mismatched content of search results, and inability to identify multiple meanings of words. These problems are inevitable in search engines using keyword matching and full-text retrieval techniques. An ontology is a structured set of terms that will be expanded and extended outward after mutual intelligent learning by machine to machine, thus containing more information that may be used to achieve crossover in different domains. Adding the concept of ontology to an intelligent information retrieval system will allow users to get more accurate results and even make the whole retrieval process more efficient.

## 2. About ontology

### 2.1. Semantic Web related concepts and architecture

The concept of the Semantic Web proposed by BERNERS-LEE, which consists mainly of documents for human reading, including data and information for computer operation. The Semantic Web is a web of actionable information -- information derived from data through semantic theory that is used to interpret symbols. Semantic theory provides an interpretation of "meaning," where logical connections of terms establish interoperability between systems. In order to make the search engine's results more accurate, the semantic web technology is introduced based on the original web, which may make today's intelligent information retrieval system more improved.

BERNERS-LEE also proposed an architecture for the Semantic Web, as shown in Figure 1.

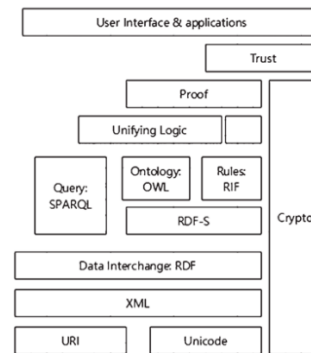


Figure 1. Architecture for the Semantic Web

The Semantic Web architecture is explained as follows:

- (1) Unicode and URI are the base layers. URI is used to identify resources on the network. And the identity of each resource must be unique and unique. Unicode is used to encode resources in any language around the world, enabling the Semantic Web to support diverse cultural elements.
- (2) XML/XMLS. Is used to describe the content and structure of resources. XML is an important branch of document construction. In terms of content, a URI is the identity of the discovery and identification entity that makes up the NS (namespace). XML documents enable language-wide control through XML Schema.
- (3) RDF+ RDF Schema data layer. This layer applies the existing rules to all types of resources and their corresponding types. Give a description. Among them, RDF supports simple semantic description, and RDF Schema is an optimization and improvement of RDF, which provides a method to define classes and attributes, realizes the dual constraints of classes and attributes, and provides a method to detect constraints.
- (4) Ontology vocabulary layer. Ontology vocabulary layer can accurately express all kinds of information and its internal relations in a specific domain according to the abstract representation method. The types of information in the domain and their internal connections.

- (5) Logic layer. In order to realize the function of reasoning, the logic layer provides a powerful logic language.
- (6) Proof layer. The proof layer analyzes and verifies the program of the language specialty and carries out corresponding work according to the actual scene of the language.
- (7) Trust layer. The purpose of the trust layer is to provide secure and secure information transfer.

## 2.2. Overview and applications of ontology

### 2.2.1. Concepts of ontology

In the AI community, ontology is defined as the set of basic terms and relations between terms that constitute the relevant domain, and the definition of rules that specify the extensions of these terms using these terms and relations, a definition proposed as early as 1991 by Neches et al. .The goal of Ontology is to capture the knowledge of the relevant domain, provide a common understanding of the knowledge of the domain, identify commonly accepted vocabularies within the domain, and give clear definitions of these vocabularies (terms) and the interrelationships between the vocabularies in terms of formal patterns at different levels. So, an ontology is much like a list of vocabulary that can be shared.

RDF describes resource objects based on the standard syntax of XML and can express certain semantics, so that machines can understand the description information. However, due to RDF(S) limitations of its expressiveness, it cannot express the relationships between words. To further improve the description capability of Web information resources, Ontology, which originally belongs to the field of philosophy, is applied to the field of computer science, especially the semantic Web. The ontology is very suitable for describing concepts and relationships between concepts.

At present, ontology plays an important role in improving the accuracy of information retrieval. By using ontology, more accurate query results can be obtained in semantic retrieval, instead of just collecting some web pages with keyword information. In order to make up for the shortcomings of incomplete information query results, ontology-based semantic retrieval can query the hidden information contained in the information through the inference system, so that the query results are more consistent with the user's questions.

### 2.2.2. Composition of Ontology

In the application of network resources, ontology analyzes specific terms in various professional fields, and builds conceptual models by classifying knowledge and finding specific relationships between information.

Common components of ontologies can be seen in Figure 2.

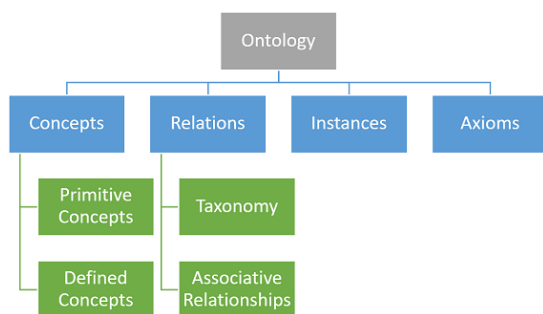


Figure 2. Components of ontology

### 2.2.3. Construction of Ontology

In different domains, the elements used are different, and the different types of projects will affect the construction of the ontology. Therefore, there is no fixed standard in the construction process of ontology. Only by combining the analysis of its own requirements with the elements in the research field, can an appropriate ontology be constructed. Among them, the construction criterion proposed by Gruber is the most widely used, which fully demonstrates what it means to build an ontology according to the current situation of a specific domain. Each researcher's research field is different, resulting in different conditions when constructing ontology. Therefore, they use methods with different characteristics to construct ontology, so that they can make choices according to their own situation. Among them, there are ontology construction methods that named METHONTOLOGY, TOVE and IDEF5 and so on are highly respected.

## 3. Existing Researches on Intelligent Information Retrieval Based on Ontology

In such an era of big data, the use of formal representations is very important, but such data is often complex. In this regard, a certain person emphasizes the importance of using clear and formal representations between objects in order to give the right meaning to data relevance. In the case of multimedia, Antonio M. Rinaldi and Cristiano Russo's research proposes a formal model that uses a semantic approach to formalize multimedia big data to implement an intelligent information retrieval system that allows better communication between objects. This model combines a top-level ontology model and a graph model represented by a labeled, attribute-based structure to consider semantic, linguistic and multimedia data simultaneously. And to facilitate machine processing, the ontology is developed using OWL. The model proposed in the article has been implemented in a NoSQL graph database populated by different knowledge sources, but the limitations of the model are that scalability and efficiency for large amounts of data are not considered.

In B. Selvalakshmi and M. Subramaniam's paper, a new semantic information retrieval system is proposed, and researchers use feature selection and classification to improve relevance scores. In the research, a new intelligent fuzzy rough set-based feature selection algorithm and an intelligent ontology and Latent Dirichlet Allocation based semantic information retrieval algorithm were used to improve the semantic information retrieval system, Not only are they able to improve relevance scores and retrieval speed, but they also have the ability to handle big data. In order to perform semantic information retrieval using an ontology matching method, researchers propose a new algorithm for preprocessing and LDA-based document classification. For the preprocessing algorithm, its role is to remove irrelevant and noisy information from documents to form an ontology, and subsequently, a new intelligent fuzzy rough set-based feature selection algorithm and an intelligent ontology and Latent Dirichlet Allocation based semantic information retrieval algorithm. The combined effect of these two algorithms reduces the retrieval time, and increases the semantic relevance by applying the ontology, so as to achieve the effect of fast and accurate retrieval.

## 4. Conclusion

Combining the two new information retrieval systems mentioned above, a more efficient and accurate framework for the search system can be derived. First, from the second retrieval system, the reader can fully feel the importance of pre-processing, which plays a role in information filtering during the formation of ontologies. Second, making good use of the top-level ontology model mentioned in the first retrieval system and a graph represented by a labeled, attribute-based structure enables the system to consider semantic, linguistic, and multimedia data at the same time, improving the accuracy of retrieval. Finally, an intelligent feature selection algorithm based on fuzzy rough sets and an ontology and Latent Dirichlet Allocation-based semantic information retrieval algorithm can be used in the system to improve the semantic information retrieval system, which not only improve the relevance score and retrieval speed, but also have the ability to handle large data.

## References

- [1] Jiang, J., Wang, Z., Liu, C., Tan, Z., & Min, L. (2010) In: The technology of intelligent information retrieval based on the semantic web. J. International Conference on Signal Processing Systems. Dalian. 1664-1667.
- [2] Berners-Lee, T., Hendler, J., Lassila, O. (2001) The Semantic Web in Scientific American. Lecture notes in computer science. 284(5): 34-43
- [3] Wei, W. (2001) Research on Semantic Web and its Core Technology. China Computer & Communication, 33(11): 6-9.
- [4] Zhang, T. (2020) Research on Ontology Construction and Semantic Retrieval in Tourism Field. Thesis of Taiyuan University of Technology, 07.
- [5] Neches, R., Fikes, R.E., Gruber, T.R., et al. (1991) Enabling Technology for Knowledge Sharing. AI Magazine, 12(3): 36-56.
- [6] Zammali, S., Arour, K., Bouzeghoub, A. (2013) Using ontologies to build tested for peer-to-peer information retrieval systems. In: 27th international conference on advanced information networking and applications. Barcelona. 1033-1040.
- [7] Wikipedia. (2022) Ontology (information\_science). [EB/OL] [https://en.wikipedia.org/wiki/Ontology\\_\(information\\_science\)](https://en.wikipedia.org/wiki/Ontology_(information_science)).
- [8] Chhavi, V., Sapna, S., Sunil, K.K. (2017) Unified Ontology for Data Integration for Tourism Sector. In: International Conference on Infocom Technologies and Unmanned Systems. Dubai. 52-156
- [9] Rinaldi, A.M., Russo, C. A. (2018) Semantic-based Model to Represent Multimedia Big Data. In: Proceedings of the 10TH International Conference on Management of Digital Ecosystems. Online. 31-38.
- [10] Selvalakshmi, B., Subramaniam, M. (2019) Intelligent ontology based semantic information retrieval using feature selection and classification. Cluster Compute 22 (Suppl 5), 12871-12881.