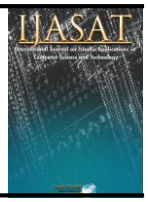
International Journal on Islamic Applications in Computer Science And Technology, Vol. 4, Issue 4, December 2016, 1-8



Developing an Ontology of Concepts in the Qur'an

Rasha Ahmed [1, a], Eric Atwell [2,b]
[1] University of Taibah, Almedina Almonoura, Saudi Arabia
[2] University of Leeds, Leeds, UK
[a] Rasha.ibrahim.cts@gmail.com, [b] e.s.atwell@leeds.ac.uk

ABSTRACT

In recent years, there is growing interest in IT for Islamic Knowledge. Researchers in religious studies have started to use ontologies to improve knowledge construction and extraction from religious texts such as the Qur'an and Hadith. An ontology can be used to describe a logical domain theory with very expressive, complex, and meaningful information. Recent research has been done in Arabic language ontology and on holy Qur'an ontology but they are still incomplete. Also there are some other issues including the process used to extract and construct an ontology that needs extra work. This paper describes our actual and ongoing work in developing an ontology. Our approach is to investigate the applicability of ontology methods of formal Knowledge Representation from Artificial Intelligence and Text Analytics research to capture and represent abstract concepts in the Qur'an. We will implement three ontology methods. The first one is to elicit or extract the abstract concepts from experts in the domain. Another approach is to semi-automatically extract concepts from text sources from the domain. The last approach is to find existing partial ontologies for the domain, and try to unify and re-use them. Finally to evaluate our general Qur'an ontology, we will investigate practical use of the ontology in a semantic search application. We have implemented the third approach, merging of existing ontologies, The experimental verification result reveals that our proposed merging methods work well, as checked by human expert. A similarity measure was applied for ontology merging, and we report a high accuracy and recall through experiments.

***Keywords*: Qur'an, Knowledge Representation, Natural Language Processing, Ontology.**

1. INTRODUCTION

Qur'an ontology was introduced in 2009 by Dukes; he represented the holy Qur'an by using a class hierarchy for the main concepts and relations between them. Al-Yahya et. al; (2010) introduced an Arabic based ontological model for representing semantic lexicons aimed to design and implement a data driven ontological model for time noun concepts in the Holy Qur'an. Othman et. al, (2013) designed a system to construct and collect Zikr and Doa al-Ma'thur in al-Qur'an and Hadith from the authentic books semi-automatically; the study claim a higher precision at the knowledge level.

2. LITERATURE REVIEW

An Ontological approach by (Harrag et. al, 2013) used data mining techniques to extract an ontology of Qur'an and prophetic narrations (Hadith) and used association rule algorithms to identify Islamic jurisprudence (Fiqh) concepts and semantic relations between them from the Sahîh Al-Bukhârî Haddith book. Another interesting research by (Khan et. al, 2013) involved semantic search of animals and birds that are mentioned in the Holy Qur'an.

Research by (Yauri et. al, 2013) to extract Qur'anic verses via ontology derived from existing ontology from the University of Leeds, and they added more concepts and more restrictions which help the proposed system to deal with more possible user queries. A recent research to answer Arabic questions about the holy Qur'an by Shmaisani (2014) addressed the semantic issues by using synonyms of keywords extracted from user query. The research built an Arabic relation extractor which uses grammatical structure to detect relations and concepts of an Arabic verse and used them to build an ontology.

We present the most related work to our contribution and review the approaches in this section. A comparative study of three ontology mapping methods (Affix, Edit distance and synonym (WordNet)), that applied to ontology, in order to compare a basic process concept. The experiment show that there is limitation in mapping method and don't present a complete mapping (Yang, et. al, 2004). A paper by (Kong, et. al, 2005) presents a study in merging heterogeneous domain ontologies for building a more complete ontology. The method was based on WordNet. The research by (Sunitha, et. al, 2013) proposes a new cluster based multiple ontology parallel merge process. The proposed method used similarity measure and parallel merge process and produces a single ontology with reduced time, which can be used by heterogeneous applications.

Different merging methodologies have been proposed. Most of them use similarity measure. Other merging systems depend on a heuristic method involving syntax and semantics; however in heuristic methods, some information may be lost during the merging process.
There are some other systems and frameworks supporting ontology merging, such as OntoMorph system (Chalupsky H., 2000), Chimaera system (McGuinness et. al, 2000), the Prompt algorithm in Protégé 2000 (Fridman, 2000) and so on.

Most recent ontology research have been using Protégée ontology tool; it has many advantages over other ontology tools; it is free, open source, it allows generation and manipulation of ontologies, and ontologies built using protégées can be accessed from Java programs through the Protégé-OWL API. Also Protégé can be used directly to unify the format of ontologies, and describe the ontologies in OWL (Lagoze et. al,2008). Previously, the work of merging; and aligning ontologies was performed manually; this task is very hard and time consuming. Today more steps in the process of merging can be done automatically or semi-automatically.

3. METHODOLOGY

Our ongoing project investigates the applicability of Ontology methods of formal Knowledge Representation from Artificial Intelligence and Text Analytics research to capture and represent abstract concepts in the Holy Qur'an. We will implement three ontology methods. One approach is to elicit the abstract concepts from experts in the domain. Another approach is to semi-automatically extract concepts from text sources from the domain. A third approach is to find existing partial ontologies for the domain, and try to unify and re-use these.

For the Qur'an, we can combine all three approaches. Qur'an scholars have written reviews of Qur'an verses which capture concepts and relationships, so we will encode these concepts in an ontology formalism. We can also apply text analytics to semi-automatically extract concepts from the Qur'an text, and/or Qur'an-related texts such as Tafsir or commentaries; we assume each word and/or each verse is a potential source of Qur'an concept(s). A third source

of Qur'an concepts, not found in many other domains, is a growing Qur'an Computing research community, including a number of researchers building partial Qur'an ontologies. We will investigate the feasibility of merging a number of existing Qur'an ontology resources with the ontology concepts captured from Qur'anic scholars' works and from text analytics of the Qur'an. The result will be a general Qur'an ontology, unifying a range of sources. For evaluation, we will investigate practical use of the ontology in a semantic search application.

3.1 Quran ontology model

Our model breaks down into the following main approaches as shown in Fig. 1:
Approach (QO1), adapt the Qur'an concepts in a book of Tafseer or commentary into a formal ontology. Approach (QO2), apply Text Analytics tools to the Qur'an to extract a formal ontology. Approach (QO3), adapt and combine two or more existing but incomplete Qur'an ontologies. Approach (QO4): merge QO1, QO2 and QO3 into a general Qur'an ontology unifying a range of sources. Approach (Q): investigate practical use of the ontology in a semantic search application. Also we can discuss it with experts in the field.

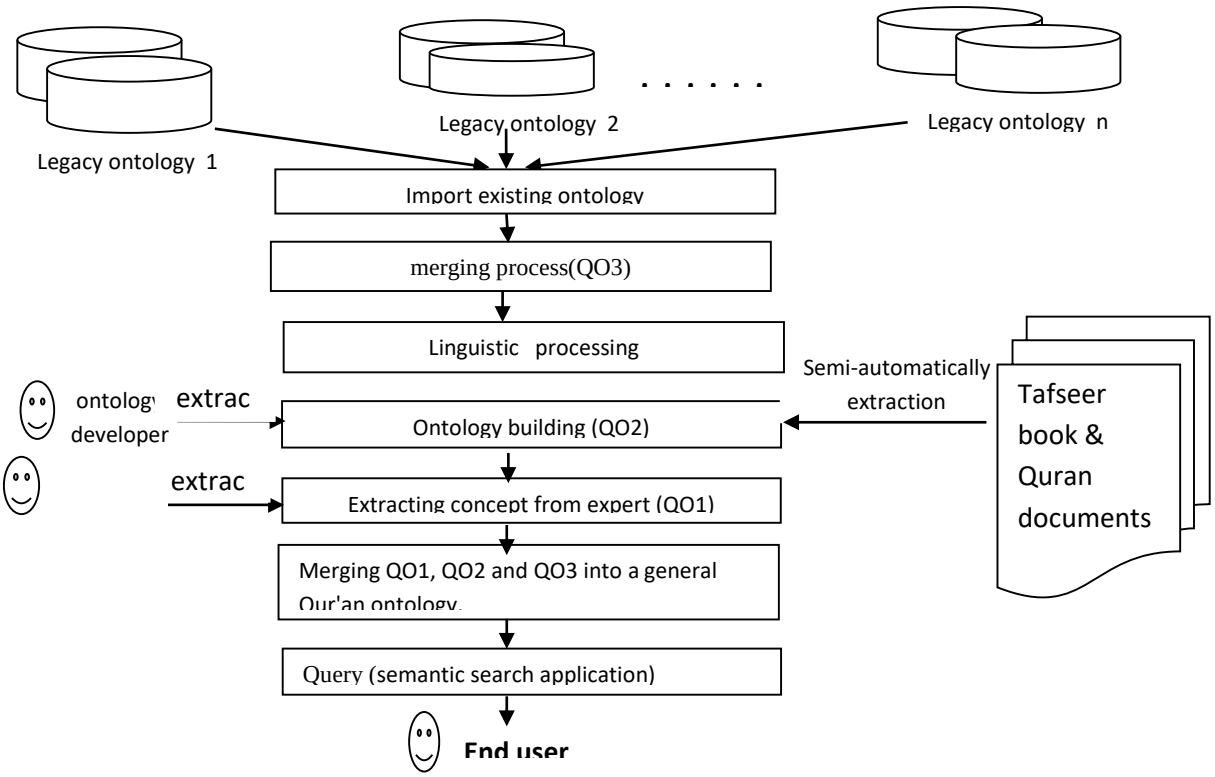


Fig. 1: Quran ontology model

3.2 Approach (QO3) combining of the legacy ontologies

As mentioned in preceding sections our project included three main approaches. The first approach we have implemented (QO3) is to collect some of the exiting Qur'an ontology (legacy ontologies), and merge them to create a single coherent ontology. Ontology merge is

a deliberate process aimed to create a new ontology from the culmination of two or more ontologies (Siham et. al, 2012).

One of our acquired ontologies describes living creatures including Animals and Birds mentioned in the Holy Qur'an (Khan et. al, 2013). The other one is an online Qur'an ontology built at University of Leeds (Dukes, 2009). Using the Java language; it includes 300 concepts and about 350 relationships linking the concepts (Yauri et. al, 2013). The other one by (Khan et.al, 2013) was built using the Protégée ontology tool, it includes 380 concepts and about 50 relationships linking the concepts.

The (Dukes, 2009) ontology includes many concepts such as Allah, Allah's-throne, event, location, holy-book, and living-creation concept, ...etc. The most important part of this ontology is living-creation classes, which is equivalent to other partial ontology (living creatures). So this is the most important part in merging process. The Fig. 2 and Fig. 3 show parts of each ontology.

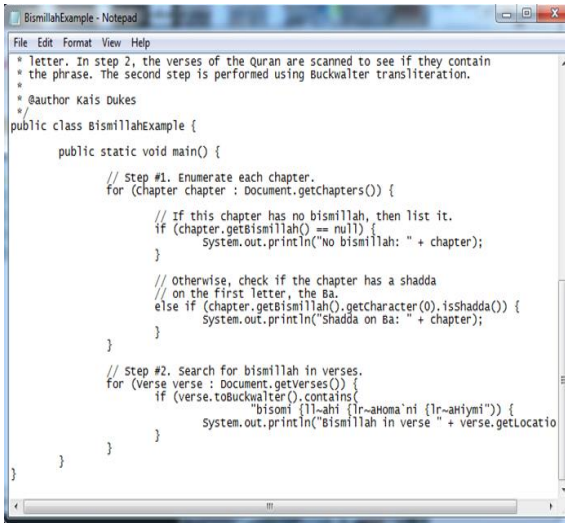


Fig.2 : (Dukes, 2009) Leeds Qur'an ontology

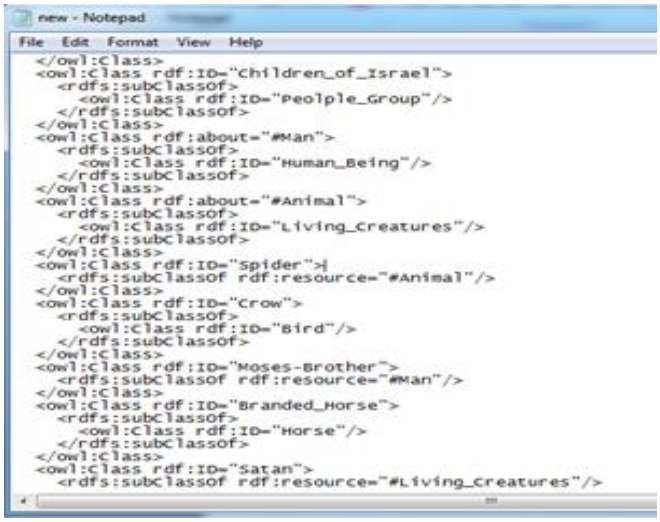


Fig. 3: (Khan et al 2013) Qur'an Ontology of Animals and Birds

3.3 Jaccard similarity of sets of words

Our approach used the Jaccard Similarity Coefficient(JSC), to measure distance among; sets of concepts of ontology-A (S1) and set of concepts of ontology-B (S2), The similarity between the two sets S1 and S2 calculates by equation1 (Eq.1).

The JSC is a statistical method used to measures similarity, dissimilarity, and distance between sets of data, and it is defined as the result of division between the number of features that are common to all divided by the number of properties (Suphakit et. al, 2013).

$$J(S1,S2) = \frac{|S1 \cap S2|}{|S1 \cup S2|} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

Where S1 and S2 are the set of concept of two ontologies.

(If S1 and S2 are both empty, define $J(S1, S2) = 1$, $0 \leq J(S1,S2) \leq 1$)

It takes the lowest value 0, when S1 and S2 are disjoint, and the highest value 1 when S1 and S2 are the same. When JSC used to merge the whole of ontology-A with whole of ontology-B the result is:

$$J_{all}(S1, S2) = \frac{|S1 \cap S2|}{|S1 \cup S2|} = 280/600=0.46$$

When JSC used to merge the living creature part of ontology-A with whole of ontology-B the result is:

$$J_{part}(S1, S2) = \frac{|S1 \cap S2|}{|S1 \cup S2|} = 280/350=0.8$$

We evaluated the similarity performance of the merging ontologies by using the precision, recall, and F-measure (Daniel, et al., 2009). It was calculated as follows.

- Precision= the number of correct resolution *100 / the number of all resolution.
- Recall= the number of correct resolution *100 / the number of expected resolution.
- F-measure= 2 • (precision • recall) / (precision + recall).

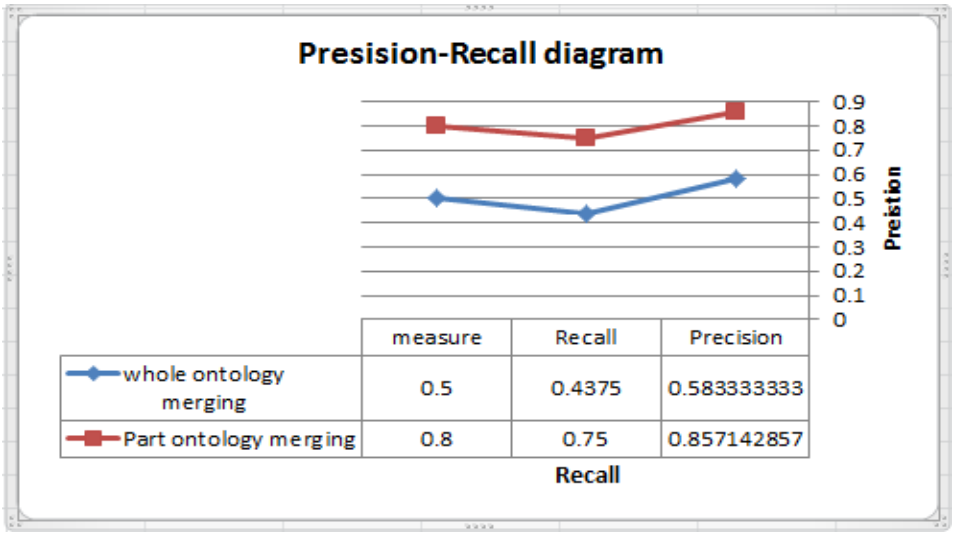


Fig. 4 similarity evaluation between two ontologies

3.4 Merging by using PROMPT algorithm

We want to take advantages of the both manually and automatically ontologies merging methods. Our ontologies have been combined semi-automatically, using the PROMPT algorithm, based on the Protégé-2000 knowledge-modeling environment; most of merging and unifying process was done automatically, but it needs some intervention from the user. The knowledge-base has been defined for ontology merging, in order to identify each step in the algorithm. We should define changes that are automatically done by PROMPT, the conflicts in the ontologies if found and propose solutions, and the suggestions that PROMPT presents to the user (Natalya, 2000). The PROMPT algorithm takes two ontologies as input and produces one merged ontology; Fig. 5 shown the algorithm.

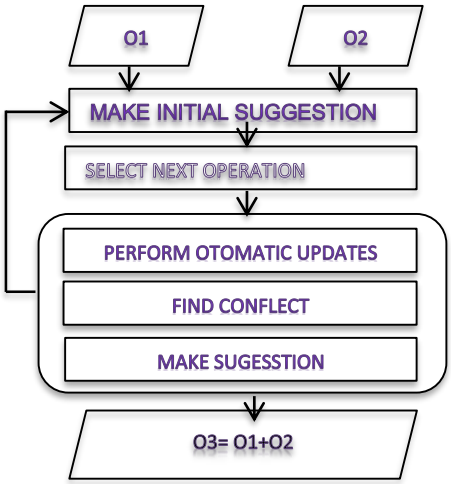


Fig. 5: PROMPT algorithm

4. DISCUSSION

Our approach used two ontology merging methods, PROMPT Tool and similarity measurement method. The experimental verification result reveals that the proposed method works well, for the two systems. The PROMPT algorithm or PROMPT tool provides good results for small and medium ontologies, however it presents less clear results for very big ontologies, and it needs more testing. Another weakness in the PROPMT algorithm is that the users may sometimes be overwhelmed by too many suggestions.

Our matching process includes only the concepts, rather than attributes. In order to match the attributes and generate table for attribute mapping, we will need some help e.g from Qur'anic WordNet. The similarity measure of concepts we are looking at here is character-level similarity (syntactic), rather than semantic "similar meaning" which requires the method to examine the words in the context. To use the semantic similarity we should use one of the two methods: Rule-based word-by-word similarity or Semantic similarity of effective words (Hamed et. al, 2015).

5. RESULTS AND EVALUATION

The two acquired ontologies were presented as input to an algorithm shown in fig.5. Ontology (A) contained about 300/350 classes/relation, in general Qur'anic topics; ontology (B) contained about 380/50 classes/relation in the topic area of living creatures. The first check was made, in order to test the consistency and to evaluate our experiment. Manual alignments have been done between the concepts and instances; we benefitted from domain experts' knowledge in this step, then we compare the merging results produced by the PROMPT system with merging results obtained from domain experts. The quality of results is evaluated by comparing the two results.

- correct resolution results obtained from human experts= 100%
- correct resolution results produced by system= 82%

The quality of PROMPT result is also measured; thus:

- The percentage of correct suggestions presented by the PROMPT tool and followed by human experts = 85%.

- The percentage of conflict resolution done by system and resolved by human experts =71%
- The percentage of total knowledge-base operation suggestions=80%

Finally we apply our similarity measure for ontology mapping and get a higher accuracy and recall through experiment

6. CONCLUSION

The interest in providing information on the web has increased and a Qur'an website augmented with an ontology has become very beneficial. Recent research has proposed several methodologies for developing and building a Qur'an ontology. Our research goal is to apply or implement the three Ontology methods mentioned earlier, and investigate the possibility of merging these methods in order to develop a single general Qur'an ontology. Finally we will evaluate the results of ontology building by using it in a semantic search application. This research will help web-based search for the Holy Qur'an and give users the ability to find their query results accurately using a general Qur'an ontology.

7. REFERENCES

AlMaayah Manal, Sawalha Majdi, Abushariah Mohammad A. M. (2014). A Proposed Model for Qur'anic Arabic WordNet, Computer Information Systems Department, King Abdullah II School of Information Technology, The University of Jordan, Amman, Jordan.

Atwell E., Brierley C., Dukes K., Sawalha M. and Sharaf A. M. (2011). An Artificial Intelligence Approach to Arabic and Islamic Content on the Internet, National Information Technology Symposium (NITS)", Riyadh, Saudi Arabia.

Al-Yahya Maha, Al-Khalifa Hend, Bahanshal Alia, Al-Odah Iman, and Al-Helwah Nawal. (2010). An Ontological Model For Representing Semantic Lexicons: An Application On Time Nouns In The Holy Qur'an. Paper Received September 30, 2010; Paper Revised December 29, 2010; Paper Accepted December 2010 The Arabian Journal for Science and Engineering, Volume 35, Number 2C.

Carl Lagoze – Cornell University, (2008), Building OWL Ontologies with Protégé (2).

Chalupsky H, (2000). OntoMorph: A translation system for symbolic knowledge. Proc. KR'OO, Breckenridge, CO, USA, 471-482.

David Robinson (2005). Muslim Societies in African History (New Approaches to African History , Chapter 1.

Daniel Jurafsky, James H. Martin (2009). Speech and Language Processing. USA: Pearson

Education International.

Fridman NOY N, MUSEN M A.(2000). PROMPT: Algorithm and tool for automated ontology merging and alignment , Proc. AAAI'OO.

Gruber, Thomas R. June.(1993). A translation approach to portable ontology specifications, Knowledge Acquisition 5 (2): 199–220.doi:10.1006/knac.1993.

Gruber Thomas R. .(1995) .Toward principles for the design of ontologies used for knowledge sharing? , International Journal of Human-Computer Studies, Volume 43, Issues 5–6, November 1995, Pages 907–928.

Hamed Hassan zadeh,Tudor Groza,Anthony Nguyen, Jane Hunter. (2015). A Supervised Approach to Quantifying Sentence Similarity: With Application to Evidence Based Medicine, journal.pone.0129392.

Harrag Fouzi, Alothaim Abdulwahab, Abanmy Abdulaziz, Alomaigan Faisal and Alsalehi Salah, (2013) . Ontology Extraction Approach for Prophetic Narration (Hadith) using Association Rules

, International Journal on Islamic Applications in Computer Science and Technology, Vol. 1, Issue 2, September 2013, 48-58.

Kais Dukes , Nizar Habash, (2011). Morphological Annotation of Quranic Arabic, School of Computing, University of Leeds, LS2 9JT, United Kingdom

Khan Hikmat Ullah, Saqlain Syed Muhammad , Shoaib Muhammad, Sher Muhammad. (2013). Ontology Based Semantic Search in Holy Qur'an, International Journal of Future Computer and Communication, Vol. 2, No. 6, December 2013.

McGuinness, D.L., Fikes, R., Rice, J. and Wilder, S. (2000). An Environment for Merging and Testing Large Ontologies, Breckenridge, Colorado.USA,483-493

Natalya Fridman Noy and Mark A. Musen,(2000), PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment

Othman Roslina, Tawil Siti Fatimah Mohd. (2013). Ontology Development for Zikr and Doa Al-Ma'thur at Wisdom Level, Taibah University International Conference on Advances in Information Technology for the Holy Qur'an and Its Sciences, December 22-25, 2013, Madinah, Saudi Arabia.

Siham Amrouch and Sihem Mostefai,(2012). Ontology Interoperability Techniques, The State of the Art, Journal of Information Organization, DLINE, Volume 2, No. 1, pp. 20-

Suphakit Niwattanakul, Jatsada Singthongchai, Ekkachai Naenudorn and Supachanun Wanapu. (2013). Using of Jaccard Coefficient for Keywords Similarity, Proceedings of the International MultiConference of Engineers and Computer Scientists 2013 Vol I, IMECS 2013, Hong Kong,

Shmaisani Hashem A. J. (2014). USING SEMANTIC APPROACHES TO ANSWER ARABIC QUESTIONS FROM THE HOLY QUR'AN, Thesis the Master's Degree in Computer Science, Philadelphia University.

Yauri Aliyu, Abdul Kadir Rabiah, Azman Azreen, Murad Masrah, (2013). Qur'anic Verse Extraction base on Concepts using OWL-DL Ontology, Research Journal of Applied Sciences, Engineering and Technology 6(23).

Sunitha Abburu, G Suresh Babu.(2013), A Cluster Based Multiple Ontology Parallel Merge Method, International Conference on Recent Trends in Information Technology (ICRTIT)

Yuan Yang', Hungzhou, Zhejiang Province, China .(2004). A comparative Study on Three Ontology Mapping ethods in the integration of Product Knowledge ,Proceedings of the 5m World Congress on Intelligent Control and Automation, China

Hyunjang Kong1, Myunggwon Hwang1, Pankoo Kim1, (2005)A New Methodology for Merging the Heterogeneous Domain Ontologies based on the WordNet, Proceedings of the International Conference on Next Generation Web Services , IEEE.