International Journal of Pure and Applied Mathematics
Volume 116 No. 10 2017, 167-171
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

ONTOLOGY MATCHING TECHNIQUES USING FOR INFORMATION RETRIEVAL

Dr.K.P.Kaliyamurthie[1], Mr.S.Srigowthem[2]
[1]Professor ,[2]Assistant Professor,
[1,2]BIST,BIHER,Bharth University,Chennai-73.
[1]kpkaliyamurthie@gmail.com, [2]srigowthem.cse@bharathuniv.ac.in

Abstract: Rapid increase in plethora of information made available on the web, designing of efficient tools and software services to retrieve accurate and relevant information essentially became a huge substantial challenge. Such is the issue with ontologies which tends to be highly heterogeneous. Limited interactions by search engines on the information or documents retrieved ,with little to negligible explanations efficacy on the queries. Variations of meanings and ambiguity in ontology or entity related interpretations, depending on the subjective reasoning of users to match their expectations or desired queries. This paper highlights some techniques that might be able to resolve the issues faced in ontology based knowledge mining.

Keywords: Information retrieval, Machine Learning, Natural Language Processing, Ontology**.**

1. Introduction

Ontologies can be defined as data model that represents knowledge as a set of entities or concepts within a domain and the relationships between these entities or concepts which can be understood by a machine.In the semantic web field[1-6], ontologies serve as basic conceptual knowledge models and gives the semantic lexical that makes the domain knowledge available for sharing or reviews among the information system of peers.However, problem arises with the increase in the size of the ontology since its heterogeneity increases exponentially[7-9], which leads to the confusion of users to select documents displayed from the user's query because the subjective interpretations of the user to accurately select the most pertinent document retrieved by the search engine.

It is required that the search engines or tools used for processing and retrieving the information or documents of the requested queries interact with the available documents to extract some relevant text knowledge and further provide some concise and precise information with accuracy and efficiency of the documents at retrieval.

More problems arise with the limitation in the retrieval of relevant projects which possibly exists in a different language alphabetic-texts or symbols.

Aspects of matching strings:

Syntactic similar:

Two strings that are identical or differ only in few characters at few positions.

Semantic similar:

Two strings that are synonyms or have close definition in a thesaurus or similar in lexical database.

Tokenize:

A string that can be separated between their tokens to identify two consecutive words.
The rest of this paper will briefly discuss some techniques that can potentially help improve the accuracy and efficiency of the search engines in selecting, grouping and retrieving the most relevant documents to a user's query in ordered and systematic manner.

Ontology based information retrieval system:

Most of the information retrievals based on ontology are related to the use of semantics for information representation. The motive is To find data's satisfying the information required from a large set of databases. To achieve this[10-15], the following three processes are implemented:
Indexationfocuses on the representation of documents and queries with sets of ranked concepts or entities that summarize the contents of the information
Search function contains the systematic and algorithmic strategy for fetching the documents that matches the query. Weighing (keyword) of relevant documents to be selected are based on a score strategy which depends on their indexation.
Query expansionis an intermediate procedure which reformulates the user's query based on the database information to enhance the quality of the outcome.

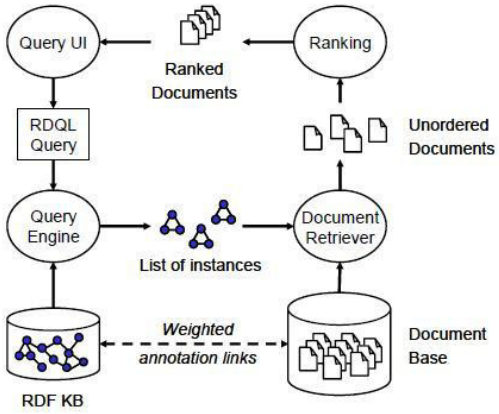

Figure 1. Ontology-based information retrieval

In order to avoid the misinterpretations of relations and ambiguities, recent retrieval systems maps the keywords to the concepts they represent. This requires general or domain conceptual structures on which to map the terms. This includes knowledge bases such as dictionaries, thesaurus, or ontologies.

Ontology matching using Machine Learning:
Ontology matching which discovers the correlation between semantically related concepts of ontologies becomes a necessity in semantic web applications.

For this, some machine learning techniques can be espoused to improve the quality of the ontology matching. Information retrieval techniques are recalled for effectively measure the similarities in the comparison of labels and concepts or entities at the most basic level.

A rule classification will be formed based on the existing database of knowledge and ontologies with relations provided by the domain experts. In a matching framework, for each pair of concepts in ontologies[16-19], the classification classifies them in to matched or not matched groups. Some things to consider are:

2. Learning

A list of semblance scores is computed for each pair of concepts by applying a list of semblance measures[20-25]. Each pair of concepts is considered as a learning object O. Each semblance measure becomes an O's attribute and its equivalent semblance is considered as an O's trademarked value.O becomes an unclassified object if two of its concepts are to be matched ontologies. An unknown value is assigned in its class.O is assigned back to an instance of training data if two concepts are in ontologies within the knowledge base.

Selection of semblances:

Similarity measures represented that are capable of dealing with different types of terminological heterogeneity based on the aspects of matching strings are selected.

Machine learning model:

Several machine learning models can be used to build a classification rule from a give learning data. Some models examples are; tree based, rule based, probability based, instance based, function based, semantic nets, etc

Classifying unclassified objects:

Unclassified objects can be classified into "unclassified" to check further semblances by further deploying the information retrieval technique.

A sample algorithm to combine the mapping results obtained from two different matcher A_1 and A_2 is given below:

Input: $A_1 = \{(e_i, e_j \equiv c_e)\}$

$A_2 = \{(e_p, e_p \equiv c)\}$

Output: $A_{final} = \{(e_x, e_y \equiv c)\}$

$\theta \leftarrow \min(m.c_2) | \, m \in A_1 \cap A_2;$

$A \leftarrow \text{WeightedSum}(A_1, \theta, A_2, (1- \theta));$

$A_{final} \leftarrow \text{GreedyChoice}(A, \text{threshold});$

Return A_{final};

Natural Language Processing based on Ontology:
Most of the ontology extractions are dealt mainly in English. This arises in a limitation, when dealing with languages other than English such as Cyrillic alphabets, Japanese, Chinese and so on that could potentially contain some vital semantic similarities between terms from two different languages within a project within the ontology, which could be conceptualized.

Machine learning is combined with term recognition and linguistics.NLP mainly deals with the interaction between human and machine through natural languages.

Term recognition could be achieved the following NLP techniques:

Tokenizer:

Splits the words used with blank space or other signs that indicates that the string is separable.

Dictionary:

It includes the lexemes and the semantics associated.

Morphology:

Processing of and splitting of a sentence into words.

Word Sense Disambiguation:

Removing the uncertainty of semantics between the senses of a word.

Chunking:

It is the process of shallow parsing that identifies parts of speech (POS tagging) and labeling the simple phrases from the tagged output. It is commonly used by Support Vector Machine (SVM).

Multiword Expressions (MWE) is Reduplicated (RMWE) to improve Chunk identifications.

Transliteration:

Text conversion of from one language script to another language.

3. Conclusion

This paper aims to integrate the discussed techniques to provide an accurate, efficient and reliable information retrieval service to the user's queries retrieval of knowledge documents. Future work aims to implement the various AI algorithms such as expert systems to aid the user extract even more detailed knowledge on the related ontology documents, research proposals or projects from the knowledge reservoir.

References

[1] Jian Ma, Wei Xu, Yong-hong Sun, Efraim Turban, Shouyang Wang, and Ou Liu "An Ontology-Based Text-Mining Method to Cluster" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, VOL. 42, NO. 3, MAY 2012.

[2] David Vallet, Miriam Fernández, and Pablo Castells "An Ontology-Based Information Retrieval Model" Universidad Autónoma de Madrid Campus de Cantoblanco, c/ Tomás y Valiente 11, 28049 Madrid.

[3] Edward H.Y. Lim, Hillman W.K. Tam, Sandy W.K. Wong, James N. K. Liu and Raymond S. T. Lee "Collaborative Content and User-based Web Ontology Learning" FUZZ-IEEE2009, Korea, August 20-24, 2009.

[4] Andreas Heß "An Iterative Algorithm for Ontology Mapping Capable of Using Training Data" VrijeUniversite it Amsterdam University College Dublin

[5] Jan Paralic, Ivan Kostial "Ontology-based Information Retrieval" Department of Cybernetics and AI, Technical University of Kosice, Letna 9, 040 11 Kosice, Slovakia. Department of Cybernetics and AI, Technical University of Kosice, Letna 9, 040 11 Kosice, Slovakia.

[6] DuyHoa Ngo "Enhancing Ontology Matching by Using Machine Learning, Graph Matching and Information Retrieval Techniques" https://tel.archives-ouvertes.fr/tel-00767318 Submitted on 20 Dec 2012

[7] Sylvie Ranwez, Vincent Ranwez, Mohameth-François Sy, Jacky Montmain, Michel Crampes, "User Centered and Ontology Based Information Retrieval System for Life Sciences" LGI2P Research Centre, EMA/Site EERIE, Parcscientifique G. Besse, 30 035 Nîmescedex 1, France. Laboratoire de Paléontologie, Phylogénie et PaléobiologieInstitut des Sciences de l'Evolution (UMR 5554 CNRS), Université Montpellier II,CC 064, 34 095 MONTPELLIER edex 05, France

[8] Dominique.Estival,Chris.Nowak,Andrew.Zschorn "Towards Ontology-Based Natural Language Processing" Human Systems Integration Group Defence Science and Technology Organisation PO Box 1500, Edinburgh SA 5111 AUSTRALIA.

[9] Michael Collins "Machine Learning Methods in Natural Language Processing" MIT CSAIL.

[10] Udayakumar R., Kaliyamurthie K.P., Khanaa, Thooyamani K.P., Data mining a boon: Predictive system for university topper women in academia, World Applied Sciences Journal, v-29, i-14, pp-86-90, 2014.

[11] Kaliyamurthie K.P., Parameswari D., Udayakumar R., QOS aware privacy preserving location monitoring in wireless sensor network, Indian Journal of Science and Technology, v-6, i-SUPPL5, pp-4648-4652, 2013.

[12] BrinthaRajakumari S., Nalini C., An efficient cost model for data storage with horizontal layout in the cloud, Indian Journal of Science and Technology, v-7, i-, pp-45-46, 2014.

[13] BrinthaRajakumari S., Nalini C., An efficient data mining dataset preparation using aggregation in relational database, Indian Journal of Science and Technology, v-7, i-, pp-44-46, 2014.

[14] Khanna V., Mohanta K., Saravanan T., Recovery of link quality degradation in wireless mesh networks, Indian Journal of Science and Technology, v-6, i-SUPPL.6, pp-4837-4843, 2013.

[15] Khanaa V., Thooyamani K.P., Udayakumar R., A secure and efficient authentication system for distributed wireless sensor network, World Applied Sciences Journal, v-29, i-14, pp-304-308, 2014.

[16] Udayakumar R., Khanaa V., Saravanan T., Saritha G., Retinal image analysis using curvelet transform and multistructure elements morphology by reconstruction, Middle - East Journal of Scientific Research, v-16, i-12, pp-1781-1785, 2013.
[17] Khanaa V., Mohanta K., Saravanan. T., Performance analysis of FTTH using GEPON in direct and external modulation, Indian Journal of Science and Technology, v-6, i-SUPPL.6, pp-4848-4852, 2013.
[18] Kaliyamurthie K.P., Udayakumar R., Parameswari D., Mugunthan S.N., Highly secured online voting system over network, Indian Journal of Science and Technology, v-6, i-SUPPL.6, pp-4831-4836, 2013.
[19] Thooyamani K.P., Khanaa V., Udayakumar R., Efficiently measuring denial of service attacks using appropriate metrics, Middle - East Journal of Scientific Research, v-20, i-12, pp-2464-2470, 2014.
[20] R.Kalaiprasath, R.Elankavi, Dr.R.Udayakumar, Cloud Information Accountability (Cia) Framework Ensuring Accountability Of Data In Cloud And Security In End To End Process In Cloud Terminology, International Journal Of Civil Engineering And Technology (Ijciet) Volume 8, Issue 4, Pp. 376–385, April 2017.
[21] R.Elankavi, R.Kalaiprasath, Dr.R.Udayakumar, A fast clustering algorithm for high-dimensional data, International Journal Of Civil Engineering And Technology (Ijciet), Volume 8, Issue 5, Pp. 1220–1227, May 2017.
[22] R. Kalaiprasath, R. Elankavi and Dr. R. Udayakumar. Cloud. Security and Compliance - A Semantic Approach in End to End Security, International Journal Of Mechanical Engineering And Technology (Ijmet), Volume 8, Issue 5, pp-987-994, May 2017.
[23] Thooyamani K.P., Khanaa V., Udayakumar R., Virtual instrumentation based process of agriculture by automation, Middle - East Journal of Scientific Research, v-20, i-12, pp-2604-2612, 2014.
[24] Udayakumar R., Thooyamani K.P., Khanaa, Random projection based data perturbation using geometric transformation, World Applied Sciences Journal, v-29, i-14, pp-19-24, 2014.
[25] Udayakumar R., Thooyamani K.P., Khanaa, Deploying site-to-site VPN connectivity: MPLS Vs IPSec, World Applied Sciences Journal, v-29, i-14, pp-6-10, 2014.