International Journal of Latest Trends in Engineering and Technology
Special Issue SACAIM 2017, pp. 122-126
e-ISSN:2278-621X

ONTOLOGY-BASED INFORMATION RETRIEVAL MODEL FOR INDIAN SPICES

Mr. Ruban S[1], Mr. Pranam[2] & Mr.Sunil Suraj D'Souza[3]

Abstract- The internet as we know is going to change. It is a massive global network which allows people to communicate and share data with each other. The Information which is shared is all saved as the huge data, hence accessing of this data whenever needed becomes important. There are many ways for retrieving these data but a retrieving method is not so efficient. The widely used information retrieval method based on the key-words is very far from user's satisfaction in recall and precision. The method to retrieve this information should be updated such that information retrieval through search engine becomes easier and more efficient. Solution to improve this information retrieval about Indian Spices E-Commerce is presented in this paper. Indian Spices E-Commerce ontology based on Information retrieval is constructed by gathering the information about the Indian spices and implemented. The Indian Spices Ontology consists of some kind of Spices of their family and their taxonomy. The taxonomy is further divided into different categories needed by the user. Indian E-commerce ontology has more functions than the keyword-based retrieval. Certain experiments are made to check the results of information retrieval.
Keywords- Information Retrieval, Ontology, Indian Spices.

1. INTRODUCTION.

The Internet an inter-connected network generally consists of huge data stored as the form of information. The stored data has become the huge information for the retrieval and has become part of the economy. Information stored in this vast network must be retrieved effectively whenever required by the users. There are many methods to retrieve this information, but whenever the information is more the effective means of recovering this information required. The commonly used keyword–based retrieval is re-framing the strings given in the search engine and finding information matched to this keywords present in that string. Matching the keywords and recovering the data may not give the effective information needed by the user, in the other words the keyword-based information recovering may not meet the user's needs. One of the means to solve this problem of recovering the required information through the web is by matching these strings with the semantic web, which will retrieve data related to the meaning rather than matching key-words by one another. In this method of information retrieval from web-engine, data can be converted into a high-level conceptual level of retrieving information this is known as semantic retrieval.

Semantic retrieval is the tool to improve accuracy by understanding the searcher user's intent and the related meaning of terms as they appear in the search able data space on the web in order to generate more relevant results to the user. Semantic-based search systems consider various points including the context of search in keywords location related the variation of key-words, generalized, synonyms, and specialized queries, matching concepts and natural language queries to give relevant search results to the users through the search engines. Major web information retrieval search engines like Google, Yahoo, Baidu and Bing incorporate some of the elements of semantic search.

In order to recover the information accurately and effectively, the concept of retrieving information has become important. It was one of the awaited problems to be solved. At Present scenario, the domain based ontology serves as the main form of the semantic web by providing necessary meaning and related data of a given domain to make information sharing and exchange easier. The criteria to overcome the disadvantages of key word based search for the Information retrieval system semantic was explained by several types of research and projects based on/in ontology-based information retrieval. The efficiency of information retrieval has made ontology technology to be widely used. The domain ontology helps to find the specific meaning and understating the meaning of these words depending on the level of taxonomy to given strings. Domain ontology basically understands the given key-words and gives back specific information. It provides users with meaning full extension and information's so that the needs of the consumer can be improved or achieved.

In this paper, we have described information retrieval of Indian spices, a web page based domain ontology. The ratio of information on ontology is found higher than the information retrieved from the general key-word based search. First, the Indian spices ontology process describes how it has been structured constructed and implemented. Second, the information retrieval including key functions and modules are discussed. Then the search string is implemented with wordnet to get the matching keywords, which will be combined with search string to get efficient result. Last, testing and evaluation are mentioned with the appropriate examples related to Indian Spices.

[1] ASST PROFESSOR, Aloysius Institute of Management and Information Technology (AIMIT), Beeri, Mangalore, Karnataka, India
[2] ST Aloysius Institute of Management and Information Technology (AIMIT), Beeri, Mangalore, Karnataka, India
[3] ST, Aloysius Institute of Management and Information Technology (AIMIT), Beeri, Mangalore, Karnataka, India

2. RELATED WORKS

Query reformulation is a procedure of adjusting or altering an underlying query to enhance search results and retrieval performance. It is an iterative procedure amongst users and Search Engine in which user draw in to discover useful data that could fulfil their inquiry objectives. This is on account of powerful query reformulation can improve the search results. It has been demonstrated by works done by [1], [2] and [5].

In the early years of query reformation deep search, Fidel [3] has distinguished three problematic circumstances that guided users to reformulate query: (1) retrieved results were too lengthy; (2) retrieved results were too small; or (3) retrieved results were not matching exact requirement. Fidel additionally has characterized query reformation into two classes of query reformation: operational and conceptual (theoretical). In 1999, Lau and Horvitz [4] has built up a strategy to characterize query reformation naturally into four restrictive sorts; speculation, new, reformulation and specialization. This scientific classification was additionally utilized by numerous different analysts to identify seek limits in look log naturally.

At that point Rieh and Xie [2] broadened the Fidel's [3] classifications from two to four classes in 2006. They analyzed query reformation by looking at the syntactic and semantic viewpoint. They found that, there are some basic methodologies being utilized by the users to perform query reformation, for example, specification, generalization, replacement and parallel movement.

The research of query reformation sort has been extended when Huang and Efthimiadis [5] distributed a broad scientific classification comprising of twelve query reformation types. Past researches particularly in recognizing the classes of reformulations have motivated momentum analysts to investigate more methods in query reformation field.

3. MATERIALS AND METHODS

*3.1 Query Set***:** Our query set contains 50 queries. The queries are all related with domain (Indian Spices domain). The queries were gathered from online blogs, and Indian Spice Board. All the queries had undergone cleaning process to remove unrelated, incomplete and redundant queries. Table I shows some of the queries in our collection.

Table 1 Queries

Query No	Initial Tested Query
Q1	Indian Spices
Q2	Usage of Indian Spices
Q3	Properties of Indian Spices
Q4	Family names of Indian Spices
Q5	Indian spices belonging to the family Zingiberaceae
Q6	Botanical Names of Indian Spices
Q7	Spices and its category
Q8	Spices Grown in Tamil Nadu Only

3.2 Search Engine

Any web engine can be used to test our approach. For our situation we picked the most mainstream web search tools like Google to be our test bed.

3.3 Ontology

We utilized Indian Spice (Spice Ontology) in our proposed strategy. Indian Spice has been made in the before stage [8]. It has been made utilizing Information provided by Spice Board of India. Every one of the actualities in Spice has been gathered from trusted sources.

3.4 Evaluation Metric

With the end goal of assessment, standard exactness and normal accuracy computation was utilized to quantify the retrieval performance.

*Methods***:** We proposed a system that consolidates the query's keyword and the ontology terms. We execute four arrangements of analysis keeping in mind the end goal to analyze the execution of the proposed strategy. Right off the bat we present the original query to the web search tool as a benchmark. Also we reformulate the original query by simply substituting it with catchphrase from the original query. Thirdly we reformulate it into ontology term alone. Ultimately we consolidate the watchword and the ontology terms. At that point we break down the outcomes utilizing the exactness and review measures.

4. IMPLEMENTATION

This area quickly portrays the usage of the proposed system. The entire test query being submitted to the search engine accordingly based on the proposed technique. Then, fifty outcomes were recorded for each of the submitted queries. The outcomes were inspected and broke down utilizing the exactness and review computation. Tests of retrieved outcomes are

appeared in Figure 1. Every single beginning query (unique query) is in natural language. The underlying queries were reformulated into new queries as per the recommended techniques.

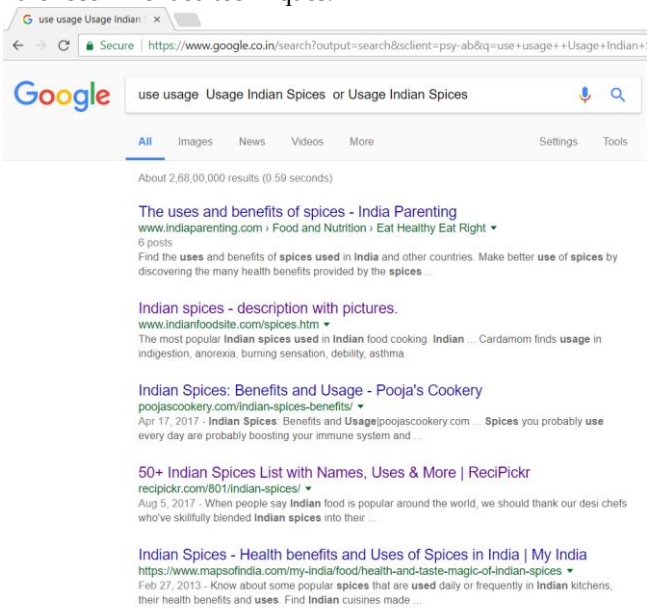


Figure I. Screen Shot Of Retrieval Results

For the first method, the first query was submitted to the web index and afterward the retrieved outcomes were caught. Case of query being submitted is "Usage of Indian Spices". For the second method, original query and keywords were submitted to the web crawler. In view of the case above, keywords submitted were "use usage Usage Indian Spices or Usage Indian Spices". Those keywords were separated from the first query utilizing Stanford Parser. Stanford Parser, a natural language parser is a program that works out the grammatical structure of sentences, for instance, which groups of words go together (as "phrases") and which words are the subject or object of a verb.

Table 2. Reformulated Queries

SL.No	Original query	Original + Keywords	Original+Keywords+Ontology Terms
Q1	Indian Spices	Indian Indian Spices or Indian Spices	Indian Spice Indian Spices or Indian Spices
Q2	Usage of Indian Spices	use usage Usage Indian Spices or Usage Indian Spices	Use Indian Spice usage Indian Spice Usage Indian Spices or Usage Indian Spices
Q3	Properties of Indian Spices	property Properties Indian Spices or Properties Indian Spices	Property Indian Spices Properties Indian Spices or Properties Indian Spices
Q4	Family names of Indian Spices	family Family Indian Spices or Family Indian Spices	family Indian spices Family Indian Spices spice family or Family Indian Spices
Q5	Indian spices belonging to the family Zingiberaceae	spice Zingiberaceae family Indian spices belonging family Zingiberaceae or Indian spices belonging family Zingiberaceae	spice Zingiberaceae family Indian spices Zingiberaceae belonging family Zingiberaceae or Indian spices belonging family Zingiberaceae
Q6	Botanical Names of Indian Spices	botanical name Botanical Names Indian Spices or Botanical Names Indian Spices	botanical name spice Botanical Names Indian Spices or Botanical Names Indian Spices
Q7	Spices and its category	category Spices category or Spices category	category Indian Spices category or Spices category
Q8	Spices Grown in Tamil Nadu Only	grow Tamil Spices Grown Tamil Nadu Only or Spices	Spices grow in Tamil Spices Grown Tamil Nadu Only or

		Grown Tamil Nadu Only	Spices Grown Tamil Nadu Only

For the third method we substitute the second query with terms from ontology. Cases of the submitted terms were "Use Indian Spice usage Indian Spice Usage Indian Spices or Usage Indian Spices". This strategy made an interpretation of the first query into SPARQL query beforehand. At that point it was utilized to query the domain ontology. Answers given by the ontology (ontology terms) was utilized to substitute the first query. We did the investigation by presenting the query one by one as per the queries number for all the recommended methods. At that point all the recovered outcomes were recorded accordingly. Tests of the reformulated questions as indicated by the recommended strategy are appeared in Table II.

5. RESULTS & DISCUSSION

The popular IR evaluations metric is the precision and recall calculation. It is respectively formulated as in equation 1 and 2:

$$Precision = \frac{Number\ of\ relevant\ links\ retrieved}{Number\ of\ links\ retrieved} \quad (1)$$

$$Recall = \frac{Number\ of\ relevant\ links\ retrieved}{Number\ of\ relevant\ links\ in\ the\ collection} \quad (2)$$

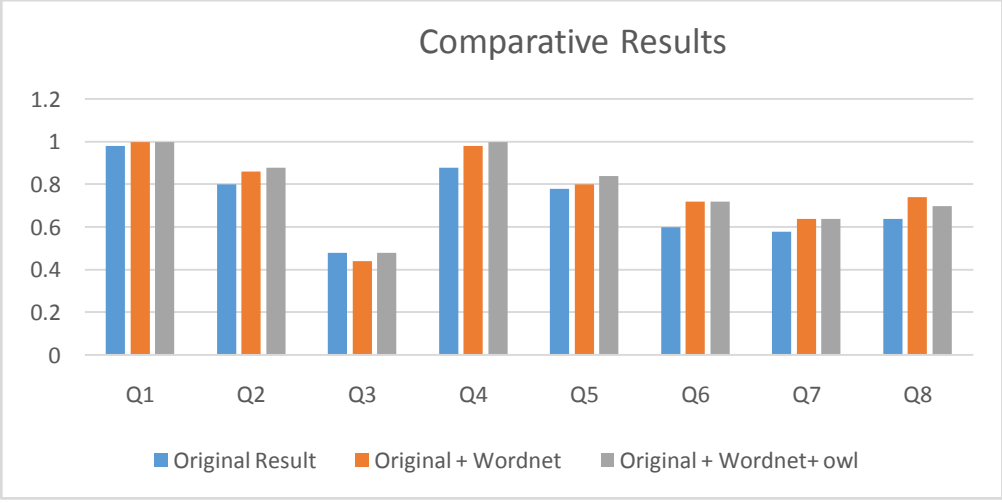
In a typical controlled experiment, both precision and recall value can be effortlessly computed. We can precisely know the quantity of relevant archives retrieved, the quantity of reports retrieved and the quantity of important records in gathering. However for this experiment, we can just figure and measure the precision value. The recall value can't be computed since this trial utilized Google as the test condition, which is inconceivable for us to know the correct number of applicable records in the whole Google's collection.

Along these lines, we logged the quantity of retrieved connections for each query submission session as indicated by the regarded strategies to speak to the recall. Despite the fact that the aggregate number of retrieved connections given by Google is said to be just as estimation figure, however at any rate it can give us a general thought of the recall value for the given query.

Table 3.Show the total and average of retrieved connections.

Queries	Original Result	Original + Wordnet	Original + Wordnet+ owl
Q1	0.98	1	1
Q2	0.8	0.86	0.88
Q3	0.48	0.44	0.48
Q4	0.88	0.98	1
Q5	0.78	0.8	0.84
Q6	0.6	0.72	0.72
Q7	0.58	0.64	0.64
Q8	0.64	0.74	0.7
Average	0.7175	0.7725	0.7825

Table 3. Precision Value Of Queries



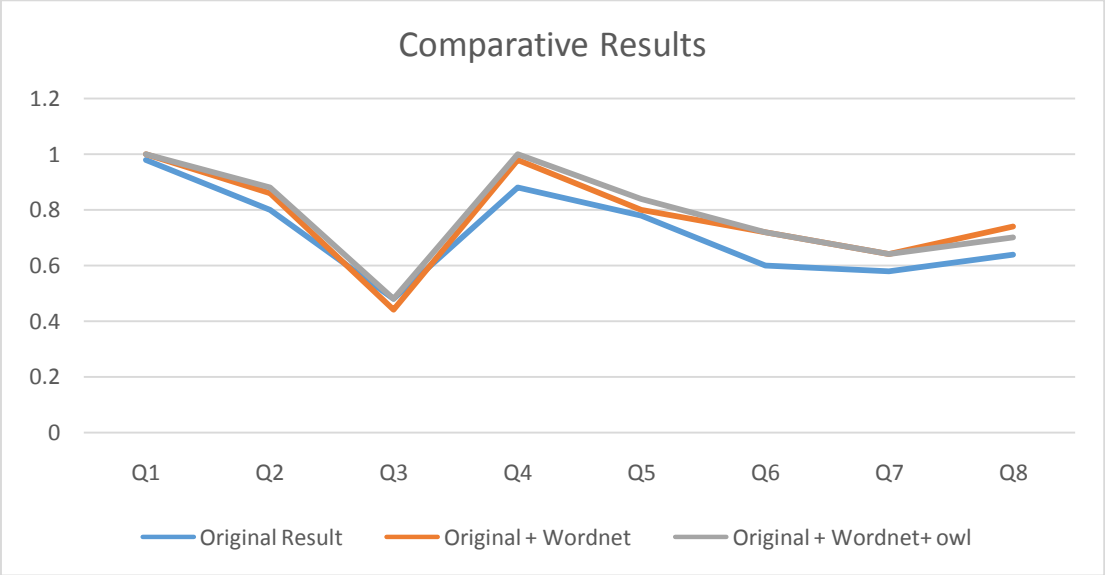


Figure 2. Google Search V/S Ontology based Search

It has been concluded that "there is significant difference between Google search and Ontology search." From Table III, it is observed that precision using ontology is 0.7825 mean for precision used for Google 0.7175.So, it is statistically confirmed that precision is improved using the proposed query refinement.

6. CONCLUSION

The goal of this works is to distinguish better query reformulation strategies in retrieving better outcomes for the predetermined area. In this paper, test information and results for three unique strategies were illustrated. This study concentrated on the utilization of area particular to ontology in reformulating new query. The outcomes demonstrated that blend of keywords and ontology terms give better precision, contrasted with keywords and natural language techniques. Despite the fact that part of research is being done to build up a unified global ontology, it is yet to wind up plainly a reality.

7. REFERENCES

[1] B. J. Jansen, D. L. Booth, and A. Spink, "Patterns of Query Reformulation During Web Searching," vol. 60, no. 7, pp. 1358–1371, 2009.
[2] S. Y. Rieh and H. (Iris) Xie, "Analysis of multiple query reformulations on the web: The interactive information retrieval context," Inf. Process. Manag., vol. 42, no. 3, pp. 751–768, May 2006.
[3] R. Fidel, "Moves in Online Searching," Online Rev., vol. 9 (1), pp. 61–74, 1985.
[4] T. Lau and E. Horvitz, "Patterns of Search:Analyzing and Modeling Web Query Refinement," Proc. User Model. Conf., pp. 119–128, 1999.
[5] E. N. Efthimiadis, J. Huang, A. Spink, and J. Jansen, "Query formulation in web search," Proc. ASIST Annu. Meet., vol. 46, 2009.
[6] A. Azizan, Z. A. Bakar, and S. A. Noah, "Analysis of Retrieval Result on Ontology-Based Query Reformulation," in IEEE 2014 International Conference on Computer, Communication, and Control Technology (I4CT 2014), 2014, no. I4ct, pp. 244–248.
[7] "Stanford Parser." Link: https://nlp.stanford.edu/software/lex-parser.shtml
[8] S. Ruban, Sunil Suraj D'Souza and Pranam "An Ontology for Knowledge Exploration in Indian Spices" in International Journal of Latest Trends in Engineering and Technology, Special Issue SACAIM 2016, pp. 388-396, e-ISSN:2278-621X
[9] An Ontology-Based Information Retrieval Model for Domesticated Plants Ruban S, KedarTendolkar, Austin Peter Rodrigues, Niriksha Shetty
[10] Teng-yang TAO and Ming ZHAO ,"An Ontology-Based Information Retrieval Model for Vegetables E-Commerce",pp.800-807, May 2012