Comparative Study of Semantic Search Techniques using RDF

Salil C. Damle
Department of Computer
Engineering
Dwarkadas J. Sanghvi College
of Engineering
Mumbai, India

Shreyas S. Kupekar
Department of Computer
Engineering
Dwarkadas J. Sanghvi College
of Engineering
Mumbai, India

Khushali Deulkar
Assistant Professor
Department of Computer
Engineering
Dwarkadas J. Sanghvi
College of Engineering
Mumbai, India

ABSTRACT
In the sea of information available to us today, the prevalent searching techniques using keywords and ranking algorithms fall short on many aspects. In such a scenario, the emergence of semantic searching techniques is attempting to fill this void. It is required by Semantic searching techniques to understand the intent of users and the meaning of the query entered using Natural Language. The query is semantically broken down and stored in the Resource Description Framework (RDF) format. Such a structured format is more useful for implementing the search. In this paper, a comparison of various searching techniques which make use of RDF is done, based on certain parameters. The main intention is to identify the searching techniques which will be best suited for different purposes.

General Terms
Semantic Search, Algorithms.

Keywords
RDF, RDF(s), Semantic Web, ontology.

1. INTRODUCTION
Today, the web is evolving from a simple keyword extraction based technique to a more advanced and useful semantic way. The intention has now changed to understanding the reasons for which a user posts a query rather than to simply search for articles and documents which have the words from the query. Semantic web [1] thus forms the next generation of internet i.e. web 3.0. It provides a standard data format and exchange protocols which delivers fast results and at the same time more relevant results to the user. From the early web solutions of HTML which could simply illustrate documents and provide linking between them, recent advancements have now come to semantic HTML [2], Extensible Markup Language (XML), Web Ontology Language (OWL) and Resource Description Framework (RDF) which can describe arbitrary things such as objects, people, places etc. in a structured way.

The Resource Description Framework [3] is a World Wide Web Consortium (W3C) specification which is used to convert a natural language query into a tabular format that can be easily shared and used universally. It was designed as an improved data model over the likes of the class diagram or the E-R model. The RDF constructs a Metadata format which is used to store a query or descriptions of web resources in a subject-predicate-object configuration also known as the triples format in the RDF nomenclature [4]. The simple directed multi-graph format and the capacity of modelling abstract information has made it an essential component for usage in Knowledge management applications.

This paper has been organised into various sections as follows: Section 2 introduces the parameters which have been used to compare the approaches. Section 3 summarizes and talks about the papers which have been used for the comparative analysis. The next section i.e. the section 4 compares the papers selected in the Section 3 on the basis of the parameters described in the section 2. The comparison observations are concluded in the section 5. Sections 6 and 7 have the acknowledgements and references respectively.

2. CLASSIFICATION PARAMETERS
This section mentions and explains the parameters on the basis of which the comparing of RDF based semantic searching techniques has been conducted. The criteria are so decided because they are best able to provide a clear idea about the approaches used and to supply a distinct classification between them. The classification parameters used in this paper are as follows: Methodology and Indexing, Evaluation of performance, Advantages and Disadvantages.

There are other parameters to classify the searching techniques which have not been discussed in this survey. Some of them include Ranking model, inferencing, and ontology construction. They haven't been included as the parameters chosen for the comparison are sufficient to classify the search techniques and some parameters are not clearly identifiable through the papers. The comparison table created using these parameters is given the section IV.

2.1 Methodology and Indexing
The ontology techniques or methods used are one of the main criteria for effectively achieving semantic web. The methodologies include crawlers based on mining tools, inference engines, annotation tools etc. The accepted ontology description languages include RDF and OWL in their structure. These languages are necessary for the description of the ontologies. This comparison has technologies which include RDF, DBpedia, and SPARQL etc.

Indexing is the process of storing the given data in an organised manner which helps in improving the speed of data retrieval operations. Without the use of indexing, a search engine will have to scan through all the documents in the database. Hence, the application of indexing to the content found during the crawling process eases up the retrieval task. Following are the types of indexing which can be found.

- Forward indexing- It is the list of documents and the words which appear on them.

- Inverted indexing- It is the list of words and the documents in which they appear.

- Indexing based on annotation weights- This indexing includes the annotation weights during the indexing process.

- Graph indexing- Graph index is used in query graphs for retrieval of answers and their authentication.

2.2 Evaluation of performance

This parameter gives the actual performance of the given approach. The performance can be decided on the basis of criteria such as the level of precision, recall, relevance etc. The raw values of parameters such as precision and recall can show the difference between existing web searching techniques and semantic searching techniques.

2.3 Advantages and Disadvantages

Advantages and disadvantages give a clear picture of where a particular approach is most useful and should be employed. Depending on this information, future researchers and developers will be able to take a better call on which approach they should choose. For example, in the paper titled 'Ontology based framework for semantic web content mining', human intervention is needed. Need of human intervention may not be desirable in certain circumstances.

3. LITERATURE SURVEY OF SELECTED APPROACHES

This RDF based Semantic web search comparative study has taken into account nine different approaches from which six approaches have been considered for the comparative analysis. The approaches not selected include Ontology Driven Semantic Search [5], Searching and browsing Linked Data with SWSE: The Semantic Web Search Engine [6], Keyword Search over Probabilistic RDF Graphs [7]. This section briefly explains each of the approaches used in this study and their results are presented in section IV.

3.1 Faceted product search powered by the Semantic Web

A multifaceted approach of searching for products using Semantic web search technology has been introduced in this paper [8]. They have proposed an E-commerce platform called XploreProducts.com. The new search engine proposes to include multiple language search, parameterised search and unified view of all the information about the searched object. In this way, the main idea is to cater to the advanced preferences of current E-commerce shoppers.

3.2 An Ontology-Based Framework for Semantic Web Content Mining

This approach talks about how semantic web can be used for easing the information retrieval procedure and to provide search results to the web user which are closer to their

requirement [9]. Extraneous web pages are not ranked in the intended approach to improve the efficiency of Web access. Application areas of this approach include Tourism, Education, and Q and A sites since the proposed framework is semi-automatic.

3.3 Mining Association Rules for Adaptive Search Engine Based on RDF Technology

A method has been proposed in this paper to create an adaptive search engine which uses mining association rules that reflect the users past search history [10]. The rules that are extricated are used to improve the performance of the meta-data based search engine. The intended method includes the creation of a document registration module and reasoning base. The intention behind this approach is to realise the humatronics concept.

3.4 Search Engine: Intelligent Web Service Search

In this paper, a new searching model has been developed for enhancing the precision of query results [11]. The intelligent web service search engine makes use of semantic web technologies for achieving the said goals. Reliability and relevance of query results has been achieved with the use of RDF methodology and a hierarchical online directory.

3.5 QuizRDF: Search Technology for the Semantic Web

This paper proposes a new information seeking system named as QuizRDF which combines traditional keyword based querying and RDF annotations of those resources [12]. A distinguishing factor of this approach is that when RDF annotations exist, the search results will be improved by a great deal. At the same time, in cases where they aren't available a search capability will still be provided, though it might not be as good.

3.6 Scaling Concurrency of Personalized Semantic Search over Large RDF Data

The paper proposes a simple and light-weight interpretation approach that achieves less memory overhead by utilizing indexing to improve concurrency and throughput [13]. The selected approach is flexible enough to permit distributed or partitioned execution. This approach has been implemented in a system called 'SKI' whose experimental evaluation provides better performance results over the prevalent techniques.

4. COMPARATIVE ANALYSIS

In this section, the comparison of the selected approaches mentioned in the preceding section has been carried out on the basis of the parameters mentioned in section 2. The comparison table includes the Paper Title, Authors and the classification criteria. Table 1 presents an overview of the techniques selected in this survey. In some approaches, gathering clear information for certain parameters was not possible and those places are marked with unclear.

Table 1.

Author	Approach	Methodology	Evaluation of performance	Advantages	Disadvantages
Damir Vandic, Jan-Willem van Dam, Flavius Frasincar [8]	XploreProducts.com platform	RDF, SPARQL	Precision, Accuracy, Recall and Specificity all above 91%	Product Name Identification, Category Mapping	Restricted only to E-Commerce domain.
S.Yasodha S.S.Dhenakaran [9]	An Ontology Based Framework for Semantic Web Content Mining	RDF	Precision, Average Precision, Relevance	The technique is applicable to any domain	Human intervention required in ontology creation
Yasufumi Takama, Shunichi Hattori [10]	Adaptive Hybrid Search Engine	RDF, RDF (S)	Precision and Recall	Adaptive Search System, Integration of Full-text search and Metadata-based search.	Unclear
N. Rajkumar, B.Gohin, Dr. Viji Vinod [11]	Intelligent Web Service Search	RDF	Precision, Relevance and Reliability	Retrieval of query result is more accurate than current search engines.	The techniques lacks domain ontology concept. Which can be implemented using annotation and growth reasoning algorithm.
John Davies, Richard Weeks [12]	QuizRDF search Engine	RDF,RDF(S), Forward Index	Evaluation done by EnerSearch	The system is applicable to both annotate metadata and traditional free text search.	Querying system not optimal. Large number of results are difficult to display on the current interface
Haizhou Fu, HyeongSik Kim, Kemafor Anyanwu [13]	Scaling Concurreny of Personalized Semantic Search over Large RDF Data	RDF, DBPedia	System is evaluated over DBPedia and Billion Triple Challenge 2009 database and the proposed system outperform rest of the similar systems.	Lightweight interpretation approach which reduces memory overhead significantly.	Overall system is optimal and does not have any discernible flaws.

5. CONCLUSION

In this paper, a comparison of several semantic web search techniques has been conducted, which have been recently proposed and made use of RDF in their approach. Various classification criteria have been introduced for RDF based semantic search approaches. Further, it also discusses various technologies used, evaluates the performance, and talks about various advantages and disadvantages of the selected approaches. The study helps in analysing the current semantic searching approaches and how future researches can try to minimize the flaws in each approach. This survey will help in making an informed decision about which approach will be best suited for specific purposes as required by researchers and developers. Semantic Search is the need of the day to improve the quality of search results and also to save the time of both professionals and laymen in their searching requirements.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] Parth R. Agarwal: "Semantic Web In Comparison to Web2.0." 2012 Third International Conference on Intelligent Systems Modelling and Simulation.

[2] Debajyoti Mukhopadhyay, Rituparna Kumar, Sourav R. Majumdar, Subhobroto Sinha: "A New Semantic Web Services to Translate HTML Pages to RDF." 10th International Conference on Information Technology.

[3] Stefan Decker, Prasenjit Mitra, and Sergey Melnik: "Framework for the Semantic Web: *An RDF Tutorial*."

[4] Eric Miller: "An Introduction to the Resource Description Framework." Bulletin of the American Society for Information Science and Technology. Volume 25, Issue 1, pages 15–19, October/November 1998.

[5] Dario Bonino, Fulvio Corno, Laura Farinetti, Alessio Bosca: "Ontology Driven Semantic Search." Dipartimento di Automatica ed Informatica Politecnico di Torino Corso Duca degli Abruzzi, 10129 Torino, Italy.

[6] Aidan Hogan, Andreas Harth, Jürgen Umbrich, Sheila Kinsella, Axel Polleres, Stefan Decker: "Searching and browsing Linked Data with SWSE: The Semantic Web Search Engine." Web Semantics: Science, Services and Agents on the World Wide Web 9 (2011).

[7] Xiang Lian, Lei Chen and Zi Huang. "Keyword Search over Probabilistic RDF Graphs." IEEE Transactions on Knowledge and Data Engineering, VOL. 27, NO. 5, MAY 2015

[8] Damir Vandic, Jan-Willem van Dam, Flavius Frasincar: "Faceted product search powered by the Semantic Web." Decision Support Systems 53 (2012) 425–437

[9] S.Yasodha, S.S.Dhenakaran: "An Ontology-Based Framework for Semantic Web Content Mining." 2014 International Conference on Computer Communication and Informatics (ICCCI -2014), Jan. 03 – 05, 2014.

[10] Yasufumi Takama, Shunichi Hattori: "Association Rules for Adaptive Search Engine Based on RDF Technology." IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, VOL. 54, NO. 2, APRIL 2007.

[11] N. Rajkumar, B.Gohin, Dr. Viji Vinod: "Search Engine: Intelligent Web Service Search." Chennai and Vivekanandha College of Technology for women, Third International Conference on Sustainable Energy and Intelligent System (seiscon 2012),VCTW, Tiruchengode, Tamilnadu, India on 27-29 December, 2012.

[12] John Davies and Richard Weeks: "QuizRDF: Search Technology for the Semantic Web." Proceedings of the 37th Hawaii International Conference on System Sciences – 2004.

[13] Haizhou Fu, HyeongSik Kim, Kemafor Anyanw: "Scaling Concurrency of Personalized Semantic Search over Large RDF Data." 2013 IEEE International Conference on Big Data.