INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN PROGRESSIVE EDUCATION & DEVELOPMENT

OPEN ACCESS JOURNAL

Vol 7, Issue 4, (2018) E-ISSN: 2226-6348

Word Classification in the Online Database of Malay-Arabic Comparable Phrases

Taj Rijal Muhamad Romli[1], Md Zahar Othman[2], Mohd Hilmi Abdullah[3], Mohd Zawawi Awang Hamat[3]

Email: taj.rijal@fbk.upsi.edu.my, zahar@fskik.upsi.edu.my, hilmi@fbk.upsi.edu.my, mohd.zawawi@pbmpu.upsi.edu.my

PBPU, Universiti Pendidikan Sultan Idris, 35900 Tg. Malim, Perak

To Link this Article: http://dx.doi.org/10.6007/IJARPED/v7-i4/4853 DOI:10.6007/IJARPED/v7-i4/4853

Published Online: 17 November 2018

Abstract

This paper will present the word classification method in developing a Malay- Arabic corpus online database. Word classification was developed in this model to explain to users about the types of words in the sentences and phrases displayed as the search results of Malay-Arabic comparable sentences. Word classification is a method of describing to users, especially to students who have a problem with identifying word classes, particularly in the Arabic language. Word classes are divided based on three main classes of the Arabic language, namely *ism*, *fi`il* and *harf*. *Ism* is divided into *ism `alam* and *ism mushtaq*. *Ism mushtaq* is divided into *masdar*, *ism al-fa`il*, *ism al-ma`ful*, *ism al-zaman* and *ism al-makan*. *Fi`il* is divided into three main components, namely *fi`il al-madhi*, *fi`il al-mudhari'* and *fi`il al-amr*. While *harf* is in *harf* component only. The method of listing each of these word class components is done using a software that separates each sentence component entry based on word class. The development of this database was made by using PHP software, MySQL Database and Linux-based Web Server System. The word class displayed will be in accordance with each phrase or sentence entry inserted, together with the translations. User is only required to enter a keyword in the search box and the system will display the search results of comparable Malay-Arabic phrases or sentences along with the word classes.

Keywords: Database, Comparable, Entry, Sentence, Word Classes.

Introduction

In the early millennium, the dictionary as a translation tool had been increasingly growing, especially the use of a computer aid named Computer Aided Translation (CAT) (Wan Rose Eliza, 2013). Dictionary stored in computer software was not only in the form of vocabularies, but also in the form of phrases and sentences. This software was developed and built so that it could translate certain phrases and sentences. It was what named as translation machine (MT). The data of these vocabularies, phrases and sentences were named corpus. These data could be accessed anywhere with the development of internet usage.

In the world of translation, the more language data stored in a computer software, the more it affects the smoothness and speed of the translation process, while simultaneously helping translators to become more productive (Hutchins, 2001). The sophistication in processing large data has become the driver in the field of communication and information. It is with this technology that makes European language a prominent language. Technology has helped the European empire to expand their languages and cultures, including the English and French languages. The United States, as the only great power state, had made English as a ' lingua franca' in the modern world. The power of the state is not what causes language development, but the development of the translation engine technology used (Yang, 2009).

Information that are quickly translated provide huge input for a country. Countries that are able to compete with this development are the ones that will conquer the world. Western countries and developed countries such as the United States, Europe, China and Japan are constantly competing to build and expand the creation of this translation machine as they see its importance and effects in the world communication and the development of universal knowledge. This paper explains the first step in developing a Malay-Arabic corpus database using limited data but built with word class features.

E-Kamus Software Development

Malay-Arabic e-Kamus is a software that displays the meaning of the word searched, together with examples of parallel and comparable sentences. Parallel sentence is a sentence that comes with its translation, while comparable sentence comes with non-translated sentences (Rusli Abdul Ghani & Norhafizah, 2001). The development of this web-based e-Kamus system was designed using PHP software, MySQL Database and Linux-based Web Server System.

PHP

PHP stands for Personal Home Page: Hypertext Preprocessor which is a scripting language used for web development (Hariss, 2004), (Meloni, 2003) and (Thomson, 2016). A web application is developed, characterised by server application that is carried out by using the Linux operating system. The PHP version used in developing this system was the version 5.0. We have chosen PHP as the programming basis because the programming language was very easy to be used in various operating system platforms such as Windows, Linux and Mac OS.

MySQL Database

MySQL is a short form for My Structure Query Language which is an open-source database (Hariss, 2004), (DuBois, 2006), (Nixon, 2014) and (Thomson, 2016). This database software is free and does not require any payment to use it. The MySQL database stores a large dictionary data and can process various applications or languages including PHP, Java and Python.

In addition, MySQL can store data in the binary form. This is important if we want to store a dictionary that has graphic and audio data. In general, MySQL is used with PHP in web development. It allows users to view information stored in the database via the internet.

Linux-based Web Server System

In order to develop a web-based dictionary system using PHP and MySQL database, we needed a web server system where the e-Kamus application would be implemented. We had

chosen the Linux platform (Love, 2013) as the operating system because this Linux-based web server was already available at UPSI. We were only required to ask for permission from the Computer Department, Faculty of Art, Computing and Creative Industry, to create an account to access the system.

Indirectly, this has cut the cost to purchase a computer as a server for web application development. Three key components in the creation of e-Kamus were PHP, MySQL and Linux-based web system. All of these three components support one another. Without integrating these three components, this web-based dictionary application would not have been produced.

Software Data Design

The following Figure 1 below shows the basic overview of the development design of the Malay-Arabic e-Kamus database. Basically, the database must be consisted of two main language components, namely Malay and Arabic. Each entry of new data will include its translation entry next to it.

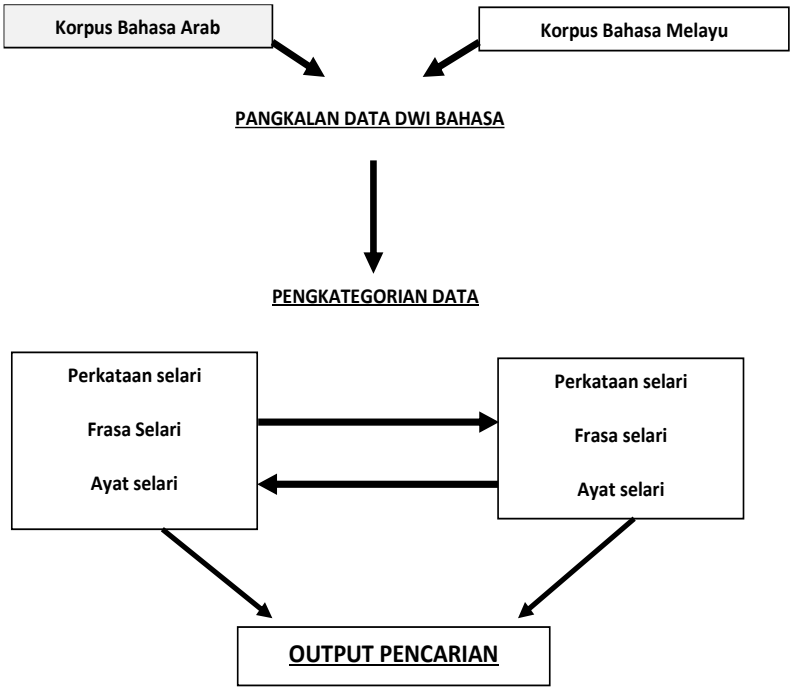

Figure 1. An overview of the e-Kamus data storage design

Software Users

This software was framed in such a way that it would be easy to be operated by the public users, researchers and language analysts, built as a web-based so that it could be accessed easily at all times and places. Divided into two forms of access. The first access is for the public, especially school and university students. In this access, students can enter a keyword in the search box.

The software will display the keyword in the form of a sentence in Malay and its Arabic translation in a concordant manner. Students can also view the translation in the forms of word level (*mu'jami*) and sentence level. The second access is specifically for language researchers, language experts and skilled translators. In this access, they are allowed to add

in relevant data and add in recommended translations based on parallel and comparable methods. They will be given a briefing first before using the software.

Data Sources

The data managed to be recorded and stored in the e-Kamus database include 6660 entries which equal to 222 pages. There are two types of sources, namely the source from parallel translation data taken from selected Quranic verses of *Surah al-Baqarah* and *al-Sajdah*. *Hadis pilihan himpunan Hadis 40* by Imam al-Nawawi, selected phrases and articles taken from students' assignments that referred to student dictionary such as Oxford's Arabic-Malay-Arabic dictionary arranged by Abdul Rauf Hj. Hassan and the data from the Primary School (KSSR) Year 2 to 6 Arabic Language textbook translated into Malay. These parallel data will be processed (input) to be included into the corpus data based on parallel method where it will be coded in pair with its translation texts according to the word, phrase and sentence levels. The second source is comparable data taken from bernama.com online open corpus, focusing on the translation assignments of the students who conducted a study on comparable data between Arabic and Malay texts.

Website and Display Function

The draft of the software that had been successfully developed was placed on the UPSI's website at the address: http://computing.upsi.edu.my/~zahar . Users may test this software by going to the above address. After the e-Kamus software has been successfully developed using PHP, MySQL Database and Linux-based Web Server System. The following is the display that has been given the function to generate e-Kamus based on five usage requirements of the software:

- A display of translation box TRANSLATE'
- A display of ' LOGIN'
- A display of ' REGISTER'
- A display of ' MALAY-ARABIC DATABASE'
- A display of ' HOME'
- A display of ' LOGOUT'

The display that serves the purpose of entering data is Register Display built to place text data with parallel and comparable translation together with word classification.

Data Entry Process

The e-Kamus website development has provided a specific site for entering and subsequently storing data sources. This site has provided a special room for sentences and the translations. Each sentence will refer to its keyword and root word. To enter data, the software operator will be given an ID username and password by the owner of e-Kamus and all he needs to do is log in to the software to perform entering, altering and removing of the data in the base. The LOGIN display is at the upper left of the software homepage as shown in Figure 2 below.

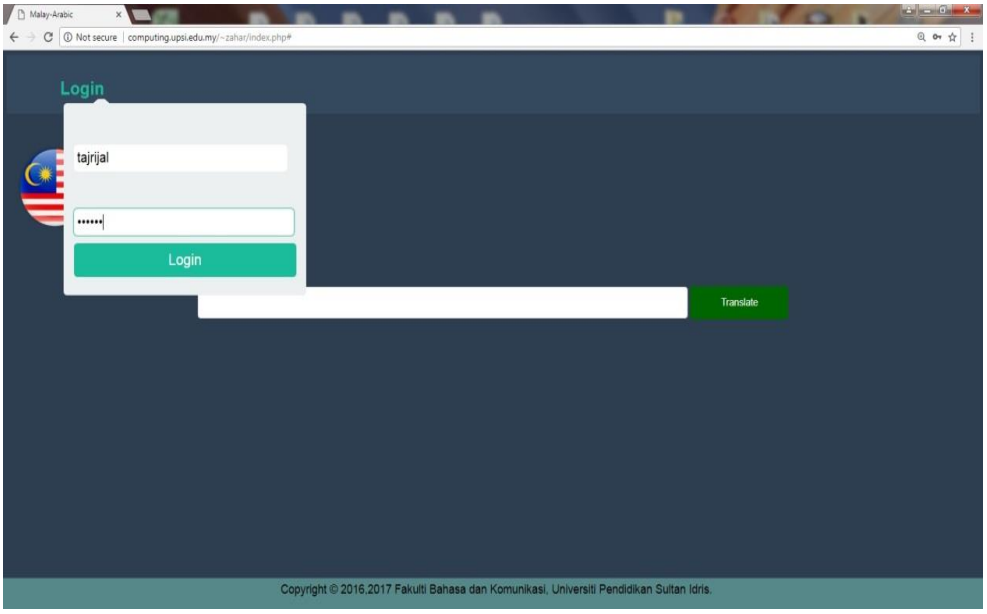

Figure 2. Login Display

Via the ' Register' display, the operator is fully responsible of entering the data using the boxes provided. The boxes are the break down of word classification with translation as the following:

i. list box of words, phrases, clauses or sentences in Malay
ii. list box of translations in Arabic
iii. keyword box for the words selected as search keyword in Malay (such as KK for Verb and KN for Noun)
iv. list box of word classification in Arabic (such as *Ism*, *fi'il Madhi* and *fi'il Mudhari'*)
v. list box of the root word of the word selected from a phrase or sentence

The operator only has to select a phrase or sentence that needs to be entered. Type in the Malay phrase or sentence in the ' Malay' box and then the parallel or comparable translation, and if there is no translation, the operator needs to translate it himself. After having chosen the best translation, the operator will select a keyword and type of word in Arabic. The main thing is to choose the root word of the sentence or phrase. If this is not done, the software will not be able to record the data.

The following Figure 3 shows a rough draft of the data entry process into the corpus software.

Figure 3. Data entry process

Data Classification

Word classification in text data is a method used to explain to users, especially to students and the general public who are having trouble identifying the Arabic word class, particularly in long texts. The word classification is divided into three main classes of Arabic words which are *ism*, *fi`il* and *harf*. *Ism* is divided into *ism `alam* and *ism mushtaq*. *Ism mushtaq* is broken down into *masdar*, *ism al-fa`il*, *ism al-ma`ful*, *ism al-zaman* and *ism al-makan*. *Fi`il* is divided into three main components, namely *fi`il al-madhi*, *fi`il al-mudhari'* and *fi`il al-amr* (Abdul Hamid, 1965), (Hassan, 1985) and (Ghalayini, 1987). While *harf* is in the harf component only. The *Ism* and *fi'il* parts are shown in Table 1 below.

Table 1
The Ism and fi'il parts

حرف	فعل	اسم	
	ماض	اسم علم	
	مضارع	مصدر	اسم المشتاق
	أمر	فاعل	
		مفعول	
		زمان	
		مكان	

Every recorded data will be classified based on the part entered in the 'option' box in Figure 4 below. If the text data has more than one word class, the operator must repeat a new record entry based on other classes. In Figure 4 below, the break down of the word classes for *masdar*, *ism al-fa`il*, *ism al-ma`ful*, *ism al-zaman*, *ism al-makan*, *fi`il al-madhi*, *fi`il al-mudhari'* and *fi`il al-amr* are directly placed in the option box to facilitate the entry process.

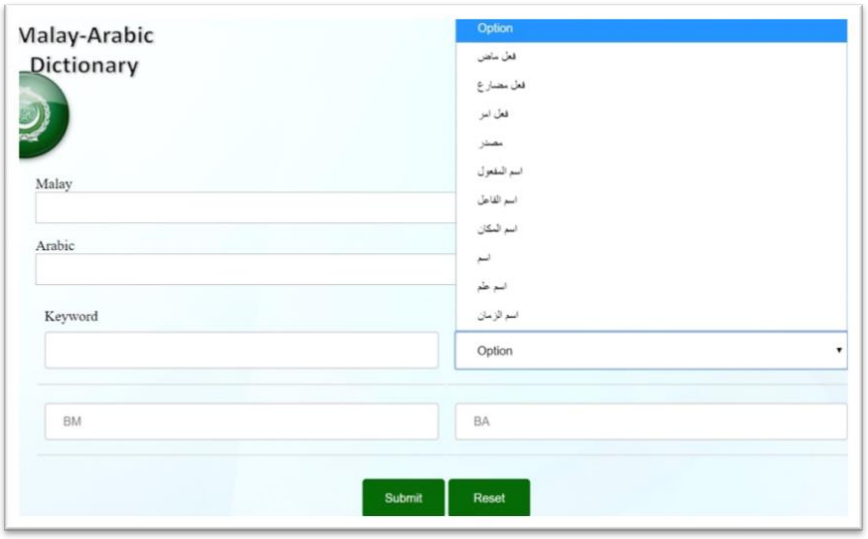

Figure 4. The 'Option' box for word classes

The display of word classes record

Each data recorded by word class can be checked by testing the ' Translate' translation search box. The following Table 2 shows the examples of data that have been stored in the e-Kamus database, the sentences that are based on the verb' lihat' . The sentences or phrases that have the word ' lihat' are displayed along with the translations, root words and types of word classes.

Table 2
Data display of 'lihat' in the e-Kamus database

Kamu melihat تَنْظُرُ	mu melihat	تَنْظُرُ	فعل مضارع
Dengan apa yang anda lihat? بِمَاذَا تَنْظُرُ ؟	anda lihat, kamu melihat	تَنْظُرُ	فعل مضارع
Saya melihat أَنْظُرُ	Saya melihat	أَنْظُرُ	فعل مضارع
Saya melihat dengan mata أَنْظُرُ بِالعَيْنِ	Saya melihat, memandang	أَنْظُرُ	فعل مضارع
Umar melihat burung atas pokok عمر يَرى العُصْفُورَ فوق الشَّجَرَةِ	melihat	يَرى	فعل مضارع
Jom kita lihat waktu-waktu makan هيَّا نَنظُر إلى مواعيدِ الطَّعام	kami melihat	نَنظُر	فعل مضارع
Sila lihat muka surat انْظُرْ إلى الصَفْحَةِ	lihatlah	انْظُرْ	فعل امر
Pelajar itu melihat buaya ينظُرُ التِّلميذ إلى التِّمْسَاح	dia melihat	ينظُرُ	فعل مضارع
Mereka berdua melihat رأَيَا	dia melihat	رَأى	فعل ماض
Lihatlah (kamu semua) kepada انظُرُوا	Kamu semua lihatlah	انظُرُوا	فعل امر
saya sedang melihat أَنْظُرُ إلى	Saya melihat	أَنْظُرُ	فعل مضارع

saya sedang melihat wajah saya dengan cermin	saya melihat	أنظُرُ	فعل مضارع
أنظُرُ إلى وجهِي بالمِرْآة			
untuk melihat	penglihatan	نَظْرٌ	مصدر
للنّظر إلى			
manusia menggunakan aku untuk melihat diri mereka	penglihatan	نظر	مصدر
يستخْدِمُنِي الناس للنظر إلى أنفسهم			

The word classes in Table 2 above are placed at the right side of the screen display, after root words. As in the last two rows of the table, the class of the word is masdar and the root word is nazr.

Word Entry Process

The following Table 3 shows the examples of how each word class in a complete sentence is listed in the e-Kamus. Each entry represents each of the word class listed. Each listed entry is accompanied by a complete sentence along with its translation.

Table 3
Examples of word classes in sentences

Record 1 – root word **اصْطَفَّ** Pelajar-pelajar **berbaris** bersebelahan kelas sebelum memasukinya untuk mendengar saranan daripada pengetua **Word class:** **فعل مضارع**	يَصْــطَفُّ الطُّلاَّبُ بجانبِ الفصــلِ قبل دخولِهَا لِيَسْتَمِعُوا إِلى التوجيهاتٍ من المديرِ Berbaris **يَصْطَفُّ**
Record 2 – root word **طالب** **Pelajar-pelajar** itu berbaris bersebelahan kelas sebelum memasukinya untuk mendengar saranan daripada pengetua **Word class:** **اسم**	يَصْــطَفُّ الطُّلاَّبُ بجانبِ الفصــلِ قبل دخولِهَا لِيَسْتَمِعُوا إِلى التوجيهاتٍ من المديرِ Pelajar-pelajar **الطُّلاَّبُ**
Record 3 – root word **توجيه** Pelajar-pelajar itu berbaris bersebelahan kelas sebelum memasukinya untuk mendengar **saranan** daripada pengetua **Word class:** **مصدر**	يَصْــطَفُّ الطُّلاَّبُ بجانبِ الفصــلِ قبل دخولِهَا لِيَسْتَمِعُوا إِلى التوجيهات من المديرِ Saranan **التوجيهات**

The data that have been successfully recorded and stored must be checked in the database. It can be tested by typing in the root word then check whether the vocabulary is listed in the database as shown in Figure 5 below.

Malay	Arabic	Root words	Delete	Update
Kamu melihat	تَنْظُرُ	Kamu melihat تَنْظُرُ		
Dengan apa yang anda lihat?	بِمَاذَا تَنْظُرُ ؟	anda lihat, kamu melihat تَنْظُرُ		
Saya melihat	أَنْظُرُ	Saya melihat أَنْظُرُ		

Figure 5. The display of the recorded data checks

The operator may add, correct and remove the saved data using the ‘ edit’ and ‘ delete’ buttons on the right side of the display.

Conclusion

This study is hoped to be the first step towards the creation of a larger Malay-Arabic-Malay corpus database software model based on sentence and its translation. A good software should be developed in such a way that it is easy to be operated by school students, researchers, language analysts and the general public. Developed as a web-based so that it is easy to be accessed at all times and places and is designed in such a way that translation search results are in the form of sentences, clauses, phrases and vocabularies. The software should provide a link to the full text in the form of an option. At the same time, provide an opportunity for linguists and skilled translators to add relevant data and add translation suggestions based on parallel and comparable methods.

Acknowledgements

Part of the text translation work in this database is the effort of the students taking the Arabic-Malay Translation Application course at UPSI which is also an assignment that is required to be completed before the final exam.

References

Hariss, A. (2004). *PHP 5/MySQL Programming*. Boston: Premier Press.

Schwartz, B**.** (2012). *High performance MySQL*. US: O'Reilly Media.

Ya'kub, E. (1985). *Al-ma'ajim al-lughawiyah al-arabiyah*. Beirut: Dar al-Malayin.

Guidere, M. (2002). *Toward corpus based machine translation for standard arabic*. Translation Journal. Vol.6. no. 1.

Hutchins, J. (2009). *Multiple uses of machine translation and computerised translation tools*. International Symposium on Data and Sense Mining, Machine Translation and Controlled Languages ISMTCL

Julie, C. M. (2003), *PHP essential, 2nd edition,* Boston Premier Press.

Thomson, L. (2016). *PHP and MySQL web development*. US: Pearson Education.

Abdul Hamid, M. (1965). *Syarh Ibn Aqil*, Jld II. Beirut: Dar al-Kitab al-‘ Arabiy.

Jabbar, M. A. (1977). *Arabic loan-words in Malay*. Bangi: UKM.

Mahmud Ismail, M. Z. (1994). *Al-nizam al-nahwiy fi al-lughat al-arabiyyah wa al-maliziyyah: dirasat fi al-tahlil al-taqabuliy*. Mesir: Univ. Iskandariah.

Mustafa Ghalayini al-Syikh. (1987). *Jami' al-durus al-arabiyyah*, Jld 2,3. Beirut: Maktabat al-Asriyyat.

Olahan, M. (2004). *Introducing corpora in translation studies*. London: Routledge.

Osman, H. K. (2006). *Kamus Besar Dewan Arab Melayu*. DBP. Kuala Lumpur

DuBois, P. (2006). *MySQL cookbook. solutions for database developers and administration 3th edition*. US: O' Reilly.

Love, R. (2013). *Linux system programming*. US: O'Reilly Media.

Nixon, R. (2014). *Learning PHP, MySQL, JavaScript, CSS & HTML5: A step-by-step guide to creating dynamic websites* (3rd Edition). US: O'Reilly Media

Abdul Ghani, R. & Mohamed Husin, N. (2001). *Yang selari dan yang setanding: peranan korpus dalam penterjemahan*, Dalam Kertas Kerja Persidangan Penterjemahan Antarabangsa Ke-8, Langkawi, Kedah.

Sinclair, J. (2005). Corpus and text. *Basic principles in developing linguistic corpora: a guide to good practice*, ed. M. Wynne (pp. 1-16). http://www.ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm. Oxford: Oxbow Books.

Steven, G. & Steve, S. (2009), *Linux Administration A beginner' s guide, fifth edition*, McGraw Hill & Osborne.

Hassan, T. (1985). *Al-lughat al-arabiyyah ma' naha wa mabnaha*. Mesir: al-Haiah al-Misriyyah al-'Ammah Lil al-Kitāb.

Tengku Mahadi, T. S., Vaezian, H. & Akbari, M. (2010). In *Design and development procedure of an English-Malay parallel corpus*, Universal Corpora for Comparative and Translation Studies (UCCT). UK: Edge Hill University.

Wan Rose Eliza, A. R. (2013). *Penggunaan teknologi dalam penterjemahan*. Asas Terjemahan dan Interpretasi. Pulau Pinang: Penerbit Universiti Sains Malaysia

Yang, Y. (2009). *Technology in a Changing World*. Technology and Language Dominance. Singapore Management University: GBI Books & Wee Kim Wee Center.

Zanettin, F. (1998). *Bilingual comparable corpora and the training of translators*. Meta: Translators' Journal, vol. 43, no. 4, p. 616-630. http://www.erudit.org/erudit/meta/v43n04/zanettin/zanettin.html