Ontology Based Semantic Search in Holy Quran

Hikmat Ullah Khan, Syed Muhammad Saqlain, Muhammad Shoaib, and Muhammad Sher

Abstract—The Holy Quran, due to its unique style and allegorical nature, needs special attention about search and information retrieval issues. Many works have been done to accomplish keyword search from Holy Quran. The main problem in all these works is that these are either static or they does not provide us semantic search. In this paper, we propose that the concepts of ontology of semantic web can be applied for carrying out semantic search in Holy Quran. For this purpose, exploratory search have been done from semantic web field of knowledge. The sample domain ontology, based on living creatures including animals and birds mentioned in Holy Quran, has been developed in protégé ontology editor tool. SPARQL Queries have been run to depict the proper role of ontology. Then certain recommendation for the project of attaining semantic search from all domains and resultantly all text of Holy Quran has been proposed. These recommendations include model and framework including creation of Quranic WordNet, integration, merging and mapping of domain ontologies under the umbrella of upper ontology. This work can be extended to other Islamic knowledge sources like Hadith, Fiqh etc.

Index Terms—Framework, information retrieval, ontology, semantic search.

I. INTRODUCTION

The last decade is referred as age of information and surely this is the era of knowledge. Every one from novice user to expert desires authenticated, reliable and relevant results of queries. There is requirement to develop an application that may give us dynamic conceptual search. With the emergence of semantic web as a modern field where structure can be formulated that may be understandable not only for humans but also for machines, domain specific ontologies are created and inference from ontologies leads us to Semantic search [1].

A large number of software is available on web about search from Holy Quran as document has been discussed in our earlier research work [2]. The software provides electronic audio and text representation facility and some of them provide keyword searching facility but lack the semantic search. Few software show remarkable topic search against a query word but these are just static search based software.

Quran has its unique style of describing the topics. At some places, some topics are explicitly mentioned while some others are meant implicitly. There are many difficulties for implementation of semantic search for Holy Quran as mentioned in [2].

The authors are with the Department of Computer Science and Software Engineering, International Islamic University, Islamabad (e-mail: Hikmat_niazi@yahoo.com, Syed.Saqlian@iiu.edu.pk, shoaibishaaqiiu@yahoo.com, m.sher@iiu.edu.pk).

A. Semantic Search

The word semantics is derived from the Greek word semantikos,"significant" [3] from semaino, "to signify, to indicate" and that from sema,"sign, mark," [4]. Linguistically, it is the study of interpretation of signs or symbols which are used for some specific contexts [5]. Semantic analysis is the process of relating syntactic structures, from the levels of phrases, clauses, sentences and the whole text, to their language-independent meanings. It causes removal of characteristics related to fussy linguistic contexts up till that level that such a project is possible. [6].

II. RELATED WORKS

Saidah Saad *et al.* [7], [8] describe ontological work for extraction key word and key phrase candidate for developing ontology of Islamic literature. It presents an algorithm for automatic extraction of the key words. They have presented a general and skelatel methodology and life cycle for building the ontology for Islamic literature. They also applied their technique on English text for mining ontologies from natural language.

Ontological work for semantic search is getting mature day by day, a short descriptions of related work earlier achieved are discussed. By using semantic reasoning based on ontology, [9] improves the search engine to understand user's query purpose. A prototype of search engine is developed and the search results are better than those of common search engine. In proposed framework in [10], query is parsed, looked for RDF triples in domain ontologies and relevant web documents, after ranking, are retrieved. SemPub [11] introduces concept space, graph of terms occurring in documents associated by the frequency and relationships between them. Thus by optimizing complex object space into more meaningful concept space, customized information retrieval is improved. The general structure of semantic search supporting is developed and some principles of followed ontology method formalization have been proposed with the help of framework of UML sequence diagram [12].

Semantic retrieval system based on ontology is given by constituting ontology entity and the language that describes the ontology too [14]. ESTER [15] supports text queries, ontological queries as well as combinations of the two by reducing queries to two fundamental operations, prefix search and join, implemented efficiently in terms of both processing time and index spaces.

III. IMPLEMENTATION

In [2], we have extensively discussed the issues that are obstacles in achievement of semantic search specific to Holy Quran. In focusing this, exploratory research has been

carried out and then sample domain ontology has been developed. Definition of such Islamic beliefs may well require very extensive care and team of different scholars should be given a task to define the concepts in such a way that it should be accepted universally. Thus all such domains have been avoided and domain of animals mentioned in Holy Quran has been selected. The Basic behind it was that it is absolutely free of such concepts that can cause differences. Additionally the concepts are largely physical and well defined. When it was selected that it also helped as the taxonomy of the animals is well defined and unambiguous.

A. Difficulties in Animal Domain

Simple animal ontology is easy to build as taxonomy whether it is general or scientific is easy to formulate and develop, but the definition of animal related concepts present in Holy Quran are difficult to define. First, it is worth-mentioning that we have only tried to declare all concepts exactly as defined in Holy Quran. There are about 167 direct or indirect references of animals in Holy Quran [16]. Main difficulties include

In Arabic for one animal many Arabic words are present. For instance, a camel has words like Abal, Buaaeer, Gamal.

Majority of animals have been mentioned regarding giving similarity to some people. This creates difficulty in creating metaphorical or abstract relationship. For example, the non believers who have no faith in Holy Quranic verses are resembled to donkey carrying books. Many animals have been mentioned regarding their behavior. For instance, ass, mule and horse have been defined as source of usage of ornament and carrying luggage. Behavior is such a difficult domain that two universities SIG are extensively working merely on the domain of animal behavior. [17].

More problematic is when animals are mentioned in some specific context. For instance, donkey gets frightened when sees a lion. As a relation of context is NLP is still a problem under consideration thus these were challenging scenarios. A separate ontology structure has been defined for context based animal referencing. There is very thin line between defining basic functionality of animal and its behavior. Thus, concept of property and class has been used to separate. This is defined as separate two classes of enumerated type in separate ontology.

B. Selection of Ontology

English Translation of Holy Quran done by Pickthall has been selected [18]. It is regarded as one of classical and authentic English translations as done by the person who has English as first language.

C. Precautions for Development of Domain Ontology of Holy Quran

Muslims believe that words of Holy Quran are divine and eternal. No alteration is it whatsoever is possible as Allah Almighty Himself has taken into His Hand. Even being static document, we still have to be very cautious about the following regarding definition of concepts in Holy Quran.

Define the concepts exactly the same manner defined in Holy Quran. No perception, addition of concept, addition of characteristics should be allowed. Even this addition is obvious ad right one.

If extensibility of ontology is required, for instance, we want to define the terms as mentioned in Holy Quran as well as in Hadith books, then either new ontology should be developed and Holy Quran ontology should be used Upper ontology. [19].

If we want to add definition of other knowledge domain then separate ontology should be developed and mapping/merging of that ontology should be carried out. For instance, if we want to improve the search of animals mentioned in Holy Quran then we can define the taxonomy of animals on scientific basis and map with our ontology.

D. Steps in Implementation

Ontology construction is a creative work. Ontology are iterative in nature and never complete. There are many steps and issues that are considered in sequence to create ontology. Reference [20] is the regarded as best source for first ontology construction and have been adopted.

Complete survey of Holy Quran was carried out to enumerate all the terms and then create Ontology covering all animals and birds mentioned in Holy Quran. This was quite extensive work as it was meant to be done extensive manual exploration work. All the verses of Holy Quran has been searched by using certain index books [16], [21], [22] and helping material The important difference of ontology with respect to conventional ontology was that as the animals have been mentioned in Holy Quran in some contextual way. That is, domain was not easy one and well defined thus, contextual analysis and depicting context as defined in Holy Quran has to shown in our Ontology.

There is no ontology present which may have defined as concepts defined in Holy Quran. This there is no question of reusing. Yes, there are many ontologies for definition of concepts regarding animals [17], [23], [24], but all these do not fulfill our requirement of definition of concepts as defined in Holy Quran as these are basically based on general or scientific taxonomy.

For extensibility purposed, we have focused of creating more classes so that in future, more characteristics regarding other knowledge subjects like science, Hadith etc can easily be made part of it. The part of class hierarchy showing classes is shown. (Fig. 1-Fig. 4)
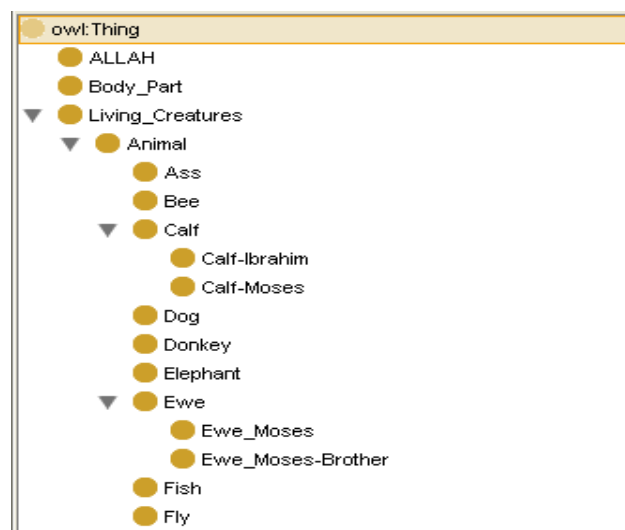

Fig. 1. Part of the sample ontology

Name	Prefix	Range	Domain	Inverse	Other Characteri...
Becomes		Snake or Swine	Moses_Nation ⊔ Object		
Bring		Calf-Ibrahim	owl:Thing		
Can_not_reate		Fly	Man		
Carry		Object	Ass		
Causes_to_Forget			Satan		
Choose		Calf-Moses or Object	Bee ⊔ Moses_Nation		
Deal_with		Elephant_Owner	ALLAH		
Eat		Goat or Object or Qu...	Bee ⊔ Children_of_Israel ⊔ Man		
Forbade_to_eat		Swine	ALLAH		
Forget		Fish	owl:Thing		
Frightens_from		Lion	Ass		
Give_example		Gnat	ALLAH		
Has_Body_part		Body_Part	Bee ⊔ Goat ⊔ Horse		
Has_Brother		Moses-Brother	owl:Thing		
Has_Quantity		owl:oneOf{}	Ewe_Moses ⊔ Ewe_Moses-Brother		
Has_Strength		owl:oneOf{"Weakes...	Object		
Has_Usage		owl:oneOf{"Orname...	Ass ⊔ Horse ⊔ Mule		
Has_Voice		owl:oneOf{"Harshes...	Ass		
Hunt		Fish	Man		
Inspire		Bee	ALLAH		
Like		Dog or Non_Believers	Ass ⊔ Non_Believers	■ Like	Symmetric

Fig. 2. Part of the object properties defined in the ontology

IV. SPARQL QUERIES AND RESULTS

To depict the accuracy and functionality of the ontology for checking that the concepts from relationship of other concepts and as well as from its characteristics can be retrieved not necessarily from its keyword only, we have used SPARQL query language [25]. Thus ensures that using ontology we can do semantics manipulation and inference, among many following sample queries were used.

Query: Which animal swallow Prophet and also lives in Sea?
Answer: Fish
SPARQL Query:
SELECT   ?animalClass
WHERE { ?animal :Swallow :Yunas.
:Yunas rdf:type :Prophet.
?animal rdf:type ?animalClass.
?animalClass rdfs:subClassOf :Animal.
?animal :Lives_in "Sea" }

Query: Which Animal is forbidden (to eat) by ALLAH
Answer: Swine
SPARQL Query:
SELECT   ?animalClass
WHERE {
:Allah :Forbade_to_eat ?animal.
?animal rdf:type ?animalClass }

Query: which animal has usage of ride:
Answer: Horse, Mule and Ass
SPARQL Query:
SELECT ?animalClass
WHERE { ?animal :Has_Usage "Ride".
?animal rdf:type ?animalClass }

V. RECOMMENDATIONS

A. Dynamic Contextual Ontology

There has been unique Endeavour in this regard that as it has revealed after exploration of Holy Quran for getting domain knowledge that most of the references are given with respect to some context and all these are very unique and very rarely relating to each other. So ontology based on this would be static one as for each context, we have created ontology specifically for that context.

B. The Fourteen Primitive Actions

Roger C. Schank established a conceptual structure to represent the conceptual information underlying a natural language sentence [26]. It uses the basic actor-action-object framework. It was the intent that these structures have only one representation for one meaning, regardless of the semantic form of the sentence being represented. It has been found out that these fourteen actions classification does not fulfill the requirement for actions mentioned in Holy Quran. For instance, Slaughter is an action that cannot fall in any of the fourteen categories mentioned above. So it is suggested that research should be done and new classification of primitive actions should be introduced.

C. Development of Quranic WordNet as Upper Ontology

It is proposed that as WordNet [27], [28] and Arabic WordNet has been constructed as an upper merged ontology for the querying from the concerned language thus Quranic WordNet should be developed. Quranic WordNet will be collection of all sort of terms used in Holy Quran. The benefits of proposed work are obvious and immense. The Quranic WordNet will be collection of Quranic concepts and will be used to define the Islamic terms which are either not defined in WordNet and Arabic WordNet and if found then these concepts have defined by non-Muslims then surely we Muslims would like to have our own definition of concepts according to our faith and belief.

D. Islamic and Quranic WordNet

This development may be extended further to develop Islamic WordNet, a huge WordNet containing definition of concepts from all spheres of Islamic knowledge. This Islamic WordNet may comprise of Quranic WordNet, Hadith WordNet and Fiqh WordNet and many more. All these will be used as upper merged ontology and domain ontologies from all these sources will be created and would be used using techniques like ontology merging, ontology integration and ontology mapping at abstract level.

VI. THE PROPOSED FRAMEWORK

The Let us finally discuss the proposed framework. There is various consideration level of proposed framework thus these are being proposed in context with certain already existing framework not only proposed but also implemented. On the basis of framework proposed in [14], the framework has been proposed that explains the layers into better modules and in a sequence. It clearly depicts that Holy Quran document will be used for annotations and that will be directly linked with Semantic Web Rule Language that will be used to give us knowledge base consisting of classification and rules. At the same level, ontology developers along with mutual consultation with domain experts will develop domain ontology. Such consultation is mandatory because proper definition of concepts is ultimate goal and no second thought can be taken into consideration. The same care has been taken into account that concepts have only been defined as defined in Holy Quran only.

As in our recommendation that this model can be extended to Islamic Documents/Books of Hadith and Fiqh thus for

such complex architecture, the following architecture is being proposed based on [10]. It is evident that query will be processed in the following order that when the user request a query. Then if it is simple keyword based query that it can be bypassed from WordNet, the sole purpose is not to complicate the process into many definition, as not required also less time will be consumed. Then parser will be used as the query can be broken down into token and parsing can be done. Next step is Reasoner, as inference for Reasoner is based on ontologies then thus Reasoner is linked with ontologies data/knowledge base which were created on the basis of Holy Quran text. As we want to do search the concepts from other Islamic sources on the basis of ontologies of Holy Quran thus other documents can be annotated for semantic search from those documents. As the number of documents is more then ranker has been proposed to rank the result on simple logic of frequency of occurrences in all those documents except Holy Quran.
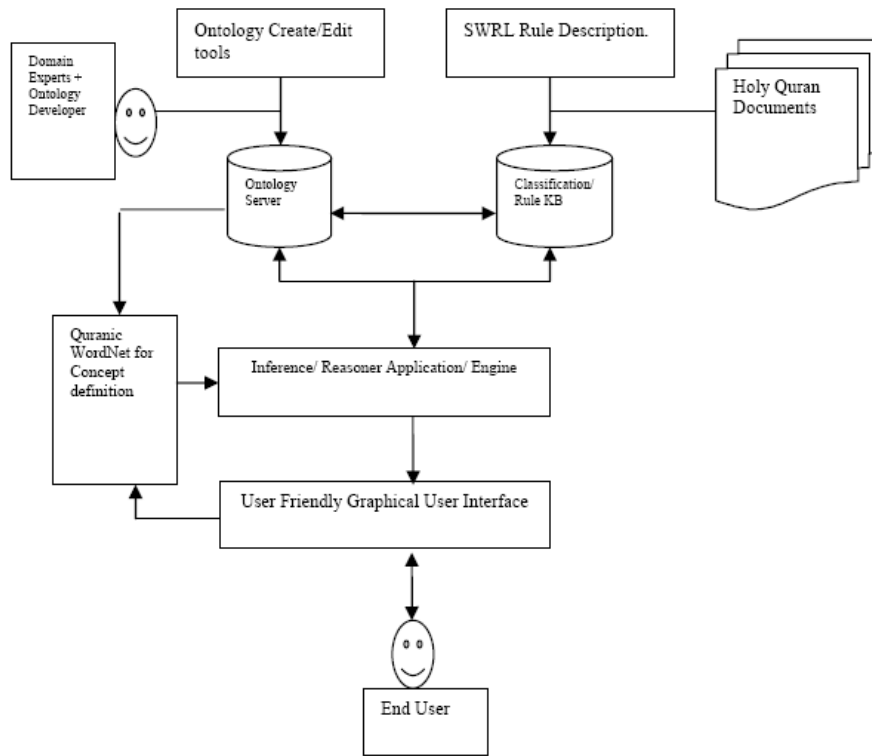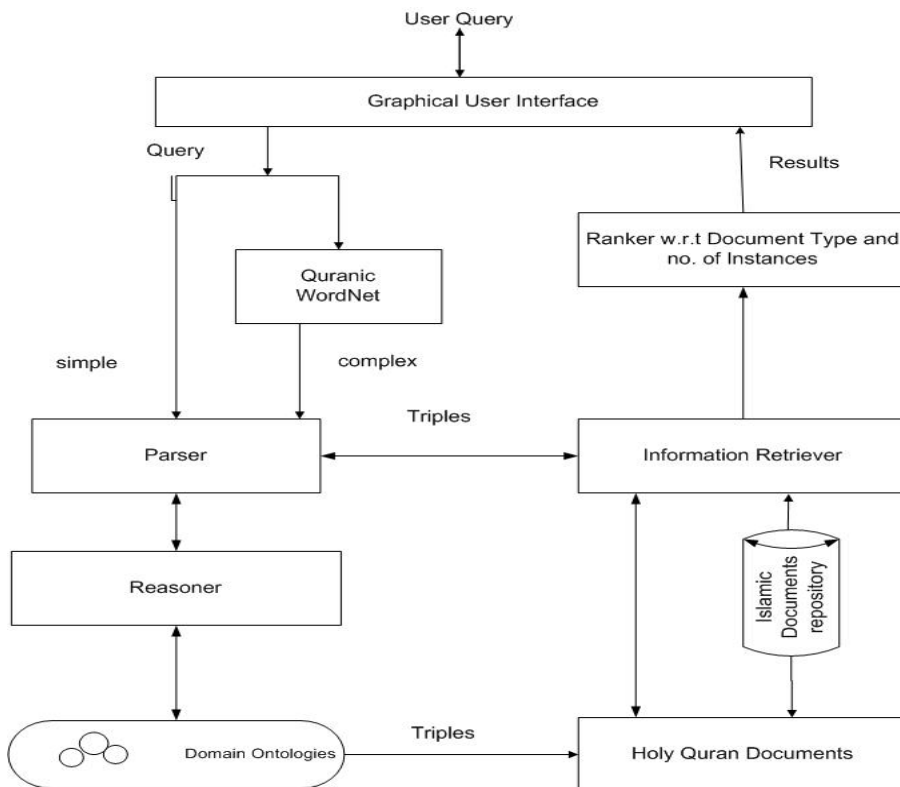

Fig. 3. The proposed framework

Fig. 4. Query processing for the proposed semantic search system

VII. Future Works

As described earlier too, but there are many tasks still to be done to accomplish the complete task of semantic search in Holy Quran. First Quranic WordNet is the requirement as it will contain definition of concepts as defined in Holy Quran. Also, as the classification of fourteen actions lack the ability to describe actions described in Holy Quran thus, research should be done and new classification of primitive actions should be introduced.

As the extensive work on sample ontology for the domain of animals and birds have been done but extensive research is required to create domain ontologies for all main domains mentioned in Holy Quran. Also Implementation of Complete Application covering the entire proposed framework needs to be accomplished. This work can be extended to Hadith and Fiqh documents. It is hoped that our recent works will also be helpful to work in such domains to include our experiences in knowledge management [29], novel algorithms [30] and use of Suffix Array based RDF Indexing using RDQL Queries [31].

VIII. Conclusion

A lot of work has been done in information retrieval domain of knowledge but no work has been done for efficient topic search from Holy Quran. Thus, this work proposes that the concepts of semantic web used can be used for implementation of semantic search in Holy Quran. It is basically exploratory work, thus initially, the causes of problems of search from Holy Quran has been explored. Then the extensive exploratory work has been carried out to get the knowledge in this domain. Then the recommendations regarding proposed model including proposed basic building block, ontology language, type of OWL, ontology editor, reasoners etc. Then the important concepts like upper ontology, ontology merging, ontology integration and ontology mapping has been discussed. WordNet is important structure giving dynamic touch to information retrieval from documents as well as web pages. But already existing WordNet cannot be used thus it is proposed that Arabic WordNet and Quranic WordNet should be built. Inline with the basic recommendation, sample ontology for domain of animals mentioned in Holy Quran has been developed and SPARQL queries of various types has successfully been run that show the Inference ability and ability of ontology that it can lead to overall big project for semantic search project for Holy Quran.

References

[1] P. Mika, "Ontologies are us: A unified model of social networks and semantics," *LNCS*, vol. 3729, pp. 522–536, Springer, Heidelberg (2005)

[2] M. Shoaib, M. N. Yasin, U. K. Hikmat, M. I. Saeed, and M. S. H. Khiyal, "Relational Word Net model for semantic search in holy quran," in *Proc. International Conference on Emerging Technologies*, Pakistan, pp. 29-34, 2009.

[3] H. G. Liddell and R. Scott. Semantikos. [Online]. Available: http://www.perseus.tufts.edu/cgi-bin/ptext?

[4] S. H. G. Liddell and R. Scott, *An Intermediate Greek-English Lexicon*, 2009.

[5] O. Neurath, R. Carnap, and C. F. W. Morris, *International Encyclopedia of Unified Science*, Chicago, IL: University of Chicago Press, 1955.

[6] Semantic Analysis and Linguistic Analysis. [Online]. Available: http://en.wikipedia.org/wiki/Semantic_analysis_%28linguistics%29.

[7] S. Saad, N. Salim, and N. Omar, "Kkeyphrase extraction for islamic knowledge ontology," *Information Technology IT Symposium*, vol. 2, August 26-28, 2008.

[8] S. Saad and N. Salim, "Build islamic ontology based on ontology learning," Postgraduate Annual Research Seminar 2007, July 3-4, 2007.

[9] Z. Q. Du, J. Hu, H. X. Yi, and J. Z. Hu, "The research of the semantic search engine based on the ontology," in *Proc. International Conference on Wireless Communication*, pp. 5403-5406, 2007.

[10] W. D. Fang, L. Zhang, Y. X. Wang, and S. B. Dong, "Towards a semantic search engine based on ontologies," in *Proc. of the Fourth International Conference on Machine Learning and Cybernetics*, Guang Zhou, vol. 3, pp. 1913-1918, 18-21 August 2009.

[11] R. Loganantharaj and V. B. Narayan, "SemPub: An ontology based semantic literature retrieval," in *Proc. of the 19th IEEE Symposium on Computer –Based Medical Systems* (CBMS'06), pp. 875-880, 2006.

[12] R. Pasichnyk and A. Sachenko, "Semantic WEB-Search Developing by Problem-Oriented Ontology Means," in *Proc. IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications 2007*, Dortmund, Germany, pp. 445-448, September 6-8, 2007.

[13] Q. Y. Sheng and G. S. Ying, "Measuring semantic similarity in ontology and its application in information retrieval," *2008 Congress on Image and Signal Processing*, vol. 2, pp. 525-529, 2008.

[14] X. H. Zhang and G. S. Ying, "Ontology-based semantic retrieval system," in *Proc. International Conference on Wireless Communication, Networking and Mobile Computing*, pp. 1-4, October 12-14, 2008.

[15] H. Bast, F. Suchanek, and I. Weber, "Semantic Full-Text Search with ESTER: Scalable, Easy, Fast," in *Proc. 2008 IEEE International Conference on Data Mining Workshops*, pp. 959-962, 2008.

[16] M. A. M. D. Abadi, and A. H. F. A. Quran, *Hewanat-E-Qurani*, Book Published by: Hijra International Publishers, Lahore, pp.4-44, 2000.

[17] Animal Behavior Ontology. [Online]. Available: http://metadata.net/WildNET/Animal.owl.

[18] J. Cardoso, "The semantic web vision: Where are we?" *IEEE Intelligent Systems*, September/October 2007, pp. 22-26, 2007.

[19] Protégé tool Features. [Online]. Available: http://protegewiki.stanford.edu/index.php/Protege4Features.

[20] N. F. Noy and D. L. McGuinness, "Ontology development 101: A guide to creating your first ontology," Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.

[21] A. Rahman, *Subject Index of Quran*, Published by: Islamic Publications (pvt) Ltd, Lahore Pakistan, pp. 12-29, 2001.

[22] S. H. K. Q. Jhangiri and J. Q. Ashariya, Published by: Maktaba Ashat Talemaat-Ul-Quran, Karachi, Pakistan, pp. 2-23, 1998.

[23] AI. RUG. [Online]. Available: http://www.ai.rug.nl/ki/2006/practicum/eerstejaars/onto/animal.owl.

[24] Animal Ontology. [Online]. Available: http://metadata.net/WildNET/Animal.owl

[25] S. Eric and S. Andy. SPARQL Query Language for RDF. [Online]. Available: http://www.w3.org/TR/rdf-sparql- query/

[26] R. C. Schank, "The fourteen primitive actions and their inferences," Stanford Artificial Laboratory, Computer Science Department, Stanford University, USA, 1973.

[27] C. Fellbaum, *An Electronic Lexical Database*, Bradford Books, 1998.

[28] R. Richardson *et al.*, "Using WordNet as a knowledge base for measuring semantic similarity between words. School of computer applications," Dublin City University, Glasnevin, Dublin 9, Ireland, pp. 123-128, 1994.

[29] T. A. Malik and H. U. Khan, "Performance measurement using distributed performance knowledge management system: empirical case study of coca cola enterprises," *International Review of Business Research Papers*, vol. 6, no. 1, pp. 250-282, Feburary 2010.

[30] T. A. Malik, H. U. Khan, and S. Sadiq, "Dynamic time table generation conforming constraints a novel approach," in *Proc. International Conference on Computing and Information Technology (ICCIT 2012) 1st Taibah University International Conference held in Al-Madinah Al-Munawwarah*, Saudi Arabia on 12-14 March 2012.

[31] H. U. Khan and T. A. Malik, "Finding resources from middle of RDF graph and at Sub-Query level in suffix array based RDF indexing using

RDQL queries," *International Journal of Comuter Theory and Engineering*, vol. 4, no. 3, June 2012.

Hikmat Ullah Khan has done his Master of Computer Science and Master in Science (Computer Science) from Department of Computer Science and Software Engineering, International Islamic University, Islamabad, Pakistan. He has keen interests in research field of Information Retrieval, Semantic Web and Semantic Cache, Web Mining, Social Network mining etc. He has already published many papers in international conferences and journals. He has served as reviewer in many international conferences including 2011 4th IEEE International Conference on Computer Science and Information Technology (IEEE ICCSIT 2011) held in China on June 10 - 12, 2011. This research work was done as part of research thesis for Master of Science (Computer Science).

Syed Muhammad Saqlain is serving as Assistant Professor in Department of Computer Science and Software Engineering, International Islamic University, Islamabad. His research interest includes Image Processing, Computer Vision and Machine Learning.

Muhammad Shoaib is PHD (CS) scholar in Department of Computer Science and Software Engineering, International Islamic University, Islamabad. His research interest includes data mining, academic social network and name disambiguation.

Muhammad Sher is serving as Dean, Department of Computer Science and Software Engineering International Islamic University, Islamabad. He has done his Ph.D. in Computer Science from TU Berlin, Germany, and his M.Sc. in Computer Science from Quaid-e-Azam University, Islamabad, Pakistan. His research interests include but not limited to Next Generation Networks, Network Security andAlgorithms, Computer Networks, Information Retrieval etc. He has been awarded with many awards, few are as follows:

Fifth Annual (ISC)2 Asia-Pacific Information Security Achievements (ISLA) Award by International Information Security System Consortium (ISC)2 Asian Advisory Board, against the Category "Senior Information Security Professional of Asia 2011," at Jakarta, Indonesia, on 26th July 2011. The award is given on the basis of research and development contribution and recognition in the field on Information Security at International Level. "Best University Teacher Award 2008," by Higher Education Commission (HEC), Islamabad, Pakistan, on 21st August 2009. The award is conferred on the basis on academic, research and creative work performance for the year 2008. "HEC-DAAD Fellowship Award 2004," for higher studies of Universities & Degree Awarding Institutions Teachers of Pakistan in Germany Programme, for three years (2004-2007). "National IT Excellence Award 2000" by NCR Pakistan, in recognition for excellence achieved the field of IT Research and Development, for the period of 1997-2000 at Convention Center Islamabad, on 25th November 2000.