

# Query Expansion Algorithm with Metadata Support for Ontology Matching

Ancy Abraham

PG Scholar

Department of Computer Science and Engineering  
Amal Jyothi College of Engineering, Kerala, India  
ancytabraham@gmail.com

Rubin Thottupuram

Head of the Department

Department of Computer Applications  
Amal Jyothi College of Engineering, Kerala, India  
rubinthottupuram@amaljyothi.ac.in

**Abstract**—To improve the precision and recall of current IR systems, search engines that work on meaning of data should be implemented. Semantic Web provides meaning to the data. Data is written in an ontology format. The core concept of semantic search engine is ontology mapping. Every information, documents, books etc available in the World Wide Web will be replaced by ontologies in the near future. So, for an efficient information retrieval, user provided search query must be expanded to perform matching. In this project, an algorithm to perform ontology matching is proposed.

**Keywords**-semantic web, semantic search, ontology matching, ontology expansion, similarity matching

\*\*\*\*\*

## I. INTRODUCTION

In our traditional web, information and documents are stored and retrieved on basis of keyword searching. If we are trying to find the humidity at New Jersey, our traditional web will provide a result of documents that relates both humidity and New Jersey. Humans have to analyze and find relevant information from the results. That is, the WWW is human understandable and it is a collection of documents. Here arises the importance of a web which works on the basis of meaning. i.e., we need a web of data rather than documents. Both human and machine should be able to use that data. A web which surrounds the syntax of documents should be replaced by web of meaning. Semantic Web is nothing but an extension of current web in which information is given a well defined meaning [1].

Semantic Web is proposed for efficient knowledge mining. To add meaning, data is written as an Ontology which currently uses OWL (Web Ontology Language). Ontology is the core concept in a Semantic Web. It is the latest form of XML, RDF, RDFS which we will see in the next section. Simply, ontology is the philosophical study of the nature of being, becoming, existence, or reality, as well as the basic categories of being and their relations [1]. To improve the precision and recall of current IR systems, search engines that work on meaning of data should be implemented. The core concept of semantic search engine is ontology mapping. Every information, documents, books etc available in the World Wide Web will be replaced by ontologies in the near future. So, for an efficient information retrieval, user provided search query must be expanded to perform matching. There are several methods for ontology reformulation. In this paper, instead of converting the query ontology into an OWL/RDF form, query string expansion is proposed. Also an algorithm to perform ontology matching by reducing time is developed.

## II. SEMANTIC WEB

Tim Berners-Lee, father of World-Wide-Web, defines semantic web as “an extension of the www in which data is given well-defined meaning, enabling computers and people to work in cooperation” [1]. The Semantic Web will give structure as well as meaning to web pages, creating an environment where software agents can use web of data for different tasks. It is an extension of the current one. Data is

given its meaning. Tim Berners-Lee proposed seven layer structures of the semantic web.

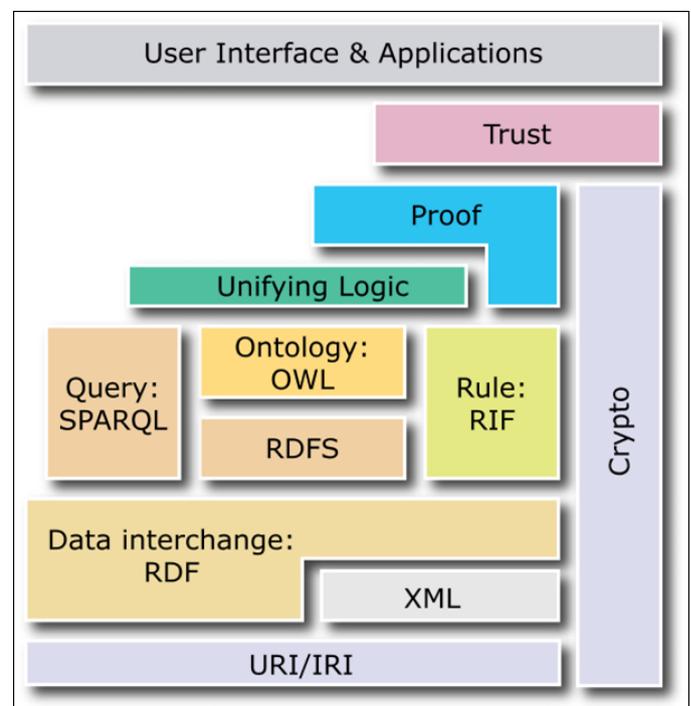


Figure 1. Semantic Cake

- URI/IRI (Unique Resource identifier / Internationalized Resource Identifier) is the basis of the whole semantic web, and it represents the resources of the web.
- XML/RDF (eXtensible Markup Language/Resource Description Framework), XML which is an extension to the HTML and is used to represent the content and structure of data & RDF describes the resources & the type of them. These are the important technologies for developing the Semantic Web. When using XML, user has the provision to create their own tags, hidden labels that annotate web pages. To provide meaning, RDF is used in form of triples,

which contains subject, object and predicate. It can be written using XML tags.

- RDFS/OWL (RDFSchema/Web Ontology Language) is used to describe ontologies, i.e. different resources and the relations between them. RDFS provides basic structure by using classes and properties. OWL defines and instantiates web ontologies. OWL extends RDFS by describing object-oriented classes, properties and their instances. (These are the main building blocks of ontology.) Also OWL defines disjoint classes, equivalent classes and cardinality which put restriction. OWL builds on top of RDF & RDFS. The full language is OWL Full which has two sub languages OWL-DL & OWL Lite.
- RIF (Rule Interchange Format) is used for the sharing of knowledge & business rules (eg: production rule systems) between different software agents.
- SPARQL (Simple Protocol and RDF Query Language) is the query language like SQL & XQuery which is used for RDF. RDF acts as an interface for interacting with an RDF database.
- Logic layer is intended to provide a consistent model for semantic web technologies. But, its implementation is nonexistent in the present scenario.
- Proof & Trust layers: Semantic web is an area of research. The trust layer provides trustworthiness to Semantic Web technologies as the proof layer tests and verifies according to logic layer.
- Crypto layer uses different encryption techniques in the lower layers of semantic cake.

### III. EXISTING METHODS

There are several methods in ontology development and semantic information retrieval. A number of methods are available for query reformulation. For the creation and management of domain ontologies, architecture for an ontology management system is proposed by Sugumaran et al. [2]. Reinberger et. al. [3] proposes an unsupervised method for the extraction of semantic relations from text. By using clustering and vector space model, Khan et al. [4] constructed ontologies automatically. Query reformulation is not necessary in semantic information retrieval; rather it is only needed to expand the query. This paper aims at query expansion which we will see soon.

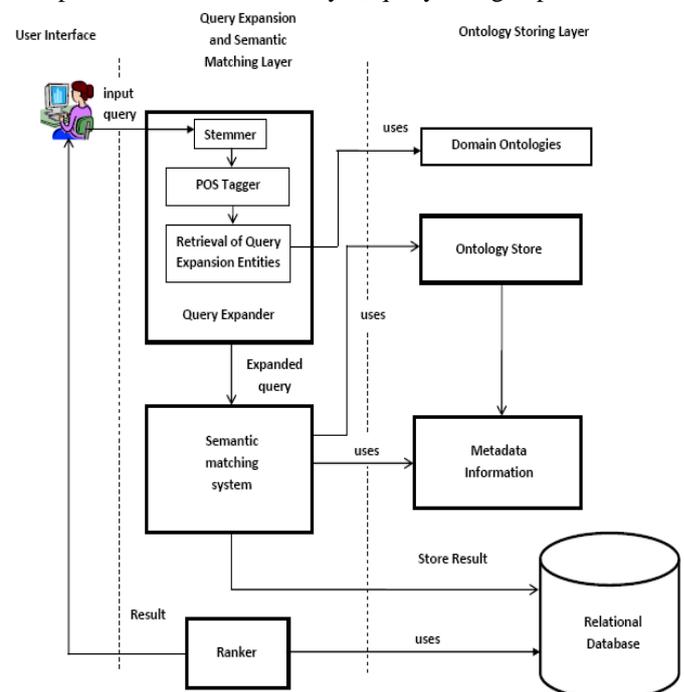
Also different methods are available for ontology matching. Jiwei Zhong et al. [5] proposed a conceptual graph matching algorithm for semantic information retrieval. But, it considered hyponymy senses only. Alireza Ensam et al. [6] developed an algorithm for semantic searching. It used semantic annotations and similarity measurements. Jung Ae Kwak et al. [7] did ontology matching using WordNet (Lexical English Database). Rajesh Thiagarajan et al. [8] use name and namespace of ontology in the calculation of similarity between ontologies. Ying Wang et al. [9] proposed a matching algorithm which is based on the structure of ontology graph. Qian Gao [10] proposed a similarity matching algorithm for ontology based semantic information retrieval model. In this paper domain and range of a relation is not considered in ontology matching. The query reformulator parses the query

which can be avoided. Matching is performed by considering hyponymy senses only.

## IV. QEAMS

### A. QEAMS Architecture

Semantic Web provides efficient knowledge mining. To add meaning, data is written as an Ontology which currently uses OWL (Web Ontology Language). Search engines which work on meaning of data will replace current document retrieval mechanisms. The core concept of semantic search engine is ontology mapping. For an efficient information retrieval, user provided search query must be expanded to perform matching. In this paper query string expansion and an algorithm to perform ontology matching are developed. The Ontology-based Semantic Information Retrieval Model is composed of user interface layer, query string expansion and



semantic matching layer and ontology storing layer. User Interface provides users and systems an interface to communicate with each other. User submits their query and result is given back to user through this layer. Query expansion layer expands the query string by finding common synonyms [11], hyponyms [7] and hypernym [11][12] with the help of lexical database WordNet[13]. Semantic matching layer uses the metadata information base to retrieve relevant ontologies from the ontology store and perform similarity matching between expanded query and retrieved ontologies. This layer applies some Semantic Matching rules to calculate the similarity between the expanded query and application ontologies. It makes use of the semantic similarity matching algorithm. After similarity calculation, the ranker is used to sort the results according to the semantic neighborhood between the expanded query and ontology. This is possible by extracting the hierarchical relations between different entities in the ontology.

B. Query Expansion Algorithm with Metada Support

- 1) Input: Query submitted by the end user
  - 2) Query Expansion
    - a) Perform Stemming.
    - b) Find Part Of Speech form.
    - c) Using domain ontologies, find synonyms, hyponyms and hypernyms common to the query entities.
    - d) Expand query using the result obtained in the previous step.
  - 3) Use metadata information to find the relevant ontologies to be matched with expanded query.
    - Metadata is the data which describes ontology present in the ontology store.
  - 4) For each ontologies retrieved from the store,
    - a) Perform Semantic Matching
      - i. Match expanded query with name of ontology
        - o If two ontologies have high semantic similarity, then name and namespace of both ontologies will be similar.
      - ii. For each entity present in the expanded query,
        - a. Match with all related senses in the ontology taken from the store.
        - b. Match all properties of ontology by their name and attribute.
        - c. Match the domain and range of similar relationships.
          - All semantic matching can be done using any methods like HSO, JCN, LIN, PATH etc[14]
            1. HSO: Two entities are semantically similar if the path distance between its WordNet synsets is less.
            2. LCH: It calculates the shortest path between two synsets by overall depth of hierarchy. It focuses on is-a relations.
            3. LESK: It calculates the similarity by number of overlapping definitions of two entities.
            4. WUP: Similarity calculation is done using depth of two synsets.
            5. RESNIK : Similarity is obtained as the information content in the lowest and most specific subclass or child entity.
            6. JCN: It finds the similarity as a probability of occurrence of a child synonym as an instance of a parent synonym. It has no upper bound.
            7. LIN: Same as JCN. But the similarity metric will be in between 0 & 1.
              - Also Levenshtein distance method can be used in similarity matching
  - b) Store similarity metric value for each ontology
- 5) After similarity matching, rank the ontologies according to similarity metric and previous history.
- 6) Display results to the end user
- 7) Output: Relevant ontologies that have sigh semantic

V. IMPLEMENTATION AND RESULTS

To find the best methods for semantic similarity calculation, a number of words are matched with all available methods. A sample is shown below. From the results, it is found that HSO,

Methods	Nuclear Family / Social Unit	Lens / Camera
HSO	6.0	6.0
LCH	2.59	1.74
Lesk	2.0	2.0
WuP	0.875	0.7
Resnik	4.78	3.44
JCN	0.0	0.08
LIN	0.0	0.38
PATH	0.33	0.14

TABLE I. WS4J METHOD ANALYSIS

WuP, Lin and Path methods have an upper limit while the rest does not have an upper bound. If the upper bound varies according to the position of the word in the hierachy, it is quite difficult to calculate the percentage of similarity for ontology. So I adopted HSO, WuP, Lin and Path methods for semantic matching.

Methods	MIN VALUE	MAX VALUE
HSO	0.0	16.0
LCH	0.0	unknown
Lesk	0.0	unknown
Resnik	0.0	unknown
WuP	0.0	1.0
Lin	0.0	1.0
Path	0.0	1.0

TABLE II. WS4J METHODS UPPER & LOWER BOUND

This paper tried to propose an algorithm to match ontologies corresponding for input query. This reduces the time for matching since it matches expanded query with related senses in the ontology. Another advantage of this model is, it eliminate the cost of query ontology creation by query expansion.

The results obtained by performing matching algorithm are given as follows. Ontologies used in the store were camera, disease, film, flower, heart, houseboat, movie, people, and professional, school, transport, travel and wine.

Java was used to implement the system and eclipse IDE was used to analyze and operate the ontology files. WordNet database and API's like JWNL, OWL and WS4J were also used.

Query	Results	Similarity
transport of flower	transport.owl	52.14%
	flower.owl	51.58%
travel	-	0.0%
body and disease	disease.owl	42.28%
houseboat and people	people.owl	11.59

TABLE III. SEMANTIC SIMILARITY

VI. CONCLUSION

To improve the precision and recall of current IR systems, search engines that work on meaning of data should be implemented. Semantic Web provides meaning to the data. Data is written in an ontology format. The core concept of semantic search engine is ontology mapping.

In this paper, an ontology matching algorithm is developed for semantic information retrieval. It considers name, namespace, classes, properties and range & domain of relationships for calculating similarity. Also this model replaces reformulated query ontology by expanded query with the use of WordNet. It helps in reducing the time taken for processing ontologies.

Domain ontologies other than WordNet can also be added for efficient knowledge mining. Another possible improvement could be retrieving the result as data rather than the file itself.

#### REFERENCES

- [1] T. Berners-Lee, J. Hendler and O. Lassila, "The Semantic Web," Scientific American, 2001.5.
- [2] Vijayan Sugmaran and Veda C. Storey. "Ontologies for Conceptual Modeling: Their Creation, Use and Management, Data Knowledge Eng," vol. 42, 2002.
- [3] Marie-Laure Reinberger, Peter Spyns, "A. Johannes Pretorius and Walter Daelamans. Automatic Initiation of an Ontology," Proc. Of ODBase04, Springer-Verlag, 2004.
- [4] Latifur Khan and Lei Wang. "Automatic Ontology Derivation Using Clustering for Image Classification," Multimedia Information Systems, pp. 56-65, 2002.
- [5] Jiwei Zhong, Haiping Zhu, Jiaming Li, Yong Yu, "Conceptual Graph Matching for Semantic Search," Proceedings of the 10th International Conference on Conceptual Structures: Integration and Interfaces table of contents, Springer-Verlag London, UK, pp. 92-196, 2002.
- [6] Alireza Ensan, Yevgen Biletskiy, "Matchmaking Through Semantic Annotation and Similarity Measurement," 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), 2012.
- [7] JungAe Kwak, Hwan-Seung Yong, "Ontology Matching based on Hypernym, Hyponym, Holonym, & Meronym sets In WordNet," International journal of Web & Semantic Technology (IJWesT), Vol.1, No.2, April 2010.
- [8] Rajesh Thiagarajan, Geetha Manjunath, Markus Stumptner, "Computing Semantic Similarity Using Ontologies," HP Laboratories HPL-2008-87.
- [9] Ying Wang, Weiru Liu, David A. Bell, "Structure based Similarity Spreading Approach for Ontology Matching," Proceedings of the 4<sup>th</sup> International Conference, SUM 2010, Toulouse, France, pp. 361-374 September 27-29, 2010.
- [10] Qian Gao. "Similarity Matching Algorithm for Ontology-Based Semantic Information Retrieval Model," Proceedings of the 10th World Congress on Intelligent Control and Automation, July 6-8, 2012, Beijing, China.
- [11] Raghuvar Nadig, J. Ramanand, Pushpak Bhattacharyya. "Automatic Evaluation of Wordnet Synonyms and Hypernyms," Proceedings of ICON-2008: 6th International Conference on Natural Language Processing, 2008.
- [12] J. Smith and D. Smith, "Database Abstractions: Aggregation and Generalization," ACM Trans. Database Systems, vol. 2, pp. 105-133, 1977.
- [13] George A. Miller. "WordNet: A Lexical Database for English," Communications Of The ACM November 1995/Vol. 38, No. 11
- [14] Giannis Varelas, Epimenidis Voutsakis, Paraskevi Raftopoulou, Euripides G.M. Petrakis, Evangelos E. Milios, "Semantic Similarity Methods in WordNet and their Application to Information Retrieval on the Web," WIDM '05 Proceedings of the 7th annual ACM international workshop on Web information and data management Pages 10-16, 2005.