

<http://www.ftsm.ukm.my/apjitm>
Asia-Pacific Journal of Information Technology and Multimedia
Jurnal Teknologi Maklumat dan Multimedia Asia-Pasifik
Vol. 7 No. 2, December 2018: 71 - 84
e-ISSN: 2289-2192

EXTRACTION OF CONCEPT AND CONCEPT RELATION FOR ISLAMIC TERM USING SYNTACTIC PATTERN APPROACH

SAIDAH SAAD
UMMU KALSOM LATIFF

ABSTRACT

Ontology Learning is a semi automation step to learn ontology from text. The identification of a term become a prerequisite for all aspects of Ontology Learning. The Ontological Learning Layer is started by identifying terms, synonyms, concepts, hierarchical concepts, relationships and rules for various domain text including Islamic-based text or Islamic Glossary. The glossary of Islamic terms translated into English has been in abundance and requires extraction of important information for a clear understanding of an Islamic term. The existence of a list of Islamic terms is to minimize the spelling diversity, to seek the concept of the term and provide guidance for a unique Islamic concept. However, this electronic form provides a serious problem to achieve machine operation. The study aims to identify and extract concepts, taxonomies, relationships and rules that can be built on the domain of the terms in Islamic glossary specific to the field of Islamic Pillars. This extraction involves the use of Hearst pattern approach. The data set used is from the Dictionary or Islamic Glossary of the International Islamic University of Malaysia (DEED 2015). The dictionary consists of Islamic terms, which are the concepts and intentions of each concepts in alphabetical order. The study used six phases involving the preparation, processing and testing phase that are merged with the methodology design of the study. A total of 41 concepts were successfully extracted based on 6 Hearst pattern, 31 manually generated rules from 19 sentences and 9 non-taxonomic relationships. The result of the study concluded that the objective of this study was achieved in the scope determined when the results of the study and testing conducted by the assessors in the domain showed positive results. Research constraints are presented to enable researchers to improve the research from time to time. Research proposals for future research have been described so that this study will be more useful and expanded to the next comprehensive guide of Muslims.

Keywords: DEED, Description Logic, NLP, POS, NP.

INTRODUCTION

The popular ontology definition used was developed by Thomas Gruber, 1993, "A Translation Approach to Portable Ontology Specification, Knowledge Acquisition". Ontology is a formal and explicit understanding of shared concepts. The ontological specification is also formally described; it contains a collection of terms and their machine-understandable connections. 'Explicit' means clearly explaining the type of concept used and the constraints of its use. 'Formal' refers to documents that machine can understand. 'Shared together' means that knowledge within the ontology needs to be agreed upon and accepted by a group or community. 'Concept' refers to an abstract model that contains related concepts and relationships that exist in some situations. Ontology facilitates the identification of concepts to particular classes and subclasses or categories of objects within a domain.

A knowledge based that classifies concepts and relationships is called a hierarchical concept (Sanderson and Croft, 1999). This knowledge based is core to any ontology where Ontology Learning requires a taxonomic hierarchy as stated by Cimiano et al (2009) and Ali &

Saad (2016). The hierarchical concept consists of conceptual classes categorized into super class and sub-class hierarchy. Definition of concepts and relationships between concepts is very important in the development of ontology as well as the need to know the symbols referring to the concepts and relationships. It also contains the taxonomy of *is-a* relationships or non-hierarchical relationships. There are two ways to create ontology, first Ontology Editor and Ontology Learning. Ontology Editor is an application software used to generate ontology manually. While Learning Ontology is a partial step of automation to learn ontology from texts.

Figure 1 shows the layers of ontology learning with examples. Terms are candidates for concepts and related relationships and they contain compound words or words. The identification of a term is a prerequisite for all aspects of Ontology Learning from the text. Synonyms search for words that have the same concept or it is the same word semantically. In the development of Ontology Learning, besides the need to acquire language knowledge of the terms used to refer to the specific concepts in the text, it also identifies the relevant synonyms. This synonym layer is to detect related terminology, as well as synonyms for terms. The terms and synonyms collected basically will form the concept. The concept of finding a conceptual definition of the terms and synonyms, the lexical signs used refer to them. Concept Hierarchy is the backbone of ontology development. It organizes and arranges the concepts identified to the hierarchy or taxonomy structure. Each concept relates to the other concepts above or below. Attributes and relationships are used to characterize the concepts in the hierarchy. Relations examine existing relationships and identify the domain and range involved. The rules result in more complex relationships and relationships between concepts and relationships.

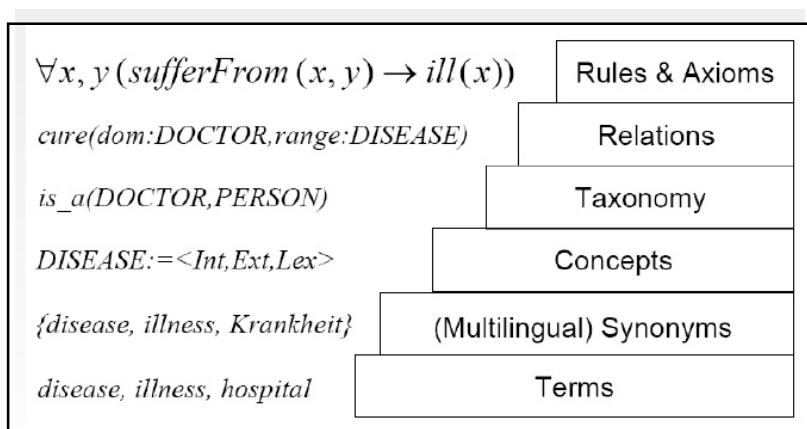


FIGURE 1. Ontology Learning Layer Cake (P. Buitelaar et al, 2005)

According to Dalloul (2013), Saad & Salim (2008) and Saad et al (2009), the Islamic domain is one of the specific domains that has been studied the concepts and hierarchies. The glossary of Islamic terms translated into English has been in abundance as in al-islam.org, islamicity.com and clarionproject.org/ glossary_islamic_terms, which it requires to access and extract essential information for a true understanding of an Islamic term. The existence of a list of Islamic terms is to minimize the diversity of spelling, to seek the concept of the term and provide guidance for a unique Islamic concept. However, this electronic form provides a serious problem to achieve machine operational. In the previous study, researchers tend to use the traditional approach of lexico-syntactic patterns proposed by Weaam & Saad (2016), Saad et al

(2009) and Hearst (1992). The Islamic terms quoted from the Qur'an have different language styles, rich languages, distinctive layers of layers and complex morphology. This condition makes the process of extraction more difficult. Ontology also needs to have extensive coverage in the domain to get the perfect model by defining meaningful and consistent generalization. Challenges in involving ontological development are necessary to maintain a balance between modelling large knowledge but to ensure that the model is simple and compact. To ensure the ontology generated should be accepted and agreed by a group of communities, involving the complex development process, as different parties will agree and different design options. The ideal solution to this problem is to use the approach automatically. This approach dramatically reduces the cost of building ontologies (Cimiano, 2006). The aim of this study is to design a conceptual relationship for the Islamic domain. In order to achieve these key objectives, several objectives have been developed as follows: Extract concepts and concept relationships using the Hearst pattern approach for the term in Islamic glossary; and generate rules manually from glossary of Islam.

METHOD

The methodology phases involved document analysis, pre-processing of documents, natural language processing, Hearst taxonomic extraction, generated rules manually, extraction of non-taxonomic relationship concepts and testing. The description of each phase is as follows:

PHASE ONE, DOCUMENT ANALYSIS

The document analysis phase involves the identification of the datasets and the selection of the domain scope of the dataset. The selected dataset is from the Dictionary of Islam that has been compiled by International Islamic University, Malaysia. This glossary contain a large number of concepts or Islamic terms and their description. These glossaries can be found at <http://www.iium.edu.my/deed/glossary/index2.html>. The choice of the domain scope of the Islamic glossary is an effort towards focusing on one section or field by creating a smaller sub-domain. This study focuses on the fundamental areas in Islam that is the Pillars of Islam. The five pillars of Islam are utter the shahadatain, five prayers, fasting in Ramadan, issuing zakat and performing Hajj in Makkah for those who are capable.

PHASE TWO, PRE-PROCESSING OF DOCUMENTS.

Referring to Saad et al (2009), to produce Islamic ontologies, some things need to be understood and Islamic documents need to go through several pre-processing approaches to provide datasets before the extraction process takes place. Things to consider for this study are;

1. Description of the meaning of the concept in the glossary will only use explicit meaning without seeing hidden meaning.
2. The phrase that needs to go through the parsing process will go through the temporary replacement process of the meaning for the preparation of the parsing process as well as "His Greatness" to be replaced by the "Greatness of God".

The approach used in providing datasets is as follows:

Capitalization. In the glossary of Islam, capital letters indicate that it is a concept, unless the word is at the beginning of the verse.

THIRD PHASE, NATURAL LANGUAGE PROCESSING

The natural language processing phase involves several methods of part of speech tagging, parsing and concept extraction. This method is a syntactic analysis used to identify the grammatical tags for each word (POS tagging) and the parsing function for extracting the Noun Phrase (NP).

PHASE FOUR, HEARST'S LEXICO-SYNTACTIC TAXONOMY EXTRACTION

Hearst (1992) has produced several syntactic lexico pattern to extract the phrase from a certain corpus. This finding involves the search of specific terms for a particular corpus, which is linked through the semantic relations and lexico-syntactic patterns most likely to arise from the method. Lexico-syntactic patterns have frequency advantages for different types of texts, and also have the most optimum overall accuracy even on pre-encoded knowledge (Hearst, 1998). Unlike study by Saad et al (2013), this study will divide and use the pattern into 6 pattern of lexico-syntactic by Hearst (1992).

FIFTH PHASE, MANUALLY RULES GENERATED

Rules are based on a subset of First Order Logic (FOL) and extensional possibilities. FOL is the reason for a statement that is solved to the subject and the predicate. Extension also involves an instance or individual of a subject. The resulting rules are manually based on the natural language in the Islamic document used.

SIXTH PHASE, EXTRACTION OF NON-TAXONOMIC RELATIONSHIPS

The discovery of non-taxonomic relationships is an important point in ontology learning. However, it is one of the under-developed areas of knowledge. This study uses a semi-automatic extraction process to obtain non-taxonomic relationships from the dataset.

SEVENTH PHASE, TESTING

Some testing stages are conducted to ensure the output of each methodology phase is appropriate. This is done by manually reviewing each methodological phase. Subsequent testing involves evaluating by assessors in the relevant field (domain expert). The assessors look at the logic and accuracy of the domain content. Figure 2 shows the methodological design for this study.

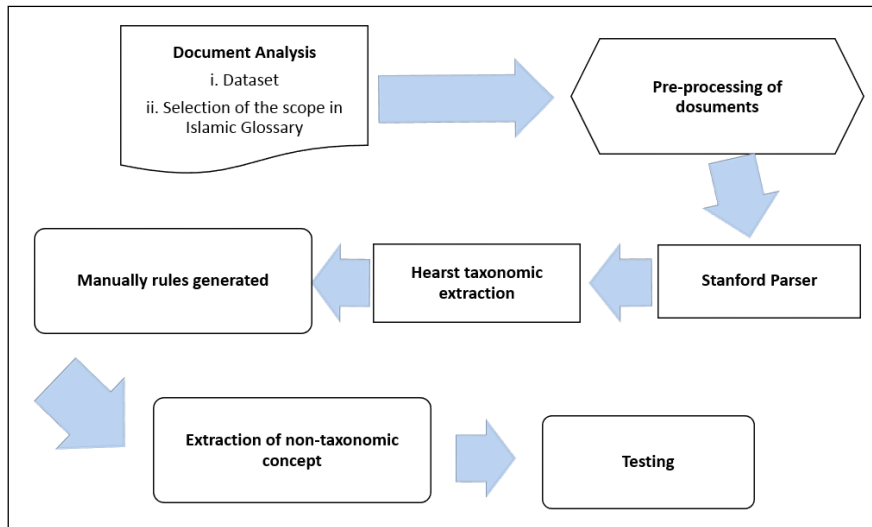


FIGURE 2. Methodology Design

RESULTS AND DISCUSSION

The results of the syntactic analysis involve the results of the POS tagging and the extraction of the NP as a preliminary document processing. The results of the Hearst pattern extraction and concepts involving regular expression match are shown. The concept extraction testing involves the assessors from the relevant field. The dataset taken from the Islamic Dictionary of Glossary which focuses on the Pillars of Islam contains a total of 232 sentences.

RESULTS OF THE SYNTHESIS ANALYSIS

This section shows the results of syntactic analysis for POS tagging, parsing, Hearst pattern extraction and noun phrase listing. The POS tagging function is to identify the syntactic class for each word in a sentence. The parser depends on the tagging of words to form the noun phrase and is represented in the tree diagram. The use of regular expression is to extract relationships that meet each Hearst Pattern.

POS TAGGING

Table 1 shows a sample of results for POS tagging for an original text sample that tells basic of the Pillars of Islam. The table shows every word marked with a certain group of tag such as NN, noun, VB, verb, JJ, adjective and so on.

TABLE 1. Sample Of Part Of Speech Tagging Results From Text

Text	Part of Speech Tagging
There are five Arkan of Islam. Shahadatain is a bearing witness. Salat is a prayers. Seeaam is a fasting. Zakat is a wealth dues. Hajj means pilgrimage.	There EX are VBP five CD Arkan NNP of IN Islam NNP . . Shahadatain NNP is VBZ a DT bearing NN witness NN . . Salat NNP is VBZ a DT prayers NNS . . Seeaam NNP is VBZ a DT fasting NN . . Zakat NNP is VBZ a DT wealth NN dues NNS . . Hajj NNP means VBZ pilgrimage NN . .

Figure 3 shows the result of the parsing in tree diagram using the depicted parser for text "The One to Whom all hearts submit in love, fear, reverence, desire, trust and sincerity". There are 3 noun phrases extracted, first: [The, One, to, Whom, all, Heart] which is the combination among DT, NN, VP, TO, VB and NNS, secondly: [all, heart], which is the combination between DT and NNS, third: [love, fear, reverence, desire, trust, and, sincerety] which is the combination between NN dan CC.

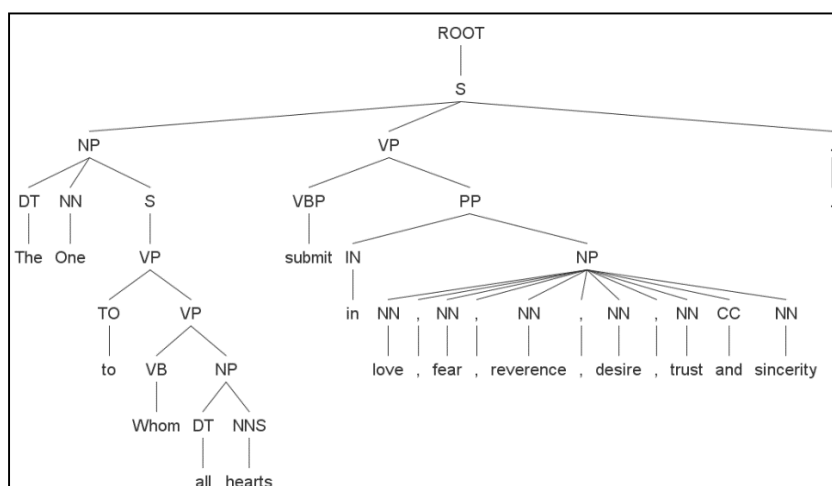


FIGURE 3. Parsing of text "The One to Whom all hearts submit in love, fear, reverence, desire, trust and sincerity."

Table 2 shows the sample extraction of noun phrase from the original text. This extracted phrase is referring to the tree diagram that has been parsing by the parser. There are some phrase names are single phrases such as Salat, worship, Allah and Saum. There are also a noun phrase formed as a result of a combination of words from the DT class, CC, JJ and so on.

TABLE 2. Sample Of Noun Phrase Extracted From Text

Text	Noun Phrase
There are five daily obligatory prayers in Islam, consisting of fixed sets of standings, bowings, prostrations and sittings in worship to Allah.	[There] [five, daily, obligatory, prayers, in, Islam, consisting, of, fixed, sets, of, standings, bowings, prostrations, and, sittings, in, worship to, Allah] [five, daily, obligatory, prayers] [Islam] [fixed, sets, of, standings, bowings, prostrations, and, sittings] [fixed, sets] [standings, bowings, prostrations, and, sittings] [worship to, Allah] [worship] [Allah]
Seeaam is a Fasting from food and drink and from sexual intercourse if you are married during daylight, from the first light of dawn until sunset.	[Seeaam] [a, Fasting, from, food, and, drink, and, from, sexual, intercourse] [a, Fasting] [food, and, drink] [sexual, intercourse] [you] [daylight] [the, first, light, of, dawn]

[the, first, light]
[dawn]
[sunset]

HEARST PATTERN AND CONCEPT EXTRACTION

The Hearst Pattern extraction process is based on 6 patterns. Regular expressions used are fixed as shown in Table 3. A regular expression is a method used for pattern matching. It is a flexible and concise method of matching the text.

TABLE 3. Regular Expressions Based On Hearst Pattern

Hearst Pattern	Regular Expressions
<i>NP such as</i> {NP,*{(and-or)} NP	(?:[\\(*\\s*,*\\-*\\w*\\.*\\'*,*\\s*\\.*)*])? [\\(*\\w*\\s*\\)*such\\)* [\\(*\\w*\\s*\\)*as\\)* (?:[\\(*\\s*,*\\-*\\w*\\.*\\'*,*\\s*\\.*)*])?
<i>Such NP as</i> {NP,*{(and-or)} NP	[\\(*\\w*\\s*\\)*Such\\)* (?:[\\(*\\s*,*\\-*\\w*\\.*\\'*,*\\s*\\.*)*])? [\\(*\\w*\\s*\\)*as\\)* (?:[\\(*\\s*,*\\-*\\w*\\.*\\'*,*\\s*\\.*)*])?
<i>NP {,NP}*{,} or other NP</i>	(?:[\\(*\\s*,*\\-*\\w*\\.*\\'*,*\\s*\\.*)*])? [\\(*\\w*\\s*\\)*or\\)* [\\(*\\w*\\s*\\)*other\\)* (?:[\\(*\\s*,*\\-*\\w*\\.*\\'*,*\\s*\\.*)*])?
<i>NP {,NP}*{,} and other NP</i>	(?:[\\(*\\s*,*\\-*\\w*\\.*\\'*,*\\s*\\.*)*])? [\\(*\\w*\\s*\\)*and\\)* [\\(*\\w*\\s*\\)*other\\)* (?:[\\(*\\s*,*\\-*\\w*\\.*\\'*,*\\s*\\.*)*])?
<i>NP {,} including</i> {NP,*{(and-or)} NP	(?:[\\(*\\s*,*\\-*\\w*\\.*\\'*,*\\s*\\.*)*])? [\\(*\\w*\\s*\\)*including\\)* (?:[\\(*\\s*,*\\-*\\w*\\.*\\'*,*\\s*\\.*)*])?
<i>NP {,} especially</i> {NP,*{(and-or)} NP	(?:[\\(*\\s*,*\\-*\\w*\\.*\\'*,*\\s*\\.*)*])? [\\(*\\w*\\s*\\)*especially\\)* (?:[\\(*\\s*,*\\-*\\w*\\.*\\'*,*\\s*\\.*)*])?

Table 4 shows the sample results of the Hearst Pattern extraction for the Islamic Dictionary glossary text sample. As for the text, "Zakat is also due on other things such as silver, animals, crops, etc.", the discovery of the noun phrase [other things] and [silver, animals, crops, etc] associated with the phrase 'such as'. This taxonomic relationship pattern corresponds to the first Hearst pattern. While for the text "Emission of impurities from the private parts: urine, faeces, wind, prostatic fluid, or other discharge.", the discovery of the phrase urine, faeces, wind, prostatic fluid and discharge is associated with the phrase 'or other'. This taxonomic relationship pattern corresponds to the third Hearst pattern.

TABLE 4. Sample Of Hearst Pattern Extraction

Hearst Pattern	Text	Extraction
<i>NP such as</i> {NP,*{(and-or)} NP	Zakat is also due on other things such as silver, animals, crops, etc.	NP(NP(JJ(other) NNS(things)) PP(JJ(such) IN(as) NP(silver, animals, crops, etc.)))
<i>Such NP as</i> {NP,*{(and-or)} NP	Such sins as Shirk, Qatl (murder), Zinah (fornication and adultery), the taking of Riba (usury), Sirq (theft), etc.	NP(JJ(Such) NNS(sins) PP(IN(as) NP(Shirk,)) Qatl (murder), Zinah (fornication and adultery), the taking of Riba (usury), Sirq (theft), etc.)
<i>NP {,NP}*{,} or other NP</i>	Emission of impurities from the private parts: urine, faeces, wind, prostatic fluid, or other discharge.	NP(NP(urine,) NP(faeces,) NP(wind,) NP(prostatic fluid,) CC(or) NP(JJ(other) discharge.))

<i>NP { ,NP}*{ , } and other NP</i>	Fasting the month of Ramadhan, celebrating the two major feasts ('Eid Al-Fitr and 'Eid Al-Adhha), performing the pilgrimage to Makkah, and other religious activities depend upon the lunar months.	NP(Makkah,) CC(and) NP(JJ(other) religious activities)
<i>NP { , } including {NP,}*{(and-or)} NP</i>	In other words, a term that indicates all that pleases Allah, including sayings and actions of the heart or limbs.	NP(NP(Allah.) PP(VBG(including) NP(NP(sayings and actions) PP(of the heart or limbs.))))
<i>NP { , } especially {NP,}*{(and-or)} NP</i>	Mathani is the often repeated Ayat of the Holy Qur'an, especially the Surat al-Fatiha, for it is always recited during Salat, in every Rak'a.	NP(NP(the Holy Qur' an,) RB(especially) NP(the Surat al-Fatiha,))

Table 5 shows the number of Hearst Pattern extraction and the number of extracted concepts. The first Hearst Pattern corresponds to 14 concepts extraction, second pattern, 9 concepts extraction, third and fourth pattern with 5 and 2 concepts extraction respectively, fifth pattern, 9 concepts extraction and sixth pattern with 2 concepts extraction. The amount of concepts extraction that corresponds to the Hearst pattern is as much as 41 concepts extraction.

TABLE 5. Number Of Hearst Pattern And Concept Extraction

Hearst Pattern	Total of Extraction	Total of Concept	Concepts
<i>NP such as {NP,}*{(and-or)} NP</i>	3	14	Zakat, Silver, Animals, Crops, Worship, Prayers, Supplications, Sacrifices, Invocations, Worshipped Worshipped Object, Fire, Idols, Fire, Animals
<i>Such NP as {NP,}*{(and-or)} NP</i>	2	9	Sin, Shirk, Qatl, Zinah, Riba, Sirq, Books, Sahih Bukhari, Sahih Muslim
<i>NP { ,NP}*{ , } or other NP</i>	1	5	Emission of impurities, Urine, Faeces, Wind, Prostatic Fluid
<i>NP { ,NP}*{ , } and other NP</i>	1	2	Religious activities, Pilgrimage to Makkah
<i>NP { , } including {NP,}*{(and-or)} NP</i>	3	9	Living thing, Insects, Plant, Tree, Disbelief in Allah's command, Refusal to accept Prophet Muhammad taught, Pleases Allah, Saying of the heart, Action of the limbs
<i>NP { , } especially {NP,}*{(and-or)} NP</i>	1	2	Al-Quran, Al-fatiha

RULES

Table 6 shows sample of rules generated manually from domain Islamic-Glossary. The rules generated are based on a subset of First Order Logic (FOL) and extensional possibilities. It is manually generated based on the natural language in the Islamic document used. Rules are formed from previously extracted concepts and relationships. The rules are written in the form of FOL, i.e Px or P (x), where P is the predicate and x is the variable that represents the subject. Symbols \equiv and \rightarrow mean 'equals' or 'mean'. The symbol \vee means 'or' while the symbol \wedge stands for 'and'. A complete sentence is merged and manipulated logically according to the same rules used in Boolean algebra.

TABLE 6. Sample Of Rules Generated

Text	Rules
Arkan is a pillars.	$Arkan(?a) \equiv Pillars(?a)$
There are five Arkan of Islam	$ArkanOfIslam(?a) \equiv Shahada(?a1) \vee Solat(?a2) \vee Zakat(?a3) \vee Fast(?a4) \vee Hajj(?a5)$
Shahadatain is a bearing witness.	$Shahadatain(?s) \equiv BearingWitness(?s)$
All Muslims must believe in and utter the Shahadatain.	$Muslim(?m) \wedge Shahadatain(?s) \rightarrow believe(?m,?s) \wedge utter(?m,?s)$
The First Shahada is Ashhadu an la illaha illal'lah. (I bear witness that there is no deity worthy of worship except Allah.) The Second Shahada is Ashhadu anna Muhammadar Rasoolullah. (I bear witness that Muhammad is the Messenger of Allah.)	$FirstShahada(?x) \equiv FirstShahada(Ashhadu\ an\ la\ illaha\ illal'lah)$ $SecondShahadah(?y) \equiv secondShahadah(Ashhadu\ anna\ Muhammadar\ Rasoolullah)$ $FirstShahadah(?x) \wedge hasMeaning(?x, "I\ bear\ witness\ that\ there\ is\ no\ deity\ worthy\ of\ worship\ except\ Allah")$ $secondShahadah(?x) \wedge hasMeaning(?x, "I\ bear\ witness\ that\ Muhammad\ is\ the\ Messenger\ of\ Allah")$ $Shahadatain = FirstShahada(?x) \wedge secondShahadah(?y)$
Salat is a Prayers. To do all the five compulsory daily prayers regularly in the exact manner as was practised by the Holy Prophet Muhammad, may Allah bless him and grant him peace.	$Salat(?st) \equiv Prayers(?st)$ $Prayers(?st) \equiv Subh(?st1) \vee Dhur(?st2) \vee Asr(?st3) \vee Maghreb(?st4) \vee Isha(?st5)$
Zakat is a wealth dues. To pay 2.5% of one's yearly savings to the poor and needy Muslims.	$Zakat(?z) \equiv Wealthdues(?z)$ $Zakat(?z) \equiv YearlySavings(?z, 2.5\%) \wedge [Recipient(?r) \rightarrow Muslim(?r) \wedge Poor(?r) \wedge Needy(?r)] \rightarrow receive(?r, ?z)$ $ZakatPendapatan(?x) \equiv Muslim(?y) \wedge hasSaving(?x,?y) \wedge YearlySavings(?y, 2.5\%)$ $\forall p\ People(?p) \wedge Zakat(?x) \rightarrow Muslim(?m) \wedge Poor(?m) \wedge payZakat(?p,?m)$

NON-TAXONOMY RELATIONSHIP

Table 7 shows sample of non-taxonomy relationship generated manually from domain Islamic-Glossary.

TABLE 7. Sample Of Non-Taxonomy Relationship Generated

Text	Verb	Non-taxonomy relationship
All Muslims must believe in and utter the Shahadatain.	Believein utter	Believein(Muslim, Shahadatain) Utter(Muslim, Shahadatain)
The One to Whom all hearts submit in love, fear, reverence, desire, trust and sincerity.	Love Fear Reverence Desire Trust Sincerety	$\forall x\ Person(x) \wedge has(x,heart) \wedge Allah(a) \rightarrow love(x,a) \wedge fear(x,a) \wedge reverence(x,a) \wedge desire(x,a) \wedge trust(x,a) \wedge sincerity(x,a)$

And to Whom all limbs submit in all forms of worship such as prayers, supplications, sacrifices, invocations, etc.	submit	Submit(worship, Allah) Legend: Worship isA prayer Worship isA supplications Worship isA sacrifices Worship isA invocations
Other excused people are required to feed a poor person one meal for each day they do not fast if they can afford it, such as the elderly people and the ones who have permanent diseases like ulcers.	feed	feed(ExcusePeople,PoorPerson) Legend: PoorPerson isA elderlyPeople PoorPerson isA PermenentDiasese PermanentDiasease isA Ulcer
During the first third of the fast you taste Allah's mercy; during the second third you taste Allah's forgiveness; and during the last third you taste freedom from the Fire.	taste	Taste(FirstThir dofFast, Allah'sMercy) Taste(SecondThir dofFast,Allah'sForgiveness) Taste(LastThir dofFast, FreedomfromFire)

TESTING

10 assessors who are lecturers from the Department of Public Education, Polytechnic of Sultan Idris Shah, assessed the extraction results. The assessor has an academic background in the field of Islamic Studies and has experience in terms of teaching courses, involvement, conferences, research, presentations, publications and others related to the field of Islamic Studies. Specifically, 5 assessors from the Islamic Studies specialization, 3 from Usuluddin specialization, Fiqh and Fatwa specialization, an 1 from Shariah specialization and 1 from the Islamic Civilization specializations. Criteria tested on the assessor are as follows:

1. Part A: Knowledge of the field.
2. Part B: Hearst Pattern Extraction
3. Part C: Rules
4. Part D: Extraction of Non-Taxonomy Relations

PART A: KNOWLEDGE OF THE FIELD.

Figure 4 shows the percentage achieved by the assessor of knowledge in the field of Islamic religion and the Pillars of Islam. There are 7 elements that specialize in assessor knowledge about the Pillars of Islam particularly. Scales 1 to 5 refer to strongly disagree, disagree, uncertain, agree and strongly agree. 73% of the assessors strongly agree to have knowledge of the Islamic religion generally and the Pillars of Islam specifically. While 27% of assessor agree to have knowledge in related fields. The chart shows no uncertain or uninformed assessor about Islam. This is because all users have an academic qualification in the field of Islamic religion. The results of analytical knowledge assessments show that the assessor has the knowledge of Islam and the Pillars of Islam specifically.

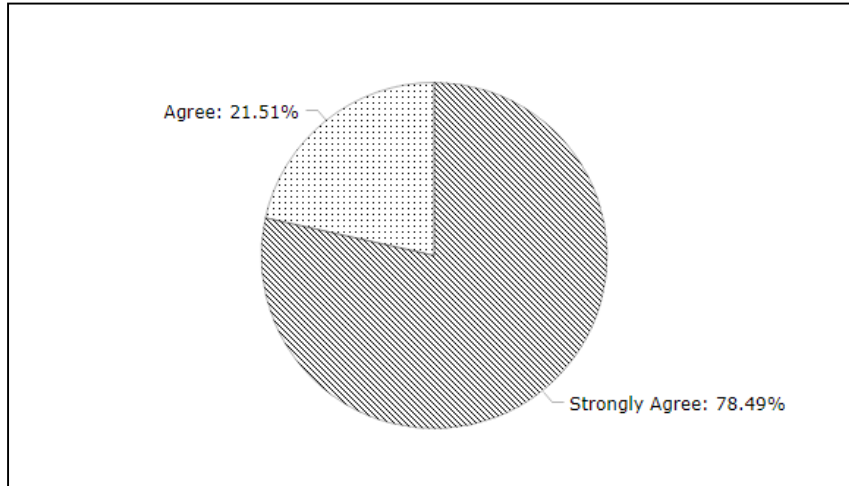


FIGURE 4. Percentage of Assessor's Knowledge

PART B: HEARST PATTERN EXTRACTION

Figure 5 shows the percentage reached by the assessor about Hearst Pattern Extraction. The first three elements are assessed based on the knowledge of the hierarchical concept and the Hearst Pattern. While 11 elements are the result of Hearst Pattern extraction compared to the original texts. 68% of the assessors strongly agree with the results of the Hearst Pattern extraction through comparison with the original texts given from the Dictionary of Glossary of Islamic Dictionary. While 31% of the assessors agreed with the results of the Hearst Pattern extraction. However, there is a 1% uncertain decision from the assessor. This is because the result of extraction is from the document that takes into account explicit meaning only. The results of the Hearst Pattern extraction analysis show that the majority of assessor understand the hierarchical concepts and Hearst Patterns described by researchers and strongly agree with the results of the Extracted Hearst Patterns.

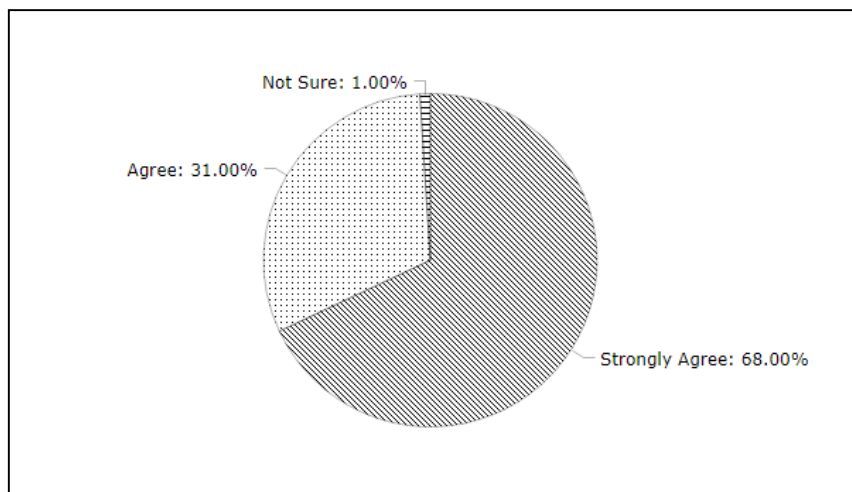


FIGURE 5. Percentage of Hearst Pattern extraction testing

PART C: RULES

Figure 6 indicates the percentage of testing result for rules generated. The first two elements are assessed based on the assessor's understanding of the formatting of rules in the form of FOL (first order logic) and the logic of the rules that conform to the domain of the Pillars of Islam in general. While nineteen elements are, results of manually rules generated compared to the original texts. 40% of the assessors strongly agree with the results of the rules through comparison with the original texts given from the Dictionary of Glossary Dictionary of Islamik. While 42% of the assessors agree with the results shown. However there are 17% uncertain results and 1% disagree from the assessor. This is because there are some opinions from the assessor about the rules that do not describe the Pillars of Islam or the religion of Islam as a whole because of the constraints of the text taken from the Glossary of Islamic Dictionary. The testing results of the rules generated show that the generation of rules should take into account the complete information of any Islamic Pillar and not constrained by the text from the original text only.

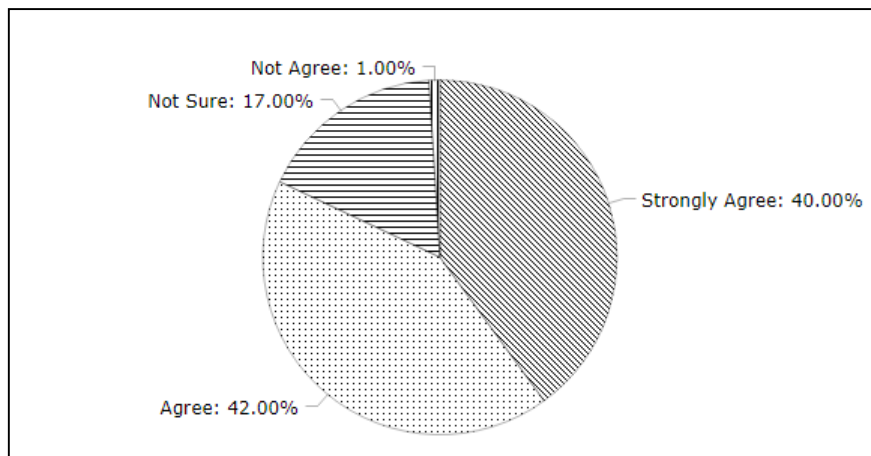


FIGURE 6. Percentage of Rules testing

PART D: EXTRACTION OF NON-TAXONOMY RELATIONS

Figure 7 shows the percentage reached by the assessor on the testing of Non-Taxonomy Relations. The first 3 elements were assessed based on the assessor's understanding of the concept of non-taxonomic relationships and the logic of the rules that conform to the domain of the Pillars of Islam in general. While 9 elements are the original texts, verbs and non-taxonomy relationships generated. 27.5% of the assessors strongly agree with the results of non-taxonomic relationships through comparison with the original texts given from the Dictionary of Glossary Dictionary of Islamic. While 56.7% of the assessors agreed with the results of non-taxonomic relationships shown. However there are 15.8% uncertain results from the assessor. As with the result of the rules testing, there are also some opinions from the assessors regarding the relationship caused by the constraints of the sentence taken from the Dictionary of Islamic Glossary. The results of the non-taxonomic relationship testing analysis show that the income process needs to take into account the overall pillar of Islam.

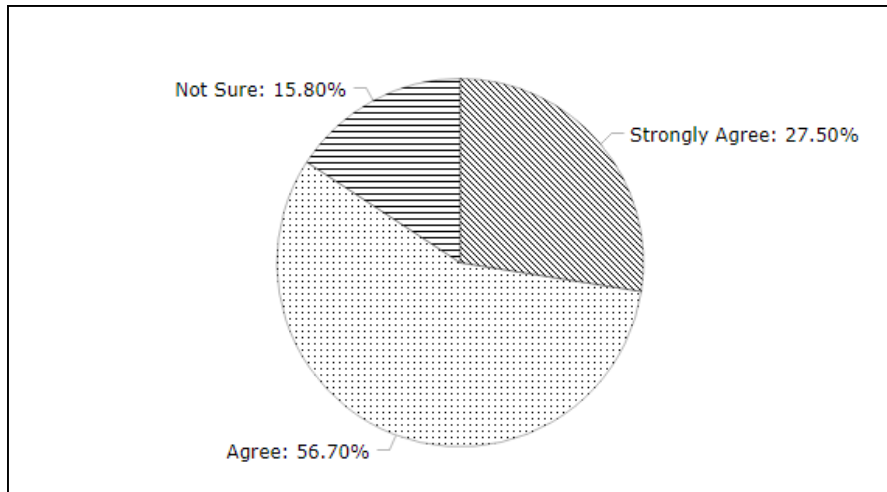


FIGURE 7. Percentage of Non-Taxonomy Relations testing

CONCLUSION

This study has successfully achieved the objectives set out that is to extract the hierarchical concepts based on the Hearst pattern and generate manual rules. Syntactic analysis has been shown and the test uses a likert scale against an evaluator with a background in line with the domain and scope of the study, the Pillars of Islam. Non-taxonomic relationship and rules have been manually generated for this Islamic Dictionary glossary data. There are 6 different types of Hearst patterns used to extract concept relationships from the dataset. Extraction results show 3 exploratory concepts based on the first Hearst pattern, 2 extraction from second pattern, 1 extraction from third and fourth patterns, 3 extraction from fifth pattern and lastly 1 extraction from sixth Hearts pattern. The number of successful concepts extracted from the Hearst cheerleaders is as many as 41 concepts. Manually rules generated from 19 sentences of 31 rules. Non-taxonomic relationships and the resulting rules have contributed to the learning of ontology for Islamic domains that focus on the Pillars of Islam.

REFERENCES

- Ali A.A & Saad S. 2016. Unsupervised Concept Hierarchy Induction Based On Islamic Glossary. ARPN Journal of Engineering And Applied Sciences. 11(13): 8505-8510.
- Buitelaar P., Cimiano P., and Magnini B. (Eds.). 2005. *Ontology Learning from Text: Methods, Evaluation and Applications*, Series information for Frontiers in Artificial Intelligence and Applications, Amsterdam IOS Press.
- Cimiano, P. 2006. *Ontology Learning and Population from Text, Algorithms, Evaluation and Applications* . Springer US.
- Cimiano, P., Mädche, A., Staab, S. & Völker, J. 2009. *Ontology Learning*. Handbook on ontologies. 245-267. Springer Dordrecht Heidelberg London New York.
- Dalloul, Y. M. 2013. *An Ontology-Based Approach to Support the Process of Judging Hadith Isnad*. Tesis Islamic University of Gaza.
- DEED, D. E. E. D. 2005. *Islamic Dictionary-Glossary*. <http://www.iiu.edu.my/deed/glossary/index2.html> [3 February 2017]
- Gruber, T. R. 1993. *A Translation Approach To Portable Ontology Specifications*. Knowledge acquisition 5(2). pp. 199-220.

- Hearst, M. A. 1992. *Automatic Acquisition Of Hyponyms From Large Text Corpora*. Proceedings of the 14th conference on Computational linguistics-Volume 2, 539-545.
- Hearst, M.A.: Automated Discovery of WordNet Relations. In: Fellbaum, C. (ed.) *WordNet: An Electronic Lexical Database*, pp. 132–152. MIT Press, Cambridge (1998).
- Saad, S. & Salim, N. 2008. *Methodology of Ontology Extraction for Islamic Knowledge Text*. Postgraduate Annual Research Seminar.
- Saad, S., Noah, S., Salim, N. & Zainal, H. 2013. *Rules and Natural Language Pattern in Extracting Quranic Knowledge*. Advances in Information Technology for the Holy Quran and Its Sciences (32519), 2013 Taibah University International Conference on, pp. 407-412.
- Saad, S., Salim, N. & Zainal, H. 2009. *Islamic Knowledge Ontology Creation*. Internet Technology and Secured Transactions, 2009. ICITST 2009. International Conference for, pp. 1-6.
- Sanderson, M. & Croft, B. 1999. *Deriving Concept Hierarchies From Text*. Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, pp. 206-213.
- Weaam & Saad. 2016. *Ontology Population From Quranic Translation Texts Based on a Combination of Linguistic Patterns and Association Rules*. Journal of Theoretical and Applied Information Technology. 86(2): 250-257.

Saidah Saad (PhD.)

Assistant Dean (Entrepreneurship and Creativity) / Senior Lecturer,
Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor.
saidah@ukm.edu.my
Off Tel: 03 - 8921 6183 / 6668
(Corresponding author)

Ummu Kalsom A. Latiff

Lecturer
Department of Information Technology and Communication
Polytechnic Sultan Idris Shah, 45100 Sungai Lang, Sungai Air Tawar, Selangor.
ummukalsom@psis.edu.my
Phone: 012-9137299

Received: 2 June 2018
Accepted: 15 August 2018
Published: 18 December 2018