International Journal of Psychosocial Rehabilitation, Vol. 23, Issue 01, 2019
ISSN: 1475-7192

An Investigation on Ontology Based Fuzzy Semantic Information Retrieval

J. Naren, D. Raja Rajeswari, Nikhith Sannidhi and Dr.G. Vithya

Abstract--- *Information retrieval, a core research area has caught the minds of many Computer Science researchers. The research survey is an amalgamation of various findings done in the effectiveness of Ontology Based Information retrieval with fuzzy logic content evaluation. The inference of the scan with many papers widens the understanding of a Researcher on Fuzzy Ontology Based Information Retrieval which has come by converting an Ordinary Ontology to a Fuzzy Ontology. The paper gives an explanatory note on several models used in Information Retrieval. Comprehensive comparative studies on algorithms in the areas revolving Ontology with Information Retrieval are also prescribed. A Framework on the existing OBIR (Ontology Based Information Retrieval) being developed for Education could be next nearest work following the literature Review remains the conclusion of the proposed work.*

Keywords--- *Fuzzy Logic, Fuzzy Ontology, Information Retrieval, Ontology Engineering, Semantic Web.*

I. INTRODUCTION

In the modern technology-driven era, many complex programs, applicat ions run on different devices to meet various needs and desires of people. As time passes, the applications generate tons of data which present over the internet is gigantic. A major and a hot research on how efficiently and quickly can the required in formation be retrieved from huge amount of data is the talk of the day. In the recent past, many information retrieval techniques and methodologies were formulated and developed. But, the techniques use a naive approach of taking the keywords from the query and comparing it with all the available documents. The successfully matched ones are retrieved and given to the user. There is a drawback in the aforementioned approach. The approach blindly retrieves required information without even understanding the meaning behind user's query. The information retrieved may not be of any relevance to the user's context. For example, if the user query is "Where do bass live?", then the search engine extracts the keyword by eliminating the questionnaire words and retrieves the information about Bass - the musical note. But the actual user's context is the Bass - which is a kind of fish.

The above problem is solved by the introducing the concept of ontology which models and simulates the real world scenario. It establishes the relationships among the concepts in the concerned domain of the user's query. Ontology development has also some limitations.. Ontology alone cannot work efficiently when there is "uncertainty" in the information and the IR systems still lack the understanding of semantics of the query. The concept of fuzzy logic deals with the degree of truthness of an event or a process rather than giving a crisp rigid '1' or '0' value.

J. Naren, Assistant Professor, School of Computing, SASTRA Deemed University, Thanjavur, India. E-mail: naren.jeeva3@gmail.com
D. Raja Rajeswari, B.Tech Computer Science and Engineering, SASTRA Deemed University, Thanjavur, India. E-mail: aarthydurai1997@gmail.com
Nikhith Sannidhi, B.Tech Computer Science and Engineering, SASTRA Deemed University, Thanjavur, India. E-mail: niksami10@gmail.com
Dr.G. Vithya, Professor, School of Computing, KL University, Vijayawada. E-mail: vithyamtech@gmail.com

Received: 16 Nov 2018 | Revised: 25 Dec 2018 | Accepted: 10 Jan 2019

Fuzzy logic and fuzzy set theory are the perfect concepts to deal with present problems faced during Information Retrieval and the concepts are merged with the aforementioned ontology-based approach which enhances the understanding of semantics and also elevates system's performance, efficiency and the degree of relevance to the user's query. Fuzzy ontology mathematically represents the relationships present in the ontology which improves the semantic understanding of the information. The paper showcases the on-going research in information retrieval domain using semantics and fuzzy ontology. A brief introduction of each concept is presented in the introduction section. Papers were surveyed and the findings are exhibited in the sections 3, 4 and 5 which are as follows-ontology based information retrieval, fuzzy logic based information retrieval

and fuzzy ontology based information retrieval. Sections 5, 6 and 7 provide a detailed comparative study among the algorithms used during Information Retrieval with normal Ontology, Fuzzy Ontology and Fuzzy Logic Evaluation. The methodologies surveyed in aforementioned topics are also tabulated under the sections respectively. In section 4, different models used for information retrieval are presented.

A. Semantic web

Semantic web is an extended version of current Web which makes the machine to understand Human Linguistics and focuses on data integration. In the current Web, the data retrieved is in the form of web pages which is a HTML document that are linked to each other through hyperlinks. Machines read the pages and extract the keywords but cannot understand the meaning of the pages. So data access gets limited as only the best fit documents are retrieved. Semantic web is concerned with the meaning of data but not its structure. Now, Search Engines would have the ability to select the pages that the user really wants. The main aim of Semantic web is to discover resources and integrate web information with less effort. Semantic web deals with understanding the hidden meaning underlying in a sentence. For example, if a sentence says "Rome wasn't built in a day", which is a famous quote, the computer should not literally interpret that a day is not enough to construct the city of Rome. Rather it should interpret that "achieving huge and complex goals in life cannot be done in a single day". The interpretation mentioned in the above is fed as knowledge to the system by which an understanding on the issue is obtained.

B. Ontology Engineering

Ontology is an amalgamation of concepts in a particular domain and emphasizes the relationships among the concepts. The branch of engineering that deals with the rules and methodologies for construction of ontology, the various tools and ways for maintaining ontologies is called Ontology Engineering [1]. Ontology includes necessary entities, attributes and the relationships among the concepts in the domain to model and simulate real world objects. It is analogous to a class diagram. Ontology also establishes communication between the users and machines via information exchange to provide better semantics and understanding to the system. Ontology is useful when ambiguity in information is found.

C. Fuzzy Logic

Today, many applications run on binary algebra. The complex programs working behind the applications run only on basic elements of algebra, which are 1s and 0s. But in Boolean algebra a certainty on the occurrence of an event is felt.

Received: 16 Nov 2018 | Revised: 25 Dec 2018 | Accepted: 10 Jan 2019

In today's modern era, an observation on "uncertainty" about the occurrence of the event usually prevails. In such occasions, Fuzzy logic dealing with values that lies between 0 and 1 emphasizes on the degree of truthness. Applications of fuzzy logic are versatile and are predominantly seen in many areas such as control systems, expert systems, artificial intelligence, medicine, defense, business and many other branches of engineering.

D. Information Retrieval

IR is the process of obtaining information for the user from a collection of resources by query processing. The searches can be based on metadata or keyword matching. Traditional search algorithms are based on keyword matching which retrieve limited information due to lack of semantics in query processing. Also a traditional algorithm either processes the metadata or does not consider the results or processing of metadata is neglected. Hence the concept of Semantic Web in Information Retrieval which links web information with semantics is integrated. Semantic Web's Integration with Information Retrieval helps in retrieving user relevant context information.

E. Fuzzy Ontology

Fuzzy ontology uses fuzzy set theory, is an extended version of ontology that handles "uncertainty" in a process or information or an event. An explanation on the extent of ambiguity or vagueness of the elements in the ontology is established [3]. The usage of Fuzzy Ontology could be found in applications such as information retrieval, decision making, data mining and also in semantic-based applications [4].

II. ONTOLOGY BASED INFORMATION RETRIEVAL

Ontology is an efficient tool in capturing and structuring the meaning from natural languages. In order to express the contents of a document precisely, a necessity to create associations between document and the contents relevant to documents in the domain model is needed. The use of an Ontology in Information Retrieval helps to define concepts and relations for knowledge representation of a particular document in domain specific terms.

Jun Hu et .al [5] proposed a Semantic Information Retrieval Approach based on a Rough Ontology. The proposed approach infers the properties of correlated query by association searches of rough ontology where properties are taken as equivalence relations to build an approximation space. Further, an algorithm of computing similarity in rough ontology is used, and approximation space is employed to compute similarity for ranking documents in semantic document indexing space.

Sajendra Kumar et .al[6] suggested that the gap between the texts based web pages and the RDF based pages of the Semantic web can be bridged using Ontology. The proposed architecture builds Semantic Index for the constructed ontology. Semantic indexing for information extraction and retrieval showed efficient results to find effective and context based information from unstructured and millions of documents.

Meenakshi et .al[7] proposed Ontology annotation knowledge representation methodology based on Concept Relevancy Ranking of Link and Page Contents algorithmic approach in Semantic enhanced Information considering the time of response, dead and redundant links. Both user query and SERP were pre-processed for domain ontology and semantic annotation.

Received: 16 Nov 2018 | Revised: 25 Dec 2018 | Accepted: 10 Jan 2019

Root words were extracted from the user query to construct a repository. The link content and page content of the SERP's were also compared with repository so that the more relevant pages would be retrieved.

III. FUZZY LOGIC BASED INFORMATION RETRIEVAL

To improve the performance and efficiency of information retrieval, the concept of semantics or the linguistics need to be properly interpreted during information retrieval process. During the process, information retrieval algorithms might encounter information that is "uncertain". If there is an encounter, an inference for deciding the uncertainty in information and its relevance with or with respect to user's query is obtained. For overcoming the above challenge, the introduction of fuzzy logic for IR techniques is done. The relevance in information is given by fuzzy logic and an assignment of a score for sorting the retrieved documents in order of relevance with respect to the user's query is established [8].

Mustapha et.al [8] emphasized on the fact that keywords alone are not enough to assess the relevance of documents that are retrieved with user's query, as most information retrieval techniques consider only keywords from queries to retrieve documents. A belief that ontologies and thesauri also play a key role in assessing the closeness between queries and the retrieved documents existed. Therefore many information retrieval techniques were surveyed and two approaches came up. The first one dealt with ontologies which assessed the relationships between terms namely synonyms, antonyms, hyponym etc. The approach used a symbolic pattern matching technique and gave a value which represented the degree to which the documents were relevant to the query using pattern matching techniques. The second approach projected the degree of relevance of fuzzy sets in queries and ontologies then implemented the concept using sub-trees. The projections were then compared for rank ordering documents.

Henrik's [9] thesis discussed the methodologies and the positive effects of introducing ontology in information retrieval. Similarities between the ontologies and the semantic relations established between the nouns were stressed. Natural Language processing technique was also employed. Weighted shared nodes technique were used to compute the similarities between the concepts present in the ontologies and the similarities, with the introduction of fuzzy logic, was used to scale up the performance of information Retrieval. Later, the similarities were used in query evaluation.

Rahul et.al [10] throws light on the importance of information retrieval and the need for improving the performance. Data generation has been numerous in today's world.

Consequently, fetching the pertinent information from the internet has become one of the major difficulties in information retrieval. In the paper, fuzzy logic was incorporated into web based information retrieval and a fuzzy inference system (FIS) was proposed. The inputs to this system were the scores obtained from Page Rank algorithm and HITS algorithm.

The system gave the relevant documents and non-relevant documents according to the user's query as the output. The system implementation was through MATLAB and the performance judgment was done by information retrieval parameters, which were, precision and recall.

Yogesh Gupta et.al [11] presented an approach which uses fuzzy logic to improve the performance in terms of precision and recall parameters. This approach aims to develop a similarity function using fuzzy logic. The function takes inputs generated from the membership function. The implemented their approach using MATLAB and they presented a detailed graphical analysis of data collected fr om CACM data corpus and their experiments produces much better results as compared to *okapi-BM25* and *Rubens'* approach in terms of recall and precision parameters.

Jagendra Singh et.al [12] stresses the importance of query expansion terms selection methods which can greatly improve the efficiency of information retrieval process. A method was proposed for fuzzy logic based query expansion which mines for additional terms from the query and assigns degree of importance which, along with fuzzy rules, was then used to calculating weight for these additional terms. The newly computed weights of additional terms and weights of the original terms are then merged to form a new query which improves the performance of information retrieval. Experiments were made on TREC and FIRE benchmark datasets and results showed significant improvement of precision rate, recall rate and f-measure.

Yogesh Gupta et.al [13] developed ranking function which uses term weighting schema. Fuzzy logic is used to describe the relevance score of the retrieved document with respect to the user's query. Experiments were carried out on CACM and CISI benchmark datasets and results shows improved performance of their ranking function as compared to the similar ranking functions developed by Rubens' and *okapi- BM25* based on parameters of information retrieval.

IV. FUZZY ONTOLOGY BASED INFORMATION RETRIEVAL

While ontology describes the knowledge representation during IR process, it fails to support information which is uncertain. Fuzzy logic is introduced while constructing fuzzy ontology. Ontology supported with fuzzy set theory can solve this complexity and also boosts the performance of information retrieval systems [14].

Balasubramanian [15] developed ontology to deal with uncertain information in information retrieval. The present techniques are not competent enough to deal with uncertain information and therefore a new technique has to be developed. Whenever we talk about uncertainty, the concept of fuzzy comes into picture. It has been observed that the both fuzzy and ontology based information retrieval yielded better results in terms of uncertainty and semantics. The algorithm which they have developed is based on matching the items present in the input queries and the domain in which the information is pertaining to. This algorithm maps the hybrid ontology with keyword. The input is a fuzzy ontology along with a keyword and the output obtained is an ontology which explains a set of keywords in the specified view. An issue that arises while developing a concept hierarchy for a particular domain is a strenuous task. To resolve this, they have used a Hybrid FOGA (Fuzzy Ontology Generation framework) for creating fuzzy ontology automatically using clustering technique from formal concept analysis. The performance is evaluated by precision, recall and f-measure values.

Maryam et.al [14] states that the present available techniques are not enough for efficient Information retrieval. They believe that ontology that represents the processing of information retrieval can improve the results.

Received: 16 Nov 2018 | Revised: 25 Dec 2018 | Accepted: 10 Jan 2019

But Ontology alone cannot solve the problem because it can fail to show information which is uncertain. Therefore, they develop an ontology which incorporates fuzzy logic as fuzzy ontology will be robust while processing "uncertain" information. They propose an approach that produces fuzzy ontology which has two degrees of uncertainty.

Zeinab et.al [16] presents an information retrieval model that can handle problems dealing with multiple fields. Their model also aims to enhance the recall rate as compared to the present available information retrieval models and retrieval of documents is done semantically. They also propose a page ranking algorithm to prioritize the relevant documents.

Nagarajan et.al [17] emphasizes the lack of semantic between information resources in information retrieval system. Semantic is expressed using ontology. Using the concept of multi-modal ontology, text and image is retrieved. The input can be text, image or ontology. Analyzing images becomes difficult due to the semantic gap between image features and human linguistics. Visual Ontology was created with respect to the image. Bag-of-visual word based image indexing was proposed where in the image is segmented into patches and each patch is given a visual word to provide semantics of visual content. Visual ontology and domain ontology is combined to create multi-modal ontology. Also fuzzy concepts with respect to cognitive spatial relationship like touch, overlap, disjoint was also incorporated to identify action on the image. The overall accuracy in identifying the image was improved compared to other systems.

Proceedings of the 12th INDIACom; INDIACom-2018; IEEE Conference ID: 42835 2018 5th International Conference on "Computing for Sustainable Global Development", 14th - 16th March, 2018 Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA)

Table I: Comparative study on fuzzy ontology based information ret rieval algorithms

Name of the paper	**Input**	**Algorithms and Technologies used**	**Measures used**	**Output**
Ontology representation and mapping of common fuzzy knowledge.	Bibliographic of Ontology set benchmark from OAEI(Ontology Alignment Evaluation Initiat ive)	Ours (Proposed algorithm on Ontology m ap ping), Harm on y, Reliability, Sigmoid, Property, Instance, Structure, Avg, Min, Max, Concept name, SupportVector Machine, Neural Networks.	Precision, Recall, f-measure.	The algorithms Harm on y, Reliability, Sigmoid, Avg outperforms Concept Name, Property, Instance and Structure and has value of three measures greater than 54%. Ours, Support Vector Machine, Neural Networks perform better than the above algorithms and the value of three measures is more than 76%. f-measure value for Ours is the highest and is greater than 80%. The Standard Deviation of Precision and Recall is less for Ours compared to all other algorithms.
Hybrid Fuzzy-Ontology Design using FCA based	1000 scientific text Documents from IEEE website on the research	Standard FOGA, Keyword matching,	Precision, Recall, f-measure.	The evaluation result shows that the Precision and f-measure value for Hybrid

Clustering for Information Retrieval in Semantic Web	area “Information Retrieval” 	 Hybrid FOGA with keyword.		FOGA with keyword is 0.89, 0.85 which is greater than other two algorithms. The Recall value is greater for Standard FOGA i.e. 0.87. It can be concluded that associating Hybrid fuzzy ontology with keyword has optimized the Ontology definition and the three measures.
Semantic Information Retrieval Model: Fuzzy Ontology Approach	Input can be a set of documents or user query.	Expansion algorithm, Ranking algorithm,	Recall, Relevance Degree, Confidence Degree, Updating Degree.	Multi-view fuzzy ontology is used to deal with multi-field topics problem. The user query is expanded using expansion algorithm which has enhanced the Recall measure. The resulted set of documents is ranked based on the relevance degree with respect to user’s query, confidence degree and updating query .
A Semantic Information Retrieval Approach Based on Rough Ontology	1,500 periodical Metadata from the Operating system Category in the Website CNKI	Rough Ontology ba se d I R, Ontology based IR(OntoSCORM), Keyword based IR(Lucene).	f-measure	Rough ontology-based technology has greater f-measure than other two technologies for all 10 test queries.

V. Conclusion And Future Scope

The significance for improving the performance of Information Retrieval has been emphasized. As fuzzy ontology, fuzzy logic and semantic web provides a catalyst to improve the efficiency of IR, various papers under these topics were surveyed to bring out the latest techniques, methodologies and problems confronted by Information Retrieval to light. Also, comparative study of algorithms have been tabulated which highlights the methodologies and the metrics currently used in Ontology Based IR, Fuzzy Ontology Based IR and Fuzzy Logic Based IR. These techniques can be employed for solving the problems of IR in Educational domain. In Educational Domain, there are many branches of study and a plethora of courses to learn. In many cases, some content in one course can appear in other course and their semantics changes as context changes. This mushrooms a huge challenge to Information Retrieval. Depending on the user’s query, the IR algorithm should properly interpret the meaning with respect to the user’s context. Future work focuses on developing a framework to puzzle out the aforementioned challenge present in the education domain.

References

[1] Y. Sure, S. Staab and S. Studer, “Ontology Engineering Methodology”, *Springer,* pp. 135-152, May 2009.
[2] S. Calegari, D. Ciucci, “Towards a Fuzzy Ontology Definition and a Fuzzy Extension of an Ontology Editor”, *ICEIS 2006,* Vol. 3, pp. 147-158, May 2006.
[3] P. Alexopoulos, M. Wallace, “Improving Automatic Semantic Tag Recommendation through Fuzzy Ontologies”, *IEEE Press,* December 2012.

Received: 16 Nov 2018 | Revised: 25 Dec 2018 | Accepted: 10 Jan 2019

[4] P. Alexopoulos, M. Wallace, K. Kafentzis, D. Askounis, "IKARUS-Onto: a methodology to develop fuzzy ontologies from crisp ones", *Knowledge and Information Systems, Springer*, Vol. 32, pp. 667-695, November 2011.
[5] J. Hu, Xinzhou. Lu, C.Guan, "A Semantic Information Retrieval Approach Based on Rough Ontology", *The Open Cybernetics and Systemics Journal*, Vol. 8, pp. 399-404, December 2014.
[6] S. Kumar, R.K. Rana, P. Singh, "Ontology based Semantic Indexing Approach for Information Retrieval System", *International Journal of Computer Applications*, Vol. 49, July 2012.
[7] S. Meenakshi, R.M. Suresh, "A Study and analysis on Web Information Retrieval System for Distributed Environment", *International Journal of Applied Engineering Research*, Vol. 11, pp. 2165-2176, 2016.
[8] M. Baziz, M. Boughanem, Y. Loiseau, H. Prade, "Fuzzy Logicand Ontology-based Information Retrieval", *Fuzzy Logic, Vol. 215, pp. 19 3-2 18, 20 07.*
[9] H.B. Styltsvig, "Ontology based Information Retrieval", *A dissertation for the Degree of Doctor of Philosophy, Roskilde University*, May 2006.
[10] R.S. Khokale, M. Atique, "Web Based Information Retrieval using Fuzzy Logic", *JSCSE 3*, Vol. 3, pp. 62-68, March 2013.
[11] Y. Gupta, A. Saini, A.K. Saxena, "A new similarity function for information retrieval based on fuzzy logic", *International Conference on Advances in Computing, Communications and Informatics, IEEE Press*, September 2014.
[12] J. Singh,A.Sharan,"A new fuzzy logic-based query expansion model for efficient information retrieval using relevance feedback approach", *Neural Computing and Applications, Springer, pp 1–24*, February 2016.
[13] Y. Gupta, A. Saini, A.K. Saxena, "A new fuzzy logic based ranking function for efficient Information Retrieval system", *Expert Systems with Applications. 42*, pp. 1223–1234, February 2015.
[14] M. Hourali, G.M. Montazer, "An Intelligent Information Retrieval Approach Based on Two Degrees of Uncertainty Fuzzy Ontology", *Advances in Fuzzy Systems. 7*, Vol. 2011, August 2011.
[15] K. Balasubramaniam, "Hybrid Fuzzy-ontology Design Using FCA Based Clustering for Information Retrieval in Semantic Web", *2nd International Symposium on Big Data and Cloud Computing, Procedia Computer Science. 50*, Vol. 50, pp. 135-142, May 2015.
[16] Z.E. Attia, A.M. Gadallah, H.M. Hefny, "Semantic Information Retrieval Model: Fuzzy Ontology Approach", *International Journal of Computer Applications 13*, Vol. 91, April 2014.
[17] G. Nagarajan, R.I. Minu, "Fuzzy Ontology Based Multimodal Semantic Information Retrieval", *International Conference on Intelligent Computing, Communication & Convergence, Procedia Computer Science 48*, pp.101 – 106, May 2015.
[18] J.Liu, et al., "Ontology representation and mapping of common fuzzy knowledge", *Neurocomputing*, Vol.215, Pp.184-195, November 2016.

Received: 16 Nov 2018 | Revised: 25 Dec 2018 | Accepted: 10 Jan 2019