International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 3 Issue: 8
ISSN: 2454-4248
139 – 148

Ontology Based Semantic Web Information Retrieval Enhancing Search Significance

Kinjal Sheth
Research Scholar
C. U. Shah University
Wadhwan City, Gujarat, India

Dr. Harshad Bhadka
Dean - Computer Science
C. U. Shah University
Wadhwan City, Gujarat, India

Dr. Ashish Jani
Associate Professor
School of Engineering
P P Savani University,
Surat, Gujarat, India

Abstract - The web contain huge amount of structured as well as unstructured data/information. This varying nature of data may yield a retrieval response that is expected to contain relevant response that is expected to contain relevant as well as irrelevant data while directing search. In order to filter out irrelevance in the search result, numerous methodologies have been used to extract more and more relevant search responses in retrieval. This work has adopted semantic search dealing directly with the knowledge base. The approach incorporates Query pattern evolution and semantic keyword matching with final detail to enhance significance of relevant data retrieval. The proposed method is implemented in open source computing tool environment and the result obtained thereof are compared with that of earlier used methodologies.

Keywords*: query pattern evaluation, structured, semantic keyword matching, unstructured*

*****

I. INTRODUCTION

The data or the information denotes the fact that there is some relevant information which is signified or coded in particular form. The digital data refers to any of the characters or symbols on which the specific operations are performed by means of a computer and it could be stored, recorded, or transmitted. Today, the search that people does with the aid of search engine such as Google, etc. search for the specific keyword in order to satisfy the queries of the users (Windsor et al. 2016).These search engines search for unnecessary pages because the main focus of these search engines to solve the queries close to accurate result and most of the time, the user did not get the required information. The problem semantic web technology plays an important role in retrieving meaningful information. With the increase in technologies, the requirement of the digital data has made the internet access such as the World Wide Web (WWW) one of the most important platform for the assortment of relevant documents (Athukorala et al. 2016). As searching and surfing documents over internet have become an integral part of people's life, there is an increasing demand of retrieving such documents from a large source of information which were relevant to the information referred. In order to cope up with the demand and the supply paradigm, the Information Retrieval (IR) gains importance day by day. Accuracy and speed were the two basic requirements of the effective retrieval methodologies (Jeffrey et al. 2011). The information retrieval process starts when a user enters a particular query into the system. The user queries are matched against the database information. In information retrieval process, the results returned may or may not match the query given, so the queries must be typically ranked. The ranking of the results is a key difference of information retrieval searching while compared to the database searching. Depending on the application, data objects may be of text data, audio, video or mind maps (Kamvar et al. 2012). The documents or the data are not directly stored in the Information Retrieval system, but they are as an alternative, represented in the system by a metadata. Some of the commonly used document retrieval systems are statistical methods and the Natural Language Processing (NLP) approaches which uses different document representations and the query structures.

The semantic search seeks to improve the search accuracy by understanding the searcher's intent and the meaning of the terms as they appear in searchable data space (Fernandez et al. 2011).Retrieval of the documents online is of wide interest in the Information Retrieval community. The document retrieval actually refers to finding such documents which were similar for a given user's query. A user's query may be of a complete sentence or just a few keywords. Among the widely used retrieval methods used by several search engines like Google, where the unskilled users provide only a few keywords to the search engine and the search engine in return, provides a list of relevant documents which are available online. Another way of document retrieval is to use the context of the query given by applying certain language models, which incorporates several contextual elements so that the document could be ranked based on the context relevant to the user's query (Campos et al. 2015). The usefulness of the keyword search process model gets limited by the phenomenon known as "keyword barrier", which states that the precision of results of any information retrieval systems cannot be guaranteed when it is based on a set of words extracted from the document through any of the syntactic techniques. Other than such semantic retrieval process models which are capable of abstracting the representations of the actions, the products and techniques involved in such retrieval processes

are yet to be designed (Denker et al. 2015). The IR systems suggested for handling the technical queries are the Boolean, the vector space, and the probabilistic models.

Ontology is an official overt description of a collective conceptualization. The ontology provides a common understanding of a particular term along with its relationship with the other terms. The Semantic Information Retrieval has become the major part of any search (JingJing et al. 2013). Dynamic Semantic Engine implements a context driven approach in which the keywords are processed in the framework of information in which they are retrieved, to solve the semantic ambiguity and also to obtain a more accurate retrieval based on the user interests. The Semantic Information Retrieval System is mainly concerned with retrieval of information from a sports ontology using a SPARQL query language. The specific information is retrieved from the ontology (Song et al. 2013). A simple semantic based information retrieval process is incomplete till the time a semantic similarity measure was developed. By knowing the actual terms of a query, gives us the key terms for which the information is to be searched, and the user will be annotating such documents which contains the key terms to provide ease to the search engine (Biancalana et al. 2013). Being a new field of research, not many works has been carried out and also there are only very few developers who has actually performed certain works on irrelevant terms. This research works paves the way to propose an adaptive updatemodel for information retrieval and to propose a semantic search model for the given user query and create rank based simultaneous history and knowledgeupdate models.

II. LITERATURE REVIEW

Jain et al. (2013) suggested a new way to the users for extracting the information from the web. As there were huge number of documents presented on the web and to retrieve the relevant information from them was a tedious task. It generated the concept of information retrieval and semantic web. The semantic markup documents were used for extracting the information from web documents. According to Vigneshwari et al. (2015), the data sets were generated with the aid of both SWETO and WordNet. The cross ontology was performed with hashing alone and without hashing and also with both semantic annotation and with hashing. The test was conducted by utilizing 50 queries and the hybrid approach involved both semantic annotation and with hashing yielded better performance while compared to other approaches. Rodriguez et al. (2014) proposed a semantic platform for cloud service annotation and retrieval from their description. The system automatically annotated different cloud services from their natural language description, which was available in a number of document formats such as XML, HTML, or PDF. The research method was additionally implemented, by considering multi ontology environment in order to be able to cope up with several domains.

Chauhan et al. (2013) implemented a domain specific semantic information retrieval system by using appropriate tools and the proposed technique of ontology based automatic query expansion. It utilized the query concept as well as the synonyms of the concepts to perform query expansion. The new terms were added only for a consistency of a similarity within a threshold. Only the most relevant document acquired the top rank. Singh et al. (2014) emphasized the concept of semantic web and several approaches used for retrieving information from the web. The Information Retrieval over the collection of those documents offered new challenges and opportunities. This study presented a framework for integrating the search that supported Inference engine. Solyu et al. (2016) presented a novel multi paradigm and ontology based VQS, named as Optique VQS for the end users with no technical knowledge and skills. It was built on a powerful OBDA framework and had a flexible and extensible architecture that allowed to combine and orchestrate different representation and interaction paradigms. The results of the usability experiment suggested that Optique VQS provided a decent level of expressivity and high usability. Muller et al. (2016) proposed that the generic concepts can be mapped between the two ontologies, or the generic concepts be completely discarded. The mapping of the generic concepts links shared the concepts without unnecessary noise. The removal of the generic concepts resulted in an information loss. related all the variables and the associated values to terms from the interoperable ontologies listed at the OBO foundry. OBO foundry topologies provided the benefit of wide coverage but it could also be selectively imported in order to create an application ontology such as the EuPath ontology. When the existing terms were not available for mapping, the new ones were created for introduction into the source ontology or just placed in the application ontology. Bansal et al. (2016) designed a novel approach of ontology based information retrieval system for classified ads. The ads database was taken for the house data. Various features were extracted using the ontology based rules which has not been dealt in the past. The results obtained were found to be quite advantageous and proved the effectiveness of the proposed algorithm. Foschini et al. (2016) introduced CDDI platforms which was not only able to grant timely and reliable retrieval of the queried context data, but also to guarantee the required scalability and trustworthiness.

Ca et al. (2014) defined and solved the problem of multi keyword ranked search over the encrypted cloud data, and established a variety of privacy requirements. Among the multi keyword semantics, the efficient similarity measure of the “coordinate matching” was chosen. In order to meet the challenge of supporting multi keyword semantic without privacy cracks, the basic idea of MRSE using the secure inner product computation was proposed. Two improved MRSE schemes were given to achieve various stringent privacy requirements in two different threat models. Nasution et al. (2016) showed that a well-known paradigm of a querying a document web is simpler for accessing by inputting the keyword. Bourgonjeet al. (2016) addressed the issue of combining the NLP, IR, and the MT procedures into a system that enabled knowledge workers to explore a collection of documents in an intuitive and efficient way. The study focused the combination of the individual components and linked the output of the methods, rather to improve the output of individual state of art procedures. The

 ISSN: 2454-4248
139 – 148

information contained in multiple documents were aggregated and presented in a way that allowed the knowledge workers to view the documents. A locational reference deriving model and associated prototype preprocessing layer that has the potential to promote critical spatial thinking by expanding the data source options was presented. The model's integrity was restricted by the credibility of the data retrieval sources, and limited to handling the vector data. Zhu et al. (2017) designed and implemented a framework of building a natural language interface to a graph based bibliographic information retrieval system. The framework allowed the user to query bibliographic information by formulating and replying the queries represented in natural language. An important step in interpreting natural language query was to recognize the bibliographic named entities in natural language queries. The framework was tested using a large empirical dataset and the experimental results showed that the method correctly interpreted 39 out of 40 natural language queries with several levels of complexities.

III. ONTOLOGY DRIVEN SEMANTIC SEARCH ENGINE

In existing approaches, a centralized database was applied for indexing information in traditional keyword search systems. These were based on queries from simple keywords, and the recall rate was high with a little accuracy. Those mentioned above was due to the disambiguation, wrong context, and the use of synonyms. These systems seldom considered the semantic content of the document to the index (Wu et al.,2012). Therefore, a distinct approach is required that examines the semantics of the document and focuses on techniques for information retrieval based on ontologies. Ontology based semantic search is described to be an information retrieval process that applies the background knowledge of the domain ontology. The objective of ontology based semantic web information search is to maximize precision and recall where,

$$Precision = \frac{Total\ number\ of\ relevant\ documents\ retrieved}{Total\ number\ of\ documents\ retrieved} \quad (1)$$

$$\text{Recall} = \frac{Total\ \text{number of documents retrieved}}{Total\ \text{number of relevant documents}} \quad (2)$$

Presently, various classification criteria are used to categorize different approaches for ontology based search along with several modifications in the methods. The following classification criteria achieve essential characteristics of search process: Indexing, ontology technology, ranking, semantic annotation, Information retrieval model, and performance improvements.

A. Indexing

Indexing is the method of storing the information for an efficient retrieval upon a search query. A Search engine maintains all the contents found during the crawling process, and stores it in an index for the easy retrieval. The purpose of indexing is to advance the matching in an index (Fu et al.,2016). Further, the extraction process requires streaming through the collected web pages. The different types of indexing are as follows:

1. Forwarded index: The list of words for each document will be stored.
2. Inverted index: The list of documents for each word will be retained in this index.
3. Graph indexing: Given a query graph, looks up in an index and retrieves the set of answers to verify the graphs that contain the query graph and returns the query results.

B. Ranking

The ranking is the method that determines the ordering of results of a search query. The search requires matching and ranking. Where in matching a subset of the elements is selected to be scored. The ranking determines the degree of matching using some perceptions of relevance. The ranking is performed after the syntactic or semantic mapping is executed. The rank will be calculated depending on the score of the web pages (Wang et al.,2014). The results from the ranked web page are indexed before returning to the web user. The different type of ranking model is as follows:

1. Syntactic Ranking Model: The search relies on term matching among the query and the engine database.
2. Semantic Ranking Model: The is based on the result significance which will be achieved by bridging the gap between the syntax and semantics that provides a focused result and better satisfaction to the user.

C. Information Retrieval Model(IR Model)

The purpose of IR model is to provide a formalization of information determination process (Chen et al.,2013). There are three types of IR models as follows,

1. Boolean model – The keyword manipulation is utilized to represent a document as a combination of keywords, and the query will be represented by a logical expression composed of words.
2. Vector model – The user queries and documents are represented as vectors in the space.
3. Probabilistic model – The mathematical model based on the theory of probability is utilized.

The ontology based semantic search executes the search engine more reasonable by adding meaning and structure to the web pages and query. The use of ontology facilitates approaches to define concepts and relations representing knowledge of a particular domain.

IV. PROBLEM DEFINITION

The relationship between documents is vital to express a text adequately. Therefore, developing a method that can retain and utilize this connection is necessary to increase the speed during the search phase. In addition, data search results returned to the users may contain damaged data or might have been distorted by the malicious administrator, software or hardware failure, and storage corruption. Hence, the mechanism should be provided for users to verify the correctness and completeness of the search results. The search time can be considerably reduced by abandoning the irrelevant categories and selecting the desired category. When Compared with all the documents in the dataset, the

 ISSN: 2454-4248
139 – 148

number of records which user aims at is minimal as the limited number of the desired documents can be further divided into several sub-categories (Cao et al.,2011). Acknowledging a large number of on-demand data users and the vast amount of data documents outsourced in the cloud, this creates the problem to meet the requirements of performance, system usability, scalability, and, to meet the practical data retrieval requirements. The vast number of documents demands the cloud server to perform result relevance ranking, instead of returning similar results. This ranked search system facilitates data users to find the most relevant information instantly, rather than sorting through every match in the collection of data (Wang et al.,2012). The ranking system should support multiple keywords search to improve the accuracy as well and to enhance the searching experience. As a conventional practice, the users provide a set of keywords as an indicator of the search interest to retrieve the required data. Each keyword in the search request can further narrow down the search result.

The direct application of these approaches to the secure largescale data would not be plausible, as they cannot support high service-level requirements like system usability, user searching experience, and accessible information discovery. Significant techniques to design an efficient encrypted data search mechanism that supports keyword semantics without privacy breaches remains a challenging open problem. Query processing System is still a challenging field of research due to the issue of extraction for matching the query, the accuracy of the answers retrieved and performance in the retrieval of results. The fuzzy semantic search engine was previously used to extract effective information (Li et al.,2010). However, semantic keyword search increased the accuracy of the information retrieval. Thus, the development of History based and knowledge update for better improvement in information retrieval is necessary.

This is the core issue taken care for enhancing the relevancy level of information extraction

V. PROPOSED METHODOLOGY

This research work is derived from the observation that user's retrieval needs are concentrated on a specific field and proposed study could speed up the searching process by computing relevance score between the query and documents belonging to the specific field with the query. As a result, exclusive documents which are classified to the field specified by user's query are assessed, and the irrelevant fields are ignored thereby enhancing the search speed. Figure 1 shows the proposed architecture flowchart of semantic keyword matching based information retrieval system, in which the matching engine contains the phenomenon as keyword matching, semantic and evaluation of query pattern. The semantic keyword matching algorithm provides a way for matching on the semantic level, and the proposed matching algorithm combines the flexibility of keyword-based retrieval and owing to the ability to query and reason on meta data typical of semantic search systems, the accuracy could be improved. The algorithm for the proposed approach is as shown in Figure 2.

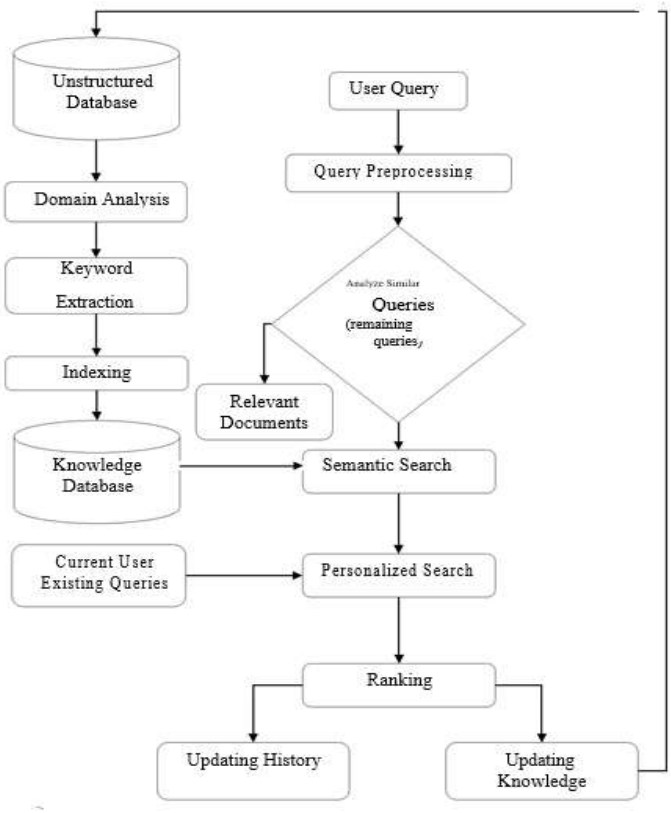

Figure1: Semantic keyword matching architecture

 ISSN: 2454-4248
139 – 148

D. Query pattern evaluation

Query formulation is significant to get required precision and recall. For web search engines usually, accuracy is more important than recall as the total number of relevant documents is immense to be considered by the user.As the number of documents is large in the case of the web, a large number of documents are often pertinent to the query which cannot be further ranked based only on the internal features of the document. Thus, the proposed algorithms are efficient, through which queries will be handled expeditiously. Furthermore, meta fields like "keywords" are often used to match the query and simplify the process.

E. Semantic Keyword search model

The objective of the suggested method is to address the semantics and descriptions of the web service as a whole which is not administered by the existing methods. In this proposed approach, the web service ontology framework is extended to establish an external database and implementing the matching algorithm to produce the precise results for web services. The figure 2 shows the proposed architecture of semantic based information retrieval system. In this study, matching engine contains the phenomenon as keyword matching, semantic and ontology algorithm. Keyword matching is utilized to return the results quickly with a limitation of understanding user's goals.The semantic matching algorithm provides a way for matching on the semantic level by combining the flexibility of keyword-based retrieval and owing to the ability to query and reason on meta data typical of semantic search systems; the accuracy will be improved.

The relevance feedback is one of the classical methods of refining search engine rankings. In this approach, search engine initially generates the initial set of rankings and users select the relevant documents within this ranking.Furthermore, based on the information in these documents a more consistent ranking is presented. These retrieved set of documents are commonly the documents for which similarity is greater than a threshold value. That is if a query has the highest similarity with any document on the lower side, the cut-off threshold can be brought down. Aforementioned, vector based model does not enforce that all retrieved documents should be an exact match of the query and allows only partial matches to be retrieved. Thus, by fixing the term-weights, documents, and query vectors in k dimension (where k is some index terms in vocabulary) and it is necessary to find the similarity between them. Hence, in this research work, a widely used measure of similarity called the cosine similarity is utilized.

The cosine similarity between two vectors $\vec{r_j}$(the document vector) and$\vec{s}$ (query vector) is given by:

$$similarity\left(\overrightarrow{r_j}, \overrightarrow{s}\right) = \cos\theta = \frac{\overrightarrow{r_j} \cdot \overrightarrow{s}}{\left|\overrightarrow{r_j}\right| \left|\overrightarrow{s}\right|}$$

$$= \frac{\sum_{k=1}^{n} A_k B_k}{\sqrt{\sum_{k=1}^{n} A_k^2} \sqrt{\sum_{k=1}^{n} B_k^2}} \quad (3)$$

Here θ is the angle between the two vectors, A_k is the term-weight for k^{th} term of the document.B_k is the term assigned to, k^{th} term of the query. Cosine similarity gives maximum value when θ = 0 or when the vectors coincide. It gives lowest value when the vectors are independent of each other. Here, $A_k B_k \geq 0$ for all values k the similarity between $\vec{r_j}$ and $\vec{s}$ varies from 0 to 1. The cosine similarity measure returns value in the range 0 to 1 which allows partial matching. Ranking of the retrieved results can be determined according to the cosine similarity score (Wu et al.,2012).

For semantic search,wordnet is used to find the synonym of the preprocessed query and then the similarity between query and documents is found using cosine similarity. These retrieved documents are provided as an input to personalized search. In personalized search, the current user's existing search query are matched using cosine similarity with the retrieved documents and ranking of the documents is done based on the similarity values. Thus, the current user query is updated in history which can be further utilized in the future. The user's present queries are matched within the retrieved documents and then the rank values are assigned to the documents. Furthermore, the history and knowledge of the database is updated. This will increase the accuracy of semantic based information retrieval.

 ISSN: 2454-4248
139 – 148

Algorithm Semantic keyword matching and query evaluation

```
D←load dataset
Do←domain
Pd←preprocessed dataset

For i=0: D do
    for j=0: Do do
        Pd←stop words and stemming
    end
end

Dk←distinct keywords
For i=0: Pd do
    For j=0: Do do
        Dk←select distinct keywords
    end for
end for
Fid←file ids
For i=0: Pd do
    For j=0: Do do
        Fid←assign file id
    end for
end for
I ←indexing
For i=0: Dk do
    for j=0: Pd do
        for k=0: Do do
            I←assign index
        end for
    end for
end for
Un ←user name
Uq←user query
Pq←preprocessed query
For i=0: Uq do
    Pq←stop words and stemming
end for
Sq←similar queries
H ←history
For i=0: H do
    Sq←analyze similar queries
end for

Sk←semantic keywords
Wn←wordnet
For i=0: Wn do
    Sk←semantic keywords
end

for
Mi ←matched index
For i=0: I do
    Mi←get matched index
end
for
Cs ←cosine similarity
for i=0: Sk do
    for j=0: Mi do

$$Cs_{ij} = \frac{\sum_{k=1}^{n} A_k B_k}{\sqrt{\sum_{k=1}^{n} A_k^2}\sqrt{\sum_{k=1}^{n} B_k^2}}$$

    end
end

Eq←existing queries
For i=0: H do
    Sk = Sk + Eq
end for


Ps ←personalized search
For i=0: Sk do
    For j=0: Mi do

$$Ps \leftarrow Cs_{ij} = \frac{\sum_{k=1}^{n} A_k B_k}{\sqrt{\sum_{k=1}^{n} A_k^2}\sqrt{\sum_{k=1}^{n} B_k^2}}$$

    end for
end for
for i=0: Ps do
    for j=i+1: Ps do
        if Psj.>Psi then
            swap (Psj, Psi)
        end if
    end
end
Ku ←knowledge updating
For i=0: Ps do
    Ku ←knowledge updating
end
```

Figure2. Algorithm for the Semantic keyword matching and query evaluation

The first step towards obtaining the semantics of the input is to transform keywords into semantic keywords. In this process, the system consults a pool of ontologies to extract the plausible meanings of each keyword, integrating the meanings that are comparable to avoid redundancy. Then, the system applies different disambiguation techniques to finally establish the significance of each keyword considering its context and the possible meanings of the rest of the keywords. Further, the documents analyzed based on the specified keywords are indexed and stored in the knowledge database. As the user interpolates the query, the query is considered as the input and is preprocessed by removing the stop words and stemming. The current user query is correlated with existing queries, and if these matches the relevant document is retrieved immediately, otherwise the present query is rendered as the input to semantic search. Besides, the semantic search is executed by utilizingWordnet to determine the synonyms of the preprocessed query. The cosine similarity is adapted to associate the query and the documents. These retrieved documents after matching the query are presented as an input to personalized search and further updated in the history. In a circumstance where the user is not satisfied with the documents, they can introduce the answer as an input to the database, and this is designated as knowledge updating.

VI. RESULTS AND DISCUSSION

The proposed technique of Query pattern evaluation, semantic keyword matching is implemented and tested on an experimental platform to examine the search efficiency, search time for determining the retrieved documents, keywords in the query, accuracy, privacy, and rank accuracy. The methodology will be implemented on IDE NetBeans 7.1 with a compiler JDK 1.7 using Front end core JAVA and back end MySQL on a database of Wamp server 2.0. The data set is given as the input and is preprocessed by employing stop words and stemming. Dataset Link provide the data set: https://archive.ics.uci.edu/ml/datasets/Twenty+Newsgroups. To compile this process POS Tagger is used.

The proposed research work is compared with the existing techniques and tabulated based on various parameters. Table

 ISSN: 2454-4248
139 – 148

1 determines the time required by Query pattern evaluation, Semantic keyword matching technique to search the documents in dataset and the values obtained is compared with existing techniques and plotted on graph as shown in Figure 3.

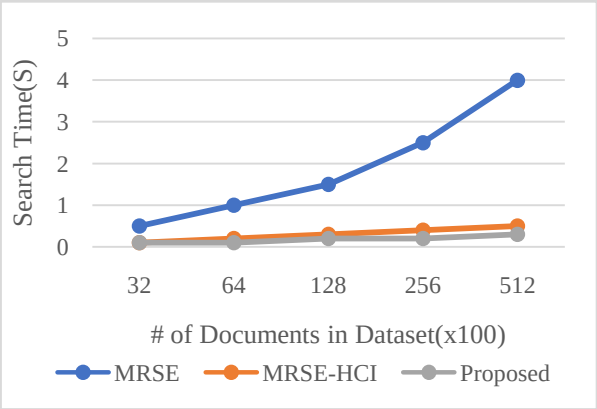

Figure 3. Search time for determining documents

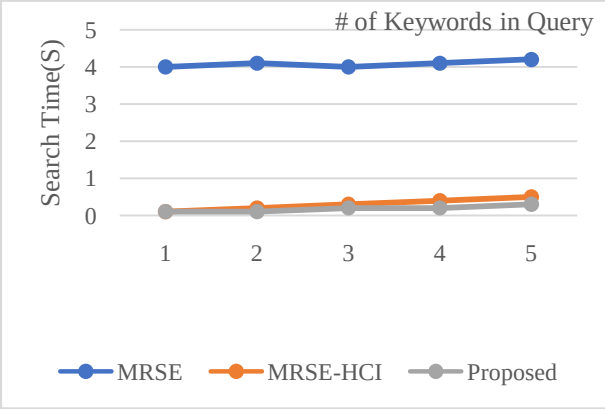

Figure 4. Search time for keywords matching

Table 1. Time for determining Documents

Number of Documents in Dataset	Search Time for determining Documents in Dataset		
	MRSE	MRSE-HCI	Query pattern evaluation, Semantic keyword matching
32	0.5	0.1	0.1
64	1	0.2	0.1
128	1.5	0.3	0.2
256	2.5	0.4	0.2
512	4	0.5	0.3

Table 2. Time for Keyword matching

Number of Keyword	Search time for Keyword matching		
	MRSE	MRSE-HCI	Query pattern evaluation, Semantic keyword matching
1	4	0.1	0.1
2	4.1	0.2	0.1
3	4	0.3	0.2
4	4.1	0.4	0.2
5	4.2	0.5	0.3

Figure4 shows time for Keyword Matching. Where X axis represents the number of searches while Y axis denotes time for keyword matching using both the existing algorithm and proposed algorithm. When the number of searches increases the time for keyword matching increases accordingly. Table 2 shows better performance of research work in terms of time for keyword matching than existing algorithms

Table 3. The ratio of documents in dataset

Number of Documents in Dataset	Ratio of Documents in Dataset		
	MRSE	MRSE-HCI	Query pattern evaluation, Semantic keyword matching
32	0.5	0.1	0.1
64	1	0.2	0.1
128	1.5	0.3	0.2
256	2.5	0.4	0.2
512	4	0.5	0.3

Table 4. The time required to retrieve data

Number of Retrieved Documents	Search time to Retrieve Documents		
	MRSE	MRSE-HCI	Query pattern evaluation, Semantic keyword matching
10	4	0.1	0.1
20	4	0.2	0.1
30	4.1	0.3	0.2
40	4.1	0.4	0.2
50	4.2	0.5	0.3

ISSN: 2454-4248
139 – 148

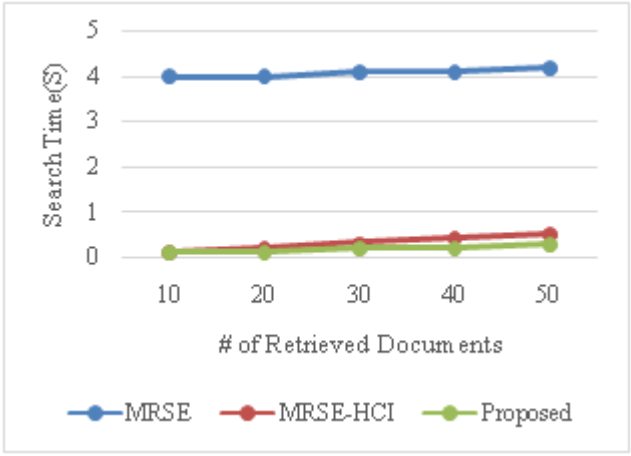

Figure 5. Search time to Retrieve Documents

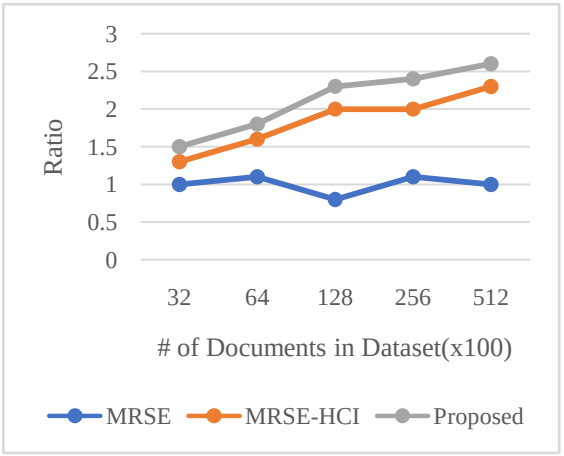

Figure 6. Ratio of documents in Dataset

Figure 5 is used to describe search time utilized to retrieve the documents and compares the results with the existing technologies used such as MRSE and MRSE-HI.It is observed that with the increase in document set size, the search time of the Query pattern semantic keyword matching algorithm increases linearly, while the search time of MRSE increases exponentially. As the Table 4 shows the search time of the system is stable with the growth of query keywords and retrieved documents. Meanwhile, the search time is subsequently less for MRSE. In addition, Figure 6, represents the represents the ratio of documents in database based on the total number of documents and the time taken to index the documents using proposed technique and is compared with different techniques.

Document collection efficiency is measured based on the number of search used for collecting the documents and ratio of documents in the dataset.Figure 8. And Table 6 shows that the precision of retrieved documents in the dataset is greater than the precision of MRSE determined by standard deviation. Document collection efficiency. X axis represents the number of documents in the dataset whereas Y axis denotes the precision of the documents in percentage.

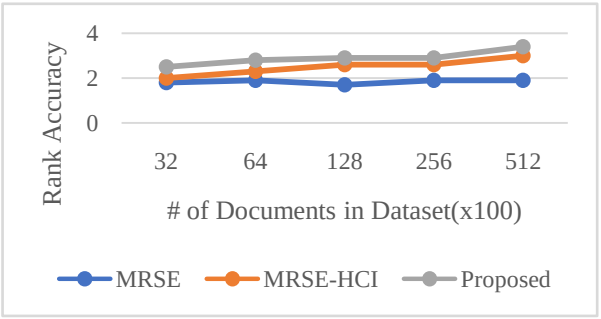

Figure 7. Rank accuracy of documents

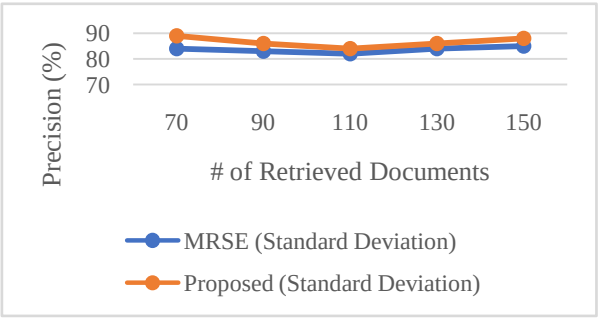

Figure 8.Precision of Retrieved Documents

Table 5. The rank accuracy of documents in database

Number of Documents in Dataset	Rank accuracy of Documents in Dataset		
	MRSE	MRSE-HCI	Query pattern evaluation, Semantic keyword matching
32	1.8	2	2.5
64	1.9	2.3	2.8
128	1.7	2.6	2.9
256	1.9	2.6	2.9
512	1.9	3	3.4

Table 6. The precision of retrieved documents

Number of Retrieved Documents	Precision of Retrieved Documents	
	MRSE (Standard deviation)	Query pattern evaluation, Semantic keyword matching
70	84	89
90	83	86
110	82	84
130	84	86
150	85	88

 ISSN: 2454-4248
139 – 148

Figure 7 describes the rank accuracy. In this test, irrespective of the number of retrieved documents, the proposed system has shown better rank accuracy than MRSE and MRSE-HCI. This mainly because of the semantic keyword matching and query evaluation introduced into search strategy From Table 5, it can be seen that the proposed technique has better accuracy compared to existing methods. Thereby, the document collection efficiency is increased and as the number of searches increased the efficiency of the proposed system increases accordingly From the result, we can conclude that proposed research work is better than MRSE and MRSE-HCI in rank accuracy, search efficiency, search time for determining the retrieved documents, and keywords matching in the query.

VII. CONCLUSION

In this research work, the search results are customized and directed to the precise information needs of individual users. The objective of the proposed semantic retrieval model has been discussed to provide better search capabilities that will yield a qualitative improvement over keyword-based full-text search, by introducing and exploiting fine-grained domain ontologies with keyword matching. This research work has shown that it is possible to develop a consistent ranking algorithm yielding measurable improvements concerning keyword-based search subjected to the quality. The architecture of the proposed system is flexible to deal with different ontologies, formal query languages, and query processing abilities of underlying data repositories. Furthermore, the system is robust to incomplete inputs, while using the retrieved background knowledge in the state that the user input is just a single keyword, the system can deal with it by exploring the implicit information description that the user had in mind when that keyword was interpolated.There is abundant opportunity for further improvement and research beyond the current results to provide a consistent model for advancement on the problems to the benefit of semantic retrieval improvements.

REFERENCES

[1] Windsor, J. J. (2016). *From Explanation to Demonstration: A Conceptual Framework for the Study and Strategic Design of Interactive Visualizations* (Doctoral dissertation, University of Alberta).

[2] Athukorala, K., Medlar, A., Oulasvirta, A., Jacucci, G., &Glowacka, D. (2016, March). Beyond relevance: Adapting exploration/exploitation in information retrieval. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (pp. 359-369). ACM.

[3] Jeffrey, L., Hegarty, B., Kelly, O., Penman, M., Coburn, D., & McDonald, J. (2011). Developing digital information literacy in higher education: Obstacles and supports. *Journal of Information Technology Education: Research*, *10*(1), 383-413.

[4] Kamvar, S. D., Haveliwala, T. H., &Jeh, G. M. (2012). *U.S. Patent No. 8,271,471*. Washington, DC: U.S. Patent and Trademark Office.

[5] Fernández, M., Cantador, I., López, V., Vallet, D., Castells, P., & Motta, E. (2011). Semantically enhanced information retrieval: An ontology-based approach. *Web semantics: Science, services and agents on the world wide web*, *9*(4), 434-452.

[6] Campos, R., Dias, G., Jorge, A. M., &Jatowt, A. (2015). Survey of temporal information retrieval and related applications. *ACM Computing Surveys (CSUR)*, *47*(2), 15.

[7] Denker, G., &Senanayake, R. (2015). *U.S. Patent No. 9,152,222*. Washington, DC: U.S. Patent and Trademark Office.

[8] JINGJING, W. (2013). *Improving Product-related Patent Information Access with Automated Technology Ontology Extraction* (Doctoral dissertation).

[9] Song, F., Zacharewicz, G., & Chen, D. (2013). An ontology-driven framework towards building enterprise semantic information layer. *Advanced Engineering Informatics*, *27*(1), 38-50.

[10] Biancalana, C., Gasparetti, F., Micarelli, A., &Sansonetti, G. (2013). Social semantic query expansion. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *4*(4), 60.

[11] Jain, V., & Singh, M. (2013). Ontology based information retrieval in semantic web: A survey. *International Journal of Information Technology and Computer Science (IJITCS)*, *5*(10), 62.

[12] Vigneshwari, S., & Aramudhan, M. (2015). Social information retrieval based on semantic annotation and hashing upon the multiple ontologies. *Indian Journal of Science and Technology*, *8*(2), 103-107.

[13] Rodríguez-García, M. Á., Valencia-García, R., García-Sánchez, F., & Samper-Zapater, J. J. (2014). Ontology-based annotation and retrieval of services in the cloud. *Knowledge-Based Systems*, *56*, 15-25.

[14] Chauhan, R., Goudar, R., Sharma, R., & Chauhan, A. (2013, March). Domain ontology based semantic search for efficient information retrieval through automatic query expansion. In *Intelligent Systems and Signal Processing (ISSP), 2013 International Conference on* (pp. 397-402). IEEE.

[15] Singh, G., & Jain, V. (2014). Information retrieval (IR) through semantic web (SW): An overview. *arXiv preprint arXiv:1403.7162*.

[16] Soylu, A., Giese, M., Jimenez-Ruiz, E., Vega-Gorgojo, G., &Horrocks, I. (2016). Experiencing OptiqueVQS: a multi-paradigm and ontology-based visual query system for end users. *Universal Access in the Information Society*, *15*(1), 129.

[17] Müller, B., Hagelstein, A., &Gübitz, T. (2016, November). Life Science Ontologies in Literature Retrieval: A Comparison of Linked Data Sets for Use in Semantic Search on a Heterogeneous Corpus. In *EKAW (Satellite Events)*(pp. 158-161).

[18] Bansal, M., & Arora, J. (2016). A Novel OBIRS System for Ontology Based Information Retrieval System.

[19] Foschini, L., Montanari, R., Boukerche, A., &Corradi, A. (2016). Scalable and mobile context data retrieval and distribution for community response heterogeneous wireless networks. *IEEE Communications Magazine*, *54*(4), 101-107.

[20] Cao, N., Wang, C., Li, M., Ren, K., & Lou, W. (2014). Privacy-preserving multi-keyword ranked search over

 ISSN: 2454-4248
139 – 148

encrypted cloud data. *IEEE Transactions on parallel and distributed systems*, *25*(1), 222-233.

[21] Nasution, M. K., Noah, S. A. M., & Saad, S. (2016). Social network extraction: Superficial method and information retrieval. *arXiv preprint arXiv:1601.02904*.

[22] Bourgonje, P., Moreno-Schneider, J., Nehring, J., Rehm, G., Sasaki, F., & Srivastava, A. (2016). Towards a Platform for Curation Technologies: Enriching Text Collections with.K

[23] Zhu, Y., Yan, E., & Song, I. Y. (2017). A natural language interface to a graph-based bibliographic information retrieval system. *Data & Knowledge Engineering*.

[24] Li, J., Wang, Q., Wang, C., Cao, N., Ren, K., & Lou, W. (2010, March). Fuzzy keyword search over encrypted data in cloud computing. In *INFOCOM, 2010 Proceedings IEEE* (pp. 1-5). IEEE.

[25] Wang, C., Cao, N., Ren, K., & Lou, W. (2012). Enabling secure and efficient ranked keyword search over outsourced cloud data. *IEEE Transactions on parallel and distributed systems*, *23*(8), 1467-1479.

[26] Cao, N., Yang, Z., Wang, C., Ren, K., & Lou, W. (2011, June). Privacy-preserving query over encrypted graph-structured data in cloud computing. In *Distributed Computing Systems (ICDCS), 2011 31st International Conference on* (pp. 393-402). IEEE.

[27] Wu, D., Cong, G., & Jensen, C. S. (2012). A framework for efficient spatial web object retrieval. *The VLDB Journal—The International Journal on Very Large Data Bases*, *21*(6), 797-822.

[28] Wang, B., Yu, S., Lou, W., &Hou, Y. T. (2014, April). Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud. In *INFOCOM, 2014 Proceedings IEEE* (pp. 2112-2120). IEEE.

[29] Fu, Z., Wu, X., Guan, C., Sun, X., & Ren, K. (2016). Toward efficient multi-keyword fuzzy search over encrypted outsourced data with accuracy improvement. *IEEE Transactions on Information Forensics and Security*, *11*(12), 2706-2716.

[30] Chen, L., Cong, G., Jensen, C. S., & Wu, D. (2013, January). Spatial keyword query processing: an experimental evaluation. In *Proceedings of the VLDB Endowment* (Vol. 6, No. 3, pp. 217-228). VLDB Endowment.

[31] Wu, D., Cong, G., & Jensen, C. S. (2012). A framework for efficient spatial web object retrieval. *The VLDB Journal—The International Journal on Very Large Data Bases*, *21*(6), 797-822.