IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-11, Nov 2018 E-ISSN: 2347-2693

Applications of Semantic Similarity Metrics

[1*]Suraiya Parveen, [2]Ranjit Bisswas

[1,2]Dept. of Computer Science and Engineering, SEST Jamia Hamdard, New Delhi ,India

Corresponding Author: husainsuraiya@gmail.com, Tel.: +00-8505802249

Available online at: www.ijcseonline.org

Accepted: 18/Nov/2018, Published: 30/Nov/2018

***Abstract*—** The objective of this work is to access the applicability of the semantic similarity in concepts of a single ontology. The measurement of semantic similarity may help not only in information retrieval, but in other applications such as semantic search and semantic clustering. The traditional key-word search technique matches the keyword with the content of the document. These techniques do not reflect the meaning or relatedness. Hence the relevance and accuracy of the retrieved documents are less. Another important application of semantic similarity measurement is in cluster analysis. The semantic clusters may be treated with same function to accomplish the perfect analysis and decision making.

***Keywords*—** Semantic Similarity, Ontology, Semantic Retrieval, Clustering

I. INTRODUCTION

Introduction should lead the reader to the importance of the study; tie-up published literature with the aims of the study and clearly states the rationale behind the investigation. It should state the purpose and summarize the rationale for the study and gives a concise background [1]. Use references to provide the most salient background rather than an exhaustive review. The last sentence should concisely state your purpose for carrying out the study or a summary of the results [2].

The purpose of the contribution statement is for you to provide a clear and concise understanding of the primary contribution provided by your manuscript. The statement should:

1. clearly articulate the ways in which the research provides insight to a consumer-relevant question;
2. situate your research within the existing knowledge on the topic; and
3. explain what the research adds to what is already known about the consumer-relevant problem.

As a last paragraph of the introduction should provide organization of the paper/article (Rest of the paper is organized as follows, Section I contains the introduction of the research paper, Section II discuss semantic similarity, Section III contain the applications of semantic similarity, Section IV discuss ontology and semantic search,.Section V explain the Information retrieval and retrieval, Section VI contain semantic information retrieval, Section VII contain ontology and semantic clusters Section VIII concludes research work.

II. SEMANTIC SIMILARITY

"Semantic Similarity relates to computing the conceptual similarity between terms which are not lexicographically similar."[4]

We are living in the age of big data and data analytics. The data is so huge that to extract relevant information from digital repositories has become very difficult. The information retrieval techniques use similarity measures to match the user queries and the document information. Accuracy and precision need very sophisticated information retrieval tools. Another way to achieve accuracy in information retrieval is by using ontology. The design of ontology is such that it models, concepts and their relationship within a domain [5].

III. APPLICATIONS OF ONTOLOGY

Ontology is a highly technical term; Figure 1 shows the major applications of ontology in Knowledge Engineering such as semantic similarity, semantic search and semantic clustering

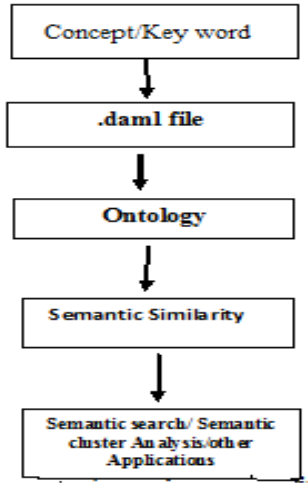


Fig. 1 Applications of semantic similarity

The Figure 1 shows the basic demonstration of support of ontology in semantic similarity measurement, semantic search and semantic cluster analysis.

IV. ONTOLOGY AND SEMANTIC SEARCH

Semantic Search is improved version of keyword search to improve the accuracy and precision. If search keyword is added to the meaning and essence, it is able to retrieve better search results. The purpose of semantic search is to apply meaning and intention to the keyword. Domain ontology provides background knowledge to achieve this. Semantic extension and retrieval are adopted for semantic-based knowledge retrieval [7]. The repeated use of ontology for IR is proving to be a more efficient method that can be faster and more efficient. Ontology is the basic building block of semantic web. It facilitates to upgrade keyword search to semantic search. It helps in specifying the meaning of the query to the system for retrieval of relevant documents related to it. Ontology is the structural representation of concepts and their relationship. If the concept is not changing ontology confine to the structure of knowledge of the domain [8]. Metadata is an elementary building block of semantic web. Metadata is data about data. When metadata is integrated with web pages, Documents, text, web can be read and interpreted by machines. Ontology has a rich and formal logic-based language for specifying the meaning of terms. The web is full of unstructured data. The ontological relationships in the ontology are based on domain knowledge to be structured. The figure 2 shows the block diagram of semantic search.

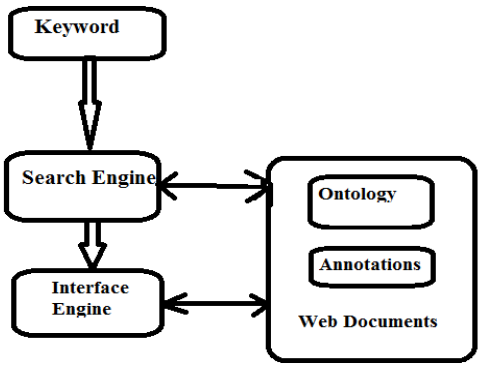


Fig. 2 Block Diagram for semantic search

Semantic search can become reality by using domain ontology. The domain ontology contain concept in the form of node and edges explain relationships. The search can be extended to the similar concepts to achieve meaningful results using ontology [8]. Ontology is most important supporting technology for semantic search. The aim of this research is to calculate the semantic similarity between the concepts of domain ontology.

V. INFORMATION RETRIEVAL

Information Retrieval (IR) is extracting relevant documents to the user needs from data repositories. These are usually Text documents of unstructured nature.

Retrieve documents with information that is relevant to the user's information need and helps the user complete a task. Figure 3 gives the block diagram for information retrieval.
xtraction of data or documents is usually the web search but Information.

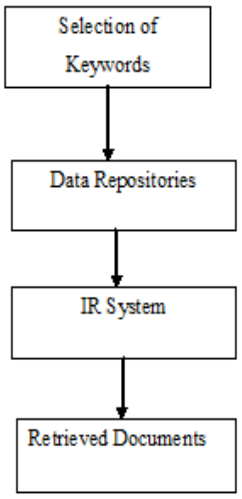


Fig. 3 Information Retrieval

Retrieval Techniques also help in other searches, such as

- E-mail search
- Searching programs or files on the desktop
- Medical Information Retrieval
- Corporate knowledge bases
- Legal information retrieval
- Library Data Base
- Multi tenant cloud environment

Accurate retrieval is difficult to achieve. Efficiency in IR is a critically important topic [9]. The retrieved documents, as a search results from the web or any other data resource has to be relevant. The efficiency of a good information retrieval system can be judged by two important features.

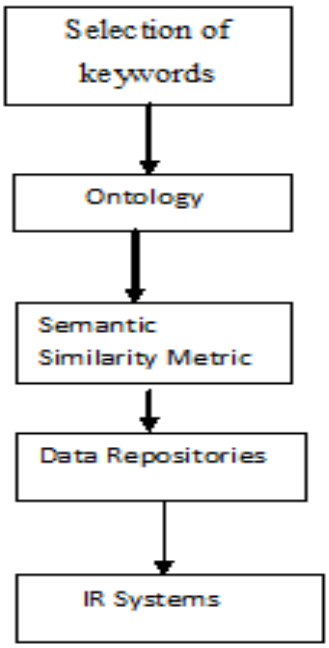
1. It must retrieve all relevant documents
2. The retrieved documents must be relevant

VI. SEMANTIC INFORMATION RETRIEVAL

Semantic retrieval improves the efficiency of existing search techniques by adding meaning to the search. Keyword search approach lacks in semantic information and cannot understand the user's query intent very well [9]. It does not

consider the related words and perform only by matching the keywords. This is the basic reason for the large number of irrelevant information as search results. The semantic search uses domain ontologies to extend the search to the related concepts [11].

Semantic Retrieval is one of the important ways to improve efficiency of retrieval. Ontology support semantic retrieval. Figure 4 shows semantic information retrieval.



Fig,4 Semantic Information Retrieval

The key to the semantic retrieval is semantic similarity measurement. Ontology and metadata are most valuable components of the semantic technology. Ontology contains the concepts within the domain, which are related to each other [10]. The relationship is demonstrated with the help of links. The concepts of ontology are related to each other, but how close or distinct they are, is essential to measure. This is the preliminary requirement not only for semantic search but also for interoperability, sharing and reuse of the ontology. Figure 4 shows the extension of the keyword search by adding semantics more relevant retrieved documents.

VII. ONTOLOGY AND SEMANTIC CLUSTERING

Clustering is a very important technique to achieve a group of similar objects. Semantic clustering means meaningful clusters. This technique is used in business, medical science, healthcare, education etc.

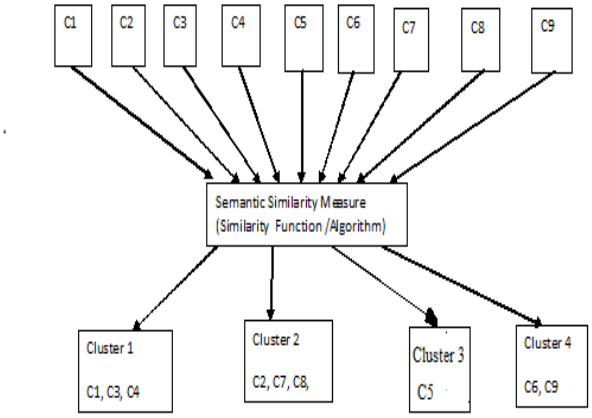


Fig. .5 Semantic Cluster

Semantic clustering can be obtained from the concepts of domain ontology using some similarity metric [7]. When the metric is applied to the concepts and similarity is quantified. The equal values or values in any specified range can be clubbed to obtain clusters. Figure 5 shows the process of semantic clustering.

The figure shows how similar concepts are obtained by calculating semantic similarity. After calculating a similarity measure of the concepts, the concepts can be grouped into clusters for analysis and interpretation of information. With the help of this technique user can quickly get the information. It is a good technique for data analysis and knowledge extraction.

VIII. CONCLUSION

The data is mostly heterogeneous, unstructured and scattered on the web or independent data repositories. The need is to determine the semantic similarity or semantic difference of concepts in an ontology. The similarity measures help in solving the critical issues in information retrieval, searching techniques and clustering.

Semantic similarity is based on ontology and carry features to extend the search. Retrieve meaningful information from the website or data repositories is a very important task. Semantic technologies such as ontology can help in retrieving the data accurately. Users query is extended with related concepts extracted by using the similarity measures and is applicable in clustering also.

REFERENCES

[1]. Tim B.L and Fischetti M., ―Weaving the Web: The Original Design andUltimate Destiny of the World Wide Web by its inventor‖, Britain: Orion Business. ISBN 0-7528-2090-7, 1999

[2]. Wang, J.Z., Ali, F.: An Efficient Ontology Comparison Tool for Semantic Web Applications. In: IEEE/WIC/ACM International Conference on Web Intelligence (WI 2005), pp. 372–378 (2005)
[3]. Hu, W., Jian, N., Qu, Y., Wang, Y.: GMO: A Graph Matching for Ontologies. In: Proceedings of Integrating Ontologies 2005, pp. 41-48. CEUR-WS.org, Aachen (2005)
[4]. Giunchiglia, F., Shvaiko, P., Yatskevich, M.: S-Match: an Algorithm and an Implementation of Semantic Matching. In: Bussler, C.J., Davies, J., Fensel, D., Studer, R. (eds.) ESWS 2004. LNCS, vol. 3053, pp. 61–75. Springer, Heidelberg (2004)
[5]. Rada, R., Mili, H., Bicknell, E., Blettner, M.: Development and application of a metric on semantic nets. IEEE Transactions on Systems, Man and Cybernetics, 17–30 (1989)
[6]. Aggarwal N. (2012) Cross Lingual Semantic Search by Improving Semantic Similarity and Relatedness Measures. In: Cudré-Mauroux P. et al. (eds) The Semantic Web – ISWC 2012.. Lecture Notes in Computer Science, vol 7650. Springer, Berlin, Heidelberg
[7]. Moldovan, D.I., Mihalcea, R.: Using WordNet and lexical operators to improve Internet searches. IEEE Internet Computing 4(1), 34–43 (2000)
[8]. Finkelstein L., Gabrilovich E., Matias Y., Rivlin E., Solan Z., Wolfman G. and Ruppin E., ―Placing search in context: The concept revisited,ACM Transactions on Information Systems, Vol. 20, No.1, pp. 116–131, January 2002.
[9]. Schenkel, R., Theobald, A. & Weikum, G. Inf Retrieval (2005) 8: 521. https://doi.org/10.1007/s10791-005-0746-3
[10]. Liu, S., Liu, F., Yu, C., Meng, W.: An effective approach to document retrieval via utilizing WordNet and recognizing phrases. In: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, Sheffield, United Kingdom, pp. 266–272 (2004)
[11]. Sapkota, K., Thapa, L., Pandey, S.: Efficient Information Retrieval using measures of Semantic Similarity (2006)

Authors Profile

Suraiya Parveen received her Bachelor of Engineering degree from Jamia Millia Islamia, New Delhi and her Master of Technology from GGS Indraprastha University, New Delhi. She is an Assistant Professor in the Department of Computer Science and Engineering at Hamdard University, New Delhi. Her research interests include Information Retrieval, Data Structure and Semantic Technology.

Prof. Ranjit Biswas is Full Professor, Dean of School of Engineering Science and Technology and Head of Computer Center, Hamdard University. He received his PhD from IIT Kharagpur and has more than 30 years of experience as an educator and researcher in various National and International Universities. His research interests include Fuzzy logic, Set Theory and Soft Computing.

