

Pengekstrakan Konsep Dan Hubungan Bagi Istilah Islam Menggunakan Pendekatan Lexico Sintaktik

Ummu Kalsom A. Latiff
Saidah Saad
Sabrina Tiun

Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia

ABSTRAK

Pembelajaran Ontologi merupakan langkah separa automasi untuk mempelajari ontologi daripada teks. Pengenalpastian sesuatu istilah menjadi prasyarat bagi semua aspek Pembelajaran Ontologi. Lapisan Pembelajaran Ontologi dimulakan dengan mengenalpasti istilah, sinonim, konsep, konsep hierarki, hubungan dan petua bagi pelbagai domain teks dan salah satu daripadanya adalah teks berbentuk Islamik atau Glosari Islam. Glosari istilah Islam yang diterjemahkan dalam Bahasa Inggeris telah wujud dalam jumlah yang banyak dan memerlukan pengekstrakan maklumat penting bagi kefahaman yang jitu tentang sesuatu istilah Islam. Kewujudan senarai istilah Islam adalah untuk mengurangkan kesamaran kepelbagaian ejaan, untuk mencari konsep istilah dan menyediakan panduan untuk konsep Islam yang unik. Walaubagaimanapun, sumber berbentuk elektronik ini tidak mampu mengesan masalah ambiguiti dalam kalangan istilah yang wujud atau dikenali sebagai machine interoperability. Kajian ini bertujuan untuk mengenalpasti dan mengekstrak konsep, taksonomi, hubungan dan peraturan atau petua yang boleh dibina berdasarkan domain istilah-istilah dalam glosari Islam khusus kepada bidang Rukun Islam. Pengekstrakan ini melibatkan penggunaan pendekatan corak Hearst. Data set yang digunakan adalah daripada koleksi Kamus atau Glosari Islam daripada Universiti Islam Antarabangsa, Malaysia (DEED 2015). Kamus ini terdiri daripada istilah-istilah Islam iaitu konsep dan maksud bagi setiap konsep mengikut turutan abjad. Kajian ini menggunakan enam fasa melibatkan fasa penyediaan, pemprosesan serta pengujian yang dirangkumkan membentuk rekabentuk metodologi kajian. Sebanyak 41 konsep berjaya diekstrak berdasarkan 6 Corak Hearst, 31 petua dihasilkan secara manual dari 19 ayat dan 9 hubungan bukan taksonomi. Hasil rumusan kajian menunjukkan bahawa objektif kajian ini telah berjaya dicapai dalam skop yang ditentukan apabila hasil kajian serta pengujian yang dilakukan oleh pakar domain dalam domain menunjukkan keputusan yang positif. Kekangan kajian dibentang bagi membolehkan para penyelidik menambahbaik kajian dari semasa ke semasa. Cadangan kajian bagi penyelidikan pada masa akan datang telah dihuraikan agar kajian ini menjadi lebih bermanfaat dan lebih diperluas bagi panduan para penyelidik seterusnya umat Islam secara menyeluruh.

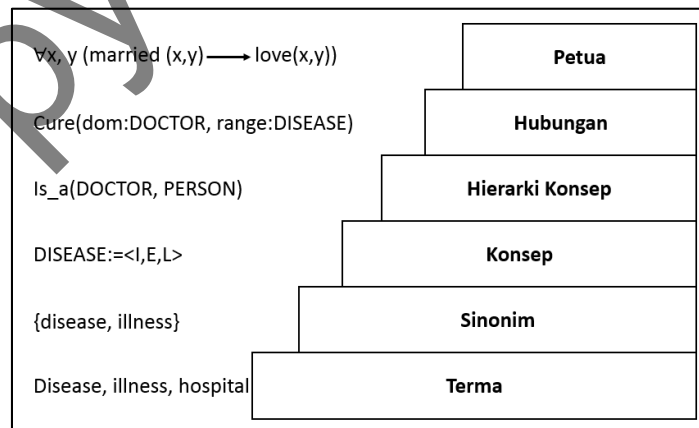
1. PENGENALAN

Definisi ontologi yang popular digunapakai telah dihasilkan oleh Thomas Gruber, (1993) “*A Translation Approach to Portable Ontology Specification, Knowledge Acquisition*”, iaitu Ontologi adalah satu persefahaman formal dan eksplisit mengenai konsep yang dikongsi bersama. Spesifikasi ontologi ini juga selain dihuraikan secara formal, ianya mengandungi koleksi istilah-istilah dan hubungannya yang boleh difahami mesin. ‘Eksplisit’ bermaksud dengan jelas menerangkan jenis konsep yang digunakan dan kekangan penggunaannya. ‘Formal’ merujuk kepada dokumen yang boleh difahami oleh mesin. ‘Dikongsi bersama’ bermaksud pengetahuan di dalam ontologi perlu dipersetujui dan diterima oleh sesebuah kumpulan atau komuniti.

'Konsep' merujuk kepada model abstrak yang mengandungi konsep-konsep yang berkaitan dan hubungan-hubungan yang wujud dalam sesetengah situasi. Ontologi memudahkan pengenalpastian konsep kepada kelas dan subkelas tertentu atau kategori objek di dalam sesuatu domain.

Proses membina asas pengetahuan yang mengklasifikasikan konsep dan hubungan ini dipanggil konsep hierarki (Sanderson dan Croft, 1999). Asas pengetahuan ini adalah teras bagi mana-mana ontologi di mana Pembelajaran Ontologi memerlukan hierarki taksonomi seperti yang dinyatakan oleh Cimiano et al, 2009. Konsep hierarki terdiri daripada kelas-kelas konsep yang dikategorikan kepada *super class* dan *sub class* hierarki. Definisi konsep dan hubungan di antara konsep amat penting dalam pembangunan ontologi selain perlu mengetahui simbol-simbol merujuk kepada konsep dan hubungannya. Ianya juga mengandungi taksonomi hubungan *is-a* atau hubungan bukan hierarki. Terdapat dua cara untuk membentuk ontology, pertama *Ontology Editor* dan kedua *Ontology Learning* atau Pembelajaran Ontologi. *Ontology Editor* merupakan perisian aplikasi yang digunakan untuk membentuk ontologi secara manual. Manakala Ontologi Pembelajaran merupakan langkah separa automasi untuk mempelajari ontologi daripada teks.

Rajah 1.1 menunjukkan lapisan pembelajaran ontologi beserta contoh. Istilah adalah calon bagi konsep dan hubungan yang berkaitan dan ia mengandungi perkataan atau perkataan majmuk. Pengenalpastian sesuatu terma menjadi prasyarat bagi semua aspek Pembelajaran Ontologi dari teks. Sinonim mencari perkataan yang mempunyai konsep yang sama atau secara semantik ia adalah perkataan yang sama. Dalam pembangunan Pembelajaran Ontologi, selain melibatkan keperluan memperoleh pengetahuan bahasa tentang istilah yang digunakan bagi merujuk konsep yang spesifik di dalam teks, ia juga mengenalpasti sinonim yang berkaitan. Tugas pada lapisan sinonim ini adalah untuk mengesan terminologi yang berkaitan, dan juga sinonim bagi istilah-istilah. Istilah dan sinonim yang dikumpulkan secara dasarnya akan membentuk konsep. Konsep untuk mencari definisi konsep bagi terma dan sinonim tadi, tanda-tanda leksikal yang digunakan merujuk kepada mereka. *Concept Hierarchy* mencari konsep hirarki, pemurnian dan lanjutan leksikal yang berkaitan. Konsep hirarki ini menjadi tulang belakang pembangunan ontologi. Ia mengatur dan menyusun konsep yang dikenalpasti kepada struktur hirarki atau taksonomi. Setiap konsep berhubungan dengan konsep lain di atas atau di bawah. Atribut dan hubungan digunakan untuk mencirikan konsep dalam hirarki. *Relations* atau hubungan mengkaji hubungan atau label yang wujud dan mengenalpasti *domain* dan *range* yang terlibat. *Rules* menghasilkan hubungan dan kaitan yang lebih kompleks di antara konsep dan hubungan.



Rajah 1.1 *Ontology Learning Layer Cake* (P.Buitelaar et al, 2005)

2. PENYATAAN MASALAH

Glosari istilah Islam yang diterjemahkan ke dalam Bahasa Inggeris telah wujud dalam jumlah yang banyak seperti di al-islam.org, islamicity.com dan clarionproject.org / glossary_islamic_terms yang mana ia melibatkan capaian dan membekalkan maklumat penting untuk kefahaman yang jitu tentang sesuatu istilah Islam. Tujuan utama kewujudan glosari ini adalah untuk mengurangkan kesamaran kepelbagaian ejaan, untuk mencari definisi istilah dan menyediakan panduan untuk istilah Islam yang unik. Walaubagaimanapun, sumber berbentuk elektronik ini tidak mampu mengesan masalah ambiguiti dalam kalangan istilah yang wujud atau dikenali sebagai *machine interoperability*. Salah satu penyelesaian kepada masalah ini adalah pengekstrakan konsep dan hubungan dalam istilah Islam bagi menjelaskan konsep dan hubungannya secara eksplisit dengan menggunakan pendekatan tertentu. Menurut kajian lepas, para penyelidik cenderung untuk menggunakan pendekatan tradisional corak *lexico sintaktik* yang telah di cadangkan oleh Hearst (1992) (Saad & Salim 2008; Saad et al. 2009). Kajian ini akan membuktikan sejauh mana keberkesanan pendekatan lexico-sintaktik dalam mengekstrak konsep dan hubungannya bagi istilah Islam.

Istilah-istilah Islam yang dipetik dari Al-Qur'an mempunyai gaya bahasa yang berbeza, bahasa yang kaya, lapisan maksud yang tersendiri dan berlapis serta mempunyai morfologi yang kompleks. Keadaan ini menyukarkan proses pengekstrakan yang akan dibangunkan. Ontologi juga perlu mempunyai liputan yang luas dalam domain bagi mendapatkan kesempurnaan model dengan menentukan generalisasi yang bermakna dan konsisten. Cabaran dalam tugas melibatkan pembangunan ontologi adalah perlu mengekalkan keseimbangan di antara memodelkan pengetahuan yang besar tetapi perlu memastikan model tersebut ringkas dan padat.

Proses pembangunan yang rumit untuk memastikan ontologi yang dihasilkan perlu diterima dan dipersetujui oleh sekumpulan komuniti kerana pada dasarnya pihak yang berlainan akan bersetuju dan pilihan rekabentuk yang berlainan. Penyelesaian yang ideal kepada permasalahan ini adalah dengan menggunakan pendekatan secara automatik. Pendekatan ini secara dramatik akan mengurangkan kos untuk membina ontologi (Cimiano, P, 2006).

3. OBJEKTIF KAJIAN

Matlamat kajian ini adalah untuk membuktikan sejauh mana keberkesanan pendekatan lexico-sintaktik dalam mengekstrak konsep dan hubungannya bagi istilah Islam. Bagi mencapai matlamat utama tersebut, beberapa objektif telah dibangunkan seperti berikut:

- i. Mengekstrak frasa nama yang sesuai bagi calon konsep menggunakan penghurai dari *Stanford Parser* bagi istilah dalam glosari Islam.
- ii. Mengekstrak konsep dan hubungan yang sesuai menggunakan pendekatan lexico sintaktik berasaskan corak Hearst (1992) bagi istilah dalam glosari Islam.
- iii. Menghasilkan petua-petua terlibat yang sesuai secara manual daripada glosari Islam

4. METODOLOGI KAJIAN

Rajah 3.1 menggambarkan fasa-fasa yang terlibat iaitu analisis dokumen, pra-pemprosesan dokumen, *Stanford parser*, pengekstrakan taksonomi Hearst, penghasilan petua secara manual,

pengekstrakan konsep bukan-taksonomi dan pengujian. Penerangan setiap fasa adalah seperti berikut:

i. Fasa Pertama, Analisis Dokumen

Fasa analisis dokumen melibatkan pengenalanpastian dataset dan pemilihan skop domain daripada dataset. Dataset yang dipilih adalah daripada Kamus-glosari Islam yang telah dikumpulkan dan disusun oleh Universiti Islam Antarabangsa, Malaysia. Kamus-glosari ini mengandungi sejumlah besar konsep-konsep atau istilah-istilah Islam beserta penerangan masing-masing (DEED 2005). Kamus-glosari ini boleh didapati di <http://www.iium.edu.my/deed/glossary/index.html>. Pemilihan skop domain dari kamus-glosari Islam ini merupakan usaha ke arah memfokuskan kepada satu bahagian atau bidang dengan mencipta sub-domain yang lebih kecil. Kajian ini memfokuskan bidang asas dalam Islam iaitu Rukun Islam. Rukun Islam mengandungi lima perkara iaitu mengucap dua kalimah shahadah, mendirikan solat lima waktu, berpuasa di bulan Ramadhan, mengeluarkan zakat dan mengerjakan Haji di Mekah bagi yang mampu.

ii. Fasa Kedua, Pra-pemrosesan dokumen.

Merujuk kepada Saad et.al. (2009), untuk menghasilkan ontologi Islamik, beberapa perkara perlu difahami dan dokumen Islamik perlu melalui beberapa pendekatan pra-pemrosesan bagi menyediakan dataset sebelum proses pengekstrakan berlaku. Perkara yang perlu dipertimbangkan bagi kajian ini adalah;

- a. Penerangan maksud bagi konsep dalam glosari hanya akan menggunakan maksud yang tersurat sahaja tanpa melihat maksud tersembunyi.
- b. Frasa yang perlu melalui proses penghurai akan melalui proses penggantian sementara makna bagi persediaan kemudahan proses penghurai seperti “KebesaranNya” digantikan dengan “Kebesaran Allah”.

Pendekatan yang digunakan dalam menyediakan dataset adalah seperti berikut;

- a. Huruf besar. Dalam glosari Islam, huruf besar menunjukkan ianya adalah konsep, melainkan perkataan di permulaan ayat.

iii. Fasa Ketiga, *Stanford Parser*

Fasa seterusnya adalah pemrosesan bahasa tabii dengan menggunakan pakej program *Stanford Parser* yang melibatkan beberapa kaedah iaitu penandaan golongan kata dan penghurai. Kaedah ini merupakan analisis sintaktik yang digunakan untuk mengenalpasti tag tatabahasa bagi setiap perkataan serta penghurai berfungsi untuk membuat pengekstrakan Frasa Nama (NP).

iv. Fasa Keempat, Pengekstrakan taksonomi lexico-sintaktik Hearst

Corak lexico-sintaktik yang dicadangkan oleh Hearst (1992) telah dipilih sebagai pendekatan yang digunakan untuk mengekstrak konsep dan hubungannya bagi domain istilah Islam. Ini berikutan menurut Hearst (1992), corak lexico-sintaktik berkebolehan dalam memodelkan pelbagai hubungan semantik umum, serta, hiponim menghasilkan keputusan yang paling tepat. Pernyataan ini disokong oleh Hearst (1998), di mana corak lexico-sintaktik ini mempunyai kelebihan kekerapan frekuensi bagi pelbagai jenis teks yang berbeza, dan juga mempunyai ketepatan keseluruhan paling optimum walaupun ke

atas pengetahuan yang tiada *pre-encoded*. Selain itu, para penyelidik yang lepas cenderung untuk menggunakan pendekatan tradisional ini seperti yang dinyatakan oleh Saad & Salim (2008) dan Saad et al. (2009). Merujuk kepada kajian lepas, seramai 5 daripada 7 orang penyelidik menggunakan pendekatan lexico-sintaktik dalam penyelidikan mereka.

v. **Fasa Kelima, Penghasilan petua secara manual.**

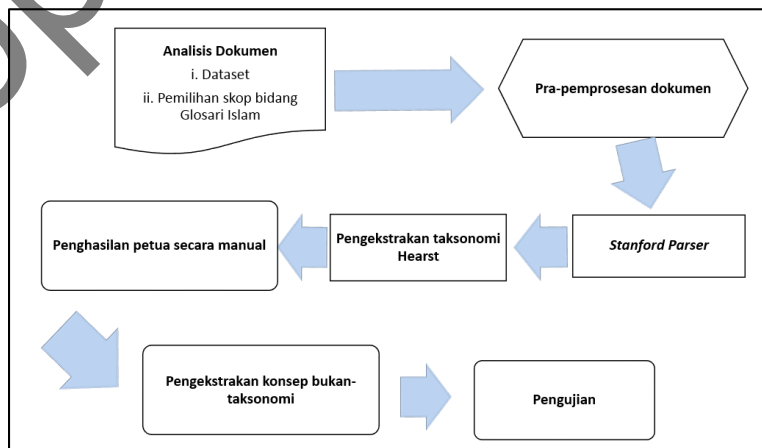
Petua adalah berdasarkan subset kepada *First Order Logic* (FOL) dan extensional yang berkemungkinan. FOL merupakan penaakulan sesebuah pernyataan yang dipecahkan kepada subjek dan predikat. Extensional pula melibatkan *instance* atau individu sesuatu subjek. Petua yang dihasilkan adalah secara manual berdasarkan kepada bahasa tabii dalam dokumen Islamik yang digunakan.

vi. **Fasa Keenam, Pengekstrakan hubungan bukan-taksonomi.**

Penemuan hubungan bukan taksonomi merupakan titik penting dalam pembelajaran ontologi. Walaubagaimanapun, ianya merupakan salah satu bidang pengetahuan yang kurang dikaji. Kajian ini menggunakan proses pengekstrakan secara semi-automatik bagi mendapatkan hubungan bukan taksonomi dari dataset.

vii. **Fasa Ketujuh, Pengujian**

Beberapa peringkat pengujian dilakukan bagi memastikan output setiap fasa metodologi adalah tepat. Ini dilakukan dengan membuat penyemakan semula secara manual bagi setiap fasa metodologi. Menurut Kaiya dan Saeki (2006), pengetahuan tentang satu-satu domain merupakan salah satu faktor penting bagi mendapatkan kejayaan yang berkualiti dan pengetahuan domain ini hanya dimiliki oleh pakar sesebuah domain. Oleh yang demikian, pengujian bagi kajian ini melibatkan penilaian daripada pakar domain dalam bidang yang berkaitan (*domain expert*). Pakar domain melihat kepada logik dan ketepatan persoalan tentang kandungan domain. Pemilihan penilaian berdasarkan pakar domain dilakukan kerana tiada penanda aras data yang sama untuk dijadikan perbandingan hasil dapatan. Pakar domain boleh dianggap sebagai pengguna akhir yang khusus yang mempunyai pengetahuan yang diperlukan tentang maksud sebenar sesuatu istilah Islam, pengertian dan huraian yang tepat dan hampir tepat kepada maksud sebenar. Rajah 3.1 menunjukkan rekabentuk metodologi bagi kajian ini.



Rajah 3.1 Rekabentuk Metodologi Kajian

5. ANALISIS KAJIAN

Keputusan analisis sintaktik melibatkan keputusan penandaan golongan kata dan keputusan pengestrakan frasa nama sebagai prapemprosesan dokumen. Seterusnya keputusan pengestrakan corak Hearst dan konsep yang melibatkan padanan ungkapan malar ditunjukkan. Pengujian pengestrakan konsep diterangkan melibatkan pakar domain yang terlibat dan bahagian-bahagian penilaian yang disusun. Dataset yang diambil dari Kamus-glosari Islamik yang memfokuskan Rukun Islam mengandungi sebanyak 232 ayat.

5.1 Keputusan Analisis Sintaktik

Seksyen ini menunjukkan hasil keputusan analisis sintaktik bagi penandaan golongan kata (POS), penghurai, pengestrakan corak Hearst dan penyenaarian frasa nama. Fungsi penandaan golongan kata adalah mengenalpasti kelas sintaktik untuk setiap perkataan dalam sebuah ayat. Penghurai bergantung kepada penandaan golongan kata bagi membentuk frasa nama dan diwakilkan ke dalam bentuk rajah pepohon. Penggunaan ungkapan malar atau *regex* (*regular expression*) digunakan bagi mengekstrak hubungan yang memenuhi Corak Hearst.

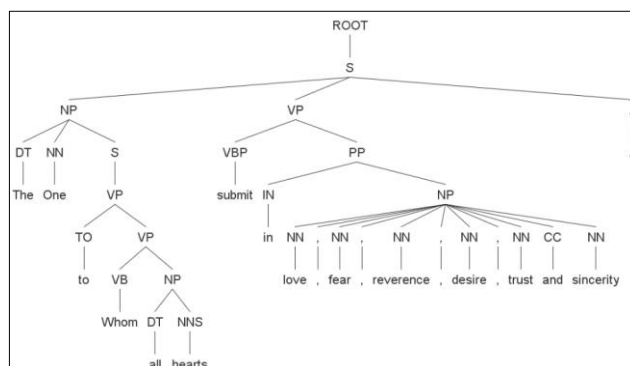
i. Penandaan Golongan Kata

Jadual 5.1 menunjukkan sampel keputusan bagi penandaan golongan kata bagi sampel teks asal yang menceritakan Rukun Islam secara umum. Didapati setiap perkataan ditandakan dengan golongan kata tertentu seperti NN, kata nama, VB, kata kerja, JJ, kata sifat dan sebagainya.

Jadual 5.1 Sampel Keputusan Penandaan Golongan Kata bagi teks

Teks	Penandaan Golongan Kata
Arkan is a pillars. There are five Arkan of Islam. Shahadatain is a bearing witness. Salat is a prayers. Seeaam is a fasting. Also spelled as Saum. To fast in the month of Ramadan. Zakat is a wealth dues. Zakat is also due on other things such as silver, animals, crops, etc. Hajj means pilgrimage. To perform the pilgrimage to the Holy City of Makkah at least once in one's lifetime.	Arkan NNP is VBZ a DT pillars NNS . . 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 . . Also RB spelled VBD as IN Saum NNP . . To TO fast VB in IN the DT month NN of IN Ramadan NNP . . Zakat NNP is VBZ a DT wealth NN dues NNS . . Zakat NNP is VBZ also RB due JJ on IN other JJ things NNS such JJ as IN silver NN , , animals NNS , , crops NNS , , etc FW . . Hajj NNP means VBZ pilgrimage NN . . To TO perform VB the DT pilgrimage NN to TO the DT Holy NNP City NNP of IN Makkah NNP at IN least JJS once RB in IN one CD 's POS lifetime NN . .

ii. Pengekstrakan Frasa Nama



Rajah 5.2 Huraian bagi teks “The One to Whom all hearts submit in love, fear, reverence, desire, trust and sincerity.”

Rajah 5.2 menunjukkan hasil huraian dengan menggunakan penghurai yang diwakilkan dalam bentuk pepohon bagi teks “The One to Whom all hearts submit in love, fear, reverence, desire, trust and sincerity.”. Terdapat tiga frasa nama yang diekstrak iaitu pertama: [The, One, to, Whom, all, Heart] iaitu gabungan di antara DT, NN, VP, TO, VB dan NNS, kedua: [all, heart], gabungan di antara DT dan NNS, ketiga: [love, fear, reverence, desire, trust, and, sincerity] iaitu gabungan di antara NN dan CC.

Jadual 5.3 Sampel Keputusan Frasa Nama yang diekstrak oleh penghurai

Teks	Frasa Nama
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]
These five Salat are one of the Arkan of Islam	[These, five, Salat] [one, of, the, Arkan, of, Islam] [one] [the, Arkan, of, Islam] [the, Arkan] [Islam]
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]

Jadual 5.3 menunjukkan sampel hasil pengekstrakan frasa nama dari teks asal. Frasa nama yang diekstrak ini adalah merujuk kepada rajah pepohon yang telah dihurai oleh penghurai. Terdapat sesetengah frasa nama merupakan frasa nama tunggal seperti *Salat*, *worship*, *Allah* dan *Saum*. Terdapat juga frasa nama yang terbentuk hasil dari gabungan perkataan dari kelas DT, CC, JJ dan sebagainya.

iii. Keputusan Pengekstrakan Corak Hearst dan Konsep

Proses pengekstrakan Corak Hearst berdasarkan enam corak utama. Penjanaan ungkapan malar atau *regex* ditetapkan terlebih dahulu seperti yang ditunjukkan dalam Jadual 5.4. Ungkapan malar merupakan suatu kaedah yang digunakan bagi pemadanan corak. Ianya suatu kaedah yang fleksibel dan ringkas bagi proses pemadanan terhadap suatu teks.

Jadual 5.4 Ungkapan malar atau *regex* berdasarkan Corak Hearst

Corak Hearst	Ungkapan malar
i) <i>NP such as</i> {NP,}*{(and-or)} NP	(?:[(* s* ,* w* ,* s* .*)]*)? [(* w* s* *such)* [(* w* s* *as)* (?:[(* s* ,* w* ,* s* .*)]*)?
ii) <i>Such NP as</i> {NP,}*{(and-or)} NP	[(* w* s* *Such)* (?:[(* s* ,* w* ,* s* .*)]*)* [(* w* s* *as)* (?:[(* s* ,* w* ,* s* .*)]*)?
iii) <i>NP {,NP}*{,} or other NP</i>	(?:[(* s* ,* w* ,* s* .*)]*)? [(* w* s* *or)* [(* w* s* *other)* (?:[(* s* ,* w* ,* s* .*)]*)?
iv) <i>NP {,NP}*{,} and other NP</i>	(?:[(* s* ,* w* ,* s* .*)]*)? [(* w* s* *and)* [(* w* s* *other)* (?:[(* s* ,* w* ,* s* .*)]*)?
v) <i>NP {,} including</i> {NP,}*{(and-or)} NP	(?:[(* s* ,* w* ,* s* .*)]*)? [(* w* s* *including)* (?:[(* s* ,* w* ,* s* .*)]*)?
vi) <i>NP {,} especially</i> {NP,}*{(and-or)} NP	(?:[(* s* ,* w* ,* s* .*)]*)? [(* w* s* *especially)* (?:[(* s* ,* w* ,* s* .*)]*)?

Jadual 5.5 menunjukkan sampel keputusan hasil pengekstrakan Corak Hearst bagi sampel teks Kamus-glosari Islamik. Sebagai contoh bagi teks, “*Zakat is also due on other things such as silver, animals, crops, etc.*”, penemuan frasa nama [other things] dan [silver, animals, crops, etc] yang dihubungkan dengan frasa ‘*such as*’. Corak hubungan taksonomi ini berpadanan dengan corak Hearst yang pertama. Manakala bagi teks “*Emission of impurities from the private parts: urine, faeces, wind, prostatic fluid, or other discharge.*”, penemuan frasa nama [urine, faeces, wind, prostatic fluid] dan [discharge] dihubungkan dengan frasa ‘*or other*’. Corak hubungan taksonomi ini berpadanan dengan corak Hearst yang ketiga.

Jadual 5.5 Sampel Pengekstrakan Corak Hearst

Corak Hearst	Teks	Hasil Pengekstrakan
i) <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.)))
ii) <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),

iii)	<i>NP {,NP}*{,} or other NP</i>	Emission of impurities from the private parts: urine, faeces, wind, prostatic fluid, or other discharge.	etc.) NP(NP(urine,) NP(faeces,) NP(wind,) NP(prostatic fluid,) CC(or) NP(JJ(other) discharge.))
iv)	<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)
v)	<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.))))
vi)	<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,))

Jadual 5.6 menunjukkan jumlah pengekstrakan Corak Hearst dan jumlah konsep yang diekstrak bagi data Kamus-Glosari Islamik yang menfokuskan kepada domain Rukun Islam. Corak Hearst yang pertama berpadanan sebanyak 3 pengekstrakan dan menemukan sebanyak 14 konsep, corak kedua, sebanyak 2 pengekstrakan dan menemukan 9 konsep, corak ketiga dan keempat dengan masing-masing 1 pengekstrakan dan masing-masing menemukan 5 dan 2 konsep, corak kelima sebanyak 3 pengekstrakan dan menemukan 9 konsep manakala corak keenam sebanyak 1 pengekstrakan dan menemukan 2 konsep. Jumlah konsep yang berpadanan dengan corak Hearst adalah sebanyak 41 konsep.

Jadual 5.6 Jumlah Pengekstrakan Konsep mengikut Corak Hearst

	Corak Hearst	Jumlah Pengekstrakan / Jumlah Ayat diekstrak	Jumlah Konsep	Konsep
i)	<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
ii)	<i>Such NP as {NP,}*{(and-or)} NP</i>	2	9	Sin, Shirk, Qatl, Zinah, Riba, Sirq, Books, Sahih Bukhari, Sahih Muslim
iii)	<i>NP {,NP}*{,} or other NP</i>	1	5	Emission of impurities, Urine, Faeces, Wind, Prostatic Fluid
iv)	<i>NP {,NP}*{,} and other NP</i>	1	2	Religious activities, Pilgrimage to Makkah

v)	<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
vi)	<i>NP { , } especially {NP,} *{(and-or)} NP</i>	1	2	Al-Quran, Al-fatihah

5.2 Pengujian

Hasil pengekstrakan dinilai oleh 10 orang pakar domain yang terdiri daripada pensyarah dari Jabatan Pengajian Am, Politeknik Sultan Idris Shah. Pakar domain mempunyai latarbelakang akademik dalam bidang Pengajian Islam serta mempunyai pengalaman dari segi pengajaran kursus, keterlibatan, persidangan, penyelidikan, pembentangan, penerbitan, sumbangan dan lain-lain yang berkaitan dengan bidang Pengajian Islam. Secara sepesifiknya, seramai 5 orang pakar domain dari pengkhususan Pengajian Islam, 3 orang dari pengkhususan Usuluddin, Fiqh dan Fatwa, seorang pakar domain dari pengkhususan Syariah dan seorang dari pengkhususan Tamadun Islam. Kriteria yang diuji ke atas pakar domain adalah seperti berikut :

- Bahagian A: Pengetahuan bidang pakar domain.
- Bahagian B: Pengekstrakan Corak Hearst
- Bahagian C: Petua
- Bahagian D: Pengekstrakan Hubungan Bukan Taksonomi

Metodologi pengujian bersama pakar domain dimulakan dengan sesi pencerahan tentang pengenalan kepada ontologi, konsep dan hubungannya, konsep hierarki, corak lexico sintaktik, petua dan formatnya, hubungan bukan taksonomi serta pengekstrakan konsep selain tujuan kajian ini dijalankan. Sesi pencerahan dilakukan terlebih dahulu bagi memastikan para pakar domain mempunyai kefahaman minimum dalam bidang semantik. Selain itu, antara item yang dimasukkan dalam borang soal selidik ini adalah tentang kefahaman para pakar domain tentang konsep dan istilah yang digunakan. Langkah seterusnya diteruskan dengan mengisi borang soal selidik mengikut bahagian-bahagian yang disusun oleh pakar domain. Penyelidik memimpin dan mencerahkan setiap item yang perlu diisi bagi meminimumkan kekeliruan dalam kalangan pakar domain. Menurut Sartor, G. et al. (2013), sebanyak 69.44% persetujuan terhadap satu-satu ontologi diterima dan menyarankan bahawa masih terdapat ruang untuk penambahbaikan hasil daripada kebolehan dalam kalangan pakar domain. Oleh yang demikian, kajian ini meletakkan sasaran minimum 69.44% sebagai peratus capaian penilaian setuju dan sangat setuju.

- Bahagian A : Pengetahuan bidang pakar domain.**

Jadual 5.7 Penilaian pengetahuan pakar domain

Pengetahuan Umum		1	2	3	4	5
1	Saya mempunyai pengetahuan tentang agama Islam					10
2	Saya mempunyai pengetahuan tentang apa itu Rukun Islam					10
3	Saya mempunyai pengetahuan tentang maksud, penerangan dan hikmah Mengucap Dua Kalimah Syahadah secara teori dan praktikal				4	6
4	Saya mempunyai pengetahuan tentang maksud, penerangan, hukum dan hikmah Solat 5 Waktu secara teori dan praktikal				4	6
5	Saya mempunyai pengetahuan tentang maksud, penerangan, hukum dan hikmah Berpuasa secara teori dan praktikal				3	7
6	Saya mempunyai pengetahuan tentang maksud, penerangan, hukum dan hikmah Membayar Zakat secara teori dan praktikal				3	7
7	Saya mempunyai pengetahuan tentang maksud, penerangan, hukum dan hikmah Mengerjakan Haji secara teori dan praktikal				5	5
Jumlah / Bilangan		0/70	0/70	0/70	19/70	51/70
Peratusan (%)		0	0	0	27	73

Jadual 5.7 menunjukkan jumlah dan peratusan yang dicapai oleh pakar domain tentang pengetahuan dalam bidang agama Islam dan Rukun Islam. Sebanyak 7 elemen yang mengkhususkan ilmu pakar domain tentang Rukun Islam secara khusus. Skala 1 hingga 5 merujuk kepada sangat tidak setuju, tidak setuju, tidak pasti, setuju dan sangat setuju. Sebanyak 73% pakar domain sangat bersetuju mempunyai pengetahuan berkaitan agama Islam umumnya dan Rukun Islam khususnya. Manakala 27% pakar domain bersetuju mempunyai pengetahuan dalam bidang berkaitan. Carta menunjukkan tiada pakar domain yang tidak pasti atau tidak mempunyai pengetahuan tentang agama Islam. Ini disebabkan semua pengguna mempunyai kelulusan dalam bidang agama Islam umumnya. Hasil analisis pengetahuan pakar domain menunjukkan bahawa pakar domain mempunyai pengetahuan tentang agama Islam dan Rukun Islam secara khusus.

b) Bahagian B : Pengekstrakan Corak Hearst

Jadual 5.8 Sampel Penilaian Pengekstrakan Corak Hearst

6	Corak ketiga: NP (NP1) or other NP Teka	Hasil ekstrak	1	2	3	4	5
6.1	Emission of excretories from the private parts: urine, faeces, wind, prostatic fluid or other discharge					3	7
7	Corak keempat: NP (NP1) and other NP Teka	Hasil ekstrak	1	2	3	4	5
7.1	Fasting the month of Ramadan, celebrating the two major feasts (Eid al-Fitr and Eid al-Adha), performing the pilgrimage to Makkah, and other religious activities depend upon the lunar calendar.					4	6
8	Corak kelima: NP (a) including NP1 + (and/or) NP Teka	Hasil ekstrak	1	2	3	4	5
8.1	Behaviour: Both men and women are not permitted to trim their hair or nails, wear perfume, harm or kill any living thing, including insects, or damage any plant or tree.						10
8.2	Kufr is Disbelief in Allah's Commands, including the refusal to accept whatever the Prophet Muhammad may Allah bless him and grant him peace, taught.				1	2	7
8.3	In other words, a term that indicates all that pleases Allah, including sayings and actions of the heart or limbs.					4	6
9	Corak keenam: NP (a) especially Teka	NP1 + (and/or) NP Hasil ekstrak	1	2	3	4	5
9.1	Mitham is the often repeated Ayat of the Holy Qur'an, especially the Surah al-Fatihah, for it is always recited during Salat, in every Rak'a.					2	8
Jumlah / Bilangan			6	0	2	43	96
Peratusan (%)			140	0	140	140	140
			0	0	1	31	68

Jadual 5.8 menunjukkan jumlah dan peratusan yang dicapai oleh pakar domain tentang Pengekstrakan Corak Hearst. Sebanyak tiga elemen pertama dinilai berdasarkan kefahaman pakar domain tentang konsep hierarki dan Corak Hearst. Manakala sebanyak 11 elemen merupakan teks asal dan hasil pengekstrakan berdasarkan corak Hearst untuk dibandingkan oleh pakar domain. Sebanyak 68% pakar domain sangat bersetuju dengan hasil pengekstrakan Corak Hearst melalui perbandingan dengan teks asal yang diberi dari dokumen Kamus-glosari Islamik. Manakala 31% pakar domain bersetuju dengan hasil pengekstrakan Corak Hearst. Walaubagaimanapun terdapat sebanyak 1% keputusan tidak pasti dari pakar domain. Ini disebabkan hasil pengekstrakan adalah dari dokumen yang mengambilkira maksud tersurat sahaja. Hasil analisis pengekstrakan Corak Hearst menunjukkan bahawa majoriti pakar domain memahami konsep hierarki dan Corak Hearst yang dijelaskan oleh penyelidik dan sangat bersetuju dengan hasil pengekstrakan Corak Hearst yang ditunjukkan.

c) Bahagian C : Petua

Jadual 5.9 Sampel Penilaian Petua

		ZakatPendapatan(?x) ≡ Muslim(?y) ^ hasSaving(?x, ?y) ^ YearlySavings(?y, 2.5%) !p People(?p) ^ Zakat(?x) → Muslim(?m) ^ Poor(?m) ^ payZakat(?p, ?m)					
3.11	The Zakat is compulsory on all Muslims who have saved (at least) the equivalent of 85g of 24 carat gold at the time when the annual Zakat payment is due.	Zakat(?z) ^ [Muslim(?m) ^ (24CaratGold(?cg) → 85g(?cg)) → saving(?m, ?cg)] → Pay(?m, ?z) ZakatEmas(?x) ≡ Muslim(?x) ^ hasGold(?x, 24 carat) ^ hasSaved(?x, >85g) !p People(?p) ^ Muslim(?p) ^ ZakatEmas(?x) → payZakat(?p, ?x)			2	6	2
3.12	Zakat is also due on other things such as silver, animals, crops, etc.	Zakat(?z) → Silver(?z1) ^ Animals(?z2) ^ Crops(?z3) Zakat(?z) ≡ Silver(?z1) ∨ Animals(?z2) ∨ Crops(?z3) ∨ ZakatEmas(?z4) ∨ ZakatPendapatan(?z5)			1	7	2
3.13	Haji means pilgrimage. To perform the pilgrimage to the Holy City of Makkah at least once in one's lifetime (if one is able to afford it).	Haji(?h) ≡ Pilgrimage(?h) Haji(?h) ^ Muslim(?m) → performIn(?h, Makkah) ^ ableTo(?m, afford)			1	5	4
3.14	Islam literally means "submission to the will of Allah".	Islam(?i) ≡ Allah(?a) ^ People(?p) ^ submissionToTheWillOf(?p, ?a)			6	3	1
3.15	The five pillar of Islam are as follows: Shahada, Solat, Zakat, Fasting and Hajj.	pillarOfIslam(?x) ≡ Shahada(?p1) ∨ Solat(?p2) ∨ Zakat(?p3) ∨ Fast(?p4) ∨ Hajj(?p5)				5	5
3.16	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) ≡ Prayers(?st) Prayers(?st) ≡ Subh(?st1) ∨ Dhur(?st2) ∨ Asr(?st3) ∨ Maghreb(?st4) ∨ Isha(?st5)				4	6
3.17	Seeam is a Fasting. Also spelled as Saum. To fast in the month of Ramadan.	Seeam(?sm) ^ Ramadhan(?r) → FastingIn(?sm, ?r)				5	5
3.18	Ibadah is acts of worship and adoration in utmost submission and obedience and with utmost love to Allah Ta'ala.	Ibadah(?x) ≡ act(?x) ^ worship(?x)			3	2	4
3.19	Love, fear, reverence, desire (hope), trust and sincerity are some forms of worship from within the heart.	WorshipFromHeart(?wh) ≡ Love(?wh1) ^ fear(?wh2) ^ reverence(?wh3) ^ desire(?wh4) ^ trust(?wh5) ^ sincerity(?wh6)				3	5
Jumlah / Bilangan			0/210	3/210	35/210	89/210	83/210
Peratusan (%)			0	1	17	42	40

Jadual 5.9 menunjukkan jumlah dan peratusan yang dicapai oleh pakar domain tentang Penilaian Petua. Sebanyak dua elemen pertama dinilai berdasarkan kefahaman pakar domain tentang format petua dalam bentuk FOL (*first order logic*) dan logik petua yang menepati domain Rukun Islam secara umum. Manakala sebanyak sembilan belas elemen merupakan teks asal dan hasil petua yang dikeluarkan secara manual ditunjukkan untuk

proses penilaian. Sebanyak 40% pakar domain sangat bersetuju dengan hasil Petua melalui perbandingan dengan teks asal yang diberi dari dokumen Kamus-glosari Islamik. Manakala 42% pakar domain bersetuju dengan hasil Petua yang ditunjukkan. Walaubagaimanapun terdapat sebanyak 17% keputusan tidak pasti dan 1% tidak setuju dari pakar domain. Ini disebabkan terdapat beberapa pendapat daripada pakar domain tentang terdapat petua yang tidak menggambarkan Rukun Islam atau agama Islam secara keseluruhannya disebabkan kekangan ayat yang diambil dari Kamus-glosari Islamik. Hasil analisis penilaian Petua menunjukkan bahawa penghasilan petua perlu mengambilkira tentang maklumat lengkap bagi mana-mana Rukun Islam dan tidak terkekang dengan ayat daripada teks asal.

d) Bahagian D : Pengekstrakan Hubungan Bukan Taksonomi

Jadual 5.10 Penilaian Hubungan Bukan Taksonomi

Pernyataan			1	2	3	4	5	
1	Saya faham tentang maksud hubungan bukan taksonomi dalam bidang ontologi setelah dijelaskan oleh penyelidik kajian ini.				3	4	3	
2	Saya merujuk kepada rajah pohon yang telah dihurai, hubungan bukan taksonomi adalah apabila ianya memenuhi corak berikut NP <verb> NP				4	4	2	
3	Saya melihat hubungan bukan taksonomi yang dikeluarkan dalam bentuk petua yang dihasilkan adalah bertepatan dengan agama Islam khususnya Rukun Islam.				2	5	3	
4	Teks	Kata Kerja	Hubungan	1	2	3	4	5
4.1	All Muslims must believe in and utter the Shahadain.	Believein utter	Believein(Muslim, Shahadain) Utter(Muslim, Shahadain)			1	6	3
4.2	The One to Whom all hearts submit in love, fear, reverence, desire, trust and sincerity.	Love Fear Reverence Desire Trust sinceraty	lx Person(x) ^ has(x,heart) ^ Allah(e) → love(x, a) ^ fear (x, a) ^ reverence(x, a) ^ desire(x, a) ^ trust(x, a) ^ sincerity(x, a)			1	7	2
4.3	And to Whom all limbs submit in all forms of worship such as prayers, supplications, sacrifices, invocations, etc.	submit	Submit(worship, Allah) Petunjuk: Worship isA prayer Worship isA supplications Worship isA sacrifices Worship isA invocations				5	5
4.4	To the Zoroastrians it creates a picture of a fire, Allah, there is none but He, King of the Universe, the Creator and Originator of all that there is.	Create	Create(Zoroastrians, PictureOfFire)			2	6	2
4.5	There are five daily obligatory prayers in Islam, consisting of fixed sets of standings, bowings, prostrations and sittings in worship to Allah.	Consist	consist (Prayer, Standing) consist (Prayer, Bowing) consist (Prayer, Prostration) consist (Prayer, Sitting)			3	5	2
4.6	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) Petunjuk: PoorPerson isA elderlyPeople PoorPerson isA PermanentDiasease PermanentDiasease isA Ulcer				6	4
4.7	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(FirstThirdofFast, Allah'sMercy) Taste(SecondThirdofFast, Allah'sForgiveness) Taste(LastThirdofFast, FreedomfromFire)			2	7	1
4.8	Zakat Al-Fitr is a small obligatory head-tax	imposedOn	imposedOn (smallHeadTax, Muslim)				7	3
	imposed on every responsible Muslim who has the means for himself and his dependants.		Petunjuk: Zakat Al-Fitr isA SmallHeadTax					
4.9	The performance of Hajj consists of ten steps: Put on the Ihram at the appointed station, then recite the Talbiyah.	Perform ConsistsOf	Perform(Hajj, TenSteps) ConsistsOf(Hajj, TenSteps)			1	6	3
Jumlah / Bilangan				0/120	0/120	19/120	68/120	33/120
Peratusan (%)				0	0	15.8	56.7	27.5

Jadual 5.10 menunjukkan jumlah dan peratusan yang dicapai oleh pakar domain tentang Penilaian Hubungan Bukan Taksonomi. Sebanyak 3 elemen pertama dinilai berdasarkan kefahaman pakar domain tentang konsep hubungan bukan taksonomi dan logik petua yang menepati domain Rukun Islam secara umum. Manakala sebanyak 9 elemen merupakan teks asal, kata kerja dan hubungan yang dikeluarkan ditunjukkan untuk proses penilaian. Sebanyak 27.5% pakar domain sangat bersetuju dengan hasil hubungan bukan taksonomi melalui perbandingan dengan teks asal yang diberi dari dokumen Kamus-glosari Islamik. Manakala 56.7% pakar domain bersetuju dengan hasil hubungan bukan taksonomi yang ditunjukkan. Walaubagaimanapun terdapat sebanyak 15.8% keputusan tidak pasti dari pakar domain. Seperti hasil penilaian petua, juga terdapat beberapa pendapat daripada pakar domain tentang terdapat hubungan yang disebabkan kekangan ayat yang diambil dari Kamus-glosari Islamik. Hasil analisis penilaian hubungan bukan taksonomi menunjukkan bahawa proses penghasilan perlu mengambilkira gambaran tentang Rukun Islam secara keseluruhan.

6. KESIMPULAN

Kajian ini telah berjaya mencapai objektif yang ditetapkan iaitu membuat pengekstrakan konsep hierarki berdasarkan corak Hearst dan menghasilkan petua secara manual. Analisa secara sintaktik telah dikeluarkan dan pengujian menggunakan skala likert terhadap pakar domain yang mempunyai latarbelakang yang seiring dengan domain dan skop kajian ini iaitu Rukun Islam telah dijalankan. Hubungan bukan taksonomi dan petua berjaya dihasilkan secara manual bagi data Kamus-glosari Islamik ini. Sebanyak 6 jenis corak Hearst yang digunakan untuk mengekstrak hubungan konsep dari dataset. Hasil pengekstrakan menunjukkan sebanyak 3 pengekstrakan hubungan konsep berdasarkan corak Hearst yang pertama, 2 pengekstrakan dari corak kedua, 1 pengekstrakan dari corak ketiga dan keempat, 3 pengekstrakan dari corak kelima dan akhir sekali 1 pengekstrakan dari corak Hearts yang keenam. Jumlah konsep yang berjaya diekstrak dari pengekstrakan sorak Hearst adalah sebanyak 41 konsep. Manakala penghasilan petua secara manual berjaya dihasilkan daripada 19 ayat sebanyak 31 petua. Hubungan bukan taksonomi dan petua yang dihasilkan telah menyumbang kepada pembelajaran ontologi bagi domain Islam yang menfokuskan kepada Rukun Islam.

RUJUKAN

- 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, IOS Press.*
- Cimiano, P. 2006. *Ontology Learning From Text.* Springer.
- Cimiano, P., Mädche, A., Staab, S. & Völker, J. 2009. *Ontology Learning.* Handbook on ontologies. 245-267. Springer.
- Gruber, T. R. 1993. *A Translation Approach To Portable Ontology Specifications.* Knowledge acquisition 5(2). 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. 1998. *Automated Discovery of WordNet Relations, in WordNet: An Electronic Lexical Database.* Christiane Fellbaum (ed.), MIT Press.
- Kaiya, H., Saeki, M. 2006. *Using domain ontology as domain knowledge for requirements elicitation.* In: Proceedings of 14th IEEE international Requirements Engineering conference (RE'06), pp 186-195.

- Saad, S. & Salim, N. 2008. *Methodology of Ontology Extraction for Islamic Knowledge Text*. Postgraduate Annual Research Seminar.
- Saad, S., Salim, N. & Zainal, H. 2009. *Islamic Knowledge Ontology Creation*. Internet Technology and Secured Transactions, 2009. ICITST 2009. International Conference for, 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, 206-213.
- Sartor, G., Casanovas, P., Biasiotti, M., Fernandez-Barrera, M. 2013. *Approaches to Legal Ontologies: Theories, Domains, Methodologies*. Springer Publishing Company. Incorporated

Copyright@FTSM