Journal of Theoretical and Applied Information Technology
10th December 2015. Vol.82. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

WORD SENSE DISAMBIGUATION BASED ON YAROWSKY APPROACH IN ENGLISH QURANIC INFORMATION RETRIEVAL SYSTEM

[1] OMAR JAMAL MOHAMED, [2] SABRINA TIUN

Faculty of Information Science and Technology
E-mail: [1]omar_jamal_m@yahoo.com, [2]sabrinatiun@ftsm.ukm.my

ABSTRACT

Word sense disambiguation (WSD) is the process of eliminating ambiguity that lies on some words by identifying the exact sense of a given word. In the natural languages, many words could yield multiple meaning based on the context. WSD aims to identify the most accurate sense for such cases. In particular, when translating one language to another, there would be a possibility to tackle an ambiguity among the translated words. Quran, which is the holy book for approximately billion Muslims, has been originally written in Arabic language. Apparently, when translating Quran to English language, several semantic issues have been caught by researchers. Such issues lies on the ambiguity of words such as '*ليلا ونهارا*' and '*يوم الحساب*', which are translated into 'day and night' and 'judgment day'. Such ambiguity has to be eliminated by determining the exact sense of the translated word. Several research efforts have been intended to disambiguate the sense of translated Quran. However, the process of identifying an appropriate method for WSD in translated Quran is still challenging task. This is due to the complexity of Arabic morphology. Hence, this study aims to propose an adaption for Yarowsky algorithm as a WSD method for Quranic translation. In addition, this study aims to develop an IR prototype based on the proposed adaption method in order to evaluate such method based on the retrieval effectiveness. In fact, the dataset that has been used in this study is a collection of Quranic content. Several pre-processing tasks have been performed in order to eliminate the irrelevant data such as stop-words, numbers and punctuation. Sequentially, two lists of senses for each ambiguity word will be created with their context. This would be performed in order to let the Yarowsky algorithm train on such example set. After that, a decision list will be constructed by the Yarowsky algorithm, which depicts the labelling sense of each word. The evaluation method that has been used in this study is the three IR evaluation metrics; Precision, Recall and F-measure. The experimental results have shown a 77% of f-measure. Such result seems to be weak in compared to the results of Yarowsky that have been applied in open domain. This is due to the lack of examples that could be extracted from Quran for both senses. Meanwhile, such result seems to be competitive in WSD of Quranic translation. Finally, it can be concluded that WSD has a significant impact on the IR system by improving the retrieval effectiveness.

Keywords: *Word Sense Disambiguation, Yarowsky Algorithm, Information Retrieval, Natural Language Processing, Quran*

1. INTRODUCTION

Languages have several kinds of ambiguity where many words can be comprehended in various aspects based on certain contexts [1]. For example, the word back in '*back home*' and '*my back*' has two meaning; the first refers to coming home and the second refers to a part of the body. Such ambiguity can be illustrated easily by the common sense of the human, while machine cannot distinguish the difference. Instead, it requires more complex computation and analysis in order to recognize the meaning, this process called Word Sense Disambiguation (WSD) [2]. Hence, assume the previous example has been processed by WSD; it would be undergone tag annotation process as '*back*/ COME *home*/ HOME. Such tag annotation process aims to provide clue for each word in order to identify the meaning [3]. In fact, WSD significantly depends on a knowledge base which considered the crucial part of it. Knowledge sources have various forms such as corpora (collection of text), dictionaries and semantic networks. Without knowledge, it would be impossible for both humans and machines to identify the meaning. On other hand, Language plays an essential role in the field of WSD where translation between languages usually causes ambiguity for example; the Italian

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

word *penna* can be translated in English as feather, pen, or author depending upon the context [4].

The Quran is the holly book of the followers of Islam (also known as Muslims). It contains teaches of the Islamic religion. The scripts of the Quran are vowelized to prevent reciters (readers) from misspelling a word, and hence changing its meaning, which tremendously changed by the vowels. However, most researches in the field of Arabic Information Retrieval (IR) did not pay much attention to the problem of searching and retrieving vowelized text. Most accomplished works even suggested removing the diacritics in the preprocessing step and unifying the content of the inverted list [5].

Current research in information retrieval (IR) primarily relies on processing tokens in text, without performing any deeper semantic analysis. Recently, utilizing natural language processing has a significant impact on enhancing the performance of information retrieval in terms of analyzing the semantic of the text. Hence, this study aims to propose an adaption for Yarowsky algorithm as a WSD method for Quranic translation. In addition, this study aims to develop an IR prototype based on the proposed adaption method in order to evaluate such method based on the retrieval effectiveness.

2. RELATED WORK

There are many approaches have been presented for word sense disambiguation for instance, Ahmed & Nurnberger [6] have proposed a WSD approach for machine translation from Arabic to English using a large parallel corpus. Basically, the proposed approach consists of two phases; first phase aims to utilize Natural Language Processing (NLP) techniques in order to analyse Arabic morphology. In fact, AraMorph toolkit which introduced by Hajic et al. [7] has been used in this phase in order to eliminate the inflectional derivation from the words. After that, each word will be measured based on its occurrence, so that similar words with different sense can be identified using the most likely method based on the occurrence. Then previous information will be used as a training dataset for Naïve Bayes classifier in order to classify word sense. In the same manner, Ahmed et al. [8] have proposed a WSD method based on statistical using parallel corpora for translating query from Arabic to English. In fact, the proposed method handle any query by translating it using AraMorph toolkit. Apparently, if the query contains ambiguous word, the proposed method will perform a matching process between any word from the query and the words from Arabic corpora using string-similarity measure. Finally, the proposed method will associate the ambiguous word with the most appropriate sense from the training data based on occurrences.

On other hand, Merhbene et al. [9] have proposed an unsupervised WSD method for Arabic using context-matching approach which aims to measure the semantic coherence score between words. The authors have firstly built the corpus by collecting online documents that contain ambiguous words. Then they have created a dictionary contains all clues for the ambiguous words. Secondly, they have applied TF-IDF on each word from the corpus in order to identify its importance. Finally, they applied a context-matching algorithm in order to match each word with the most likely sense. Moreover, Hammo et al. [10] have proposed a novel retrieval method from Quran based on WSD. In fact, the proposed method has been adopted to facilitate the process of query expansion from Quran. The authors have used WordNet as a synonym dictionary in order to match query words with its potential sense.

Eventually, Shoaib et al. [11] have proposed a relational model for semantic search of Quran. Basically, the authors started with analyzing the query itself in order to generate possible synonyms, in other word, the verse that has been queried will be match with corresponding or relevant verses. This has been performed by using WordNet as a synonym dictionary. Then, the authors have created a tree relational model in order to classify each terms with its level as child or parent node. This has been performed in order to distinguish between terms and topics in Quran which have different levels of semantic.

Tiun et al. [12] have addressed the role of WSD in terms of retrieving English translated Quranic verses. Basically, the authors have used Yarowsky algorithm as an unsupervised machine learning in order to provide sense for ambiguous words. The authors have used a dataset of translated Quranic verses that has been introduced by Kassis et al. [13]. However, Yarowsky algorithm has been used in this article with training set of data by identifying manually the sense of each word. Using the semantic and syntactic features with Yarowsky could lead to minimize the dependency on the training for identifying the sense. Ali et al. [14] have addressed the difficulties that lie on the semantic and syntactic aspects when translating Quran from Arabic to English. Tense is an obvious

Journal of Theoretical and Applied Information Technology
10th December 2015. Vol.82. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

syntactic problem that translators usually encounter in translating the Holy Quran.

3. PROPOSED METHOD

The proposed method consists of six main phases as shown in Fig 1 which are *Corpus*, *Preprocessing*, *Creating sense lists*, *Applying Yarwosky algorithm*, *Developing an IR prototype* and *Evaluation*. Corpus phase describes the process of the data collection and the type of data that has been used in this study. Whereas, preprocessing phase aims to illustrate the tasks that have been took a place before carrying out the proposed method including eliminating the unwanted and irrelevant data. While creating sense lists phase aims to construct the example training set in order to let the proposed method train upon it which leads to extract patterns. Applying Yarwosky algorithm phase aims to carry out the proposed method in order to classify word's sense based on the training example set. Furthermore, an Information Retrieval prototype will be developed in order to evaluate the proposed method. Finally, evaluation phase which aims to evaluate the proposed method based on the common information retrieval metrics; Precision, Recall and F-measure.

3.1. Corpus

The dataset that has been used in this study is The Quran Concordance which has been introduced by Kassis et al. (1983). This dataset contains all the indexing English words in Quran translation. Table 1 shows the description of such corpus.

Table 1. Corpus Description

Details	Number
Details	Number
Capacity	7.98 MB
Number of pages	1485

3.2. Preprocessing

The pre-processing phase aims to turn the data into an appropriate form by removing the irrelevant data such as punctuation (*e.g. #$*%&*), numeric data (*e.g. 23493*) and stop-words (*e.g. to, of, in,* etc.). This irrelevant data hinders the processing of text so that it has to be eliminated. In addition, preprocessing phase aims to tokenize the data, in other word, dividing words from text into clusters of consecutive morphemes, one of which typically corresponds to the word stem.

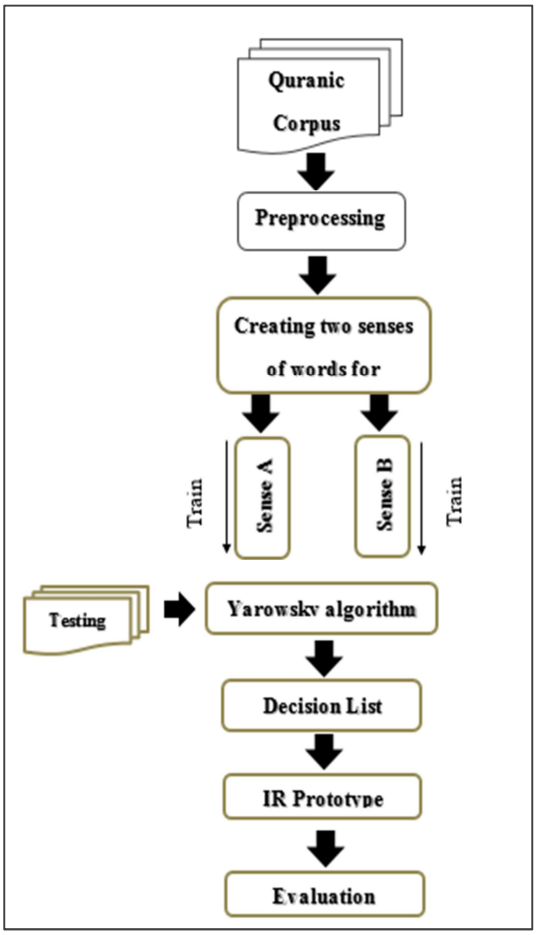

Figure 1. Proposed Method Phases

3.3. Creating Two Senses lists For Training

This phase aims to identify a small number of seed collocation representative of each sense. In this phase, two lists of each word that could yield two senses will be created. This has been performed by identifying multiple examples for a word with sense A, and multiple examples for the same word with sense B. Such procedure is crucial due to its role of training the Yarowsky algorithm. Table 2 shows a sample of example set for both senses the word 'Day'.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2. Example Of Generating Two Senses Of Giving Word

The word 'Day'	
Sense A	**Sense B**
Covers the **day** with the night which is in haste to follow it	The **Day** the Trumpet is blown and you come in droves.
And a sign for them (human beings) is the night. We strip it of the **day** and they are in darkness	An unbeliever will be brought on the **Day** of Resurrection and will be asked
Hast thou not seen how God merges the night into the **day** and merges the **day** into the night	On the **Day** the sky is like molten brass and the mountains like tufts of colored wool

As shown in Table 2, the word 'Day' has to senses; first, it refers to the sequence of day and night, second it refers to the judgement day. In fact, in the sense A, all the verses that contain the word 'Day' with the meaning of day and night will be inserted to a list. As well as, all the verses that contain the word 'Day' with the meaning of judgment day will be inserted to another list. Hence, Yarowsky algorithm can be adopted to train on such lists.

3.3.1 Training Mechanism

The training mechanism that Yarowsky algorithm would follow can be illustrated as addressing the neighboring words. First, it aims to address the two words; previous and forward words that surround the targeted word (required word in terms of sense identification). If the algorithm fails to identify the sense of given word based on the previous and forward words, an extension of the sequence of words will be addressing for both the previous and forward sides. The number of words that would be addressed is ranged from 2-10. Table 3 depicts this task.

Table 3. Mechanism Of Training

No. of neighbor words	Mechanism
One word	Previous + **Target Word** + Forward
Two words	Previous + Previous + **Target Word** + Forward + Forward
Three words	Previous + Previous + Previous + **Target Word** + Forward + Forward + Forward
Four words	Previous + Previous + Previous + Previous + **Target Word** + Forward + Forward + Forward + Forward
Five words	Previous + Previous + Previous + Previous + Previous + **Target Word** + Forward + Forward + Forward+ Forward + Forward
Six words	Previous + Previous + Previous + Previous + Previous + Previous + **Target Word** + Forward + Forward + Forward+ Forward + Forward + Forward
Seven words	Previous + Previous + Previous + Previous + Previous + Previous + Previous + **Target Word** + Forward + Forward + Forward+ Forward + Forward + Forward + Forward
Eight words	Previous + Previous + Previous + Previous + Previous + Previous + Previous + Previous + **Target Word** + Forward + Forward + Forward+ Forward + Forward + Forward + Forward + Forward

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Nine words	Previous + Previous + Previous + Previous + Previous + Previous + Previous + Previous + **Target Word** + Forward + Forward + Forward+ Forward + Forward + Forward + Forward + Forward + Forward
Ten words	Previous + Previous + Previous + Previous + Previous + Previous + Previous + Previous + Previous + **Target Word** + Forward + Forward + Forward+ Forward + Forward + Forward + Forward + Forward + Forward + Forward

3.4. Applying Yarowsky Algorithm

Yarowsky algorithm is a semi-supervised machine learning technique which adopted to train on a small portion of example set [15]. Apparently, Yarowsky algorithm is totally relying on the predefined examples that have been constructed in the Training phase where a small number of seed collocation representative of each sense have been identified for each word. Then a decision list will be created by the algorithm in order to label each word with the exact sense. The process of labeling each word can be described in the following equation:

$$Sense\ (w_i) = log\left(\frac{\Pr(Sense_A|\ Collocation_i)}{\Pr(Sense_B|\ Collocation_i)}\right)$$

In fact, Yarowsky works by identifying a value for each word in a giving sentence. For instance, assume a sentence S with a sequence of words w={w_1,w_2,…,w_3}. One of these words will be the targeted word that Yarowsky aims to identify its sense, this word called T. So that, Yarowsky will assign a value V for the other words. Such value refers to the degree of multiple senses which can be determined as follow:

$$Sense\ (W_T) = Max\big(V(W_i)\big)$$

In order to illustrate the above equation, given a verse of 'night perpetual day judgment god enlightenment hearken', the target word T that intended to provide its sense is day where it yield two meanings which are 'day and night' and 'judgement day'. Therefore, Yarowsky will assign each word with a value V. This value indicate the rate of sense. Table 4 shows each word with its sense.

Table 4. Assigning Values For Each Word By Yarowsky

Word	Sense	Value
night	A	3.2
Perpetual	B	0.2
Judgment	B	3.7
God	B	0.5
Enlightenment	A	2.2
hearken	B	0.3

As shown in Table 4, each word has been annotated with a value that refers to the sense. Hence, Yarowsky considers the maximum value as the sense of word. The maximum value was obtained by the word 'Judgement' by achieving 3.7 thus; the sense of the word would be B according to the highest value. Although this process of selection is useful but in this case the selection was incorrect where the actual sense is A. This is due to Quran commonly contains various semantics in the same verse. Therefore, our study suggest an adaption that may enhance the results. Such adaption can be viewed as a summation process rather than selecting the maximum. Based on the previous example, the values of sense A will be sum as well as the values of sense B. Hence, the highest value would be an indication for the sense as follows:

$$Sense\ A = 3.2\ night + 2.2\ Enlightenment = 5.4$$

$$Sense\ B = 0.2\ Perpetual + 3.7\ Judgement + 0.5\ God + 0.3\ hearken = 4.7$$

As shown in the above equations, Sense A has achieved a value greater that Sense B thus; the sense of word day will be assigned as A.

3.5. IR Prototype

In order to evaluate the use of Yarowsky algorithm, a prototype has been implemented to evaluate the effectiveness of retrieval. In fact, the prototype enables the user to enter a Quranic verse and then retrieve the relevant verses based on the sense of words. The prototype utilize the Yarowsky algorithm by generating the decision list in order to identify the sense of words from the typed verse (that has been entered by the user). Then based on the sense, the relevant verses will be retrieved. For example, suppose the user entered a Quranic verse as the following:

'Those who (in charity) spend of their goods by night and by day, in secret and in public, have their reward with their Lord: on them shall be no fear, nor shall they grieve'.

The word 'day' has a sense of 'day and night' therefore; the prototype will retrieve all the verses that contain the word 'day' with the same sense.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 5 shows the results of retrieving verses that relevant to the previous verse.

Table 5. Sample of Prototype results (Sense A)

No.	Sense of 'day and night' verses
1.	Thou causest the night to gain on the **day**, and thou causest the day to gain on the night; Thou bringest the Living out of the dead, and Thou bringest the dead out of the Living; and Thou givest sustenance to whom Thou pleasest, without measure.
2.	A section of the People of the Book say: \"Believe in the morning what is revealed to the believers, but reject it at the end of the **day**; perchance they may (themselves) Turn back
3.	Or else did they feel secure against its coming in broad **day**light while they played about (care-free)
4.	They celebrate His praises night and **day**, nor do they ever flag or intermit
5.	And He it is Who makes the Night as a Robe for you, and Sleep as Repose, and makes the **Day** (as it were) a Resurrection

As shown in Table 5, the prototype has retrieved all the verses that contain the word 'day' with the sense of 'day and night'. However, suppose the user has entered a verse as follow:

"The day it arrives, no soul shall speak except by His leave: of those (gathered) some will be wretched and some will be blessed"

In such verse the word 'day' has different sense which is 'judgement day' thus, the prototype will retrieve all the verses that contain the word 'day' with the same sense as shown in Table 6.

Table 6. Sample Of Prototype Results (Sense B)

No.	Sense of 'judgement day' verses
1.	There is not a population but We shall destroy it before the **Day** of Judgment or punish it with a dreadful Penalty: that is written in the (eternal) Record.
2.	And the curse shall be on thee till the **day** of Judgment
3.	Master of the **Day** of Judgment
4.	Till the **Day** of the Time appointed
5.	He said: \"Give me respite till the **day** they are raised up

As shown in Table 6, all the verses that contain the word 'day' with the sense of 'judgement day' have been retrieved.

3.6. Evaluation

The common information retrieval evaluation metrics Precision, Recall and F-measure have been implemented to evaluate any information retrieval system [16]. It is computed using the contingency which measure the level of truth of certain hypothesis. Such table is represented as follow:

Table 7. Contingency table

	True	False
Positive	TP	FP
Negative	TN	FN

(TP) is the set of correct retrieved items, (FP) is the set of incorrect retrieved items, (TN) is the set of truly non-retrieved items, and (FN) is the set of falsely non-retrieved items. Based on the previous explanation, the equations of Precision, Recall and F-measure are calculated as follows:

$$Precision\ P = \frac{|TP|}{|TP| + |FP|}.$$

$$Recall\ R = \frac{|TP|}{|TP| + |FN|}.$$

$$F = \frac{2 \times |TP|}{(|TP| + |FN|) + (|TP| + |FP|)} = 2 \times \frac{P \times R}{P + R}.$$

4. RESULTS

In fact, the evaluation has been performed using four words which are *day*, *charity*, *pray* and *right*. In addition, the evaluation has been done using two methods; the basic Yarowsky algorithm and the proposed adaption of Yarowsky. As mention earlier, the effectiveness of retrieval has been measured using precision, recall and f-measure. Table 8 depicts such results.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 8. Experimental Results

Words	Method	Precision	Recall	F-measure
Day	Yarowsky	0.81	0.41	0.47
	Adaption of Yarowsky	0.96	0.54	0.65
Charity	Yarowsky	0.72	0.6	0.62
	Adaption of Yarowsky	0.8	0.77	0.77
Pray	Yarowsky	0.56	0.33	0.37
	Adaption of Yarowsky	0.66	0.36	0.39
Light	Yarowsky	0.81	0.41	0.47
	Adaption of Yarowsky	0.90	0.45	0.50

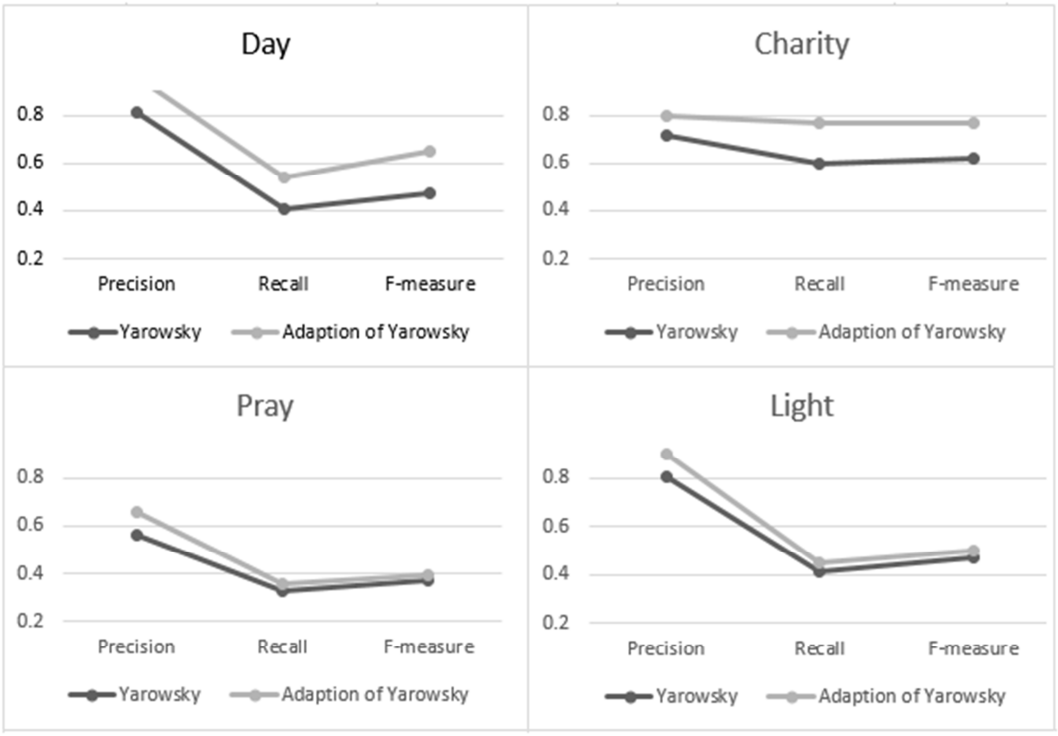

Figure 2. Comparison Between The Basic Yarowsky And The Proposed Adaption Of Yarowsky

As shown in Table 4.8 and Figure 4.1, the highest results have been achieved by the word 'Charity' by obtaining 80%, 77% and 77% of precision, recall and f-measure respectively. Subsequently, the second highest results have been achieved by the word 'Day' by obtaining 96%, 54% and 65% of precision, recall and f-measure respectively. Afterwards, the word 'Light' has achieved 90%, 45% and 50% of precision, recall and f-measure. Finally, the lowest results have been achieved by the word 'Pray' by obtaining 66%, 36%, and 39% of precision, recall and f-measure respectively.

5. DISCUSSION

In fact, the experimental results have shown that the Yarowsky algorithm achieved low results compared to the results of Yarowsky that has been

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

applied in open-domain word sense ambiguity applications such as the study of Bergsma et al. [17]. This is due to multiple reasons. (i) First, the application of Yarowsky in this study has been performed for specific-domain which is Quran. This leads to limited examples that could be extracted from the verses which leads to reduce the training examples for Yarowsky. (ii) Second, the successive words in Quran, that have been analyzed in order to identify the sense of words, have a weak impact on distinguishing the sense of words. This is due to Quran use complex semantic. (iii) Third, Yarowsky algorithm requires that the two senses of a particular word should be the same in terms of syntactic. For example, the word day has two senses meanwhile; both of these senses are noun. This hinders the process of examining many words that yield senses that are syntactically different. However, our experimental results can be seen as competitive by comparing it with WSD approaches for Quran. This has been proved by the enhancement of retrieval that have been accomplished using Yarowsky algorithm in this study compared to other studies such as [12, 18].

6. CONCLUSION

This research has proposed an adaption of Yarowsky algorithm as a WSD method for Quranic English translated. The experimental results have shown that the proposed adaption of Yarowsky has enhanced the effectiveness of retrieval based precision, recall and f-measure. However, the main drawback of this study lies on the lack of training set that can be extracted from Quran due to its complex morphology.

7. ACKNOWLEDGEMENT

This research project is funded by Malaysian Government under research Grant ERGS /1/2013/ ICT07/UKM/03/1.

REFERENCES

[1] Roberto Navigli, "Word sense disambiguation: A survey," *ACM Comput. Surv.,* vol. 41, pp. 1-69, 2009.doi:10.1145/1459352.1459355Retrieved from.

[2] Mark Stevenson and Yorick Wilks, "Word sense disambiguation," *The Oxford Handbook of Comp. Linguistics,* pp. 249-265, 2003Retrieved from.

[3] Nancy Ide and Jean Véronis, "Introduction to the special issue on word sense disambiguation: the state of the art," *Computational linguistics,* vol. 24, pp. 2-40, 1998Retrieved from.

[4] Tim Berners-Lee, James Hendler, and Ora Lassila, "The semantic web," *Scientific american,* vol. 284, pp. 28-37, 2001Retrieved from.

[5] Tim Buckwalter, "Buckwalter {Arabic} Morphological Analyzer Version 1.0," 2002Retrieved from.

[6] Farag Ahmed and A Nurnberger, "Arabic/English word translation disambiguation approach based on naïve Bayesian classifier," in *Computer Science and Information Technology, 2008. IMCSIT 2008. International Multiconference on*, 2008, pp. 331-338.doi:Retrieved from.

[7] Jan Hajic, Otakar Smrz, Tim Buckwalter, and Hubert Jin, "Feature-based tagger of approximations of functional Arabic morphology," in *Proceedings of the Fourth Workshop on Treebanks and Linguistic Theories (TLT 2005)*, 2005, pp. 53-64.doi:Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.93.880&rep=rep1&type=pdf.

[8] Farag Ahmed, Andreas Nürnberger, and Marcus Nitsche, *Supporting arabic cross-lingual retrieval using contextual information*: Springer, 2011.

[9] Laroussi Merhbene, Anis Zouaghi, and Mounir Zrigui, "Ambiguous Arabic words disambiguation," in *Software Engineering Artificial Intelligence Networking and Parallel/Distributed Computing (SNPD), 2010 11th ACIS International Conference on*, 2010, pp. 157-164.doi:Retrieved from.

[10] Bassam Hammo, Azzam Sleit, and Mahmoud El-Haj, "Effectiveness of query expansion in searching the Holy Quran," 2007Retrieved from.

[11] Muhammad Shoaib, M Nadeem Yasin, UK Hikmat, M Imran Saeed, and Malik Sikandar H Khiyal, "Relational WordNet model for semantic search in Holy Quran," in *Emerging Technologies, 2009. ICET 2009. International Conference on*, 2009, pp. 29-34.doi:Retrieved from.

[12] Sabrina Tiun, Hamed Zakr, Masnizah Mohd, Norazlinda Zainal Abidin, and Ahmad Irfan Ikmal Hisham, "Word Sense Disambiguation for English Quranic IR System," 2013Retrieved from.

[13] Hanna E Kassis, Fazlur Rahman, and Fazlur Rahman, *A Concordance of the Qur'an*: University of California Press Berkeley, 1983.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[14] Abobaker Ali, M Alsaleh Brakhw, Munif Zarirruddin Fikri Bin Nordin, and Sharifah Fazliyaton Shaik, "Some Linguistic Difficulties in Translating the Holy Quran from Arabic into English," *International Journal of Social Science and Humanity,* vol. 2, p. 588, 2012Retrieved from.

[15] Mirella Lapata, Bayesian Word Sense Induction, WSD Supervised, and WSD Unsupervised, "Natural Language Understanding," 2014Retrieved from.

[16] Mikalai Yatskevich, "Preliminary evaluation of schema matching systems," 2003Retrieved from.

[17] Shane Bergsma, David Yarowsky, and Kenneth Church, "Using large monolingual and bilingual corpora to improve coordination disambiguation," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, 2011, pp. 1346-1355.doi:Retrieved from.

[18] Sarah Abdul-Ameer Mussa and Sabrina Tiun, "A Novel Method of Semantic Relatedness Measurements for Word Sense Disambiguation on English AlQuran," *Bulletin of Electrical Engineering and Informatics,* vol. 4, 2015Retrieved from.