International Journal of Education and Human Developments, Vol. 5 No 2; May 2019
ISSN 2415-1270 (Online), ISSN 2415-1424 (Print)
Published by Center for Global Research Development

Multi- Relational Latent Lexicology -Semantic *"LEXICOLSEM"* Analysis Model For Extracting Qura'nic Concept

Asma Abdul Rahman
Associate Prof Dr (PhD)
Islamic Science University of Malaysia
Bandar Baru Nilai
Negeri Sembilan, Malaysia

Abstract

Al Quran is a divine text which represents the purest and most authentic form of the classical Arabic language. In order to understand the meaning of each verse, a deep knowledge of Arabic linguistic is essential. Therefore, our scholars have made their efforts by engaging themselves in the works of explaining al Quran's words, interpreting its meanings into Arabic and other languages. Currently, more people are interested in knowing the content of al-Quran, especially for non-Muslim, after 9/11 tragedy. Thus, a flexible model that can represent Qur'anic concept is required for people to understand the content of the Quran. In this research, we propose a Multi-Relational Latent Lexicology –Semantic Analysis Model (LEXICOLSEM) based on a combination of Arabic Semantic and six multiple relations between words, which are synonym, antonym, hypernym, hyponym, holonym and meronym, to precisely extract Qur'anic concept. The existing literatures focus only on very limited relationships between words which could not extract the in-depth concept of Qur'anic without considering the importance Arabic Semantic. Therefore, the objectives of this research are:(1) to analyse and categorize Quranic words according to Arabic Semantic patterns,(2) to propose a new model for extracting Quranic concept using LEXICOLSEM,(3) to investigate semantic relationships between Qura'nic words, and (4) to validate the proposed model with Arabic linguistic, and Qura'nic experts. This research will be conducted qualitatively through content analysis approach a new innovative technological technique. It is expected that the model will come out with a precise analysis for extracting Qur'anic concept. This will be very significant in enhancing the overall Quran's understanding among the society in Malaysia and Muslim's world for sustainable society.

Keywords: Multi-Relational, Latent, Lexicology-Semantic, Model Extracting, Qura'ni.

Introduction

In linguistic study, a concept is referred as a mental structure, which totally depends on relationships between linguistic representatives amongst words and phrases. However, the interchangeable relationship could happen due to the inherent conciseness of natural language. Therefore, many researchers try to solve this problem using ontology approach and Latent Semantic Analysis (LSA) (Chang & Meek, 2013; Ozcan & Aslandogan, 2004), but it needs some considerations before it can be implemented on Arabic, especially in al-Quran, which represents a special classical Arabic language.

Problem Statement

In linguistic study, a concept is referred as a mental structure, which totally depends on relationships between linguistic representatives amongst words and phrases. However, the interchangeable relationship could happen due to the inherent conciseness of natural language.

Therefore, many researchers try to solve this problem using ontology approach and Latent Semantic Analysis (LSA) (Chang & Meek, 2013; Ozcan & Aslandogan, 2004), but it needs some considerations before it can be implemented on Arabic, especially in al-Quran, which represents a special classical Arabic language.

The uniqueness of Arabic Syntax- Semantic makes the combination with LSA and ontology more significant to extract Quranic concept. It is because Lexicology -Semantic analysis is a basic step in various applications including text mining, information retrieval (IR), machine translation, automatic summarization, and Arabic learning systems (Desouki, 2011). LSA on the other hand has limitation, such as difficulty in differentiating fine-grained relations between lexical semantics, synonyms, antonyms and hypernyms. Due to that limitation, various studies have been carried out to overcome this weakness of LSA introducing Multi-Relational Latent Semantic Analysis using synonym, antonym and hypernym (Chang & Meek, 2013) but it has not yet been applied on Arabic Text such as al-Quran. Furthermore, Quranic ontology studies are lacking of accuracy of language because it depends on the translated version of al-Quran and uses aplastic noun (Azman Ta'a, Abidin, Abdullah, Ali, & Ahmad, 2013; Hikmat Ullah Khan, Muhammad Saqlain, Shoaib, & Sher, 2013; Maha Al-yahya & Hend Al-khalifa, n.d.). The existing literature shows that there are still gaps and problems in extracting Qur'anic concept as have been mentioned previously, due to inappropriate approaches that had been used ignoring the importance of Arabic Syntaxlogy and LSA in deriving the meaning of Qur'anic word. Therefore, in order to address the above mentions problem, we propose a new model in extracting Quranic concept based on Arabic Syntaxlogy and Multi-Relational Latent Semantic Analysis, using combination of six multiple relations between words, which are synonym, antonym, hypernym, hyponym, holonym and meronym.

Objectives of the Research

1. To analyze and categorize Quranic words according to Arabic Syntax- Semantic patterns.
2. To investigate semantic relationships between Quranic words.
3. To propose a new model for extracting Quranic concept using LEXICOLSEM.
3. To validate the proposed model.

Research Questions

1. How to analyze and categorize Quranic words according to Arabic Syntax-Semantic patterns?
2. What is result of investigate semantic relationships between Quranic words?
3. How to propose a new multi-relational Latent Syntax-Semantic Analysis for extracting Quranic concept?
4. What is the subject matter expert's opinion about the proposed model?

Literature Reviews

Arabic is considered one of the syntax logically complex languages. Each word can be derived from roots which have, in most cases, three letters by applying templates construct stems and then attaching them to prefixes and suffixes to obtain a very large number of different surface forms (Amr El-Desoky Mousa, Ralf Schl¨ uter, 2012). Besides that, a single root can be transformed into different word with different pattern, vocalism and pronunciation (Bassam Al-Salemi; Mohd. Juzaiddin Ab Aziz, 2011). The uniqueness of Arabic Syntaxlogy makes the combination with LSA and ontology more significant to extract Quranic concept.

Generally, LSA attempts to reveal the hidden conceptual relationships among words and phrases based on linguistic usage patterns. Usually it will be presented in taxonomic structure consisting of a hierarchy word and its relationship, such as synonym, antonym, hyperonym/hyponym, meronym/holonym (member, substance, and part), entailment, cause, attribute, and similarity. LSA has multiple uses in various fields such as an instrument of text summarization and summary evaluation (Steinberger & Ježek, 2004), an approach to Source-Code Plagiarism Detection and Investigation (Cosma, 2008), and Term Prediction instrument (Zhao & Callan, 2010). The following diagram illustrates the relationship between the component parts of the semantic model used by Martin Bryan (2003):

However, LSA has limitation such as difficulty in differentiating fine-grained relations between lexical semantics, synonyms, antonyms and hypernyms. Due to that limitation, various studies have been carried out to overcome this weakness of LSA by introducing the notion of polarity. The recent LSA research combines multiple relations between words by constructing three relationships which are known as Multi-Relational Latent Semantic Analysis using synonym, antonym and hypernym (Asma Abdul Rahman 2007-2017).

The above model encodes the raw data in a 3-way tensor to encode multiple word relations, which are synonym, antonym & Hypernym. Each slice captures a particular relation and is in the format of the document-term matrix in LSA. The tensor decomposition method was applied to generalize the representation and discover unseen relations between words.

Although this recent work tries to explore concept-based information access via ontology approach and LSA on natural language (Ozcan & Aslandogan, 2004) it has not yet been done in al-Quran, which represents the purest and most authentic form of the classical Arabic language(Asma Abdul Rahman, 2003-2019). Muslims believe that words of Holy Quran are divine and eternal. No alteration is whatsoever possible as Allah Almighty Himself has taken into His Hand.

Ontology on the other hand, is defined as a description of the concepts and relationships that can exist for an agent or a community of agents (Ozcan & Aslandogan, 2004). The use of ontology facilitates identification of concepts and their linguistic representatives, given a key concept. It is similar to a dictionary or glossary, but with greater detail and structure.

The existing researches show two types of approaches in Qur'anic language computational models based on ontological approach. First, traditional approach which is based on models of Roman language done by Hikmat Ullah Khan, Muhammad Saqlain, Shoaib, & Sher (2013) using English Translation of Holy Quran by Pickthall, and Azman Ta'a, Abidin, Abdullah, Ali, & Ahmad, (2013) using Syammil Al-Quran Miracle the Reference. These two studies lack of accuracy of language because it depends on the translated version of al-Quran.
Second, modern approach is based on an authoritative and rich source of Arabic language, i.e. the Holy Quran. This study uses lexicon ontology development based on the Unified Process for Ontology (UPON), an ontological engineering approach (Maha Al-yahya & Hend Al-khalifa, n.d.). The study focuses mainly on nouns from the "time" semantic field, which is aplastic noun - having no forms except one.

In the above model, semantic dimensions need to be determined, and then words will be organized into a hierarchical classification with general concepts at the top, and specific at the bottom. Words in the hierarchy are associated with components via ontological relations. This classification structure of the ontology implies that the deeper word moves into the hierarchy, the more arguments the componential formula will have, and therefore the meaning narrows. In contrast, words at higher levels have fewer arguments in their componential formula, and therefore the meaning broadens. This model however did not include the root of words which is the appropriate to Arabic language.

The existing literature shows that there are still gaps in extracting Qur'anic concept because the aspect of Arabic Syntaxlogy and LSA does not fully utilize. Therefore, in order to address the above mentions problem, we propose a new model in extracting Qur'anic concept based on Arabic Syntaxlogy and Multi-Relational Latent Semantic Analysis, using combination of six multiple relations between words, which are synonym, antonym, hypernym, hyponym, holonym and meronym.

Methodology

This research is focusing on designing a multi-relational LSA model in extracting Quranic concept based on combination of Arabic Syntaxlogy and six multiple relations between words, which are synonym, antonym, hypernym, hyponym, homonym and meronym. The researcher will used text analysis method as follow:

For more specifically, this approach will be conducted in the following techniques as follows:

1st Phase: Syntax logical Root Analysis

This phase will be conducted in following steps:

1. The researcher will identify digital document file of Qur'anic text. The text's authentication will be verified by expert.
2. The digital document will be analyzed using corpus analysis using Wordsmith version 2.1,
3. Each word in text will be categorized into two families of verb and noun, excluding particles.
4. The researcher will identify root of each word in these two categories, then categorize the words belong to similar root in another sub-family

5. To enhance the confidence of investigation, the result then will be triangulated with three Arabic referred dictionaries:

1. Lisan al-Arab by Ibn Manzuur,
2. Mu'jaam al Waseet by Majma' Luhghah al-Arabiyyah in Cairo, and
3. al-Mu'jam al-Asasi by al-Munazzamah al-Arabiyyah lil Tarbiyyah wa Thaqafah wal 'Ulum

2nd Phase: To Propose A New Model For Extracting Quranic Concept

Based on previous 1st phase result, the researcher will follow the next steps:

1.Create and propose a model using six multiple relations between words in Latent Semantic Analysis as follow:

i.Synonym : a word or phrase that means exactly or nearly the same as another word or phrase in the same language
ii.Antonym: a word opposite in meaning to another
iii.Hypernym : a word with a broad meaning constituting a category into which words with more specific meanings fall
iv.Hyponym : a word of more specific meaning than a general or superordinate term applicable to it
v.Holonym: A concept of which this concept forms a part
vi.Meronym: A term that denotes part of something

2.Related connections of word to conceptual meaning then will be identified
3.Linking the related connections to key concept in tree-form drawing.
4. Describing the design of extracting Quranic concept based on ontological approach in Latent Semantic Analysis.

3rd Phase: Investigating Semantic Relationships between Quranic Words
The proposed model will be validated by testing a concept of "Sight" in Quran as a sample. (an expected analysis is shown as below)

4th Phase: Validating the proposed model
The result then will be evaluated by two groups of experts:
1. An Arabic linguist who will examine Syntax logical Root Analysis and the merging process with multi-relational latent semantic analysis, and
2. Quranic experts who examine holistically the concept of Sight in al-Quran, according to Islamic perspective
the evaluation will be conducted in focus group interview. Any suggestion or feedbacks will be taken into consideration to improve the model.

Finding and New Result Novelty

1. Policies for government agencies to help them curb and solve the educational system in teaching and learning Arabic linguistics training and practices in tricking consumers into thinking that particular product by apply a new innovative method for whole level of students and society. This is a new invention thus; no comparable product existed in the market. Impact on human being and socio economic. The "(LEXICOLSEM)" have improved knowledge and linguistic skills in the production of higher quality of human capital. Furthermore, the researcher was sales from text and reference, more formula books, CDs and Courses, training, workshop generate additional income for.
2. Harmonized society and community where Muslim and non-Muslim can seat together enjoying their communication with multiple lingual; that Muslim be it Malaysian or international could be unity of the nation local or foreign without any doubt or miscommunication.
3. Academically closing the gap on communication between society and community related research, providing the aspect of language, culture and society perspective.
4. As academics, we aim to contribute in closing the gap on miscommunication related research by looking from the aspect of language, culture and society perspective. The research will result in publications and we target to bring about the knowledge into seminars nationally and internationally and symposium to share with universities and industry players and those who are interested.

Acknowledgment

This research is in line with government's effort to empower the integration of Naqli and Aqli agenda as National Education Policy that had been announced by Minister of Education, YAB Tan Sri Dato' Hj Muhyiddin Bin Hj Mohd Yassin, at "Majlis Perutusan Tahunan 2014 Kementerian Pendidikan Malaysia"

Conclusion

We have to find that the research output will useful to attract foreign direct investors, mainly Asian countries, and other potential interested investor from Middle East, and Europe to understand the opportunity in linguistics education, to join in the development of Malaysia and other countries of Renewal Energy (SCORE) specifically in two identified clusters Linguistics Education with a new innovative method, cultures -Hub and Tourism.

References

Amr El-Desoky Mousa, Ralf Schl¨ uter, H. N. (2012). INVESTIGATIONS ON THE USE OF MORPHEME LEVEL FEATURES IN LANGUAGE MODELS FOR ARABIC LVCSR, 5021–5024.

Asma Abdul Rahman, (2003-2007), LINGUISTICS STUDIES, Publishing Section USIM, Bandar Baru Nilai, Negeri Sembilan.

Asma Abdul Rahman, (2007-2017), A'LIYAT TAHLIL AL-NASS AL-QURANI AL-LUGHAWWY: DIRASAH LUGHAWIYYAH DALALIYYAH (Formula Book), Publishing Section USIM, USIM, Bandar Baru Nilai, Negeri Sembilan.

Asma Abdul Rahman, (2019), MODERN LINGUISTICS STUDIES, Publishing Section USIM, USIM, Bandar Baru Nilai, Negeri Sembilan.

Azman Ta'a, Abidin, S. Z., Abdullah, M. S., Ali, B. B. M., & Ahmad, M. (2013). AL-QURAN THEMES CLASSIFICATION USING ONTOLOGY, (074), 383–389.

Chang, K., & Meek, W. Y. C. (2013). Multi-Relational Latent Semantic Analysis. Cosma, G. (2008). An approach to source-code plagiarism detection investigation using latent semantic analysis. IEEE Transactions on Computers. Retrieved from http://eprints.dcs.warwick.ac.uk/401/1/cs-rr-440.pdf

Desouki, S.-G.-. (2011). An Application Oriented Arabic Syntaxlogical Analyzer, (27), 7–19. Hikmat Ullah Khan, Muhammad Saqlain, S., Shoaib, M., & Sher, M. (2013). Ontology Based Semantic Search in Holy Quran. International Journal of Future Computer and Communication, 2(6), 570–575. doi:10.7763/IJFCC.2013.V2.229.

Maha Al-yahya, & Hend Al-khalifa. (n.d.). AN ONTOLOGICAL MODEL FOR REPRESENTING SEMANTIC LEXICONS : AN APPLICATION ON TIME NOUNS IN THE HOLY QURAN, 35(2), 21–35.

Ozcan, R., & Aslandogan, Y. A. (2004). Concept Based Information Access Using Ontologies and Latent Semantic Analysis, 1–16.

Steinberger, J., & Ježek, K. (2004). Using Latent Semantic Analysis in Text Summarization. In Proceedings of ISIM 2004, 93–100.

Zhao, L., & Callan, J. (2010). Term necessity prediction. Proceedings of the 19th ACM International Conference on Information and Knowledge Management - CIKM '10, 259. doi:10.1145/1871437.1871474.

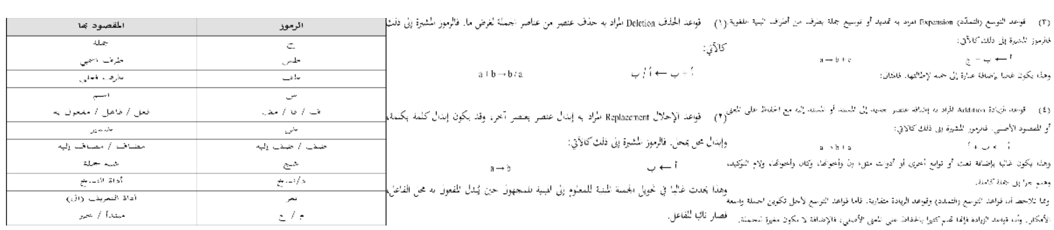

Figure 1: Multi-Relational Latent LEXICOLSEM formula Analysis by Asma Abdul Rahman

* A sample of proposed model in using LEXICOLSEM to extract Qura'nic concept

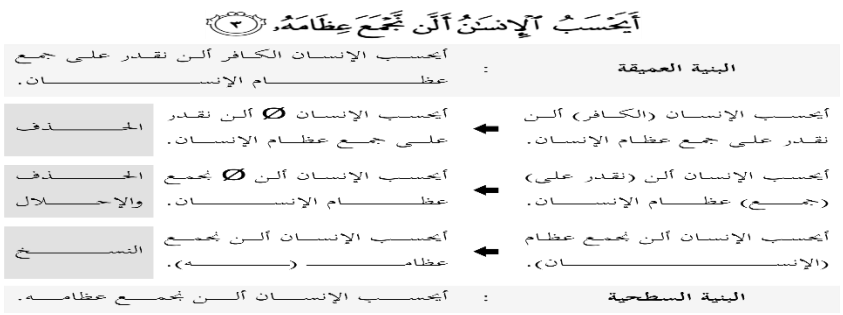

Figure 2: Extracting Quranic LEXICOLSEM concept from various multiple relationships between words by Asma Abdul Rahman

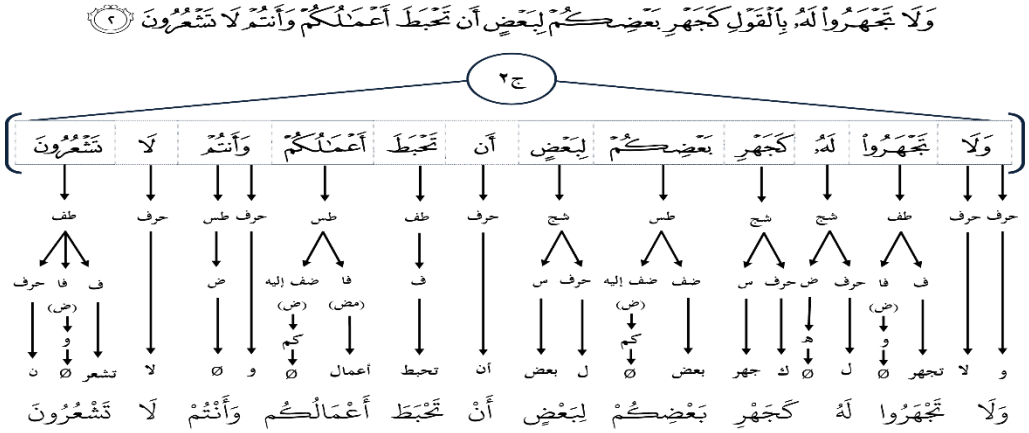

Figure 3: Model of LEXICOLSEM in Quranic Ontology by Asma Abdul Rahman

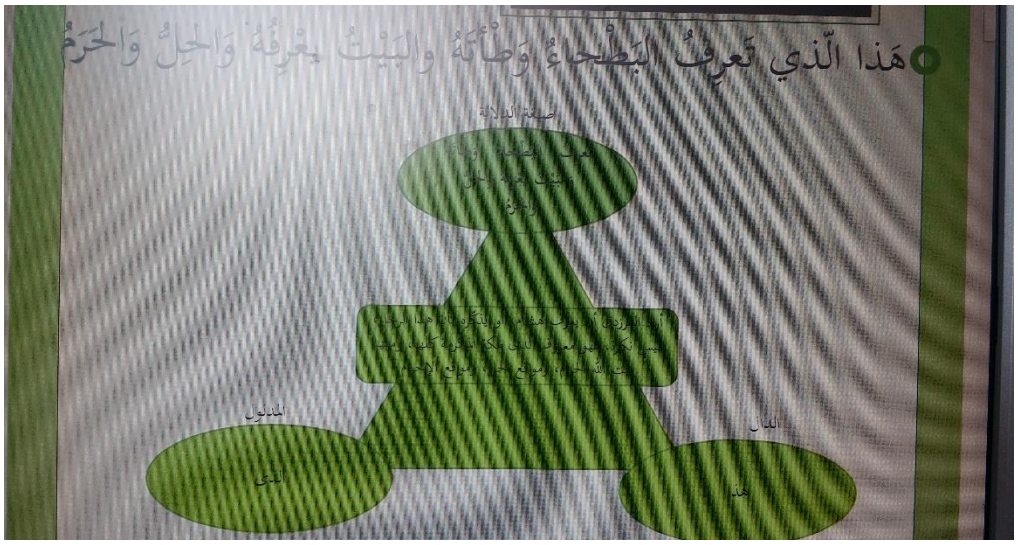

Figure 4: How to analyze Quranic verses based on LEXICOLSEM by Asma Abdul Rahman