

ISWSE: Islamic Semantic Web Search Engine

Hossam Ishkewy

Computers and Systems Engineering Department,
Faculty of Engineering, AIAzhar University
Cairo, Egypt

Hany Harb

Computers and Systems Engineering Department,
Faculty of Engineering, AIAzhar University
Cairo, Egypt

ABSTRACT

Despite the enormous number of documents and the huge amount of information contained in the Internet, the information search process is still difficult and unsatisfactory. The main reason behind this is that the search engines use the traditional search which depends on the keywords not on the concepts as semantic search engines. Tim Berners-Lee invented the Semantic Web in 2001 and until now many researches about semantic web search engines have appeared, however the Semantic Web researches in Arabic language are still few especially with regard to the Islamic information domain. Total number of Muslim countries is about 57 countries and a population of Muslims around the world is about 1.62 billion by more than 23% of the world's population. This large number believes on one book (The Holy Quran) and needs to retrieve information from it accurately and easily. The Holy Quran, which God revealed to the Prophet Muhammad as Muslims believe, has been worshipped for 1,400 years. The Holy Quran is written in the classical Arabic language. This ancient language has undergone a lot of modifications. For this reason, the Holy Quran became not fully understood by the majority of Muslims. To solve this problem, we present in this paper, the ISWSE as an Islamic Semantic Web Search Engine which searches in the Holy Quran. The proposed system is based on Islamic Ontology and uses Azhary as a lexical ontology for the Arabic language. The paper also introduces experimental results and performance evaluation for the proposed system.

Keywords

Semantic Web, Ontology, Search Engines, Information Retrieval, Islamic Sharia.

1. INTRODUCTION

Current Web contains billions of documents and has many administrative problems and limitations. The main limitations of the current Web are such as the Web content lacks a proper structure regarding the representation of information, ambiguity of information resulting poor information interconnection, inability to deal with enormous number of users and content ensuring trust at all levels. The machines are incapable to understand the provided information due to lack of a universal format and lacking of automatic information transfer [1], so the Web content is still readable only by humans. To solve these problems, Tim Berners-Lee introduced the Semantic Web as a conceptual model of Web that makes the contents to be read and used by human and intelligently by machines [2]. He mentioned that "The Semantic Web (SW) is not a separate Web but an extension of the current one, in which information is given well defined meaning. Adding semantics to the Web involves allowing documents which have information in machine readable forms, and allowing links to be created with relationship values"[3].

Tim Berners-Lee introduced in 2006 the fourth version of the Semantic Web Architecture containing eight layers [4] as shown in figure 1.

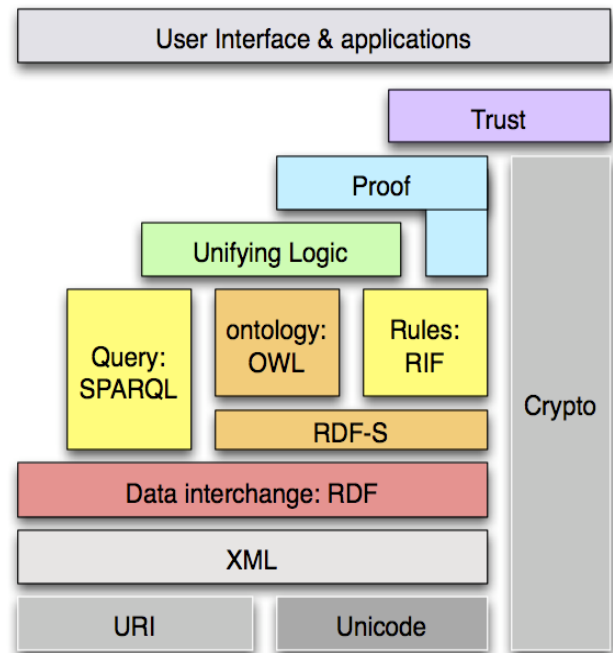


Figure 1: The Semantic Web Architecture [4]

The Resource Description Framework (RDF) and Ontology are considered to be the most important layers of this structure. The RDF is considered to be the key infrastructure to construct the Semantic Web enabling semantic interoperability. It is a standardized basis to encode, exchange and reuse the structured metadata. It is predicted that when RDF be used in a large scale on the Web, content and relationship between different resources will be described better, and this will help search engines to easily find resources on the Web and enable content to be rated [5]. Ontology is the backbone of the Semantic Web. Ontology became an interesting topic for researches and an important key in developing of the Semantic Web as it provides a sharable domain that facilitates the data understanding between people and different applications. Ontology has many definitions one of them defined by Gruber as "an explicit specification of a conceptualization" [6]. The designing of a semantic search engine related to the Quran is extremely important because Muslims need the Quran in all their affairs. But the Quran is not an easy book due to the diversity of topics that the Quran talks about. Finding accurate information directly from the Quran is very difficult. The reason behind this is not only the difficulty of Quran words, but also for its large size. Total number of pages of the Quran is 604 pages and has 77,439 words (Jalaluddin Al-Suyuti [7]).

This paper is arranged as it follows. Section 1 introduces the Semantic Web and the problem definition. Section 2 discusses some related works. Section 3 discusses our proposed ISWSE system architecture and its implementation issues. Section 4 evaluates the proposed system. The paper is concluded in Section 5.

2. THE RELATED WORKS

Searching the Web is a challenge and it is estimated that half of the complex queries go unanswered [1]. Traditional Web search engines such as Google and Yahoo are not able to provide relevant search results because they suffer from the fact that they do not know the meaning of the terms, expressions used in the Web pages and the relationship between them. A Semantic Web search engine such as Hakia [8], Swoogle [9], DuckDuckGo [10] stores semantic information about Web resources and is able to solve complex queries considering as well the context where the Web resource is targeted. Semantic search integrates the technologies of Semantic Web and search engine to improve the search results gained by current search engines and evolves to next generation of search engines built on Semantic Web [11]. Singh and Sharan [11] introduced a comparative study between traditional and semantic based search engines. The authors approved that semantic based search engine performance is higher than the keyword based search engine. Wang and others [12] classified the research of semantic search according to objectives, methodologies, and functionalities into document-oriented search, entity and knowledge-oriented search, multimedia information search, relation-centered search, semantic analytics, and mining-based search categories. Habernal and Konopík [13] presented an approach to build a semantic Web search using natural language which includes preprocessing, semantic analysis, semantic interpretation, and executing a SPARQL query to retrieve the results. They performed an end-to-end evaluation based on a domain dealing with accommodation options. They used a corpus of queries obtained by a Facebook campaign. The proposed system work with written texts in the Czech language. The authors approved that semantic search engines is one of the best search tools that can be relied upon to deal with the changes in the web structure. Vasnik and others [14] presented a semantic and context based optimized Hindi search engine. They described three models for query enhancement based on lexical variance using HindiWordNet(HWN), user context and combination of both techniques. The authors showed the ability and success of their semantic search engine to retrieve relevant results with high degree of accuracy when using the third model. Kanojia and Sharma [15] developed USSE (University Semantic Search Engine) based on a Semantic Web. They analyzed the input query on the basis of noun to find out the exact meaning of the provided input query. They also provided a new platform in which machine can understand information and process it without need of human interaction. Thomas and others [16] presented GeneView as a semantic search engine for biomedical knowledge. GeneView was built upon a comprehensively annotated version of PubMed abstracts and openly available PubMed Central full texts. It uses a multitude of state-of-the-art text mining tools optimized for recognizing mentions from 10 different entity classes and for automatically identifying protein-protein interactions (PPI). Among other entities, GeneView currently contains 32.8 million genes, 73.3 million chemicals, 914 000 Single Nucleotide Polymorphisms (SNP) and 3.9 million PPIs.

Riad and others [17] presented a new majority voting technique that combines the two basic modalities of Web images textual and visual features of the image in a re-annotation and search based framework. The proposed framework considered each web page as a voter to vote for the relatedness of keyword to the web image. The proposed approach was not only pure combination between image low level feature and textual feature but it also to consideration the semantic meaning of each keyword that expected to enhance the retrieval accuracy. Khan and others [18] proposed that the concepts of ontology of semantic web can be applied for carrying out semantic search in Holy Quran. For this purpose, they developed exploratory sample domain ontology, based on living creatures including animals and birds mentioned in Holy Quran. They then proposed certain recommendation of attaining semantic search from all domains and all resultant text of Holy Quran. These recommendations included model and framework of Quranic WordNet, integration, merging and mapping of domain ontologies under the umbrella of upper ontology. Their recommendation also extended to other Islamic knowledge sources like Hadith, Fiqh etc. Sherif and others [19] described the semantic Quran dataset, a multilingual RDF representation of translations of the Quran. The dataset was created by integrating data from two different semi-structured sources and it was aligned to an ontology designed to represent multilingual data from sources with a hierarchical structure. The resulting RDF data encompasses 43 different languages which belong to the most under represented languages in Linked Data, including Arabic, Amharic and Amazigh. The authors designed the dataset to be easily usable in natural-language processing applications with the goal of facilitating the development of knowledge extraction tools for these languages. They presented the ontology devised for structuring the data. The authors also provided the transformation rules implemented in their extraction framework. Al-Yahya and others [20] proposed a computational model for representing Arabic lexicons using ontologies. The ontology development is based on the UPON (Unified Process for ONtology) ontological engineering approach [21]. The ontology was limited to Time nouns which appeared in the Holy Quran and it was consisted of 18 classes and contained a total of 59 words. Abidin and others [22] explored the representation and classification of Holy Quran knowledge by using ontology. The ontology model for Quran was developed according to the Quran knowledge themes as described in Syammil Quran Miracle Reference. For example Iman (Faith) and Akhlaq (Ethics) main classes were chosen as the research scope for constructing the ontology. Saad and others [23] presented an approach for the automatic generation of ontology instances from a collection of unstructured known documents as the Holy Quran. The presented approach was stimulated based on the combination of natural language processing techniques, Information Extraction (IE) and Text Mining techniques.

3. THE ISWSE SYSTEM

The purpose of building the system is to overcome the difficulties faced by the Muslims to reach accurate and quick information from the Holy Quran. Not only the ordinary Muslim can use this system, but experts can also rely upon to prepare lectures, tutorials, research and books. The ISWSE is an Islamic Semantic Web Search Engine searching in the Holy Quran. It is built based on Islamic Ontology and uses Azhary [24] as a lexical ontology for the Arabic language. The ISWSE is a prototype for the semantic search engine in the Holy Quran. The ISWSE will be extended to include the Hadith and Sharia books. The ISWSE system architecture as

shown in the figure 2 is composed of different modules and processes such as Ontology Building, Ontology Extending, Verses Annotating, Query Preparing, and Searching.

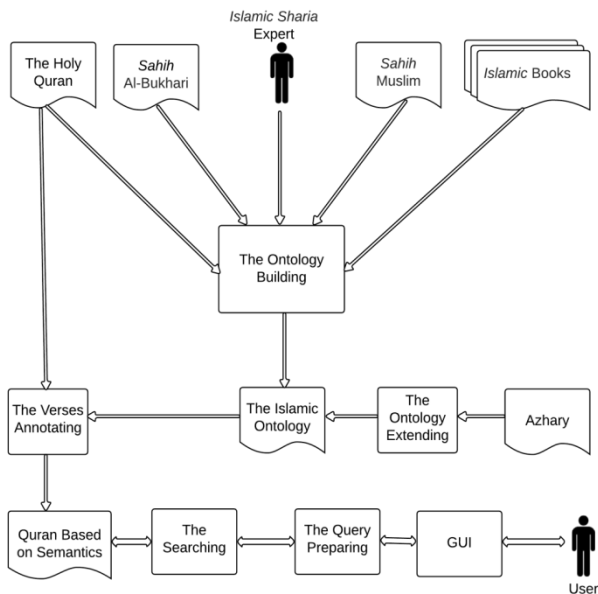


Figure 2: The ISWSE System Architecture

3.1 The Ontology Building Process

The Ontology Building is a completely manual process using protégé [25]. This process has been done through a comprehensive examination of the sources of Islamic Sharia such as the Holy Quran, the Sunnah (Sahih Al-Bukhari and Sahih Muslim) of the Prophet Muhammad and some important Sharia books (Holy Quran interpretation, Sunnah Explanations, Jurisprudence and Biography of the Prophet). It has also been achieved by taking tips and advices from some experts in the Islamic Sharia domain. The output of this process is the Islamic Ontology which includes the Islamic concepts in a hierarchal classes form. The Islamic Ontology contains only 3662 classes as a prototype ontology but it can be extend to contain all the Islamic concepts. The main classes of the ontology are Ethics (الأخلاق), Transactions (المعاملات), Permitted and Forbidden (الحلال والحرام), Government (الحكم), Stories and History (القصص و التاريخ), Work (العمل), Worships (العبادات), Punishments (العقوبات), Belief (العقيدة), Jihad (الجهاد), Manners (الأداب), Religions (الديانات) and Science and Arts (العلوم و الفنون) as shown in the figure 3. Each main term has its own hierarchal classes as shown in the figure 4.

3.2 The Ontology Extending Module

This Module extends the ontology classes to get the related subclasses, super classes, equivalent classes, associative classes, composite classes and disjoint classes. This Module use the Azhary Arabic lexical ontology which groups Arabic words into sets of synonyms called synsets, and records a number of relationships between words such as synonym, antonym, hypernym, hyponym, meronym, holonym and association relations. The user may use keywords which are related to the Islamic ontology classes by one or more of these relations.

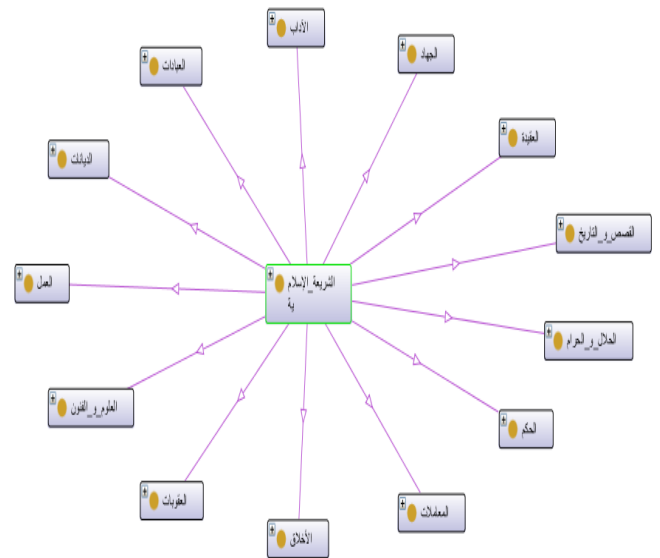


Figure 3: The Islamic Ontology Main Classes

3.3 The Verses Annotating Module

The Verses Annotating Module creates instances of the ontology classes in the Holy Quran verses. The output of this Module is a RDF statements (Quran based on semantics ontology) contains subject, predicate and object. The subject is an Islamic ontology instance, the predicate is a relationship called IS_IN (يوجد في) relation and the object is the verse number prefixed by the Surah (chapter of the Quran) number. The following example shows a statement of the Holy Quran verses. The statement means that 1319 الذبح is an instance of the الذبح where 1319 الذبح is the subject, IS_IN is the predicate, and the verse 37_105 is the object.

```
<"الذبح" rdf:about="http://www.ISWSE.org/#1319الذبح">
  <يوجد_في
  rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >37_105</يوجد_في>
</الذبح>
```

3.4 The Query Preparing Module

The Query Preparing module improves the system response time by converting the RDF triples to hash map. This module is also responsible for suggesting other equivalent query forms if it is necessary such that "الاستئجار" instead of "الأجرة" which have the same Arabic meaning.

3.5 The Searching Module

The searching Module goal is to receive the user query from the GUI, query the ontology to find the related verses, return the results to the GUI. Apache Jena [26] is used in this querying.

3.6 The Graphical User Interface

The Graphical User Interface GUI is used only to receive the user query and return the results.

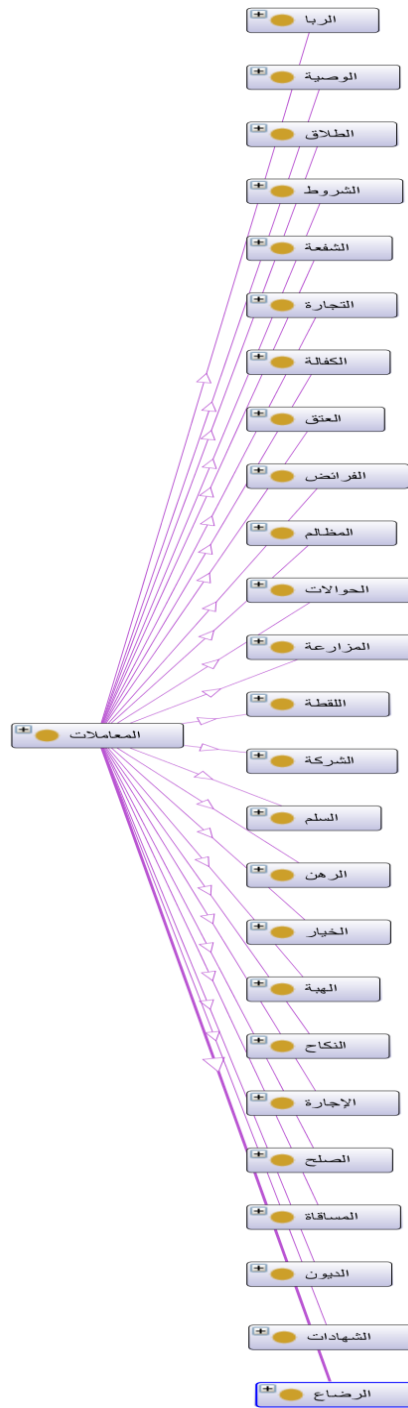


Figure 4: The Transactions Classes

4. THE EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

As a proof of concept, an experiment has used the Quran Prophets Stories as the most detailed part of the Islamic Ontology which contains 1153 concepts. This experiment is executed on 1487 annotated verses. Figure 5 shows the ISWSE GUI of the experiment and table 1 translates the GUI labels from Arabic to English. Figure 6 and figure 7 demonstrate the ontology of the prophet Musa story using protégé and Web Ontology Language (OWL). The average precision and recall of the proposed system are 98.5% and 97% respectively for 30 executed queries. The detailed precision and recall calculations are listed in the table 2.

5. THE CONCLUSION

We have created the Islamic ontology which contains the most important concepts of the Islamic Sharia. We have presented the Islamic Semantic Web Search Engine (ISWSE) which searches in the Holy Quran based on the Islamic ontology. We have explained the ISWSE system architecture. We have proved the system efficiency using the average precision and recall as a performance metrics. This work will be extended to complete the Islamic ontology. Also, the searching domain will convey the Sunnah (Sahih Al-Bukhari and Sahih Muslim) of the Prophet Muhammad and Sharia books (Holy Quran interpretation, Sunnah Explanations, Jurisprudence and Biography of the Prophet).


```

<owl:Class rdf:about="http://www.ISWSE.org/#منفذ">
  <rdfs:subClassOf rdf:resource="http://www.ISWSE.org/#قصة_موسى">/>
</owl:Class>
<owl:Class rdf:about="http://www.ISWSE.org/#الشيطان">
  <rdfs:subClassOf rdf:resource="http://www.ISWSE.org/#قصة_موسى">/>
</owl:Class>
<owl:Class rdf:about="http://www.ISWSE.org/#تتبع_أخت_موسى_لموسى">
  <rdfs:subClassOf rdf:resource="http://www.ISWSE.org/#قصة_موسى">/>
</owl:Class>
<owl:Class rdf:about="http://www.ISWSE.org/#تسع_معجزات">
  <rdfs:subClassOf rdf:resource="http://www.ISWSE.org/#قصة_موسى">/>
</owl:Class>
<owl:Class rdf:about="http://www.ISWSE.org/#نون_البقرة">
  <rdfs:subClassOf rdf:resource="http://www.ISWSE.org/#قصة_موسى">/>
</owl:Class>
<owl:Class rdf:about="http://www.ISWSE.org/#مؤمن_آل_فرعون">
  <rdfs:subClassOf rdf:resource="http://www.ISWSE.org/#قصة_موسى">/>
</owl:Class>
<owl:Class rdf:about="http://www.ISWSE.org/#الفراغ">
  <rdfs:subClassOf rdf:resource="http://www.ISWSE.org/#قصة_موسى">/>
</owl:Class>
<owl:Class rdf:about="http://www.ISWSE.org/#سوء_العذاب">
  <rdfs:subClassOf rdf:resource="http://www.ISWSE.org/#قصة_موسى">/>
</owl:Class>

```

Figure 7: The ontology of the prophet Musa story (OWL)

Table 2: The detailed precision and recall

Story	Query	Rel. & Ret.	Rel. Verses	Ret. Verses	Recall	Precision
محمد	إينات الرسالة	10	10	10	1.00	1
محمد	عزوة الخندق	16	17	17	0.9412	0.941176
محمد	تكذيب قريش	45	46	45	0.9783	1
سليمان	الهدد	9	9	9	1	1
سليمان	بلقيس	20	20	20	1	1
سليمان	تسخير الحن	3	3	3	1	1
إبراهيم	الرؤيا	6	7	6	0.8571	1
إبراهيم	صفات إبراهيم	12	13	12	0.9231	1
إبراهيم	معجزة	2	2	2	1	1
موسى	معجزة	30	31	31	0.9677	0.9677419
موسى	قوم فرعون	105	110	105	0.9545	1
موسى	السحرة	30	30	30	1	1
صالح	أصحاب الحجر	51	55	51	0.9273	1
هود	التكذيب	7	7	7	1	1
لوط	النصح	13	14	15	0.9286	0.8666667
شعيب	أصحاب الأيكة	39	41	39	0.9512	1
يوسف	المكر	14	16	14	0.875	1
يوسف	زليخة	17	17	17	1	1
يوسف	العريز	9	9	9	1	1
يوسف	السجن	18	18	18	1	1
آدم	إبليس	33	35	33	0.9429	1
آدم	إغواء إبليس للناس	10	11	10	0.9091	1
آدم	الملائكة	12	12	15	1	0.8
يونس	الحوت	4	4	4	1	1
عيسى	ميلاد عيسى	18	18	18	1	1
عيسى	الحواريين	6	6	6	1	1
عيسى	النصارى	6	6	6	1	1
زكريا	دعاء زكريا	6	6	6	1	1
داود	الحكم بين الخصوم	7	7	7	1	1
موسى	هارون	54	56	54	0.9643	1

6. REFERENCES

- [1] G.Sudeepthi , G. Anuradha and M.Surendra Prasad Babu, "A Survey on Semantic Web Search Engine" , IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1, March 2012.
- [2] Haytham T. Al-Feel, Magdy Koutb and Hoda Suoror, "Semantic Web on Scope: A New Architectural Model for the Semantic Web", Journal of Computer Science 4(7): 613-624, 2008.
- [3] Tim Berners-Lee, James Hendler and Ora Lassila, "The Semantic Web", Scientific American, May 17, 2001.
- [4] T. Berners-Lee, "Artificial Intelligence & the Semantic Web", World Wide Web Consortium, 2006. Available: [http://www.w3.org/2006/Talks/0718-aaai-tbl/Overview.html#\(14\)](http://www.w3.org/2006/Talks/0718-aaai-tbl/Overview.html#(14)).
- [5] Liyang Yu, "Introduction to the Semantic Web and Semantic Web Services", book, 2007.
- [6] T. R. Gruber, "A Translation Approach to a Portable Ontology Specification", Knowledge Acquisition, vol. 6, pp. 199-221, 1993.
- [7] الإلتقان فى علوم القرآن جلال الدين السيوطى, Available: http://www.maabernew.com/books/Itkan_Sayouti.pdf [Accessed: Sep. 5, 2014].
- [8] Hakia search engine, available at: <http://www.hakia.com/>
- [9] Swoogle search engine, available at: <http://swoogle.umbc.edu/>
- [10] DuckDuckGo, available at: <http://duckduckgo.com/>
- [11] Jagendra Singh and Aditi Sharan, "A Comparative Study between Keyword and Semantic Based Search Engines", International Conference on Cloud, Big Data and Trust, 2013.
- [12] Wei Wang, Payam M. Barnaghi and Andrzej Bargiela, "Search with Meanings: An Overview of Semantic Search Systems", International journal of Communications of SIWN, Vol. 3, June 2008, pp. 76-82.
- [13] Ivan Habernal and Miloslav Konopík , "SWSNL: Semantic Web Search Using Natural Language", Expert Systems with Applications, 40, pp 3649–3664, 2013.
- [14] Nandkishor Vasnik, Shriya Sahu and Devshri Roy, "TALASH: A semantic and context based optimized hindi search engine", International Journal of Computer Science, Engineering and Information Technology (IJCEIT), Vol.2, No.3, June 2012.
- [15] Ankit Kanojia and Varsha Sharma, "University Search Engine Based on Semantic Approach", International Journal of Computer Theory and Engineering, Vol. 4, No. 4, August 2012.
- [16] Philippe Thomas, Johannes Starlinger, Alexander Vowinkel, Sebastian Arzt and Ulf Leser, "GeneView: a comprehensive semantic search engine for PubMed", Nucleic Acids Research, pp 1–7, 2012.
- [17] Alaa Riad, Hamdy Elminir and Sameh Abd-Elghany, "Web Image Retrieval Search Engine based on Semantically Shared Annotation" , IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, March 2012.
- [18] Hikmat Ullah Khan, Syed Muhammad Saqlain, Muhammad Shoab and Muhammad Sher, "Ontology Based Semantic Search in Holy Quran", International Journal of Future Computer and Communication, Vol. 2, No. 6, December 2013.
- [19] Mohamed Ahmed Sherif and Axel-Cyrille Ngonga Ngomo, "Semantic Quran:A Multilingual Resource for Natural-Language Processing", Semantic Web Journal, 2009.
- [20] Maha Al-Yahya, Hend Al-Khalifa, Alia Bahanshal, Iman Al-Odah and Nawal Al Helwah, "An Ontological Model for Representing Semantic Lexicons: An Application on Time Nouns In the Holy Quran", The Arabian Journal for Science and Engineering, Volume 35, December 2010.
- [21] A. De Nicola, M. Missikoff, and R. Navigli, "A Software Engineering Approach to Ontology Building", Information Systems, 34(2009), pp. 258–275.
- [22] Azman Ta'a, Syuhada Zainal Abidin, Mohd Syazwan Abdullah, Abdul Bashah B Mat Ali, and Muhammad Ahmad, "AL-Quran Themes Classification using Ontology", Proceedings of the 4th International Conference on Computing and Informatics, ICOCI 2013.
- [23] Saidah Saad, Naomie Salim and Hakim Zainal, "Islamic Knowledge Ontology Creation", Internet Technology and Secured Transactions, 2009.
- [24] Hossam Ishkewy, Hany Harb and Hassan Farahat, "Azahry: An Arabic Lexical Ontology", International Journal of Web & Semantic Technology (IJWesT), Vol.5, No.4, October 2014.
- [25] Protégé, <http://protege.stanford.edu/>
- [26] Apache Jena, <https://jena.apache.org/index.html>