MJRET

Multidisciplinary Journal of Research in Engineering and Technology

Journal homepage: www.mjret.in

ISSN:2348 - 6953

Open Access

REVIEW ON ONTOLOGY BASED TECHNIQUES IN INFORMATION RETRIEVAL SYSTEMS

Komal Shivaji Mule[1], Prof. Arti Waghmare[2]
Computer Dept.
[12]Dr. D. Y. Patil School Of Engineering & Technology, (Affiliated to Savitribai Phule Pune University)
Pune, India
[1]physicskomal7@gmail.com, [2]arti.waghmare@dypic.in

Abstract: Information Retrieval [IR] is a method used for searching documents, searching information contained by documents, and for searching metadata about documents, and also for relational databases searching and the World Wide Web (WWW). There stands similarity in the usage of the terms such as data retrieval and document retrieval as well as information retrieval and text retrieval, but each of these has its own body of literature survey, philosophy, and technologies. The ontologies have been established for local information sharing and are widely used as a means for conceptually structuring domains of interest. The aim of semantic web research is to permit the enormous range of web-accessible information and services to be more efficiently exploited by either humans or automated tools. To assist this process, RDF and OWL have been established as standard formats for the sharing and integration of data and knowledge. Different methodologies are there for information retrieval based on ontology from which review of some methodologies is presented in this paper.

Keywords: *Information Retrieval, Ontological Concepts,indexing technique*

1. INTRODUCTION

1.1. What is Information Retrieval

Information over the World Wide Web is escalating day by day. Information is in variety of formats. Retrieving the relevant data is getting difficult for clients [7]. Information retrieval is the art of retrieving information from the available records. During last decade the field of Information Retrieval is dealt not only by information specialists but also by ordinary folks. Information specialists and ordinary folks alike are beginning to inundate in information. The next generation of Information retrieval provides vividly enhanced capabilities to learn, anticipate, and succor daily information gathering, analysis and management needs of individuals and groups over time span measured in decades.

Traditional Information retrieval is based on keyword, syntax matching. This technique gives simple, quick, easy, performance based on simple syntax matching. The major loophole is lack of finding context from associated content, which require effective knowledge expression, processing and understanding. Most important challenge in Information retrieval out of many is finding the context from associated content. This focuses on presenting domain information semantically and representing knowledge in comprehensible fashion. The mentioned challenge demands an Information retrieval based on semantic where semantic Information retrieval is actually semantic relations reflected using ontology.

In Information retrieval systems challenging composition of science and also engineering in various interesting unresolved issues can span different areas of CS. The areas like system architecture, distributed structures, different algorithms, compression, information retrieval, UI, machine learning mechanism, etc. Scale very larger than various other systems. Engineering issues are equivalent to the product of parameters such as document indexing, queries, query latency, index freshness or update rate, information kept back regarding each document, scoring complexity or cost or retrieval algorithms. With the budding World Wide Web the use of metadata will turn out to be necessary. In conceptual querying a set of concepts are surveyed. The over-all idea is to restrict universal knowledge ontology to specified set of conceptions, spread out this with relations and related concepts, and thus providing a structure called instantiated ontology, this is for navigation and further investigation of the concepts. The set of documents can be conceptually investigate by extracting the set of concepts appearing in the documents and also by providing means for navigation and retrieval contained by the set of extracted concepts [1].

One of the methods of information retrieval is ontology. In next section i.e. section 2 it is explained that “what is ontology”. In section 3 Ontology based techniques for information retrieval are described.

2. ONTOLOGY

In philosophy, ontology lessons the nature of being and existence. The word ‘ontology’ is imitative from the Greek words “onto”, which wealth being, and “logia”, which wealth written or spoken discourse. In computer science (CS) and material science, reuse of knowledge is facilitated by the use of obvious ontology, as contrasting to implicit ontology, i.e., information coded into software systems. Henceforth, suitable ontology languages are desired to realize explicit ontologies with respect to three vital aspects: 1) Conceptualization; 2) Vocabulary; 3) Axiomatization [13].

Actually, Semantic Web is an extension of the current one in which it represents more meaningful information for humans and computers. It enables the explanation of contents and services in machine-readable form, and it also enables annotating, publishing, discovering, advertising and combining services to be automatic. It was advanced based upon Ontology, which is considered as the strength of the Semantic Web. On other hands, the current Web is converted from being machine-readable to machine understandable. One purpose of the Web is to build a source of reference for data on several subjects, though the Semantic Web is projected to build a web of meaning. Ontology is the foundation of terminologies and effective communication on the Semantic Web. “Ontology offered explicit, formal condition of a shared conceptualization of a domain” [22].

Ontology contains concepts, relations, instances and axioms. Formal description of ontology is as shown in figure. 1 below.

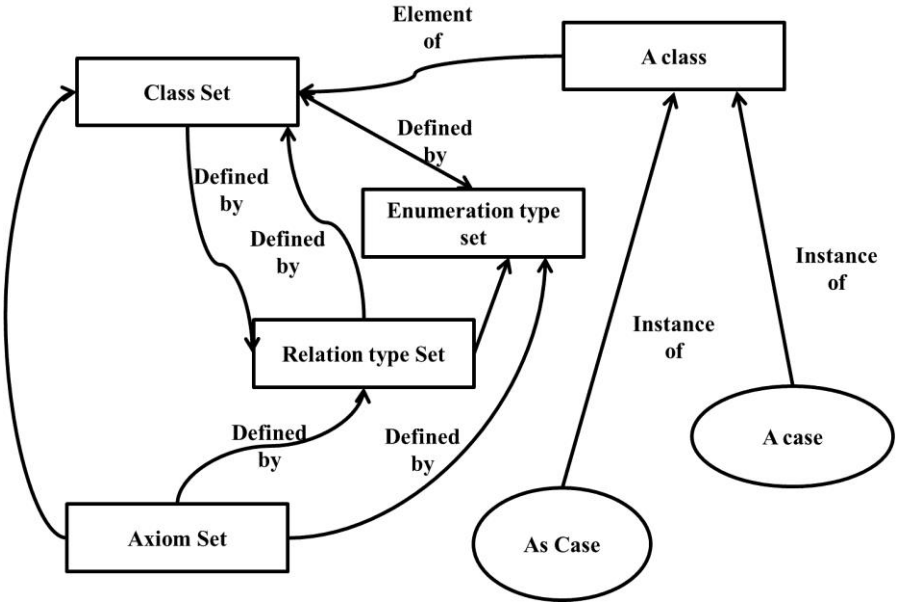

Fig.1 Ontology conceptual model

Concept is a group or a class entities or "things" in certain areas. Relation describes the relationship between concepts or between attributes properties. Concept can be defined as hierarchical form of classification system, the various concepts in the classification system linked by category relations. Relations can be described in different aspects characteristics by its property. Instance is the "things" by concept expressed. Ontology and its instance are referred to as Knowledge base. Axiom is used to express a concept or an instance constraint.

The languages which signify ontology are Resource development framework Schema (RDFs), Web Ontology Language (OWL), Ontology Interface Layer (OIL). The Resource Description Framework (RDF) is a structure that allows the encoding, interchange and reuse of metadata. RDF is a technique for data and metadata illustration. RDF statements comprises triples (Subject, Property, and Object). RDF data model is nothing but a directed graph whose nodes represents the subjects and objects and arcs represents the properties. RDF data model nodes are labeled by means of URIs describing resources or literals (i.e. strings or numbers) or can be unlabeled, called as blank nodes. Blank nodes are generally used to group properties. Edges are always represented by URIs showing a relationship between the subject and the object. Ontology has a good ability of concept hierarchy along with logical inference. Thus our efforts would be focusing on various ontology based information retrieval techniques which will be foundation for our research work based on effective IR.

3. ONTOLOGY BASED TECHNIQUES FOR INFORMATION RETRIEVAL

In paper [7], Grau et. al present freshly started EU project 'Optique', which supports for a next generation of the well-known Ontology-Based Data Access (OBDA) approach to address the data access problem in big data. By using Ontology-Based Data Access data is accessed in less time complexity as well as cost for accessing data is reduced. Ontology-Based Data Access strives for query formulation that optimizes the query to manipulate the results in less amount of time.

In this paper [9] Author has proposed a general framework for word sense disambiguation by using knowledge latent in Wikipedia. Specifically, they exploit the rich and growing Wikipedia corpus in order to achieve a large and robust knowledge repository consisting of key phrases and their associated candidate topics. Key phrases mainly derivative from Wikipedia article titles and anchor texts associated with wiki links. The disambiguation of a given key phrase is based on both the ordinariness of candidate topic and context-dependent similarity where unnecessary (and potentially noisy) context information is pruned. But this is only for Wikipedia pages or articles. The technique of forming descriptors in this paper helps in forming the ontology due to its specific concept formation.

Paper [11] defines an aggregating models based on environment and a request method to evaluate document's significance annotated using concepts of ontology. Selection of documents is shown in a semantic map to provide graphical signals that make explicit to what degree they match the user's query; this interface(human or machine) favors a surplus interactive exploration of data corpus. Regeneration leads to optimization and mathematical questions but also raises significant issues regarding feedback to users to enable them to continue to understand the IRS procedure and productively interact with it.

The paper [12] discusses the improvement of a novel information illustration system addressed in the Semantic Web and Ontology. The Retrieval Status Values (RSVs) are calculated from a relationship measurement between the concepts of an ontology. The proposed technique in this paper is to break down the RSV computation into a three stage aggregation process. The new techniques varies from other representation techniques that are based on a more sophisticated semantic illustration of information, objectives to go ahead the documentation level and objectives deliberate to be unstated and handled by machine. The current work on an Ontology based information retrieval(OBIR) allowance that will let users reformulate their query over graphically selecting the documents they value and those in which they have no interest.

Paper [16] proposes Query Classification Algorithm (QCA). Classifier classifies user queries into the proposed categories for ranking purpose. Next to the query classification procedure, input query is labeled with one or more categories sorted according to their scores Domain ontology is used as a controlled vocabulary. The formation of domain ontology is also significant to the description and use of an enterprise architecture framework. The course of classification queries based on the ontology is offered to improve accuracy value for retrieving information.

A novel ontology-based approach to information retrieval (IR) is presented in [18]. The system based on a domain knowledge presentation schema in form of ontology. New properties registered inside the system are linked to concepts from this ontology. In this a way resources may be retrieved based on the associations and not only based on limited or exact term corresponding as the use of vector model believes. In this paper a package is developed with three different methods for document retrieval, they are 1) Vector representation, 2) Latent semantic indexing method (LSI), 3) Ontology-based method, used in the Webocrat system. After the experiments author finds that document retrieval method using ontology is giving better recall-precision. Author says that this technique can be enhanced by using sophisticated inference mechanism for finding analogous concepts in query. Figure 2 illustrates information retrieval using ontology [18].

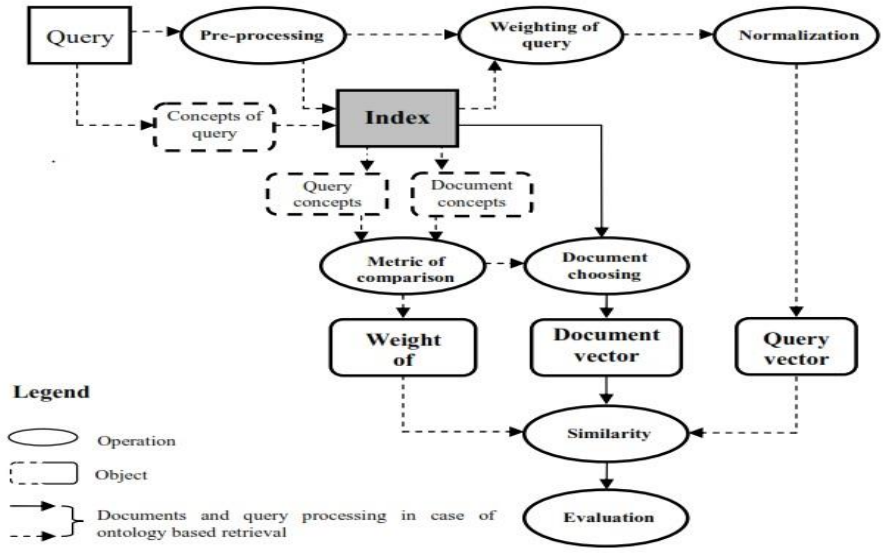

Fig.2 The process of ontology based document retrieval

Parameters / Reference	Support for Information retrieval	Use Of Ontology	Recall & precision	Clustering	Dictionary
Reference [1]	✓	✓	✓	✓	✓
Reference [7]	✓	✓	×	×	×
Reference [9]	✓	✓	✓	×	✓
Reference [11]	✓	✓	✓	×	✓
Reference [12]	✓	✓	✓	×	✓
Reference [16]	✓	✓	✓	×	✓
Reference [18]	✓	✓	✓	×	×

Table 1: Taxonomy Chart

4. CONCLUSION & FUTURE SCOPE

Data in today's world is increasing day by day. There are many techniques for retrieving data. This paper describes the review of ontological techniques used for information retrieval. Ontology has been proved one of the efficient techniques for information retrieval. Ontology overcomes the drawbacks of other Information retrieval system. In future more sophisticated Information retrieval systems can be designed to maintain the Precision and recall of search.

ACKNOWLEDGEMENT

I wish to express my sincere thanks to the guide Prof. Mrs. Arti Waghmare and Head of Department, Prof. Arti Mohanpurkar, as well as our principal Dr. Uttam Kalwane , also Grateful thanks to our PG Coordinator Prof. Roshani Raut and last but not least, the departmental staff members for their support.

REFERENCES

[1] *Rajeswari Mukesh, Sathish Kumar Penchala, and Anupama K. Ingale. Ontology Based Zone Indexing Using Information Retrieval Systems. S. Unnikrishnan, S. Surve, and D. Bhoir (Eds.): ICAC3 2013, CCIS 361, pp. 181–186, 2013.*

[2] *MyoMyo ThanNaing. Ontology-Based Web Query Classification for Research Paper Searching. International Journal of Innovations in Engineering and Technology February 2013*

[3] *Evgeny Kharlamov, Ernesto Jiménez-Ruiz, et. Al. Optique: Towards OBDA Systems for Industry 2013.*

[4] *E. Kharlamov, M. Giese, et al. Optique: Semantic Access to Big Data The Case of Norwegian Petroleum Directorate's FactPages 2013.*

[5] *Ahmet Soylu, Martin Giese, Ernesto Jimenez-Ruiz. OptiqueVQS – Towards an Ontology-based Visual Query System for Big Data October 28-31, 2013.*

[6] *Lovelyn Rose, S. and K.R. Chandran. Normalized Web Distance Based Web Query Classification, Journal of Computer Science 8 (5): 804-808, 2012.*

[7] *B. Cuenca Grau, M. Giese. Towards Query Formulation and Query-Driven Ontology Extensions in OBDA Systems 2011*

[8] *Chenliang Li, Aixin Sun, and Anwitaman Datta. A Generalized Method for Word Sense Disambiguation based on Wikipedia 2011.*

[9] *Olena Medelyan, Ian H. Witten and David Milne. Topic Indexing with Wikipedia 2011.*

[10] *Mohammad Mustafa Taye. Understanding Semantic Web and Ontologies: Theory and Applications. June 2010, ISSN 2151-9617*

[11] *He, G., An, L.: Ontology Language OWL Research Study. In: International Conference on Study Management and Service Science (MASS), pp. 1–4 (2011)*

[12] *Sylvie Ranwez, Vincent Ranwez, Mohameth-François Sy, Jacky Montmain, Michel Crampes. User Centered and Ontology Based Information Retrieval System for Life Sciences. 2010*

[13] *TroelsAndreasen, HenrikBulskov. Conceptual querying through ontologies. 2009 Elsevier*

[14] *Eduard Dragut, Fang Fang, Prasad Sistla, Clement Yu. Stop Word and Related Problems in Web Interface Integration. August 24-28, 2009.*

[15] *Almendros-Jimenez, J.M.: An RDF Query Language based on Logic Programming. In: 3rd Int. Workshop on Automated Specification and Verification of Web Systems, vol. 200, pp. 67–85. ScienceDirect (2008).*

[16] *Jes´us M. Almendros-Jim´enez. An RDF Query Language based on Logic Programming. Electronic Notes in Theoretical Computer Science 200 (2008) 67–85.*

[17] *Ian Horrocks. Ontologies and the Semantic Web 2008.*

[18] *Jan Parali, Ivan Kostial "A Document Retrieval Method Based On Ontology Associations"Udc:004.65*

[19] *Zhang, J. Ontology and the Semantic Web. Proceedings of the North American Symposium on Knowledge Organization. Vol. 1. (2007).*

[20] *Apache: Lucene. Retrieved October 2007 from http://lucene.apache.org.*

[21] *Li Ding, Pranam Kolari, Zhongli Ding, Sasikanth Avancha, Tim Finin, Anupam Joshi. Using Ontologies in the Semantic Web: A Survey. July, 2005*

[22] *Jan Paralic, Ivan Kostial. Ontology-based Information Retrieval 2003.*

[23] *Open GIS Consortium, Inc.: OpenGIS Web Map Service Implementation Specification. OpenGIS Project Document 01-068r3, Open GIS Consortium, Inc. (2002).*

[24] *B. He and K. Chang. Statistical schema matching across web query interfaces. In SIGMOD, 2003.*

[25] *Fischer Nilsson, J.: Concept descriptions for text search. In: Information Modeling and Knowledge Bases XIII, pp. 296–300. IOS Press (2002)*