Multidisciplinary Informatics Journal
Vol. 1, No. 2, December 2018, Pg. 111 – 118

SEMANTIC RETRIEVAL: ISSUES AND MOTIVATION

Azilawati Azizan[1], Zainab Abu Bakar[2] and Nurazzah Abd Rahman[3]

[1]Faculty of Computer & Mathematical Sciences, Universiti Teknologi MARA, Perak Branch, Tapah Campus, 35400 Tapah Road, Perak, Malaysia
[2]Al-Madinah International University, 40100 Shah Alam, Selangor, Malaysia
[3]Faculty of Computer & Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor Malaysia
Author Correspondence, e-mail: azila899@perak.uitm.edu.my

Received: 20 September 2018 / Accepted: 26 October 2018 / Published online: 15 December 2018

ABSTRACT

Currently, challenges in Information Retrieval are very closely related to the Semantic Web technology. Therefore, in this paper, we reported the current issue relating to semantic retrieval with main focus on ontology-based search. In order to support the presented issue, a retrieval evaluation on current semantic search engine was done to identify the current state of the retrieval performance on the Web. Several semantic search engines were selected and tested with a list of verified queries in the focused domain. The results showed that the precision of the retrieval still needs an improvement. Finally, based on existing publication and the demand from real-world applications, we proposed an initial conceptual framework for implementing semantic retrieval on the Web.

Keywords: Semantic Web; Ontology-Based Information Retrieval; Semantic Search

1. INTRODUCTION

The growth of the content on the Web is keep increasing from day to day, and it makes the Web to bear the burden of information overloaded. This situation leads to difficulties in searching for the right resources. Hence, we can expect the search engine will face a hard time to maintain quality retrieval results in the future. As this situation happens, the need for having an advanced accessibility and uniform data representation is very critical.

Therefore, the current challenge for information retrieval system is to put more semantic value into its structure. In simple words, we must try to make the machine 'understand' the

content of documents and 'understand' the user query so that it will be able to link them in a better way. But the question is how to make the machine or computer understand in all those aspects?

2. MOTIVATION

The main drawback of conventional retrieval system is the retrieved result is not concerned about the user search's intent. Mostly it is based on the keyword representation of user's query. The result is normally high in recall, but low in precision. This means the retrieval system returns hundreds of links for users to check and filter which link pages are relevant and fulfil their needs. Normally on the first round of the search activity ended with frustration until the user re-query for several times. The user also needs to reformulate queries several times for getting other related information for the same concept [1] [2].

Semantic search is seen potential to overcome this problem. Semantic search is about search by meaning. Current search techniques are not smart enough to extract the meaning of data; hence it ends by giving irrelevant result to a user's query. In fact the search should permit complex query and can do reasoning to retrieve relevant information [3]. A human can easily interpret the web content via representation by the web browser. But the computer or machine is totally unable to understand anything about the content. Figure 1 show how human see a webpage compared to computer see the same webpage.

Fig.1. Webpage and HTML codes

The ability of search system (machine) to assist user in aggregate and segregate all related information needed by human is very limited. With this condition, it is very difficult to make

the machine replace the human task of interpreting the web content into meaningful knowledge or information. Somehow, information on the Internet is expanding tremendously, thus it is impossible for humans to process it manually.

3. SEMANTIC SEARCH

3.1 Semantic Web

Initial Web is developed as a global document repository with a very easy way to access, publish and link documents. Web content is intended for direct human processing. In its current form, machine–based approaches are impossible, unless the content is transformed into machine-readable format. Tim Berners Lee [4] the inventor of WWW has devised the idea of the Semantic Web (SW) and the W3C (World Wide Web Consortium) has recommended RDF (Resource Document Framework) as the data model. RDF is a simple but very expressive data model that can support the interoperability between applications. RDF represents the knowledge in the form of triple [5]. The W3C also recommended SPARQL (Protocol and RDF Query Language) as its query language.

3.2 Semantic Search Approach

Semantic search is an application of Semantic Web. Semantic search is about search by meaning. Current search techniques are not smart enough to extract the meaning of data, hence it ends by giving irrelevant result to user's query. In fact, the search should permit complex query and can do reasoning to retrieve relevant information.

Many publications showed that most researchers employ ontology in their implementation and it is proven that ontology is capable to provide better understanding of a concept and the relationship between entities in document content understanding, query understanding and even ranking purposes. Ontology is semantically rich description to support search and it has become a current trend of representing domain knowledge on the Web. Ontology has been adopted in various ways for achieving better retrieval result. Most researchers employ ontology in the indexing process [6]. Some employ it at ranking process [7], other employ it for query expansion [8] and formulation purposes. In practice, ontologies can be very complex (with several thousands of terms) or very simple (describing one or two concepts only). It is seen that many works related to retrieval on the SW are using ontology as the main component to model their knowledge, but there is also a statement made by W3C that the use of ontology is not a compulsory in the SW.

Currently there are four approaches can be applied to semantic search [9]. As seen in Figure

3, one of the approaches is using contextual analysis, where it emphasizes on how to disambiguate queries. Another approach is reasoning. This type of approach can infer additional information from existing facts in the system. The third approach is to apply natural language understanding, which aim to identify the entity in a sentence. Last approach is ontology, where it can enrich the retrieval of specific domain related. This approach (ontology) is the most used by many researchers to develop the semantic retrieval system. And many semantic searches engine mix and match between those four approaches in various ways to give the best search experience to their user.

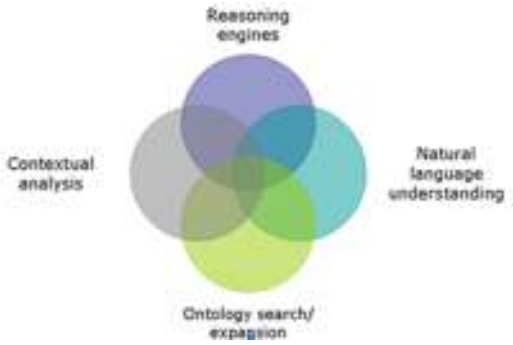

Fig.2. Approaches of Semantic Search

3.3 Ontology

There are many definitions of ontology being published, but the most popular cited is the definition given by Gruber; ontology is an explicit specification of a conceptualization [10]. In other words, ontology is a formal specification of a concept in a specific domain.

Basic activities related to ontology are; define the class / concept, arrange the concept in a taxonomy hierarchy (superclass - subclass) and define the relationship together of the value permitted. Ontology development is a complex and mostly it is a domain oriented process. To support this activity, many ontology development tools have been developed by researchers such as Protégé (Protege, n.d.), TopBraid Composer (TopQuadrant, n.d.), Ontolingua [11] and many more. Among those tools, Protégé is the most popular being used by many people and it is also a domain independent tool [12].

3.4 Why Use Ontology?

A vast amount of data on the web is in structured, semi-structured and unstructured form, so there is a need to standardize these data in a formal way so that we can benefit a lot of things with it. Ontology can be the mechanism to solve this problem [13]. Other than that, there are several other reasons for using ontology such as for knowledge sharing, logic inference, and

knowledge reuse [14]. Ontology is said as the backbone of the semantic web, because it may provide machine processable semantics of data and resources can be linked together.

However, ontology development is not an easy task because it is a collaborative approach. It needs several parties to involve such as the knowledge domain expert, the web developer and the software engineer. They must know how to model the knowledge ontologically. Other issue related to ontology development are ontology is really domain dependent, ontology alignment is difficult to manage and till now, construction of ontology is done manually or semi-automatic. So, it's a big challenge to make the Semantic Web success.

4. EVALUATION OF SEMANTIC SEARCH ENGINE

Author in [15] highlighted that the user need to filter and analyse several pages of the retrieved results before they can reach to the relevant information. This shows that the retrieval result is still low in precision whereby it still needs the user to segregate the results.

In order to get the current performance of retrieval on the Web, we made a simple retrieval evaluation on several existing web search engines. We identified more than fifteen search engines claim that they are semantic search engine [16]. For this research, we only select 5 semantic search engines (SSE) to be evaluated; that are Bing, Google, DuckDuckGo, Cluuz, and SenseBot.

The selection of the SSE is based on review done by several search engine reviewer on the Web [17], [18] and also based on their (SSE) popularity and website traffic provided by Alexa (an independent company acquired by Amazon –provide commercial web traffic data). Our domain area for this research is agriculture and focusing on durian fruit. Table 1 shows the list of queries being used for the evaluation.

Table 1. List of queries used in retrieval evaluation

No	Selected Queries
Q1	Durian species name
Q2	List of durian varieties
Q3	Durian cultivation information
Q4	Durian tree disease information
Q5	Where to buy durian
Q6	When is the durian season in Malaysia
Q7	How to choose a good durian
Q8	How durian flower is pollinated

5. RESULTS AND DISCUSSION

Figure 3 present a graph of precision and recall for the selected SSE. We identified that most of the search engines have low precision on query Q3, Q4 and Q5. The results also showed that most search engines achieved perfect precision only at recall 1 until 3, and then the precision starts to lower down. From this initial evaluation, it indicates that the precision of web retrieval is still not satisfying. Producing high precision in the retrieval is depending on various factors, for example retrieval is improved when query processing techniques is improved, or indexing is enriched, or the document processing is enhanced.

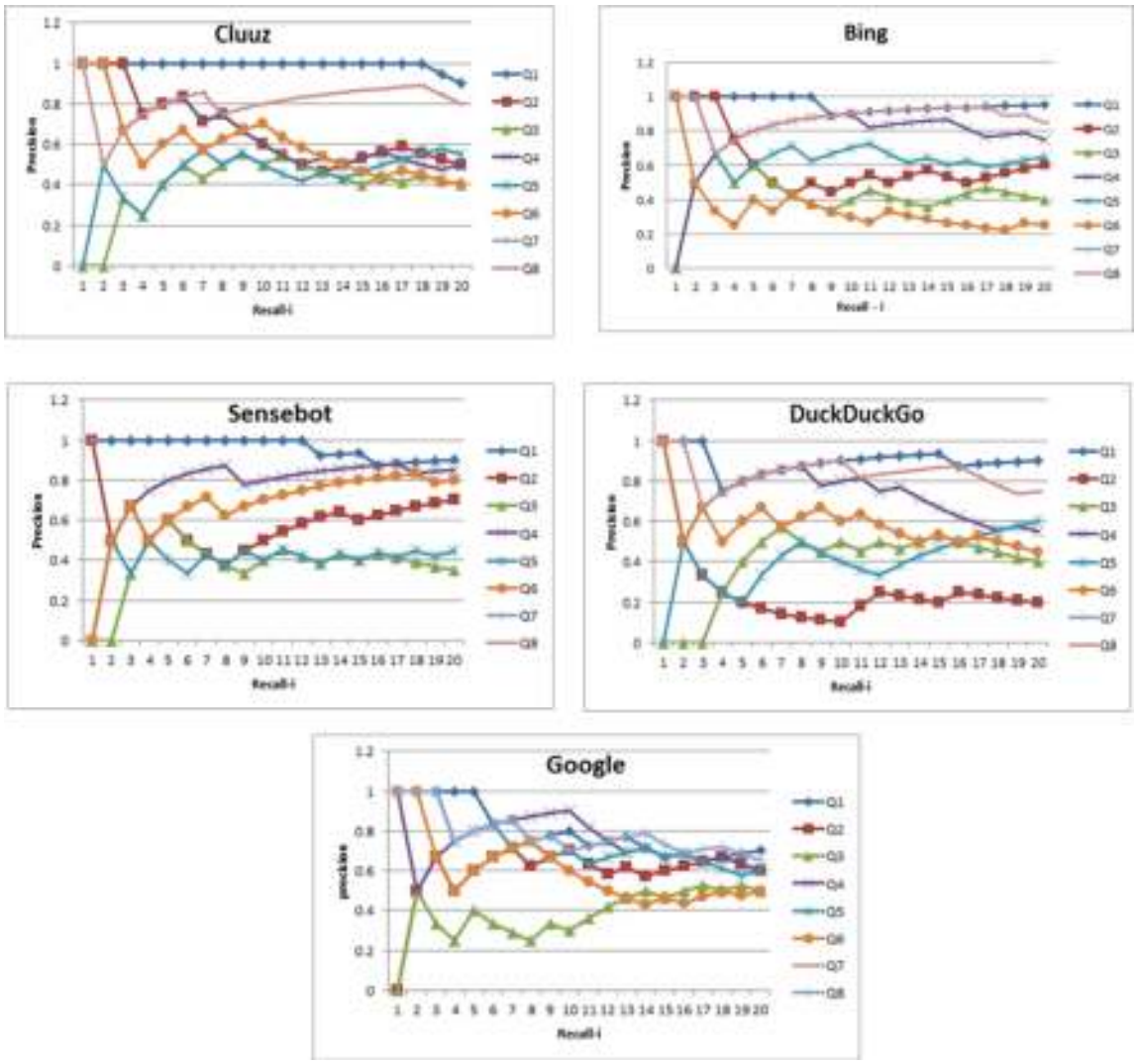

Fig.3. Precision and recall of the selected semantic search engine

Based on existing publication and the demand from real-world applications, we proposed an initial conceptual framework (Figure 4) for implementing semantic retrieval on the Web. This model was adopted from [19]

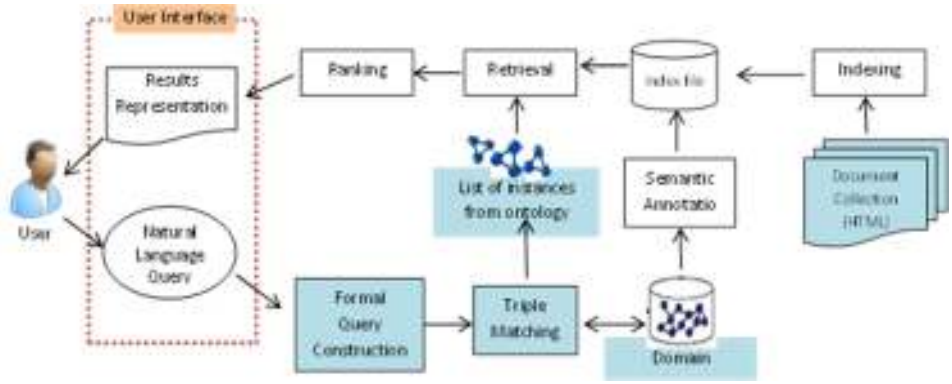

Fig.4. Precision and recall of the selected semantic search engine

6. CONCLUSION

Web is increasing very fast, and has created data deluge situation, which leads to difficulties in searching. Semantic search is believed to have the potential to solve the problems. There are four approaches that can be applied for developing semantic information retrieval system, that are contextual analysis, ontology, natural language understanding and reasoning. Most application mix and match between those four approaches in various ways to get better results. However, there is still no publication could state which combination of the four approaches is better. We found many studies in semantic information retrieval test the ontology adaptation, and all of them agree that basis for getting best result is relying hardly on the perfectness of the ontology used.

7. REFERENCES

[1] R. Alfred *et al.*, "Ontology-based Query Expansion for Supporting Information Retrieval in Agriculture," in *The 8th International Conference on Knowledge Management in Organizations Springer Proceedings in Complexity*, 2014, p. pp 299-311.

[2] Z. A. Bakar, K. N. Ismail, and N. Fooladi, "Effectiveness of Query Formulation Based on Durian Characteristics," in *IEEE Student Conference on Research and Development (SCOReD)*, 2012.

[3] M. Unni and K. Baskaran, "Overview of approaches to semantic web search," vol. 2, no. 2, pp. 345–349, 2011.

[4] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific American*, pp. 29–37, 2001.

[5] E. Terzi and A. Vakali, "Knowledge Representation , Ontologies , and the," in *APWeb 2003, LNCS 2642*, Springer-Verlag Berlin Heidelberg 2003, 2003, pp. 382–387.

[6] S. Kara, Ö. Alan, O. Sabuncu, S. Akpınar, N. K. Cicekli, and F. N. Alpaslan, "An

ontology-based retrieval system using semantic indexing," *Inf. Syst.*, vol. 37, no. 4, pp. 294–305, Jun. 2012.

[7] B. Aleman-Meza, I. B. Arpinar, M. V. Nural, and A. P. Sheth, "Ranking Documents Semantically Using Ontological Relationships," *2010 IEEE Fourth Int. Conf. Semant. Comput.*, pp. 299–304, Sep. 2010.

[8] J. Wu, I. Ilyas, and G. Weddell, "A Study of Ontology-based Query Expansion," in *Technical report CS-2011-04*, 2011.

[9] G. Sudeepthi, G. Anuradha, P. M. Surendra, and P. Babu, "A Survey on Semantic Web Search Engine," vol. 9, no. 2, pp. 241–245, 2012.

[10] T. R. Gruber, "Translation Approach to Portable Ontology Specifications by A Translation Approach to Portable Ontology Specifications," in *Technical Report KSL 92-71 Revised April 1993*, 1993, no. April 1993.

[11] A. Farquhar, R. Fikes, and J. Rice, "The Ontolingua Server: a tool for collaborative ontology construction," *Int. J. Hum. Comput. Stud.*,vol. 46,no. 6,pp.707–727, Jun. 1997.

[12] M. R. Khondoker and P. Mueller, "Comparing Ontology Development Tools Based on an Online Survey," vol. 958, 2010.

[13] V. Jain and S. K. Malik, "Using Ontologies in Web Mining for Information Extraction in Semantic Web : A Summary," 2010, pp. 3–6.

[14] X. H. Wang, T. Gu, D. Q. Zhang, and H. K. Pung, "Ontology based context modeling and reasoning using OWL," *IEEE Annu. Conf. Pervasive Comput. Commun. Work. Proc. Second*, pp. 18–22, 2004.

[15] N. Bansal and S. K. Malik, "A Framework for Agriculture Ontology Development in Semantic Web," *2011 Int. Conf. Commun. Syst. Netw. Technol.*, pp. 283–286, Jun. 2011.

[16] A. Azizan, Z. A. Bakar, N. K. Ismail, and M. F. Amran, "Interface Features of Semantic Web Search Engine," in *2013 IEEE Conference on e-Learning, e-Management and e-Services*, 2013.

[17] S. Bowen, "Top 10+ Semantic Search Engines: Best Alternative to Google Search Engine to Get More Accurate Results." Available at: http://www.webgranth.com/top-10-semantic-search-engines-best-alternative-to-google-search-engine-to-get-more-accurate-results, 2011.

[18] N. Messieh, "Top 7 Semantic Search Engines As An Alternative To Google." Available at: http://www.makeuseof.com/tag/top-7-semantic-search-engines-alternative-google-search/, 2010.

[19] D. Vallet, M. Fernández, and P. Castells, "An Ontology-Based Information Retrieval Model," in *Proceedings of ESWC Conference*, 2005.