

# PEMBANGUNAN E-KAMUS ISLAM MENGGUNAKAN UNGKAPAN NALAR BERDASARKAN TERJEMAHAN AL-QURAN DALAM MAKNA TEKS BAHASA INGGERIS

Nursyuhada binti Razali  
Saidah Binti Saad

Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia

## ABSTRAK

Pembangunan e-Kamus Islam adalah sebagai satu pelantar atas talian untuk pengguna mencari makna perkataan Islam dengan mudah. Pembangunan e-kamus ini melibatkan pengekstrakan data daripada Al-Quran terjemahan makna di dalam bahasa Inggeris. Data yang telah diekstrak akan dibentuk dalam maklumat yang lebih bermakna (mempunyai nilai semantik) dan dikelas dalam bentuk hierarki data yang disebut taksonomi. Corak petua Hearst digunakan untuk mendapatkan hierarki data daripada maklumat yang berjaya di ekstrak. Corak lain yang boleh menjana e-kamus dan taksonomi data juga dikenalpasti dan diekstrak menggunakan kaedah ungkapan nalar. Hasil analisis pengekstrakan frasa nama menggunakan tatabahasa RegexpParser berjaya mengekstrak frasa nama bagi pembangunan e-kamus bahasa Inggeris bagi domain Al-Quran.

## 1 PENGENALAN

Projek ini berfokus kepada pembangunan e-Kamus Islam secara atas talian yang melibatkan pengekstrakan data dari terjemahan Al-Quran. Seterusnya dibangunkan sebuah kamus atas talian yang berfungsi dan berfokus kepada istilah Islam di dalam Bahasa Inggeris seperti Shayatin iaitu syaitan dan seumpamanya. Pengekstrakan ini adalah untuk mendapatkan istilah-istilah yang berkaitan daripada dokumen tak berstruktur untuk dianalisis menggunakan Pemrosesan Bahasa Tabii (NLP). Seterusnya mengekstrak kepelbagaian tandaan kata (part-of-speech - POS) yang membentuk frasa katanama dengan menggunakan ungkapan nalar (regular expression) seperti Det\*Adj\*N+PP\*N\*.

Data yang telah diekstrak dibentuk dalam bentuk hierarki data yang dinamakan taksonomi. Semantik merupakan kajian tentang makna sesuatu perkataan, frasa, ayat dan ujaran. Menurut Kamus Dewan Bahasa Edisi Keempat, definisi semantik memberi erti kajian mengenai makna perkataan dan penambahan makna sesuatu kata. Nilai semantik perlu dititikberatkan kerana terdapat perkataan-perkataan Islam yang tidak dapat diberi makna kerana keadaannya yang bersifat abstrak, namun makna suatu istilah itu dapat difahami melalui hubungan diantara istilah

yang berkaitan seperti Muhamad dikaitkan dengan rasul, Mekah dikaitkan dengan Rumah Allah dan baitullah. Begitu juga istilah Sabar, AlBirr yang dikaitkan dengan perlbagai makna di dalam Al-Quran. Bidang semantik juga mempunyai istilah-istilah seperti sinonim, antonim, polisemi dan hiponim yang boleh memberi makna kepada sesuatu perkataan. Oleh itu, perkataan-perkataan ini boleh difahami maksudnya dengan merujuk kepada perkataan ini yang mempunyai hubungan dengan perkataan lain disekelilingnya.

Data yang diekstrak akan disusun secara teratur dalam bentuk hierarki yang disebut taksonomi agar pengolahannya dapat dilakukan dengan efisien. Menurut Jiawei Han, Micheline Kamber, Jian Pei (2012) konsep hierarki didefinisikan sebagai urutan pemetaan dari sekumpulan konsep peringkat rendah yang lebih spesifik hingga konsep peringkat lebih tinggi atau umum. Konsep ini dilihat boleh memperkayakan lagi hasil carian oleh pengguna kerana wujudnya perkaitan antara istilah-istilah yang ada. Oleh itu, corak Hearst digunakan untuk mendapatkan bentuk hierarki ini. Kunci ideanya adalah untuk memanfaatkan corak leksiko-sintaksis tertentu untuk mengesan hubungan dalam teks. Contohnya, corak seperti "subscript y ialah subscript x" atau "subscript x ialah subscript y lain" dan menunjukkan hubungan hipernimis bentuk x ialah y. Apabila, hubungan antara perkataan berjaya dilakukan, ini akan menjadikan kamus ini unik dan berfungsi kepada pengguna.

## 2 PENYATAAN MASALAH

Terdapat pelbagai e-kamus Islam di atas talian seperti *Islamic Term Dictionary* yang memberi banyak maklumat mengenai istilah-istilah dalam Islam yang dilihat amat membantu pengguna mempelajari ilmu Islam. Tetapi terdapat satu kekurangan pada e-Kamus Islam yang telah wujud iaitu tiada hubungan antara istilah seperti yang dibincangkan dalam pengenalan. E-kamus tersebut hanya mempunyai makna istilah satu hala sahaja. Yang dimaksudkan satu hala ialah hanya terdapat istilah dan makna istilah sahaja di mana maklumat di dalam kamus itu tidak berkembang dan tidak bertopik. Hal ini, menyukarkan proses pencarian oleh pengguna sehingga perlu menelaah istilah satu persatu dan melambatkan proses pencarian maklumat.

Seterusnya, kesukaran memahami istilah Islam yang abstrak kerana kebanyakan perkataan Islam berasal daripada Bahasa Arab dan juga salah satu istilah mungkin mempunyai pelbagai

makna yang ditafsir oleh penterjemah berdasarkan pengalaman dan latar belakang penterjemah. Bahasa tanpa unsur semantik ialah kosong dan tidak bermakna. Kebanyakan istilah mempunyai makna yang lebih kurang sama tetapi tidak dapat dikenal pasti maksud sebenarnya. Sebagai contoh, ‘destinasi’ dan ‘perhentian terakhir’ membawa makna yang sama, istilah yang berbeza. Perkara ini yang menjadi masalah dalam menghasilkan kamus ini.

### **3 OBJEKTIF KAJIAN**

Tujuan kajian ini adalah untuk mengekstrak terjemahan Al-quran dan glosari kamus untuk dijadikan kamus elektronik yang baharu dan mempunyai fungsi yang lebih baik daripada kamus sedia ada di internet. Objektif ialah untuk mengekstrak frasa nama berdasarkan ungkapan nalar yang menjadi konsep utama bagi Al-Quran serta membentuk data hierarki bagi istilah Al-Quran.

### **4 METOD KAJIAN**

#### **4.1 Fasa Perancangan**

Fasa ini melibatkan proses mengenalpasti masalah, objektif, persoalan kajian dan menentukan skop. Langkah seterusnya adalah kajian susastera yang melibatkan pengumpulan, pencarian dan pembacaan jurnal dan kajian lepas bagi mencetus idea dan inspirasi. Contoh topik yang berkaitan yang dikaji ialah mengenai pengeskrakan frasa nama serta penggunaan ungkapan nalar. Penggunaan internet untuk mendapatkan maklumat berkenaan seperti beberapa laman sesawang iaitu Google Scholar dan rujukan buku berkaitan. Maklumat dikumpul, distruktur ,disintesis dan akhir sekali dipersembahkan secara kritis dan kreatif dalam fasa analisis. Fokus pencarian lebih merujuk kepada kajian pengekrak konsep, frasa kata nama dan juga corak Hearst (Saad et.al, 2009) dan kajian lanjutan dari Corak Hearst (Stephen Roller, Douwe Kiela, and Maximilian Nickel) yang dilakukan oleh pengkaji terdahulu.

#### **4.2 Fasa Analisis**

Fasa ini melibatkan beberapa analisis dan tafsiran mengenai maklumat yang telah dikumpulkan dalam fasa perancangan. Fasa ini juga tertumpu kepada menganalisa set data yang akan digunakan iaitu Terjemahan Al-Quran dalam Bahasa Inggeris untuk meningkatkan kefahaman tentang kajian

yang akan dijalankan. Selain itu, menganalisa perisian yang akan digunakan untuk membangunkan projek sama ada bersesuaian ataupun tidak.

### 4.3 Fasa Reka Bentuk

Fasa ini melibatkan beberapa proses iaitu pra-pemprosesan dokumen, penandaan golongan kata dan pengekstrakan frasa nama dan maksud dan perbandingan frasa nama. Pra-pemprosesan dokumen melibatkan beberapa fasa iaitu membuang tanda baca yang menggunakan kaedah *join()*, pentokenan melibatkan *word\_tokenize* dan pangkasan. Seterusnya, pengekstrakan frasa nama dan maksud. Perkataan-perkataan yang bersesuaian dikeluarkan daripada teks asal yang berbentuk seperti contoh AsSalat (IqamatasSalat). Ungkapan nalar iaitu '`\w+ \w+ \w+ \(. * ? \)`' adalah untuk mengekstrak tiga patah perkataan dan diikuti dengan semua ayat atau perkataan di dalam kurungan seperti 'the Sacred Months (the Ist 7th 11th and 12th months of the Islamic calendar)'. Terdapat 188 item yang telah berjaya dikumpulkan. Frasa-frasa ini masih belum dikenalpasti sama ada frasa nama ataupun tidak. Hanya frasa di hadapan kurungan yang akan dikenalpasti. Kemudian penandaan golongan kata menggunakan *pos\_tag* untuk menanda golongan perkataan dengan mudah.

Seterusnya mengekstrak frasa nama menggunakan penghurai berasaskan peraturan *RegexParser* iaitu "`NP: {<DT>?<JJ>*<NN>}`". Terdapat sesetengah frasa nama merupakan frasa nama tunggal seperti Salat, Allah dan Sakeenah. Terdapat juga frasa nama yang terbentuk hasil dari gabungan perkataan dari kelas DT, CC, JJ dan sebagainya. Seterusnya membuat perbandingan frasa nama yang telah diekstrak dengan teks. Apabila frasa di hadapan kurungan didapati merupakan frasa nama maka ia akan mengambil keseluruhan perkataan dan ayat dan di simpan di dalam list yang baru. Sebanyak 85 frasa nama dapat dikumpulkan daripada perbandingan yang telah dilakukan. Pemboleh ubah 'nouns' mempunyai 117 frasa merupakan frasa nama yang diekstrak daripada frasa nama di depan kurungan dan di dalam kurungan. Oleh kerana hanya frasa nama di hadapan sahaja dibandingkan maka tidak semua 117 frasa nama dibandingkan.

```

self.__hearst_patterns = [
    (
        '(NP_\\w+ (, )?such as (NP_\\w+ ?(, )?(and |or )?)++)',
        'first'
    ),
    (
        '(such NP_\\w+ (, )?as (NP_\\w+ ?(, )?(and |or )?)++)',
        'first'
    ),
    (
        '((NP_\\w+ ?(, )?)+(and |or )?other NP_\\w+)',
        'last'
    ),
    (
        '(NP_\\w+ (, )?include (NP_\\w+ ?(, )?(and |or )?)++)',
        'first'
    ),
    (
        '(NP_\\w+ (, )?especially (NP_\\w+ ?(, )?(and |or )?)++)',
        'first'
    ),
]

```

Rajah 1 Corak Hearst

Rajah 1 merupakan enam ungkapan nalar corak Hearst untuk mengenali hiponim yang terdapat di dalam terjemahan Al-Quran. Namun tiada hasil yang didapati daripada terjemahan ini kerana tiada ayat yang boleh dipadankan dengan corak Hearst iaitu *such as*, *as*, *include* dan *especially*. Hal ini kerana, terjemahan Al-Quran merupakan ayat kompleks yang menjurus kepada penceritaan sesuatu peristiwa. Terjemahan ialah dimana ayat-ayat Al-Quran itu diterjemahkan dari bahasa Arab ke bahasa ibunda seseorang. Contoh bahasa Arab ke bahasa Melayu atau bahasa Arab ke bahasa Inggeris/ English. Oleh sebab itu, tiada penerangan atau ulasan kepada ayat-ayat tersebut supaya pembacanya lebih memahami atau mengetahui cerita disebalik surah berkenaan. Maka, sukar untuk dipadankan dengan corak Hearst bagi mengekstrak frasa nama dan hiponimnya.

#### 4.4 Fasa Pengujian

Bahagian ini membentangkan tentang pengujian yang dinilai oleh tiga orang yang terdiri daripada pensyarah dan pelajar dari Fakulti Pengajian Islam (FPI), Universiti Kebangsaan Malaysia (UKM). Mereka mempunyai latar belakang akademik dalam bidang Pengajian Islam iaitu

Hadith/Terjemahan al-Quran dalam Bahasa Inggeris, Al-Quran dan Al-Sunnah dan Terjemahan Arab-Melayu-Arab serta mempunyai pengalaman dari segi pengajaran kursus, keterlibatan, persidangan, penyelidikan, pembentangan, penerbitan, sumbangan dan lain-lain yang berkaitan dengan bidang Pengajian Islam. Kriteria yang diuji ke atas pakar domain adalah seperti berikut :

- i. Bahagian A: Pengetahuan bidang pakar domain.
- ii. Bahagian B: Pengekstrakan frasa nama beserta maksud

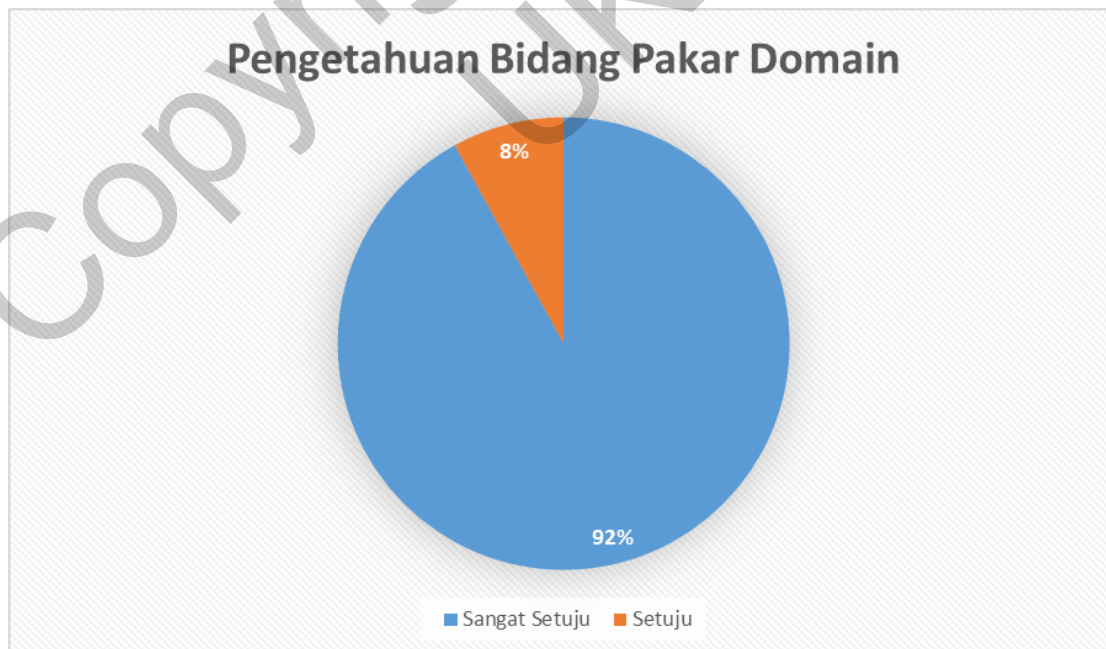
Pelan pengujian dimulakan dengan sesi penerangan tentang pengenalan projek iaitu pengekstrakan frasa nama beserta tujuan kajian ini dilakukan. Sesi ini dijalankan untuk memastikan para penilai memahami konsep pengekstrakan frasa nama dan nilai semantik. Selain itu, borang soal selidik dipecahkan kepada dua bahagian iaitu bahagian A dan bahagian B. Bahagian A menumpukan tentang penilaian pengetahuan bidang pakar domain dalam konsep dan istilah yang digunakan. Manakala bahagian B menumpukan kepada pengekstrakan frasa nama beserta nilai semantik yang telah diambil daripada Surah At-Taubah. Seterusnya pengujian diteruskan dengan pakar-pakar domain mengisi borang soal selidik mengikut bahagian-bahagian yang ditetapkan. Penyelidik memimpin dan pencerahkan setiap item yang perlu diisi bagi meminimumkan kekeliruan dalam kalangan pakar domain. Skala 1 hingga 5 merujuk kepada sangat tidak setuju, tidak setuju, tidak pasti, setuju dan sangat setuju. Penilai akan menilai sama ada frasa nama tersebut mempunyai maksud yang tepat ataupun tidak. Terdapat 45 frasa nama yang diekstrak dan diletakkan dalam borang soal selidik ini. Persetujuan terhadap satu-satu ontologi dapat diterima jika mencapai peratusan sebanyak 69.44% dan disarankan bahawa masih terdapat ruang untuk penambahbaikan hasil daripada kebolehan dalam kalangan pakar domain (Sartor, G. et al. (2013)). Oleh yang demikian, kajian ini meletakkan sasaran minimum 69.44% sebagai peratus capaian penilaian setuju dan sangat setuju. .

## 5 HASIL KAJIAN

Jadual 1 merupakan frasa nama yang telah diekstrak daripada ungkapan nalar yang disebut dalam fasa reka bentuk.

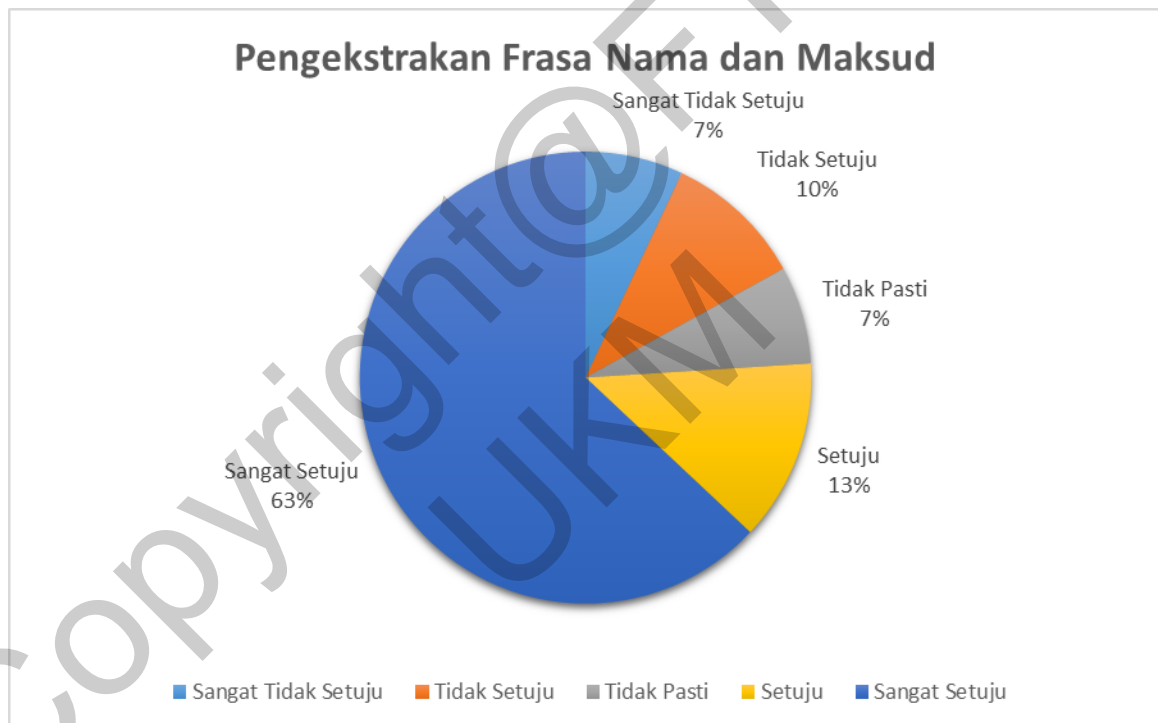
Jadual 1 Sampel Frasa Nama

Teks	Frasa Nama
<b>the Sacred Months</b> (the 1st 7th 11th and 12th months of the Islamic calendar)	[sacred months] [1st 7th 11 <sup>th</sup> ] [Islamic]
<b>AsSalat</b> (IqamatasSalat)	[assalat] [iqamatassalat]
<b>Najasun</b> (impure)	[najasun]
<b>His Messenger</b> (Muhammad SAW)	[messenger] [muhammad saw]
<b>His Sakeenah</b> (calmness tranquillity peace etc.)	[sakeenah] [calmness tranquillity peace etc]
<b>AlMasakin</b> (the poor)	[almasakin]
<b>the Prophet</b> (Muhammad SAW)	[prophet] [muhammad saw]
<b>people who are Fasiqoon</b> (rebellious disobedient to Allah)	[fasiqoon] [rebellious disobedient] [allah]
<b>the Injeel</b> (Gospel)	[injeel] [gospel]
<b>the Ansar</b> (Muslims of AlMadinah)	[ansar] [muslims] [almdinah]
<b>the Zalimoon</b> (cruel violent proud polytheist and wrongdoer)	[zalimoon] [violent proud polytheist]



Rajah 2 Carta Pai Pengetahuan Bidang Pakar Domain

Rajah 2 menunjukkan pencapaian peratusan pengetahuan pakar domain tentang Terjemahan Al-Quran dan frasa nama yang diwakili dalam rajah carta pai. 92 peratus para penilai sangat bersetuju bahawa mereka mempunyai pengetahuan berkaitan agama Islam umumnya, terjemahan Al-Quran dan frasa nama manakala sebanyak 8 peratus penilai setuju dengan pengetahuan mengenai frasa nama. 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, terjemahan Al-Quran dan frasa nama secara khusus.



Rajah 3 Carta Pai Penilaian Pengekstrakan Frasa Nama dan Maksud

Rajah 3 menunjukkan pencapaian peratusan pengekstrakan frasa nama dan maksud yang diwakili dalam rajah carta pai. Sebanyak 63% penilai sangat bersetuju bahawa frasa nama dan maksud tersebut adalah tepat. Selain itu, sebanyak 13% penilai setuju dengan maksud frasa nama manakala sebanyak 7% penilai menjawab tidak pasti dengan maksud frasa nama tersebut. Di samping itu, sebanyak 10% penilai tidak bersetuju dan 7% sangat tidak bersetuju dengan maksud



frasa nama tersebut. Frasa nama yang mempunyai maksud yang tidak tepat adalah kerana dalam terjemahan Al-Quran tidak semua perkataan atau ayat di dalam kurungan adalah maksud tetapi ia juga merupakan penceritaan lebih lanjut tentang peristiwa yang sedang diceritakan atau rujukan ayat. Kesimpulannya, hasil analisis pengekstrakan frasa nama dan maksud menunjukkan para penilai sangat bersetuju dengan hasil yang ditunjukkan dengan mencapai peratusan persetujuan sebanyak 76%.

Analisis ini menunjukkan teks surah ini mempunyai banyak frasa nama yang boleh diekstrak untuk dikumpulkan menjadi kamus. Hasil kajian mendapati berjaya mencapai satu objektif tetapi masih banyak lagi penambahbaikan perlu dilakukan kepada kod-kod yang telah dilakukan agar dapat mengekstrak frasa nama dengan lebih baik.

Hasil analisis dan pengujian menggunakan kaedah skala likert kepada pakar domain dalam bidang yang sama menunjukkan hasil yang positif. Secara kesimpulannya hasil pengujian yang dilaksanakan menunjukkan kajian ini berjaya mencapai objektif yang telah ditetapkan dimana mencapai lebih daripada kadar minimum peratusan persetujuan yang disasarkan iaitu 76% . Selain daripada itu, hasil pengujian ini akan menentukan sebahagian kebolehgunaannya. Walaubagaimanapun beberapa penambahbaikan perlu diambilkira, termasuk cadangan dan pendapat yang diberikan oleh pakar domain yang mempunyai pengetahuan dalam domain ini bagi memperluaskan manfaat dan memantapkan kajian ini.

## **6 KESIMPULAN**

Kesimpulannya, projek ini berjaya mencapai 76% dalam keseluruhan objektif yang ditulis. Antara kekangan yang dikenalpasti ketika menjalankan kajian ini adalah istilah-istilah yang di ambil daripada terjemahan Al-quran mempunyai gaya bahasa yang lebih kompleks dan mempunyai lapisan makna yang bertingkat-tingkat. Hal ini menyukarkan proses pengekstrakan kelak kerana perlu mempunyai kefahaman yang tinggi mengenai istilah-istilah tersebut.

Kajian terhadap penggunaan corak Hearst dalam terjemahan Al-Quran ini didapati tidak berhasil mendapatkan padanan konsep dan hiponim. Cadangan untuk penggunaan corak Hearst

adalah menggunakan dataset yang lain seperti glosari Islam kerana mempunyai lebih banyak konsep yang boleh dikembangkannya untuk menjadi sebuah kamus Islam.

Cadangan penambahbaikan bagi kekangan yang timbul adalah seperti mendapatkan bekerjasama dengan pakar yang lebih berkemahiran dalam Bahasa Arab seperti pensyarah di Fakulti Pengajian Islam bagi mendapatkan hasil yang jitu. Di samping itu, pembangun perlu menimba ilmu dan memahami konsep pengaturcaraan dengan lebih mendalam dengan bantuan dari buku, internet, pensyarah dan rakan-rakan agar dapat menjalankan kajian ini dengan lebih baik dan lancar.

Antara cadangan yang diperoleh daripada adalah untuk menggunakan terjemahan Al-Quran daripada Quran.com kerana menggunakan ejaan istilah yang betul. Sebagai contoh ejaan sebenar al-Muhajirun tetapi di dalam terjemahan yang digunakan menggunakan ejaan al-Muhajiroon. Selain itu, sebahagian istilah dalam Al-Quran perlu dikekalkan seperti Rasul, Ansar, Kata nama khas seperti nama para Nabi, Maryam dan nama tempat perlu dikekalkan. Kelemahan yang dikenalpasti akan diperbaiki dalam kajian kajian akan datang. Cadangan penambahbaikan amat penting bagi membantu menyelesaikan masalah yang dihadapi. Akhir sekali, terdapat pelbagai ilmu yang telah dipelajari ketika kajian ini dijalankan

## 7 RUJUKAN

(P39304), Z. I. 2015. Model pengisian ontologi domain herba ubatan daripada pelbagai sumber teks berasaskan teknik petua dan corak leksikal. Fakulti Teknologi dan Sains Maklumat, UKM.

Introduction to Stemming - GeeksforGeeks.2018, June 20. Retrieved July 21, 2021, from GeeksforGeeks website: <https://www.geeksforgeeks.org/introduction-to-stemming/>

Martin Gibbs. 2017, September 22. Semantic Data Model: Definition & Example. Retrieved from <https://study.com/academy/lesson/semantic-data-model-definition-example.html>.

NLP | How tokenizing text, sentence, words work - GeeksforGeeks. 2019, January 28

Python program to check if a word is a noun - GeeksforGeeks. 2020, December 9. Retrieved July 21, 2021, from GeeksforGeeks website: <https://www.geeksforgeeks.org/python-program-to-check-if-a-word-is-a-noun/>

Python RegEx. (2021). Retrieved July 21, 2021, from W3schools.com website: [https://www.w3schools.com/python/python\\_regex.asp](https://www.w3schools.com/python/python_regex.asp)

- Regex-based lexical analysis in Python and Javascript - Eli Bendersky's website. (2013). Retrieved July 21, 2021, from Thegreenplace.net website: <https://eli.thegreenplace.net/2013/06/25/regex-based-lexical-analysis-in-python-and-javascript>
- Snow, R., Jurafsky, D., & Ng, A. (n.d.). *Learning syntactic patterns for automatic hypernym discovery*.
- Specification of Requirements/Lexico-Syntactic Patterns - Ontology-Lexica Community Group. (2013). Retrieved July 21, 2021, from W3.org website: [https://www.w3.org/community/ontolex/wiki/Specification\\_of\\_Requirements/Lexico-Syntactic\\_Patterns#:~:text=Lexico%2Dsyntactic%20patterns%20\(LSP%2C,relations%20in%20natural%20language%20text](https://www.w3.org/community/ontolex/wiki/Specification_of_Requirements/Lexico-Syntactic_Patterns#:~:text=Lexico%2Dsyntactic%20patterns%20(LSP%2C,relations%20in%20natural%20language%20text).
- TextBlob: Simplified Text Processing — TextBlob 0.16.0 documentation. (2020). Retrieved July 21, 2021, from Readthedocs.io website: <https://textblob.readthedocs.io/en/dev/>
- The Noble Quran - English translation. (2021). Retrieved July 21, 2021, from Iium.edu.my website: [https://www.iium.edu.my/deed/quran/nobelquran/index\\_t.html](https://www.iium.edu.my/deed/quran/nobelquran/index_t.html)
- Ummu Kalsom A. Latiff, Saidah Saad dan Sabrina Tiun. 2018. Pengekstrakan Konsep Dan Hubungan Bagi Istilah Islam Menggunakan Pendekatan Lexico Sintaktik. Laporan Teknik. FTSM, UKM. PS-FTSM-2018-011. Pg 1-15.
- user993563. 2011, December 20. Get the string within brackets in Python. Retrieved July 21, 2021, from Stack Overflow website: <https://stackoverflow.com/questions/8569201/get-the-string-within-brackets-in-python>