KNOWLEDGE EXPLORATION: SELECTED WORKS ON QURAN ONTOLOGY DEVELOPMENT

[1,3]FAROOQUI N.K. AND [2]Dr. MOHAMMAD FAUZAN NOORDIN

[1]Department of Computer Science, Kulliyah of Information and Communication Technology, International Islamic University (IIUM), Kuala Lumpur, Malaysia.
[2]Associate Professor, Department of Information Systems, Kulliyah of Information and Communication Technology, International Islamic University (IIUM), Kuala Lumpur, Malaysia.
[3]Lecturer, Department of Information Systems, College of Computer Engineering and Sciences, Salman Bin Abdulaziz University, Al Kharj, Saudi Arabia.

E-mail: [1]farooquinajam@gmail.com, farooqui.najmuddin@live.iium.edu.my, [2]fauzan@iium.edu.my

ABSTRACT

This paper presents key features and challenges ahead for the development and knowledge retrieval of Quran ontology. Recent studies have made significant advances towards the development of Quran ontology. In the recent past, there have been numerous studies conducted on the application of semantic technologies on Quran. Contribution of this paper is its focus on finding the direction of knowledge exploration in Quran. Several studies on Quran ontology development help us in analyzing its linguistic features. Another dimension where Quran excels is the Knowledge of it. There are few studies that concentrate on retrieval of knowledge from Quran. In this literature review, we have included studies that can help us in developing semantic application for knowledge exploration from Quran. This paper devises challenges mainly focusing towards exploration of knowledge in the Quran and summaries research in this area, discuss key features and open research issues.

Keywords: *Islam, Knowledge management, Ontology, Semantic web, Semantic technologies, Quran.*

1. INTRODUCTION

Recently there is growing interest in Quran semantic studies. Last six years there has been published work from across the world including the University of Leeds. The complex structure and contextual complexity of the Quran text has an appeal for semantic research. These are various studies concentrating different factors and applying different methods. Researchers have followed mix approaches, such as developing different models, analyzing and applying semantic basics to Quran text, experiment with ontologies and evaluating with queries and developing tree structure of concepts in Quran. One such significant work such as Quran Arabic WorldNet model was proposed by (Manal AlMaayah1, Majdi Sawalha2, 2014). They defined semantic relations with synonyms

and autonyms. There is another very important field growing where researchers want to study Quran according to topics. They arrange the verses according to the themes. One such study by (Ta, Abidin, Abdullah, Ali, & Ahmad, 2013) elaborates the method to extract meanings of words. They studied how; meanings of words are described in English in the general sense. They further discuss that, meanings are extracting upon, words position in a sentence and its relationship with other words in the fields.

Recent studies not only focus on Quran studies but also studies have been done on Islam, Information Technology and related issue. For example, a complete book on ICT and Islam (Mohammed Fauzan Noordin, 2009) discuss similarities in Information technology and Islam. It further elaborates how Islamic concepts are

usable and applicable to information technology. It further discuss applicability of Islamic believes and issues such as a knowledge society, accountability in e-governance & e-commerce and privacy & security. Another related study was published by (M.F. Noordin & Othman, 2006) proposed a design for information retrieval from Quran. Security and ethics are other important areas, where Islamic ethics and security issues are far ahead of Information Technology. Application of Islamic concepts in security and ethics supported by crime figures in Malaysia was published by (Mohamad Fauzan Noordin, 2013).

This paper studies recent Quran semantic research. We try to find out the current trends and technology being applied. We also sort out the past outcomes of this research. This paper includes research towards the development of Quran ontology. We consider here studies that are helpful in knowledge exploration from the Quran. This study focus on research works that try to analyze and understand knowledge in Quran by applying through semantic technologies.

Goal of our study is to find the possibility of helpful features, which are applicable in knowledge exploration from Quran. In the recent past, there have been reviews of research conducted on applications of semantic studies of Quran. Our work focus on knowledge rather than analyzing language as previous studies focuses on.

One such review was conducted by (Alrehaili & Atwell, 2014). The paper reports on a survey of recent Quran ontology research project. They compared their studies on a criteria such as (1) Coverage Area (ontology for animals or particular domain) (3) Coverage portion (complete Quran/some chapter etc. (4) Underlining format (xml, PDf, owl) (5) Underlining technology (technology for building ontology). (6) Availability. (7) Concept number (Concrete /Abstract) (8) Relation type (Antonym/synonym). (9) Verification method (Scholar /Domain expert). The paper did not

solve all these issues but highlight the critical factors.

Rest of paper is organized as follows. **Section 2** describe in detail, recent studies conducted, **section 3** presents observation about the studies, section **4** write the challenges, **section 5** discuss open research issues and last **section 6** concludes the paper.

2. RECENT STUDIES ON QURAN ONTOLOGY DEVELOPMENT

Growing interest among the Muslims and non Muslims about the Quran Knowledge and its comparison with the other knowledge domains has created a wave of Quran Ontology development initiatives by different researchers around the world. Following are some important studies done, on the Quran Ontology and knowledge extraction.

2.1 Quran Arabic WordNet (Manal AlMaayah1, Majdi Sawalha2, 2014)

Paper proposes an Arabic WorldNet for defining easier semantic relation specification and relations. Paper analyzed the requirement of Quran WorldNet. They found that Q corpus has 77,430 words, 114 chapter and 6,243 verses. The complex structure, contradiction and similarities in the synonyms. taking inspiration from studies such as EuroWordNet, Global WorldNet (Fellbaum, Christiane, 2012), (Black, Elkateb, Pease, Rodriguez, & Alkhalifa, 2006) lexical resources for modern standard Arabic. (Trad, Koroni, Mustafa, & Almaghrabi, 2012) compared between Arabic and English documents. Authors prepared the Quranic WorldNet which is (1) preprocessed (2) synset, where synonyms are made in-groups and semantic relation are defined by (a) synonyms (b) Autonyms (c) glossary (d) Similarity. The model is a conceptual Quran WorldNet.

2.2 Quran Verse Extraction for various concepts was studied by (Aliyu Rufai Yauri, Rabiah Abdul Kadir, 2013)**, and received about 95% accuracy**. Users can search concepts such as list of Halal Food, and related verses from

Quran will be extracted from Quran ontology. Leeds Quran Ontology was reused. It has about 300 Concepts and 350 relationships. Some concepts and relationships were added as required. Protégé was used to build Quran ontology and to add more relationships. The components of Quran ontology are concepts, sub concepts, equivalence, property and inverse. Protégé, in built reasoner was used. Manchester, OWL syntax have been used for writing OWL abstract syntax. The ontology is represented in RDF, which is a formal language for describing structured information. Manchester OWL syntax which is influence by both OWL abstract syntax and DL style syntax. Both simple and complex queries were prepared such as '*is halal some food*' or '*Where Muslims face during Prayers?*' "*Which angel Allah used to communicate with Prophet Mohammed*? The system follows the same logic concepts to retrieve "Jibreel". Outcomes were calculated with two factors, *Average Precision* and *Average Recall*.

2.3 Thematic Classification   by (Ta et al., 2013). Two approaches were presented for ontology building, *class based* and *instance based*. In order to understand the true meaning of words, they referred to papers written in 1972 by Ulmann..He described that there are two ways in which meanings of words are considered. First it may be Analytical or operational. The analytical approach defines meaning by analyzing componental features of words, and the operational approach studies the words. According to field theory of semantics (conceptual spheres), proposed meaning of a word is considered within a given view of the world. It is dependent on its relation to other words in the same semantic field (conceptual area), word meaning is established by the position within the field, and the relationship it has with other words in its field.

The classification of concepts in their ontology is based on *Syammil Al Quran Miracle the Reference*. The classification was done, as Surah, Ayah and Juz for each theme, like Iman / Akhlaq. Both Class based ontology and instance based ontology have two levels of classes; one is

conceptual level and lexical level. The factors were considered to compare two approaches are (1) Expandability 2) Encoding bias 3) Minimal ontological commitment. Minimum encoding bias referred to "design decisions should be independent of the implementation features("Kalfoglou, Y. (2001). They take help of . Quran expert to validate the Classification and division of themes and sub themes. They validate whether searching on a particular theme, retrieve correct class in protégé.

2.4 An Experience of Developing Quran ontology with contextual information support (Iqbal, Mustapha, & Yusoff, 2013).

This paper discusses contextual information inclusion in the Quran ontology. Mixed methodology for construction of the ontology was applied A tabular form of concept, instances, data properties and object properties is given clears the depth of knowledge being covered in the ontology. Evaluation of the ontology was carried out by DL Queries. The system gives satisfactory results as per the system expectation.

2.5 Quran Ontology with Competency was presented by (Iqbal et al., 2013). Mixed methodology to build Quran ontology applied. Competency questions were used, to define the scope of the ontology that is being considered.

2.6 Extract Knowledge from Quran, Ontology Semantic Approach was used by (Aliyu Rufai Yauri, Rabiah Abdul Kadir, Azreen Azman, 2013)**.**

A Quran ontology model was proposed and presented a query reformation against the Quran ontology stored and annotated in the knowledge base. The answers are verse related to user query. Model, submits query that is then reformed to suite the knowledge base. Leeds Quran ontology was reused. Jena API was used in Extraction module. Reformation of the query involves tokenization, stop words removal, lemmatization and part of speech tagging. A

ranking algorithm was to rank the best suited triple in order to translate it into SPARQL query language.

2.7 Class based and instance based ontologies for Quranic words were presented by (Al-Khalifa, Al-Yahya, Bahanshal, Al-Odah, & Al-Helwah, 2010).

The componential analysis of Quranic words discussed in detail. The main theme of discussion is how the meaning of words was extracted in the general English and factors which affect the contextual meaning of the words. The class based ontology approach; two levels of classes are developed. Top level represents concepts, and lower level represents lexicons. OWL 2 was used as it has a new feature called '*punning*'. In the instance-based ontology approach, top-level classes are same, while second level represents instances or individuals. Both the ontology approaches were compared, and instance-based ontology approach is found easily scalable.

2.8 Modeling Quranic knowledge, its difficulties and differences with other knowledge, was studied in detail by (Baqai, Basharat, Khalid, Hassan, & Zafar, 2009).

Various challenges for semantic modeling of Quran Knowledge were discussed such as a structural organization of Quran, Thematic complexity of Quran, Organic Unity and coherence in the complex structure of Quran. The study reveals that the structure and themes of the verses are not arranged in particular order. The linguistic and contextual arrangement is far more complex than any ordinary book, and there is a need to study linking of knowledge in Quran and other sources like Hadith Knowledge. Basic layered architecture of semantic web were modified into Knowledge Modeling tool for retrieval of Quran knowledge and present, framework for 'semantic knowledge modeling and retrieval of Quran Resources' using Protégé OWL API, semantic authoring of Quran knowledge. Two main categories were created

one for Quran ontology and other for Hadith ontology that is further categorized into a number of primitive and defined classes for each domain. Both DL (Descriptive Logic) based reasoning, and Rule based Reasoning were used. The contextual relationship was written with a relation 'isContextuallyRelatedTo". Automatic, contextual relations are difficult to populate. Hence an index mechanism was used, to make this context relation authentic. Knowledge retrieval is done with the help of SPARQL Query and Jena API. The relevance of the output documents, are calculated if, its more than 50 % documents are returned to the user. Otherwise query is reformatted in order to get better results. They used weights for each match, in order to get desired results.

2.9 Quran Search for Concept Tool and Website (Abbas, 2009)

This work is the most comprehensive and dedicated towards improvement for better than word searching. She elaborated that concepts can be concrete concepts such as name, places or names of cities or concepts can be general concepts are concepts such as main pillars of Islam. This can also be called as "Abstract concept". To understand the abstract concept in Quran 'Tafsir' is quit useful, but referring does not cover the complete picture of the message with their context. For this set to the central theme is compulsory. The use of thematic approach helps to understand the message before and avoid misuse of phrases by picking them by out of context. Various thematic classification of Quran and the way in which the themes were arranged were discussed. OWL is found more suitable to build ontology and define relationships. Ontology tools were integrated with the keyword searching tool, already there in python programming language which partially integrates RDF and OWL with Sparta, Tromp and Seth. Different search tools already available on the web studied and summarized them. Quran language is used to obtain tool available online. NLTK WorldNet dictionary was used to generate synonyms. She developed searching and tree, Tree helps organized learning. Using

Google app engine they developed same for the website.

An evaluation was carried out using recall and precision for all Quranic tools available online, results were compared with their own tool and better results were achieved. It is found that the tree of concepts of Quran is the first of its kind available online. This research can be extended to further studies.

Following table summarizes above recent researches …

Table 1: Key features of recent Quran ontology Tools

Research	Aim	Features	Results
Quran Arabic WordNet	To propose WordNet for Quran.	Use of synonyms and autonyms.	Model for WordNet of Quran useful for future research.
Verse extraction on various concepts	To extract verses according to the concept.	Reuse of Leeds Quran ontology.	Conceptual reverse extraction tool with 95% accuracy.
Thematic classification	To classify/arrange Quran ontology according to topic/theme	Use of contextual understanding of techniques from English language and logic.	Theme wise ontology of Quran with contextual features.
Quran ontology with contextual information	To develop contextual Quran ontology.	Efficient use of DL queries.	Contextual Quran ontology.
Semantic approach to extract knowledge from Quran.	To extract verses related to query.	Reuse of Leeds Quran ontology and reforming of query being processed.	Tool that retrieve verses related to query.
Class and instances based ontologies for Quran words.	To develop and compare two approaches class based, and instance based.	Use two levels of classes to induce the context.	Comparative features of two approaches.
Quran knowledge modeling and retrieval.	To model Quran knowledge and retrieve documents related to the verse.	Study of semantic complexity of Quran. Use both rule based and descriptive based reasoning.	Semantic modeling of Quran text and contextual retrieval of text.
Quran research tool "Qurainy"	Semantic conceptual organization of Quran.	Integration of protégé and python. Integration of tools from different sources such as Google I/O, protégé, and python.	Online and public usable tool for searching concept in Quran.

3. OBSERVATIONS

In-depth reading and analysis of the recent research work on Quran studies reveals following results.

- A protégé, OWL, RDF is the choice of ontology researchers.

- Study of basic semantic concepts application and ontology engineering can enhance the Quran ontology development processes.

- Considering the source of Quran text or its translation is very important it will affect broader view of the context.

- Most of the studies depend on SPAROL queries. Some apply descriptive logic. Rule based logic is mostly rejecting, otherwise used in combination. The reason may be; there are few straight forward rules to understand complex texts.

- Large collection of synonyms can play significant roles, particularly when the knowledge in the ontology has multiple words representing the same thing, and other ways round where one thing is being referred with multiple names.

- *Tafsir* can play a major role but cannot be decisive, as it may not reflect correct meaning that was intended for your specific application. Intended meaning can only be acquired by connecting the sentences or words with the central theme, current chapter and topic being discussed.

- Integrating, Protégé with other programming languages and tools such as Python, Sparta, or Seth can enhance the results.

- Conceptual studies should always be arranged in Tree structure, so as to convey correct meaning.

4. CHALLENGES

- Most studies concentrate in the lingual aspect of Quran natural language processing researchers studies them. Quran presents a sea of knowledge. There is an urgent need to capture the knowledge.

- Thematic organization of Quran presents sea of opportunity for the knowledge researcher. Quran has knowledge about each aspect of life.

For example following verses in Quran has deep knowledge embedded, a researcher needs to analyze, understand its meaning, and compare with current scientific knowledge. Once we match this knowledge using methods such as ontology matching, we will be in a position to understand and better search future scientific knowledge….

o "Have those who disbelieved not considered that heaven and earth were joined entity, and we separated them and made them from water every living thing? Then will they not believe." Chapter 21: Verse 30. [Translation *Sahih Muslim]*.

Above sentence (verse) from (http://tanzil.net/#21:30), inform the reader that *earth and heaven were a one entity, it was separated and everything is made of up water*.

o "He released two seas, meeting side by side." Chapter 55: verse 19.

"Between them is a barrier [so] neither of them transgresses." Chapter 55: verse 20.

[Translation *Sahih Muslim]*. http://tanzil.net/#55:19 and http://tanzil.net/#55:20.

Above two verses, revealed knowledge that, two seas meet other but they cannot mix each other or their water cannot be merged

because there is some unseen barrier in their water.

- Selection of right ontology designing or reusing methodology. There are online and offline ontologies available for download. Leeds Quran ontology, for example, is a good resource for research available at http://corpus.quran.com/ontology.jsp. There are numerous other studied that develop Quran ontology for a particular chapter of Quran or specific types of verses. For example, Quran ontology for salat was developed by (Saad, Salim, Zainal, & Muda, 2011). OWL DL ontology was used by (Aliyu Rufai Yauri, Rabiah Abdul Kadir, 2013) to extract particular verses from Quran. There is a strong challenge here to reuse the existing ontology for knowledge extraction or build a complete Quran ontology with aim of extracting knowledge.

- Use of large but optimal semantic basic concepts, such as defining attributes and instances so that, required meaning can be acquired, or knowledge can be properly engineered. Ontology is generally application specific. All-in-one Ontology is still not practical. There are reasons for this limitations, such as all properties required cannot be embedded in one ontology, secondly affecting computational efficiency and practical applicability of such ontology are difficult. Initial ontology matching tools concentrated on features such as 'SAMEAS'. Ontology matching tool match two elements of the ontology using *sameas*. Further, they use a set of synonyms to match two entities. In order to achieve better computational results, one needs to limit the number of synonyms being applied. Same are cases for other features and attributes which we define in ontologies. Optimum number of properties and attributes in ontology that serve the purpose of knowledge creation for which ontology is being developed.

- Selection of tools, combining tools with other add-on features or third party adds-ins, so as to get the maximum benefits out of the arrangement.

- Along with *Tafsir*, adding central theme and chapter consideration is of importance.

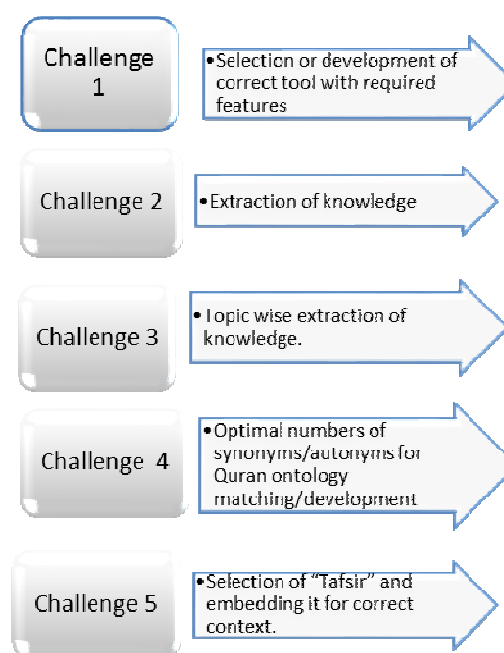Following figure shows, key challenges in developing Quran ontology …

Challenge 1	• Selection or development of correct tool with required features
Challenge 2	• Extraction of knowledge
Challenge 3	• Topic wise extraction of knowledge.
Challenge 4	• Optimal numbers of synonyms/autonyms for Quran ontology matching/development
Challenge 5	• Selection of "Tafsir" and embedding it for correct context.

Figure 1: Challenges in Quran Ontology Development and Challenges

5. OPEN RESEARCH ISSUES

Despite so much focus on application of semantic technologies on Quran following, are open research issues focusing particularly on knowledge exploration in Quran.

5.1 Selection of Tool(s) and Features

Several research tools such as protégé along with optional features are available for the researchers. Several plug-in and power of scripting provide flexibility and power. Still there is a large scope to use these features on

Quran datasets particularly with relation to knowledge extraction.

5.2 Extraction of knowledge from Quran

Quran presents a large collection of knowledge. Topic wise extraction of knowledge is a challenge because arrangements of verses are not according to the topic. Current studies limit itself to a particular domain or up to certain specific verses. They are limited to question answer type systems. A large scope for developing a full system on complete data set of Quran.

5.3 Retrieving Contextual meaning in Quran

Contextual search, contextual matching in Quran data set is a challenge because of the structure of knowledge present. There are methodologies to apply context. In order to understand Quran like human with its correct context, one needs to know background knowledge such as Islamic history, Hadith and basic principles of Islam. Better organization of such background knowledge can extract contextual knowledge. One other factors which can retrieval correct contextual knowledge are correct use of synonyms and autonyms in order to understand the text correctly. Here is a large scope for research as these two issues can bring contextual information retrieval from Quran data set.

6. CONCLUSION

This paper studies the important studies on semantic studies of Quran. It further, observes the key areas where researchers have concentrated, their choices of methodology and tools and their results. This paper contributes a collection of finding of Quran semantic studies at one place so that researchers hereafter can find the appropriate choices and their advantages at one place. Considering the challenges discussed in this paper researchers can find the direction for ontology designing methodology, selection of tools and evaluation technique.

REFERENCES

[1] Manal AlMaayah1, Majdi Sawalha2, M. A., "A Proposed Model for Quranic Arabic WordNet", In *LRE-REL2 Proceedings of the 2 nd Workshop on Language Resources and Evaluation for Religious Text*May 312014.pp:16-20.

[2] Ta, A., Abidin, S. Z., Abdullah, M. S., Ali, B. B. M., & Ahmad, M., "AlQuran themes classification using ontology", In *Proceedings of the 4th International Conference on Computing and Informatics, ICOCI* August 28-30,2013, pp. 383–389.

[3] Noordin, M. F *ICT and Islam* Kuala Lumpur: IIUM Press, Kuala Lumpur, Malaysia. Retrieved from http://iiumpress.iium.edu.my/bookshop/wwwgooglecom-266, 2009, pp. 173 pages

[4] Noordin, M. F., & Othman, "Ran Information Retrieval System for Quranic Texts: A Proposed System Design". In *2nd International Conference on Information & Communication Technologies IEEE,* Vol. 1, 2006, pp: 1704–1709 , doi:10.1109/ICTTA.006.1684642

[5] Noordin, M. F Application of Privacy, Security and Ethics in Islamic Concerned ICT, *14*(11), . 2013, pp:1548–1554. doi:10.5829/idosi.mejsr.2013.14.11.2035

[6] Fellbaum, Christiane, and P. V. "Challenge for a multilingual wordnet." Language resources and evaluation". *Language Resources and Evaluation*, 2012, pp:313–326.

[7] Black, W., Elkateb, S., Pease, A., Rodriguez, H., & Alkhalifa, M "Introducing the Arabic WordNet Project". In *Proceedings of the 3rd International WordNet Conference (GWC-06)* . January 2006, pp:295–299.

[8] Trad, R., Koroni, R., Mustafa, H., & Almaghrabi, A. "Evaluating Arabic WordNet Ontology by expansion of Arabic queries using various retrieval models". *Tenth International Conference on ICT and Knowledge Engineering*, 2012, pp:155–162. doi:10.1109/ICTKE.2012.6408547

[9] Alrehaili, S. M., & Atwell, E. "Computational ontologies for semantic tagging of the Quran : A survey of past approaches", In *LREC ProceedingsNinth International conference on language resource and evaluation.*.26-31 May, 2014, pp: 21-26

[10] Aliyu Rufai Yauri, Rabiah Abdul Kadir, A. A. and M. A. A. M., "Quranic Verse Extraction base on Concepts using OWL-DL Ontology", *Research Journal of Applied Sciences, Engineering and Technology,6*(23), 2013, pp: 4492–4498.

[11] Kalfoglou, Y. "Exploring Ontologies", In: Handbook of Software Engineering and Knowledge Engineering: vol. 1 2001, pp:863-887.

[12] Iqbal, R., Mustapha, A., & Yusoff, Z. M. "An experience of developing Quran ontology with contextual information support". *Multicultural Education & Technology Journal*, *7*(4), 2013, pp:333–343. doi:10.1108/METJ-03-2013-0009

[13] Aliyu Rufai Yauri, Rabiah Abdul Kadir, Azreen Azman, M. A. A. M "Ontology semantic approach to extraction of knowledge from holy quran". In *5th International Conference on Computer Science and Information Technology*. Ieee. December15, 2013, pp: 4492-4498. doi:10.1109/CSIT.2013.6588804

[14] Al-Khalifa, H. S., Al-Yahya, M., Bahanshal, A., Al-Odah, I., & Al-Helwah, N., "An approach to compare two ontological models for representing quranic words". In *Proceedings of the 12th International Conference on Information Integration and Web-based Applications & Services - iiWAS* November 2010, pp: 674-678 doi:10.1145/1967486.1967593

[15] Baqai, S., Basharat, A., Khalid, H., Hassan, A., & Zafar, S. "Leveraging Semantic Web Technologies for Standardized Knowledge Modeling and Retrieval from the Holy Qur ' an and Religious Texts", In *7th International Conference on Frontiers of Information Technology* 2009, pp: 42.

[16] Abbas, N. H. Quran *"Search for a Concept" Tool and Website*, School of the computing university of Leeds, Leeds. 2009.

[17] Saad, S., Salim, N., Zainal, H., & Muda, Z. "A process for building domain ontology : an experience in developing Solat ontology"*, International conference on electrical and Informatics*. July17-19, 2011,pp: 1-5.