 ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

A Survey on Semantic Web Search and Technologies

[1]K.Palaniammal, [2]Dr.S. Vijayalakshmi
[1,2]Thiagarajar College of Engineering, Madurai, TamilNadu, India

Abstract
This paper present an overview of the existing literature on semantic web search and technologies. Semantic search is the field of Web search, which is distinct from the traditional information retrieval methods. The purpose of the semantic search is to assist users to denote their search intentions and assist search engines to understand the meaning of users' queries in terms of semantic web technologies. We understand this concepts from a broader sense, in that semantic web search refers to using semantic technologies for information retrieval. In this paper, we surveyed semantic web search methods and engines are analyzed and concluded based on some common perspectives.

Keywords
Web Search, Semantic Web Search, Information Retrieval, Ontology, Survey, Semantic Web Survey, Multimedia Search, Hybrid Semantic Search.

I. Introduction

The Semantic Web provides a number of technologies to improve human and computer collaboration on the internet [54]. One important issue centre's on the management of documents and, particularly, the semantically supported document retrieval called semantic search [1]. Semantic search is an application of semantic web in the field of web search, has shown significant potential in the function of improving the performance of retrieval. Compared with the traditional search engines that focus on the frequency of word appearance, semantic search engines are more likely to try to understand the meanings hidden in retrieved documents and users' queries, by means of adding semantic tags into texts, in order to structuralize and conceptualize the objects within documents [28]. Currently the researches regarding semantic search are in the beginning stage, as the traditional search engines such as Google®, Yahoo® and so forth still dominate the markets of search engine [2]. The Semantic web is being to be developed to overcome the following problems for current web search [37].

- The web content lacks a proper structure regarding the representation of information.
- Ambiguity of information resulting from poor interconnection of information.
- Automatic information transfer is lacking.
- Usability to deal with enormous number of users and content ensuring trust at all levels.
- Current web search has a limited recall and precision due to synonyms, homonyms, spelling variants, multiple languages and spelling mistakes.
- Current web provide no means to specify the relation between a resource and a term. For example, sell/buy.
- Incapability of machines to understand the provided information due to lack of a universal format.

The term ''Semantic Web'' refers to the World Wide Web Consortium's vision of the Web of data, it is "an extension of the current Web in which information is given a well-defined meaning, better enabling computers and people to work in cooperation'' [54]. In this paper, we will discuss and evaluate six important techniques of semantic search that relate to our paper.

II. Overview of Existing Semantic Search and Technologies

This paper makes a survey over the primary literature regarding semantic web search and technologies and we review their characteristics respectively. By classifying the literature into six categories, namely;

A. Semantic web based search engines and technologies
B. XML based search engines and technologies
C. Ontology based search engines and technologies
D. Hybrid semantic search engines and technologies
E. Multimedia semantic search engines and technologies
F. Natural language search engines and technologies

A. Semantic Web Based Search Engines and Technologies

Today, a number of semantic search approaches have been published and their application areas are diverse. However, that aims to capture data relationships and make resulting data queryable [2-3].

Sheth et al. [4] proposed a Semantic Content and Retrieval Engine (SCORE). Two types of metadata are extracted from documents, which are syntactic metadata describing non-contextual information about documents and semantic metadata describing domain specific information about documents. Based on the semantic metadata, documents are associated with context ontologies by means of automatic classification technologies. Users send queries to the system by specifying contexts and attributes values of metadata. This approach has the following limitations: The keyword-based matching style could produce incomplete results. The performance of the approach relies on the quality of ontologies and queries. The approach could be challenged by large-scale knowledge bases. They do not propose any means for ranking retrieved results. They do not propose any means for evolving the domain ontologies to make them consistent with the dynamic domain knowledge. They do not propose any means for dealing with the perceptual differences between ontology designers and users toward the ontologies.

Duke et al. [5] proposed a semantic search and browse tool – Squirrel. By means of this system, once a user submits a query, the query terms will be matched with concepts (topics) from PROTON – lightweight general purpose ontology. The query terms and matched topics are sent to Web search engines. The documents that match the query terms and topics are classified according to their types. Under each topic, documents are ranked following the combination of the degree of relevance to query terms and the level of interest to user profile (a list of topics). The limitations of Duke et al.'s approach are as follows: The performance of this approach relies on the quality of ontologies and queries. The expanded query terms could burden Web search engines and cause query flooding. The keyword-based query term-ontological concept matching could lead to a mismatch or incomplete matching. User profile-based document ranking cannot work when users search for documents with new topics. They do not propose any means for ontology evolution.

Schreiber et al., [34] proposed MultimediaN E-Culture project is one of the semantic search systems, demonstrates how the novel Semantic Web and presentation technologies can be deployed to provide better indexing and search support within large virtual collections of cultural heritage resources. To search semantic paths, this system checks all RDF literals in the repository for matches to the given keyword and traverses the RDF graph until a resource of interest is found. Finally, the results are clustered based on the paths from the matching literals to their result. This research has some similarity with other semantic approach, but it lacks the ability to assign weights to properties and resources and the ability to identify the search-ending threshold. These limitations are the most important issues with this system, because they are two of the determinants for expanding the semantic network. Moreover, information travels in one direction only in this limited system: always from the query object in the triple to the corresponding subject.

Fernández et al. designed and implemented new semantic information retrieval systems. Thus, they do not fully exploit the information indexed and functionalities provided by traditional, large-scale web search engines [33]. Alkhalifa et al. in [53] presented semantic annotation tool called AraTation for Arabic news in the Web. The tool is capable of extracting news named entities. Location ontology is used for that tool. The reported performance of the tool achieves an average precision of 67% and recall of 82% on a set of ten locations over 25 Web documents.

B. XML Based Search Engines and Technologies

A number of studies have discussed semantics in XML keyword search. Most of them focus on how to effectively connect matches of keywords in a meaningful way [49].

Elghandour et al. [35] propose a system that helps users decide which indexes to build for an XML document collection. The system is tightly coupled to the query optimizer using database statistics to perform cost estimation. Chen et al. [36] propose an XML indexing approach which allow every fully specified XML query to be answerable using a single lookup. In contrast to the above work, PDPA does not attempt to index all or part of the structure of the XML document but rather allows users to plug in different types of pruning properties. These properties are then only used if they are found to be highly selective for a particular query and document collection.

XRank [38] connects keyword matches by the LCA nodes that contain at least one occurrence of all keywords after excluding the occurrences of keywords in sub-elements that already contain all keywords. XSEarch [39] introduces the concept of interconnection. Two matches are interconnected if the paths from these two nodes and to their LCA may not contain distinct nodes with the same labels except for themselves. A similar concept has also been proposed in [40]. Xu et al. [41] introduce the notion of the SLCA which extends the definition of LCA. An SLCA is a LCA that does not have further LCA nodes among their descendants. The concept of MLCA [42] is very similar to that of SLCA with the difference that MLCA imposes constraints on the node labels. Recently, coherency ranking [43, 44] is proposed for XML keyword search. It does not use any heuristics but relies on a specially designed ranking method to filter results. Another line of work is to figure out the desired result types. XReal [45] exploits the underlying statistics of different node types to compute the possibility that the node types are desired.

XBridge [46] proposes an estimation-based approach to compute the promising result types for a keyword query, it considers the value and structural distributions of the data. All these methods cater for the structural properties of XML documents, but they are not aware of the structures in keyword queries. Recently, query analysis has attracted some attention. XSeek [47] classifies the keywords in a query into two categories: search predicates and return nodes. Similarly, XReal [45] identifies the ambiguity that a keyword can appear both as an XML tag name and as a text value. However, these studies do not go further and discuss the relationships between keywords. Pu and Yu [48] discusses the problem of query cleaning (e.g. segmentations and correction) for keyword queries in a relational database context, but their aim is to improve the quality of keyword queries, and they do not consider the characteristics of XML data.

C. Ontology Based Search Engines and Technologies

Ontology-based search engines and technologies refer to the semantic search engines and technologies designed to denote users' query intentions by means of ontology concepts [3]. Ontology-based information retrieval systems are discussed in many studies from different aspects [6-15, 51].

Hogan described survey about search and browsing over RDF Web data [8]. Early prototypes using the concepts of ontologies and semantics on the Web include Ontobroker [6] and SHOE [7], which can be seen as predecessors to standardization efforts such as RDFS and OWL, describing how data on the Web can be given in structured form, and subsequently crawled, stored, inference and queried over.

SWOOGLE offers search over RDF documents by means of an inverted keyword index and a relational database [9]. Swoogle calculates metrics that allow ontology designers to check the popularity of certain properties and classes. In contrast to SWSE [8], which is mainly concerned with entity search over instance data, Swoogle is mainly concerned with more traditional document-search over ontologies. The challenge of their approach is: (i) The large-scale semantic networks may make it difficult to improve the efficiency of this approach; (ii) The performance of the search approach relies on the quality of queries. (iii) Incompleteness of search result is another issue because the search could miss some results when searching in a huge semantic tree, (iv) Halting problem – all relations in semantic have the feather of uncertainty, which means that some statements returned from semantic search cannot be proven by logic in practice.

WATSON provides a similar effort to provide keyword search facilities over Semantic Web documents, but additionally provides search over entities [10-11]. However, they do not include components for consolidation or reasoning, and seemingly instead focus on providing APIs to external services.

Sindice is a registry and lookup service for RDF files based on Lucene and a MapReduce framework [12]. Sindice originally focused on providing an API for finding documents which reference a given RDF entity or given keywords – again, document-centric search. More recently however, Sindice has begun to offer entity search in the form of Sig.Ma [13]. However, Sig.ma maintains a one-to-one relationship between keyword search and results, representing a very different user-interaction model to that presented herein.

The Falcons Search engine offers entity-centric searching for entities (and concepts) over RDF data [14]. They map certain key word phrases to query relations between entities, and also use class hierarchies to quickly restrict initial results.

Castells et al. [15] presented an information retrieval framework using an ontology in order to improve the accuracy. Basically, documents are annotated and an ontology is constructed based on the annotations. Then, the annotations for each document are weighted by an adaptation of the tf*idf measure. For a user

 ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

query, the annotations matched to the query are retrieved from the ontology, and then the documents containing the annotations are selected as the query results. The result documents are ranked by an adapted vector-space model which assigns higher scores to the documents containing many high weighted annotations. In this work, the relationship between the annotations was used to find the relevant document through the ontology query processing, but the differences among the weights of the relationships were not considered. Wei et al. [51] employ probabilistic topic models to learn relationships between a set of concepts. Particularly, they employ the LDA model to learn concept hierarchies.

D. Hybrid Semantic Search Engines and Technologies

Some search engines are employed with semantic web technology and traditional method to improve the precision of text search.

Rocha et al. [18] built a hybrid search engine combining traditional text search methodology and a spread activation algorithm over the Semantic Web. Three different measures are counted for weighting relationships between nodes in a semantic graph: 1) a cluster measure which assesses the similarities between nodes; 2) a specificity measure which assesses the differentiation or specificity between nodes; and, 3) a combined measure which integrates the former two measures. A spread activation algorithm is utilized to find related nodes based on the principle that from some starting nodes which have an "initial activation value" namely the weight of a relation to the searching task, the instance is chosen which has the highest value and is matched with constraints; and then its neighbours are activated. Constraints have three basic categories – concept type constraints on which activated nodes must be given types, fan-out constraints on which the spread must stop when more than one node are connected to the processed nodes, and distance constraint which specifies the depth of spread from initial instance. Rocha et al.'s approach may meet the following challenges: The approach could be challenged by large-scale semantic networks. The performance of the search approach relies on the quality of queries. Spread activation algorithm has the issue of query incompleteness and query flooding. There is no semantic interpretation of the activation value flowing through the network. Some inferences in the graph are uncertain, which means some statements returned from semantic search cannot be proven by logic in practice.

Han and Chen [17] presented a hybrid Web search methodology – HWS, combining traditional text search and semantic search, to improve the performance of traditional search engines. Three algorithms are adopted in the search engine, namely: BAS which is used to mine associations from existing user profile ontologies; BCH which is used to construct class hierarchies by means of a hierarchy clustering method; and MCH which merges classes into class hierarchies. A ranking algorithm utilized in HWS concerns all entities and relations, and contextual similarities between two entities. Han and Chen's methodology has the following disadvantages: Class hierarchies cannot be modified by either users or designers, and so cannot adapt to changing knowledge in real environments and satisfy users' requests for structure change. The semantic search is based on traditional keyword-based search, which has the issue of query flooding for each keyword in one query and thus could increase the cost of search time and resources, and increase the browsing time of users. The class hierarchies cannot represent more complicated relationships between entities in domain knowledge. The performance of the search approach relies on the quality of queries. The approach could be challenged by large numbers of documents.

Donghee Yoo [19] presented a hybrid query processing method for personalized information retrieval on the Semantic Web. This paper has suggested three requirements for personalized information retrieval on the Semantic Web. First, the contents on the Semantic Web should be expressed in machine-readable format using ontologies. Second, the machine should be able to assess the individual requirements of each user. Third, effective query processing based on individual requirements should be executed with a reasonable retrieval time. Drawback of this approach is, the method uses a limited set of objective or subjective terms.

E. Multi-media Semantic Search Engines and Technologies

Multimedia search engines and technologies refer to the semantic search engines and technologies designed in the purpose of retrieving audios, videos, and images [3].

Multimedia searching has become an important research field, particularly for understanding E-commerce and entertainment Web search. However, there is a major need for new multimedia Web search techniques and systems. People continually raise the bar of expectation for sophisticated Web search engines. Due to the rapidly increasing usage of digital multimedia data, a single piece of information can often be richly conveyed using multiple correlated media, thus enabling users to concurrently receive information from multiple sources [20]. Text is currently one of the most intuitive method for searching images since most of the available search engines still rely on text, and users usually think of topics, keyword, or other high-level concepts (in the form of text) to write queries. Current Web search engines, such as Google Image (http://images.google.com), MSN image (http://search.msn.com/images), and Yahoo! Image (http://images.search.yahoo.com), has been fixed at the point of text analysis relevant to the multimedia material. However, the major limitation of this approach is the tedious and ineffective nature of manually annotating every (temporal) segments of a video or every region of interests in an image. Moreover, the process of automatic annotation by mapping low-level features into high-level semantic concepts is generally difficult as it needs machine learning and interpretations. For example, systems can use modelled domain knowledge to make sense and interpret the semantic meaning of video data by observing visual, aural and text features [21].

Hong et al [52] presented a novel concept named Mediapedia which aims to construct multimedia encyclopediaby mining webknowledge. The Mediapedia distinguishes itself from traditional encyclopedia in its multimedia presentation, full automated production, dynamic update and the flexible frame work where each module is extensible to potential applications. In this approach employed the AP algorithm in producing the exemplars from image pool, while using LSA to associate exemplars to Wiki pages and utilizing document lattice model to perform Wiki pages summarization and assembled them for multimedia encyclopedia. This study can be deemed as an attempt at constructing Mediapedia by leveraging on web knowledge. The experimental results, however, were not as good as expected and should be improved further. This may be aroused by the assumption that the distribution of images from Flickr can automatically make a trade off between ''typicality'' and ''diversity''. However, this is not the truth for all the concepts. Improvement can be made by taking into account the tags in producing the exemplar, and leveraging the images embedded in Wikipedia to facilitate better association and so on. Analternative approach to tackle the problem is to start this Mediapedia from Wikipedia, by identifying different senses of the concept by Wikipedia first and then associating them with images /audios. Kherfi, Ziou, and Bernardi provided a comprehensive discussion on the issues, techniques and systems of image retrieval from the Web and reviewed some prototypes [22]. Lie presented Content-based audio retrieval (CBAR) leads to a more accurate classification than what can be achieved by CBIR systems. For example, audio stream can be classified into music, speech, and noise, as well as some semantic details like cheering, applause, and laughter [23]. This raises a challenge for bridging the gaps between low-level features and high-level semantic.

Video conveys a rich semantic presentation through the synchronized audio, visual and text presentations over a period of time. In the early days of content-based video retrieval (CBVR) research, most efforts have simply borrowed/extended systems and algorithms from image, text and sound retrieval as these types of media have been commonly used much earlier than that of video. Although this strategy has gained some degree of success, it is not always sufficient for all applications as each video genre has its own syntax, semantic, rules, and formats. Since video understanding is strongly dependent on the context and domain, CBVR needs to support different requirements from users and applications, as well as the associated browsing and search strategies. For example, color-texture-shape (i.e. image) based indexing is generally not effective when users need to search particular news video events, such as 'war in Iraq'. Based on the level of content abstraction that can be automatically extracted, current video retrieval systems can be distinguished by their usage of low-level features or high-level semantic. Yoshitaka and Ichikawa [25] surveyed query methods that utilize low-level features from image such as shape, spatial relation, color and texture; and video, including object motion, spatio-temporal relations. They have compared these features-based queries with semantic-based and knowledge-assisted retrievals including 'query-by-subject' (i.e. keyword) and 'query-by-subject/object' (i.e. derivation knowledge). The emerging applications and services from these types of retrievals include: finding images that are visually similar to a chosen picture or sketch; summarizing videos with thumbnails of key frames; finding video clips of a specific event, story, or person; and producing a two-minute skim of an hour-long program [24].

Multimedia retrieval on the Web is a major challenge involving multi-disciplinary fields, including CBIR, CBAR, and CBVR, as well as text retrieval. Multimedia has recently emerged on the Web and the search capabilities on each media is not equivalently mature. Text is still currently the most widely used method to search for multimedia as text is the oldest type of media on the Internet. To fully support multimedia search, however, Web search engines should not be based on text alone, as some users and applications may not know what the usable keywords are. As different Web domains can use certain keywords for various purposes, a single keyword could be easily misinterpreted. For example, when users search for ''bat'' images, the search engines may return animal (bat) and baseball (bat) images. Unless the Website that contains the image can give some additional hints, or the image itself has the right caption; users cannot simply search on ''bat and not baseball'' to get just the animal (bat) images. On the other hand, if query based on similar images is supported, it can effectively remove baseball (bat) as they have distinguishable features and characteristics, such as shape, color and texture.

F. Natural Language Search Engines and Technologies

Natural language search engines and technologies refer to the semantic search engines and technologies designed to seek answers for natural language questions by means of semantic technologies [3].

Lopez developed a very recent system PowerAqua is an ontology-based NLI system which surpasses traditional systems by managing multiple ontology sources and high scalability [26]. Since its NL processing module remains the same as in the previous AquaLog system. AquaLog is a portable NLIKB system which handles user queries in a natural language (English) and returns answers inferred from a knowledge base. The system uses GATE1 libraries (namely the tokenizer, the sentence splitter, the POS tagger, and the VP chunker) [27].

Cimiano proposed ORAKEL [28] is an ontology-based NLI system. It accepts English factoid questions and translates them into first-order logic forms. This conversion uses full syntax parsing and a compositional semantics approach. ORAKEL can be ported into another domain but such porting requires a domain expert to create a domain-dependent lexicon. The lexicon is used for an exact mapping from natural language constructs to ontology entities. A possible drawback of ORAKEL's approach is that the system can neither handle ungrammatical questions nor deal with unknown words.

Damljanovic presented FREyA system [29] is an NLIKB system that combines syntactic parsing with ontology reasoning. It derives parse trees of English input questions and uses heuristic rules to find a set of potential ontology concepts (for mapping from question terms to ontology concepts) using GATE and the OntoRoot Gazetteer [30]. The NLP-Reduce system [31] does not involve any advanced linguistic and semantic tools and depends on matching the query words to the KB instances. Its core part is a query generator which is responsible for creating SPARQL query given the words and the lexicon extracted from the KB. The major strength of the system is its good portability as it does not depend on any complex NLP query processing.

The ontology-based NLI system called QACID introduced in Ferrandez [32] covers a cinema/movie domain. Its target language is Spanish. It consists of two main components, the user query formulation database, and textual-entailment engine. Whereas the former component serves mainly for development and system training purposes, the latter is intended for an unknown query processing. The core of the QACID system is a database of query formulations. The database contains a set of 54 clusters, each cluster represents one type of question and it has a representative query pattern which was derived from a set of training data. Each cluster is also associated with one SPARQL query. As pointed out in the QACID evaluation, the system is not able to answer unknown ontology concepts and therefore it fails if the user poses a query using terms that are not present in the lexicon.

III. Some Common Issues of Current Information Retrieval and Semantic Web Search

From this survey we identify fundamental drawbacks of current information retrieval and semantic search technologies such as [50]:

A. Identify User's Perceptions

A user perception plays an important role in semantic search engines. In the semantic search engines are not able to satisfy user's requirements due to the difference between each person's subjective perceptions regarding the objective world [4, 17].

B. Ignorance of Evolving Knowledge Structure

Knowledge structures in the semantic search engines cannot be

ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

frequently updated to suit the change of users' requirements and the change of external environment [5, 34].

C. Low Precision and High Recall

Some search engine cannot show their significant performance in improving precision and lowering recall [18-19, 53].

D. Lack of Semantics and Lack of Evaluations

Since many semantic search methodologies are only in the phase of conceptual model development, many of them have not been tested via the reasonable number of experiments [4, 33, 45, 48].

E. Ambiguity/Polysemy

A polysemous word has more than one meaning. When searching for documents with a word such as ''play'', related to a theatre piece, a search engine can return unrelated results such as, for example, a set of games for children [32].

F. Lack of Synonym Relations

Words are synonymous if they have the same meaning. Words ''irritated'' and ''annoyed'' are very closely related; however, when searching for one of these words, found items will hardly contain the other word [31-32].

G. Lack of Consensus

The lack of consensus in the use of tags, especially as granularity is concerned, makes a traditional tagging system quite inefficient. To describe a particular item, different users may consider terms at different levels of generality/ specificity. For example, a user can tag a photograph as ''bird'', and another user can tag the same photo as ''eagle'' [21-25, 28].

IV. Conclusion

In this paper we have given brief survey to semantic web search and technologies, which is currently one of the hottest research topics in both semantic web and information retrieval. The review also made for various characteristics of semantic web respectively. In addition, the most pressing research issues are analyzed [16]: (i) how to automatically translate natural language queries into formal ontological queries, (ii) how to automatically add semantic annotations to Web content, or alternatively how to automatically extract knowledge from Web content, and (iii) how to create and maintain the underlying ontologies. This may be done either (a) manually by experts, e.g., in a Wikipedia like manner, where different communities may define their own ontologies, or (b) automatically, e.g., by extraction from the Web, eventually coming along with existing pieces of ontological knowledge and annotations (e.g., from existing ontologies or ontology fragments, and/or from existing annotations of Web pages in micro formats or RDF a), or (c) semi-automatically by a combination of (a) and (b). Clearly, the larger the degree of automation, the larger is also the potential size of ontologies that can be handled and the smaller are the costs and efforts for generating and maintaining them. So, for the very large scale of the Web, a very high degree of automation is desirable. A closely related another important research challenge is, the evolution and updating of and mapping between the ontologies that are underlying semantic search on the Web, where it is similarly desirable to have a very high degree of automation.

References

[1] Mangold, Christoph,"A survey and classification of semantic search approaches", International Journal of Metadata, Semantics and Ontologies 2.1 (2007), pp. 23-34.

[2] Dong, Hai, Farookh Khadeer Hussain, Elizabeth Chang, "A survey in semantic search technologies", Digital Ecosystems and Technologies, 2008. DEST 2008. 2nd IEEE International Conference on. IEEE, 2008.

[3] Grimes.S: Breakthrough analysis: Two + nine types of semantic search. Information WeeK, United Business Media, Manhasset, NY, USA.2010.

[4] Sheth, A., Bertram, C., Avant, D., Hammond, B., Kochut, K., Warke, Y.,"Managing semantic content for the Web", IEEE Internet Computing 6 (2002) pp. 80-87.

[5] Duke, Alistair, Tim Glover, John Davies,"Squirrel: An advanced semantic search and browse facility", The Semantic Web: Research and Applications. Springer Berlin Heidelberg, 2007, pp. 341-355.

[6] S. Decker, M. Erdmann, D. Fensel, R. Studer,"Ontobroker: ontology based access to distributed and semi-structured information, in: DS-8: IFIP TC2/ WG2.6 Eighth Working Conference on Database Semantics, Kluwer, B.V., Deventer, The Netherlands, The Netherlands, 1998.

[7] J. Heflin, J. Hendler, S. Luke, SHOE: A Knowledge Representation Language for Internet Applications, Tech. Rep. CS-TR-4078, Dept. of Computer Science, University of Maryland, 1999.

[8] Hogan, Aidan, et al.,"Searching and browsing linked data with SWSE: the semantic web search engine", Web semantics: science, services and agents on the world wide web 9.4 (2011): 365-401.

[9] L. Ding, T. Finin, A. Joshi, R. Pan, R.S. Cost, Y. Peng, P. Reddivari, V.C. Doshi, J. Sachs, Swoogle: A search and metadata engine for the Semantic Web, in: 13th ACM Conference on Information and Knowledge Management, Springer, 2004.

[10] M. Sabou, C. Baldassarre, L. Gridinoc, S. Angeletou, E. Motta, M. d'Aquin, M. Dzbor, WATSON: a gateway for the Semantic Web, in: ESWC 2007 Poster Session, 2007-06.

[11] M. d'Aqui n, M. Sabou, E. Motta, S. Angeletou, L. Gridinoc, V. Lopez, F. Zablith, What can be done with the Semantic Web? An overview watson-based applications, in: SWAP, 2008.

[12] E. Oren, R. Delbru, M. Catasta, R. Cyganiak, H. Stenzhorn, G. Tummarello, Sindice.com: a document-oriented lookup index for open linked data, Int. J. Metadata Sem. Ontol. 3 (1) (2008) pp. 37–52.

[13] G. Tummarello, R. Cyganiak, M. Catasta, S. Danielczyk, S. Decker, Sig.ma: live views on the Web of data, in: Semantic Web Challenge, 2009.

[14] G. Cheng, Y. Qu, Searching linked objects with falcons: approach, implementation and evaluation, Int. J. Semantic Web Inform. Syst. 5 (3) (2009) pp. 49–70.

[15] Castells, P., Fernandez, M., Vallet, D.,"An adpatation of the vector-space model for ontology-based information retrieval", IEEE Transactions on Knowledge and Data Engineering, 19(2), pp. 261–272 (February), 2007.

[16] Fazzinga, Bettina, Thomas Lukasiewicz,"Semantic search on the Web", Semantic Web 1.1 (2010), pp. 89-96.

[17] Han, Lixin, Guihai Chen,"The HWS hybrid web search", Information and Software Technology 48, No. 8 (2006), pp. 687-695.

[18] Rocha, Cristiano, Daniel Schwabe, Marcus Poggi Aragao,"A hybrid approach for searching in the semantic web", Proceedings of the 13th international conference on World Wide Web. ACM, 2004.

[19] Yoo, Donghee,"Hybrid query processing for personalized information retrieval on the Semantic Web", Knowledge-Based Systems 27 (2012), pp. 211-218.

[20] Tjondronegoro, Dian, Amanda Spink,"Web search engine multimedia functionality", Information Processing & Management 44, No. 1 (2008), pp. 340-357.

[21] Duan, L.-Y., Xu, M., Chua, T.-S., Qi, T., Xu, C.-S, "Content analysis: A mid-level representation framework for semantic sports video analysis", In MM'2003: Proceedings of the eleventh ACM international conference on multimedia, Berkeley, USA, pp. 33–44, 2003.

[22] Kherfi, D. Ziou, A. Bernardi,"Image retrieval from the World Wide Web: Issues, techniques and systems", ACM Computing Surveys, 36 (1) (2004), pp. 35–67.

[23] Lu, Lie, Rui Cai, Alan Hanjalic,"Towards a unified framework for content-based audio analysis", In Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on, Vol. 2, pp. ii-1069. IEEE, 2005.

[24] Chang, Shih-Fu, Alexandros Eleftheriadis, Robert McClintock,"Next-generation content representation, creation, and searching for new-media applications in education", Proceedings of the IEEE 86, No. 5 (1998), pp. 884-904.

[25] Yoshitaka, Atsuo, Tadao Ichikawa,"A survey on content-based retrieval for multimedia databases", Knowledge and Data Engineering, IEEE Transactions on 11, No. 1 (1999), pp. 81-93.

[26] Lopez, V., Fernandez, M., Motta, E., Stieler, N., "PowerAqua: Supporting users in querying and exploring the semantic web", Semantic Web, 1–17, 2011.

[27] Lopez, V., Uren, V., Motta, E., Pasin, M.,"AquaLog: An ontology-driven question answering system for organizational semantic intranets", Web Semantics, 5(2), pp. 72–105, 2007.

[28] Cimiano, P., Haase, P., Heizmann, J., Mantel, M., Studer, R.,"Towards portable natural language interfaces to knowledge bases – The case of the ORAKEL system. Data and Knowledge Engineering, 65(2), pp. 325–354, 2008.

[29] Damljanovic, D., Agatonovic, M., Cunningham, H.,"Natural language interfaces to ontologies: combining syntactic analysis and ontology-based lookup through the user interaction", In L. Aroyo, G. Antoniou, E. Hyvonen, A. ten Teije, H. Stuckenschmidt, & L. Cabral, et al. (Eds.), The semantic web: research and applications (pp. 106–120). Berlin/Heidelberg: Springer, 2010.

[30] Cunningham, H. et al.,"Text processing with GATE (Version 6)", University of Sheffield Department of Computer Science, 2011.

[31] Kaufmann, E., Bernstein, A., Fischer, L.,"NLP-reduce: A "naive" but domain independent natural language interface for querying ontologies", In 4th European Semantic Web Conference (ESWC 2007) (pp. 1–2), 2007.

[32] Ferrandez, Oscar, Izquierdo, R., Ferrandez, S., Vicedo, J. L.,"Addressing ontology-based question answering with collections of user queries. Information Processing & Management, 45(2), pp. 175–188, 2009.

[33] Fernández, M., Cantador, I., López, V., Vallet, D., Castells, P., Motta, E.,"Semantically enhanced Information Retrieval: An ontology-based approach. Web Semantics: Science, Services and Agents on the World Wide Web, 9(4), pp. 434-452, 2011.

[34] Schreiber, G., Amin, A., Aroyo, L., Assem, M., Boer, V., Hardman, L., et al.,"Semantic annotation and search of cultural-heritage collections: The MultimediaN E-Culture demonstrator. Web Semantics: Science Services and Agents on the World Wide Web, 6(4), pp. 243–249, 2008.

[35] Iman Elghandour, Ashraf Aboulnaga, Daniel C. Zilio, Fei Chiang, Andrey Balmin, Kevin Beyer, Calisto Zuzarte,"An XML index advisor for DB2, In: Proceedings of ACM SIGMOD, pp. 1267–1270, 2008

[36] Zhiyuan Chen, Johannes Gehrke, Flip Korn, Nick Koudas, Jayavel Shanmugasundaram, Divesh Srivastava, Index structures for matching xml twigs using relational query processors, Data Knowl. Eng. 60 (2) (2007) pp. 283–302.

[37] Madhu, G., Govardhan, D. A., Rajinikanth, D. T., "Intelligent Semantic Web Search Engines: A Brief Survey. International Journal of Web And Semantic Technology. (IJWesT) Vol. 2, No. 1, 2011.

[38] L. Guo, F. Shao, C. Botev, J. Shanmugasundaram, Xrank: ranked keyword search over xml documents, in: Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, ACM, 2003, pp. 16–27.

[39] S. Cohen, J. Mamou, Y. Kanza, Y. Sagiv, Xsearch: a semantic search engine for xml, In: Proceedings of the 29th International Conference on Very Large Data Bases, Morgan Kaufman, 2003, pp. 45–56.

[40] G. Li, J. Feng, J. Wang, L. Zhou,"Effective keyword search for valuable lcas over xml documents", In: Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management, ACM, 2007, pp. 31–40.

[41] Y. Xu, Y. Papakonstantinou,"Efficient keyword search for smallest lcas in xml databases", In: Proceedings of the ACM SIGMOD International Conference on Management of Data, ACM, 2005, pp. 537–538.

[42] Y. Li, C. Yu, H.V. Jagadish, Schema-free xquery, in: Proceedings of the Thirtieth International Conference on Very Large Data Bases, Morgan Kaufman, 2004, pp. 72–83.

[43] Termehchy, M. Winslett,"Effective, design-independent xml keyword search, in: Proceeding of the 18th ACM Conference on Information and Knowledge Management, CIKM '09, ACM, 2009, pp. 107–116.

[44] Termehchy, M. Winslett,"Using structural information in xml keyword search effectively", ACM Trans. Database Syst. 36 (2011) 4:1–4:39.

[45] Z. Bao, T.W. Ling, B. Chen, J. Lu,"Effective xml keyword search with relevance oriented ranking", In: Proceedings of the 25th International Conference on Data Engineering, IEEE, 2009, pp. 517–528.

[46] J. Li, C. Liu, R. Zhou, W. Wang,"Suggestion of promising result types for xml keyword search, In: Proceedings of the 13th International Conference on Extending Database Technology, ACM, 2010, pp. 561–572.

[47] Z. Liu, Y. Chen,"Identifying meaningful return information for xml keyword search, In: C.Y. Chan, B.C. Ooi, A. Zhou (Eds.), Proceedings of the ACM SIGMOD International Conference on Management of Data, Springer, 2007, pp. 329–340.

ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

[48] K.Q. Pu, X. Yu,"Keyword query cleaning", Proceedings of the VLDB Endowment 1 (2008) pp. 909–920.
[49] Liu, X., Chen, L., Wan, C., Liu, D., Xiong, N. N., "Exploiting structures in keyword queries for effective XML search. Information Sciences, 2013.
[50] Fuentes-Lorenzo, Damaris, Norberto Fernández, Jesús A. Fisteus, and Luis Sánchez,"Improving large-scale search engines with semantic annotations", Expert Systems with Applications (2012). 40 (2013), pp. 2287–2296.
[51] Wang, Wei, Payam M. Barnaghi, Andrzej Bargiela. "Probabilistic topic models for learning terminological ontologies", Knowledge and Data Engineering, IEEE Transactions on 22.7 (2010): pp. 1028-1040.
[52] Richang Hong, Zheng-Jun Zha, Yue Gao, Tat-Seng Chua, Xindong Wu,"Multimedia encyclopedia construction by mining web knowledge", Signal Processing 93 (2013), pp. 2361–2368.
[53] Saleh, L. M. B., Al-Khalifa, H. S.,"AraTation: an Arabic semantic annotation tool. In Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services (pp. 447-451). ACM, 2009.
[54] Berners-Lee, T., Hendler, J., Lassila, O.,"The Semantic Web. Scientific American, 2001.
[55] [Online] Available: http://swoogle.umbc.edu/.
[56] [Online] Available: http://watson.kmi.open.ac.uk/WatsonWUI/.
[57] [Online] Available: http://sindice.com/.
[58] [Online] Available: http://sig.ma.
[59] [Online] Available: http://iws.seu.edu.cn/services/falcons/.
[60] [Online] Available: http://www.gate.ac.uk.

Palaniammal.K Completed Master's degree in Computer Science from Alagappa University in 2007, received Master of Philosophy in Computer Science from Madurai Kamaraj University in 2008, and doing Full time Ph.D Computer Science at Thiagarajar College of Engineering Madurai. Her papers were published in International journals, International Conferences and IEEE explorer. Her research interests include Information Retrieval, Semantic Web, Ontology, Web Search, Data mining and Artificial Intelligence.

Vijayalakshmi.S Completed Master's degree in Computer Applications from Anna University of Technology Chennai in 1994, received Master of Philosophy in Computer Science from the Madurai Kamaraj University, in 2004 and received Ph.D., degree from Anna University, Chennai in 2011. At present, she is Working as an Assistant Professor in Thiagarajar College of Engineering. Her papers were published in many International Journals and Conferences. Her research interests include Database Management System, Data mining, Artificial Intelligence, and Web Development.