Journal of Theoretical and Applied Information Technology
15th April 2017. Vol.95. No 7

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

QUESTION ANSWERING SYSTEM SUPPORTING VECTOR MACHINE METHOD FOR HADITH DOMAIN

[1] NABEEL NEAMAH, [2] SAIDAH SAAD

[1] Faculty of Information Sciences and Technology, UKM, Bangi, Malaysia
[2] Faculty of Information Sciences and Technology, UKM, Bangi, Malaysia
Email: [1] nabelnabel1987@yahoo.com, [2] saidahukm@gmail.com

ABSTRACT

Retrieving accurate answers based on users' query is the main issue of question answering systems. Challenges such as analyse the need of users' query and extract accurate answers from large corpus are increase the difficulty of developing effective question answering system. This work aims to enhance the accuracy of question answering system for hadiths using useful methods. Pre-processing methods like tokenization and stop-word removal is used to identify the main concepts of users' query. Answering processing methods and techniques like N-gram, WordNet, CS, and LCS are used to update and enrich the extracted concepts of users' query based on the formal representation of hadiths answers or documents. Support Vector Machine (SVM) and Name Entity Recognition (NER) methods are conducted to classify Hadiths documents based on relevant subjects and questions types in order to reduce the searching scope of answers documents. Documents in Hadith corpus are classified according to proposed question types, and related subjects as four main classes which are: when for pray, where for pray, when for fasting, and where for fasting. The SVM classification of documents is accomplished supporting NER methods to identify the places (where) and time (when) features that included in the documents. The proposed question answering system is tested using 132 Hadiths documents about Fasting and Pray that are selected from Al-Bukhari source. The findings revealed that the average answers accuracy using CS technique is 67%, the average answers accuracy using LCS technique is 66%, the average answers accuracy using combination of CS and LCS techniques is 70%, and the average answers accuracy using CS, LCS, and SVM is 80%. SVM enhance the system accuracy up to 10% more than using other methods without classification processes. The main contribution of this research is using SVM method to reduce searching scope of Hadiths documents based on various subjects and question types beside effective analysis of query need using NLP methods. SVM provides more accurate answers than extracting answers using only similarity techniques such as CS and LCS.

Keywords: *Question Answering System, Hadiths, Pre-processing, Answers Processing, SVM, NER.*

1. INTRODUCTION

Nowadays, there are large amount of information which is published through various sources like web pages and online books. These sources contain huge volume of information that is related to variant fields of topics i.e. education, religions, health, and sport [1,2]. Thus, there is difficulty in retrieving accurate information from the various sources to satisfy the users daily need of information. Thus, systems like information retrieval and Question Answering Systems (QAS) were developed to provide accurate answers and information based on users queries or investigations [1,3]. The main aim of QAS is to process and analyse the needs of users' queries or questions and retrieve accurate answers for these queries or questions. Users searching for specific information usually face difficulty in retrieving exact useful answers that reflect their questions. This is occurring due to fact that there are large number and amount of answers documents that published through various online sources like web pages, books, and articles [4,1]. Consequently, users spend searching time and efforts to retrieve their needs of information to support their daily activities. Furthermore, users may type their questions using lack or unsuitable concepts comparing to the standard concepts that are used in answers documents. These vague queries will reduce the opportunities of similarity between questions and answers, eventually retrieving answers for the users that are not useful [5].

In the context of developing QAS for Hadiths, non-Arabic speaking users face the difficulty of providing the right concepts based on English language due to their lack of knowledge of the Arabic language as well as familiarity with Hadiths concepts using English language. Furthermore, the large numbers of Hadiths that published via various sources such as internet and printed books could increase the difficulty of extracting the right Hadiths that match with users' need of finding specific Hadith related to a specific topic.

According to [6], and [7] there are three main processes of QAS which are; 1) pre-processing phase that involve many methods like tokenization and stop-word removal in order to extract the main concepts of users questions, 2) answering processes phase to analyse the needs of users questions based on the extracted concepts, in this phase there are many methods can be adopted such as N-gram and WordNet, and 3) classify the answers documents in order to manage the searching processes for answers. Documents classification is usually conducted using machine learning techniques such as Support Vector Machine (SVM). Thus, useful answers based on the questions can be retrieved effectively. The main aim of this research is to develop question-answering systems to extract accurate answers of hadiths documents based on users queries. The proposed QAS is based on two main phases; (1) identify and enhance the concepts of users' queries using suitable pre-processing and answers processing methods, and (2) classify the hadiths documents according to proposed question types, and related subjects using suitable matching learning method in order to reduce the searching scope of documents.

1.0 RELATED WORKS

This section presents overview of question answering methods, related works of question answering systems, and related works of question answering system for hadiths. The review of related works in this section is important to support the methods selection of proposed QAS in this work.

1.1 Overview of Question Answering Methods

According to [8], the questions or query analysis is the process of identifying user's needs of answers based on their proposed questions. There are two main processes involve the questions analysis which are; (1) extracting the main concepts of the typed questions that indicates the needs of the provided query, and (2) updating the extracted concepts by replacing it by more effective concepts that match the actual concepts which are used in the answers' documents. [9] and [10] mentioned that based on the type of query, the most suitable analysis methods can be identified. The restricted query is usually analysed depending on two parts which are the question type and the body of questions. For example, the question: "what is the capital city in Malaysia" is divided into two parts; (1) question of What type that indicate answers about objects like facts and places, and (2) body of question ("the capital city in Malaysia"). On the other hand, the open questions are analysed as one part or block. Bhaskar et al. [11] argued that whatever the question types, there are many processes and methods can be adopted to provide effective analysis of the provided query. Remove stop words can be used to remove the unimportant keywords such as 'and', 'the', and 'has'. Stop words are words which are filtered out before or after processing of natural language data (text) [12]. Tokenization is used to divides the text sequence into sentences and then the sentences into tokens [13]. So, in English language, words are bounded by whitespace and optionally preceded and followed by parentheses, quotes, or punctuation marks.

The N-gram method works on listing the given sentence (query) into single words or sequence of words [14]. The n-grams typically are collected from a text or speech corpus. An n-gram of size 1 is referred to as a "unigram"; size 2 is a "bigram" (or, less commonly, a "digram"); size 3 is a "trigram". Larger sizes are sometimes referred to by the value of n, e.g., "four-gram", "five-gram", and so on" [15]. "Name Entity Recognition" (NER) aims to identify the entities in text as objects like persons' objects and places objects [16].

All of the above methods are aim to analyse the main concepts the represent the need of user query. However, these concepts could be typed by users using weak or lack concepts that not matched with the formal concepts that used in answers documents. The query concepts could be replaced and enriched using ontologies. QAS conceptual ontologies are important to design prepare and structure the formal concepts

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

synonyms of QA domains. The WordNet is the most popular built-up ontologies that adopted by QAS. The WordNet considered as an effective build-up ontology that contains concepts and synonyms related to several subject and domains [17]. WordNet is developed based on English language and its main aim is to replace the concepts of users' query by more effective concepts or synonyms according to formal concepts that used in answers documents [18]. WordNet was developed as database that contains concepts and synonyms of many domains such as sport subjects. The relationships between query concepts and WordNet synonyms are extracted at words level (i.e. replace word by word) [19,20]. In other meaning, WordNet is not able to replace many words by one or many words. Replacement of word by another word is called SynSets. Each SynSets in WordNet may have many synonyms. Therefore, WordNet can give many replacement suggestions for any concept in the query and consequently extract the proposed answers based on the similarity of various suggestions. After the query analysis using several methods and concepts enrichment through QAS ontology, the similarity between users' query and document of answers can be conducted in order to extract the proposed answers for users' query. Cosine similarity (CS) and Longest Common Subsequence (LCS) techniques are used widely to measure the similarity between users' queries and system documents [21]. The techniques of CS and LCS increase the opportunity of extract accurate answers depend on users' questions after questions analysis.

a. Cosine Similarity (CS) is an effective technique for measuring the similarity between two texts based on the term weights in each text [22,21,19]. CS is used widely in the implementation of QA systems and information retrieval. Mainly, CS is applied to measure the similarity between the words or concepts that belong to two different texts, and consequently estimate the relationship strength between these two texts. CS is conducted depending on the value of angle cosine between two vectors or texts. The value of angle is 0 and cosine value is 1 when two vectors are completely same, and the value of cosine is 0 when the two vectors are completely different [20]. CS obtained the weight of the terms (wij, wlj) that belongs to the two texts before measuring the similarity between these two texts according to computed weights. The CS between two sentences (si and sl) can be computed using the following Equation 1:

$$cs\ (s_i, s_l) = \frac{\vec{s_i} \cdot \vec{s_l}}{||\vec{s_i}|| \cdot ||\vec{s_l}||} = \frac{\sum_{j=1}^{m} w_{ij} w_{lj}}{\sqrt{\sum_{j=1}^{m} w_{ij}^2 \cdot \sum_{j=1}^{m} w_{lj}^2}}, i, l = 1, \dots, n \qquad (1)$$

where $\vec{s_i} \cdot \vec{s_l}$ is the number of similar words between both sentences, the $||\vec{s_i}|| \cdot ||\vec{s_l}||$ is the total number of the weights of words in |si| and |sl| sentences.

b. Longest Common Subsequence (LCS) is a technique that is used to measure the similarity between two strings (i.e. sequence of terms) [23]. LCS takes the first string as input and starts looking for the strings that have the same pattern of input string. There are algorithms of LCS technique focus on find the exact pattern of input while other algorithms focus on find the most similar pattern of input [24]. Based on LCS processes, given String S= s1…sm and T = t1…tn, S is the subsequence of T if for each 1 ≤ j ≤ m, 1 ≤ i1 < i2< …< im ≤ n, sj = j i t. Given a set of sequences S+ = {S1, S2, …, Sk}, the LCS of S+ is the longest [23].

Furthermore, to increases the productivity rate of QAS, machine learning methods can be applied to manage and to extract the answers' documents effectively based on users' queries [25]. Using machine learning, it is possible to generate systems that includes thousands features of questions and documents and then classify these questions and documents based on these features automatically. Any text classification algorithm can be employed such SVM to classify the texts based on the purpose of information that included in this text [26, 27]. For example, the text that talk about places is referring to where questions and the text that talk about date and time refer to when questions and so on. SVM is a useful technique for data classification and it is easier to be implemented than other classification methods such as Neural Networks [28]. A classification task usually involves separating data into training and testing sets. Each instance in the training set contains one "target value" (i.e. the class labels) and several "attributes" (i.e. the features or observed variables). The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes. Given a training set of instance-label pairs (xi , yi), i = 1, . . . , l where

$x_i \in R^n$ and $y \in \{1, -1\}$, the support vector machines (SVM) [29, 30, 31] require the solution using optimization technique.

1.2 Related Works of Question Answering System

Liljenback [32] was developed QA system to provide answers that asked frequently by universities students in the context of their learning activities. The main aim of proposed system is to deliver accurate answers for students based on their daily needs of knowledge related to their learning. The answers documents of proposed QA system consist of large learning sources such as books and articles. The student's questions were analyzed using many NLP methods such as tokenization, stop word removal, and bi-gram. Liljenback [32] developed his own ontology to provide many searching synonyms depend on concepts of students' query. The proposed system was tested using many frequent questions of students involving about 100 students from many European universities. Then, the results of proposed system were compared with known system called. After that, [32] concludes that the proposed methods are effective for the queries of 16 concepts or less. The researcher recommended N-gram to retrieve accurate answers based on long questions. On the other hand, the proposed QA system of [33] focuses on extracting accurate answers that related to subject of weather forecasting. The users' can type their questions using search engine. There are three major methods involved in the proposed system of [33] which are as the following:

a. The shallow parsing method were used to perform the named entity tagger in order to identify the most suitable answers (i.e. candidates' documents) based on the labels of users' questions.

b. Normalizer: in this stage, there are many NLP methods were used to analyse and users' questions; tokenization, stop-word removal, and similarities techniques.

c. Inference rules: in this phase the relationships between concepts of users' query was analysed in order to generate the overall meaning of question rather than extract the documents based on the similarity matching with each concept. The experimental test of proposed QA system of [33] involves many standard queries about weather subject. The average of answers accuracy of proposed system was 90.9%, which considered as effective results.

Moreover, Shiyan et al. [34] compared between two ontologies of QA system, the first one is the WordNet and the second ontology is called SUMO. The comparison was conducted through apply each ontology on QALL-ME which is QA system for tourism domain. WordNet and SUMO are used to enhance the concepts of questions before extract the results. CS and LCS techniques were used to measure the similarity between proposed questions after concepts enhancements and answers documents. The significant results show that the use of WordNet provides accurate answers more than the use of SUMO. Furthermore, the installation and applying of WordNet is simpler than SUMO. In The same context, the using of WordNet is helpful to provide accurate answers better than QA system without WordNet ontology. In addition, Tian et al. [35] measured the similarity between proposed query and answers documents using CS and LCS techniques involving WordNet ontology. The significance results of Tian et al [35] shows that the CS is effective to measure the matching between the small size queries i.e. 1-16 keywords while the LCS is effective to measure the matching of large query size.

In the context of document classification based on questions types, Panicker et al. [36] applied SVM and KNN methods to classify the document according to answers based on questions types (i.e. document is answer about when question if this documents contains places indicators). The proposed methods were tested on system called TREC. The results show that the document classification is effective to retrieve accurate answers with lower time than QA system without classification processes. On the other side, Eka and Azhari [37] focused on the features that effect on the queries processing techniques in order to provide accurate answers at real time. Eka and Azhari [37] found that the machine learning methods such as SVM and Neural network are important to minimize the searching scope of answers for both QA domain types (open and restricted). Table 1 summarizes the related works directions.

Journal of Theoretical and Applied Information Technology
15th April 2017. Vol.95. No 7

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

Table 1: Summary of Related Works of QAS

Author	QAS Based Language	Method	Main Aim
Liljenback (2007) [32]	English	Pre-processing method such as bigram	Developed QA system that contains learning documents to provide accurate answers of the frequent students' questions.
Chung et al. (2004) [33]	English	Pre-processing methods such as Normalization and NER	provide answers about queries that related with weather forecasting issues
Tian et al. (2010) [35]	English	WordNet, CS, and LCS	Measure the similarity between users' query concepts and WordNet conceptual ontology through CS and LCS techniques.
Shiyan et al (2009) [34]	English	N-gram and WordNet	Update the concepts of users' queries before retrieve the answers
Panicker et al (2012) [36]	English	SVM and KNN	Classify the documents based on Questions types; when for time, where for places, and what for facts
Eka and Azhari (2012) [37]	English	SVM and Neural network	Large document classification based on its subjects

2.3 Related Works of Question Answering System for Hadiths

Through the review of the literature of QAS implementations on Hadith domain, there are no researches conducted to develop QAS based on Hadiths dataset in English language. However, there are many research are developed to handle the Arabic dataset of Hadiths.

Jbara [38] develop QAS to minimize the difficulty of retrieve accurate answers from Hadiths documents depend on the documents subjects. There are two main stages accomplished; (1) documents analysis using NLP processes such as stop-word removal, tokenization, normalization, and N-gram, and (2) classification stage using SVM to identify the best classification of Hadiths documents i.e. pray or fasting. The dataset of Jbara study is consists of 13 books that contain 1321 Hadith documents based on Arabic language such as the Book of faith, the book of fasting, and the book of praying. The main aim of Jbara work is to identify the relevant or irrelevant hadiths documents to specific subjects. The classification was accomplished using SVM methods through classify the features of each documents. The methodology of Jbara [38] consists of five main phases which are; (1) identify the main features of hadiths documents, (2) classify the identified features according to various subjects, (3) use 90% of documents as training set and 10% as training test based on system, (4) the accuracy of training set is calculated using F-score, and the accuracy of training set is measured automatically by system, and (5) compare the classification accuracy of each method with each other. The significance results of [38] show that the classification accuracy based on human reviewing was recorded 80% and the system accuracy was recorded 90%.

Bdair [39] mentioned that the main challenge facing the question-answering research is the questions analysing and extracting to generate accurate and correct answers. Their research focuses on retrieve accurate Hadiths about Islam subjects based on user's inquiries about Islam concepts. The basic technique to manage and extract user's questions is unigram and bigram which used to generate different possible matching words options of user's query. On the other hand, the words synonym extracting via WordNet using CS and LCS techniques conducted to provide accurate similarity matching between user question and Hadiths documents based on Hadiths standard concepts. The main idea of these techniques is to produce new list of similarities sentences to maximize the accuracy of similarity matching. After using synonym technique to generate similarities sentences that are compatible with Hadiths domain concepts, the similarity technique will retrieve the matching answers in Hadiths schema to display it as answers for user's queries. The accuracy percentages of [39 result was recorded as 81.4% based on real experimental tests by users. Siddiqui et al. [40] developed a system to automatically extract the chain of narrators from a Hadith through NLP and Classification techniques. For the classification purposes [40] adopted two main classification methods which as NB and kNN methods. The main aim of classification methods is to classify the trust of Hadiths based on the chain of narrators. The classification performance shows that the average of accuracy results is 73% of the classified data. Siddiqui et al. [40] apply many pre-processing methods such as tokenization and normalization in order to analyse and extract the chain of narrators. Table 2 summarizes the directions of related works of Hadiths QAS.

Table 2: Summary of Related Works of Hadiths QAS

Author	QAS Based Language	Method	Main Aim
Jbara (2010) [38]	Arabic	Pre-processing methods (normalization, tokenization, and remove stop-words). a. SVM for documents classification	Decrease the difficulty of retrieve accurate answers from Hadiths documents depend on the users' queries
Bdair (2014) [39]	Arabic	WordNet, CS and LCS	Retrieve accurate Hadiths about Islam subjects based on user's inquiries about Islam concepts
Siddiqui et al (2014) [40]	Arabic	NB and KNN	Extract the chain of narrators from a Hadith through NLP and Classification techniques

Based on the above table 1, it can be concluded that the work of [40] was conducted for extract the chain of narrators from a Hadith documents. Thus, this study considered as irrelevant to this research context (this research aims to classify the hadiths documents according to related subjects). Jbara [38] and Badair [39] are more familiar with this research aim. However, Jbara work was conducted to classify the hadiths documents based on its subjects while Badair work was conducted to analyse and enhance the concepts of users' questions in order to retrieve more accurate answers. Thus, Badair [39] work was focused on the methods of questions processing and analysis while [38] was focused on SVM as a classification method.

For this research, Badair and Jbara methods can be integrated as two main phases which are; (1) processing and analyse users' questions based on NLP methods, (2) classify the documents according to subjects and questions types using SVM method.

2. RESEARCH METHODS

The research method involves two important directions that affect the question-answering system's accuracy. Firstly, the users could not have the effective skills to provide their questions in the right way. For example, the query typing of "what is Malaysia capital?" is better than typing "give me cities in Malaysia". Therefore, the question-answering system could address this challenge using NLP methods. Secondly, the document classifications according to queries types (i.e. When questions) using machine learning methods could improve the accuracy of provided answers. The NLP and machine learning methods selection of question-answering system for Hadiths domain take into account many points based on question-answering system aspects and research scope. These points are as the following:

a. The selected domain: there are two main domains of question-answering system which are open domains and close domains. The Hadiths domain considered as close domain because it is focuses on information based on same context rather than tourism (i.e. open domain) which focuses on many fields such as weather and hotels bookings. However, Hadith domain contains large information about various related fields e.g. Pray, Zakat, Fasting, Alshadteen, and Pilgrimage. Thus, this research focuses on two main fields which are Pray and fasting due to its relation with daily activities of people. The other fields like pilgrimage are done once in a lifetime.

b. Type of users' query: there are two main types of questions which are open and restricted questions. In this research the restricted type is adopted in order to provide more accurate

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

answers. The restricted type of questions helps the users to manage their query using define keywords such as "what", "when", "where", and "how" questions. On other hand, the documents or answers can be managed according question type which increases the opportunity to retrieve accurate answers from large documents. Specifically, this research focuses on "when" and "where" questions due to nature of selected Hadith fields. Usually, the fasting subject is related to time (i.e. "When") and the Pray is related to time and places (i.e. "where" and "when").

c. Architecture of question-answering systems: the question-answering systems involve two main directions in order to increase the opportunities of provide accurate answers; (1) query analysis, (2) and documents or answers management. The query analysis can be accomplished effectively using many methods such as tokenization, stop-word removing, and N-gram. On the other side, one of most effective methods to classify the documents based on specific indicators' (i.e. questions types) such SVM method.

d. Another important process of question-answering system is the concepts similarity which support the answers retrieving based on similarity measuring between query and system documents. Also, the similarity is important to enhance the users' query supporting ontology that contains standard or right concepts of the related domain. The Cosine Similarity (CS) and Long Common Similarity (LCS) are the most common similarity measuring techniques in question-answering systems and the WordNet is used widely for the purpose of extracting and replacing the weak concepts in queries with the right concepts.

e. In order to understand the various methods effects on the results accuracy, the results of proposed QAS will be generated according to four methods which are; CS, LCS, combination of CS and LCS, and combination of CS, LCS, and SVM. The details of these methods will be explained in chapter 4.

Consequently, Figure 2 illustrates the methodology of this research according to selected methods of question-answering system. The methodology can be described as three main phases which are pre-processing phase, answering processing phase and, and classification phase. In the pre-processing phase, there are many methods are used to analyse the main concepts of users' questions; stop word removal method to remove the un-important concepts (i.e. 'the', 'is', and '?'), tokenization method to split the main concepts in query as single words. In the answering processing phase, there are many methods are used to enhance the extracted concepts from query; N-gram is used to list the extracted concepts as 1-gram, 2-gram... n-gram. The WordNet tool is applied to enhance the concepts that listed using N-gram through replace the user concepts by standard concepts that match with documents concepts. LCS and CS techniques are used to measure the similarity between enhanced query and answers documents.

In the classification phase, SVM method is adopted to classify the documents according to proposed question types, and related subjects. Thus, the dataset could be segmented as four main classes which are: when for pray, where for pray, when for fasting, and where for fasting. The SVM classification of documents is accomplished supporting NER methods to identify the places (where) and time (when) features that included in the documents. The answers of query extracted concept from the most suitable class of documents depend on the query question type and subject. The pray and fasting documents as answers about when and where questions are classified based on several attributes to identify the documents relation with question type. For example, Mecca is the name of a place and the places are indicator of where questions. The document classified as an answer for when or where questions based on the total of question type indicators that appear in that document. For instance, 'X' document is answer about when question because the when indicators number in X is 5 while the where indicators number is 1. Technically, SVM has been trained to classified the documents into four labels based on indicators above, as the following (Figure 1):

a. L0: when

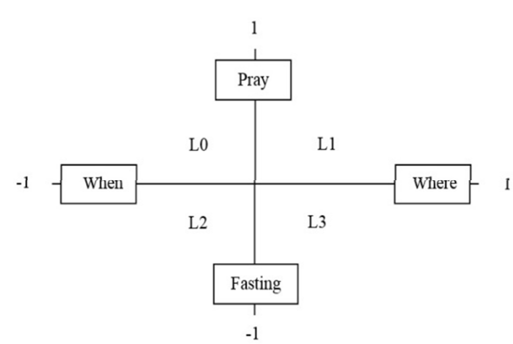

Figure 1: Documents Spaces

for pray. b. L1: where for pray. c. L2: when for fasting. d. L3: where for fasting. Positively labelled data points in document space (R_2) can be denoted as four space which are; (1) L0= (-1, 1), L1 = (1, 1), L2= (-1, -1), and L3= (1, -1). Any document that not belongs to any of these spaces considered as negative document (i.e. out of space range). Now, we define the feature (the indicators of places and time concepts) of each class ("when", "where", "pray", and "fasting"). The when and where features that contained in documents are identified and extracted automatically using NER methods. NER can extract the time and place features according to sentences context. Thus, the extracted features using NER can compared with the set of defined features in the system.

In order to clarify the effects of SVM on the enhancement of answers accuracy, SVM will be compared with the traditional processes of answers extraction using CS and LCS techniques. Thus, in the next section provides the comparisons between CS and LCS techniques with and without SVM.

Figure 2: Proposed Research Method (at the end of this paper)

3. EXPERIMENTAL DATA

According to [41], one of the most popular evaluation methods of question-answering system is F-score method. The following formulas represent the calculation of answers accuracy based on precision and recall:

a. "Recall = T/ (T+ (Hadith-T)). Where T is the true answers and Hadiths is the total number of corpus documents."

b. "Precision= T/ (T+ (N.H-T)). Where N.H is the total number of Hadiths that retrieved as candidates' answers."

c. "F1= 2*(P*R)/P+R, where P is answers precision and R is recall Hadiths".

The evaluation of the proposed question-answering system is calculated based on two methods which are; (1) calculate F-score based on comparisons between the human reviewing of provided answers and the systematic answers that provided by the proposed system, and (2) the system measure the accuracy internally based on F-score of relevant or irrelevant answers.

The main methodology that used to evaluate the accuracy of the proposed question-answering system based on external evaluation by experts are as the following:

a. Generate the answers from the system based on users' query.

b. Review the answers carefully.

c. Determine the correct and mismatch answers.

d. Calculate the precision, recall.

e. Measure the test accuracy (F-measure).

f. Calculate the average of all tests.

On the other hand, the system calculates the F-measure internally. If the system retrieves all the documents in a corpus then the system will retrieve all relevant documents in the corpus, in this case the recall will be perfect. Since there are only small proportions of documents in a collection that are truly relevant to the given query, retrieving everything will give a very low precision. A combination of good precision and good recall is needed. In the best case we would like to retrieve all the relevant documents and to discard non-relevant documents. Thus, the system evaluates the accuracy of retrieved documents as the following procedures:

a. Compute the number of features of where, when, fasting, and pray in query and documents after stop word removal.

b. Compute the intersection between various counted features in query. For example, query considered as when for pray if it at least one feature of when and one feature of pray.

c. Repeat step b on the retrieved documents from system.

d. Here, the document considered as relevant to query if the analysed features of both sides are matched, while the document considered as irrelevant to query if the analysed featured of both sides are un-matched.

e. Calculate the recall, precision, and F-score based on the number of relevant and irrelevant retrieved documents.

The Hadiths corpus of proposed system consists of many Hadiths documents which are related to fasting and pray subjects. The number of Hadiths documents is 132 Hadiths (82 Hadiths related to pray subject, and 50 Hadiths related to fasting subject). On the other hand, this research focus on provide Hadiths answers about when and where questions. The number of Hadith documents that related to when question and pray subject is 17. The number of Hadith documents that related to where question and pray subject is 11. The total number of Hadiths that related to when and where questions connecting with pray subject is 28. The number of Hadith documents that related to when question and fasting subject is 14. The number of Hadith documents that related to where question and fasting subject is 0 (i.e. fasting subjects cannot be related with places). The total number of Hadiths that related to when and where questions connecting with fasting subject is 14. Table 3 summarizes the Hadiths documents classifications.

Table 3: Summary of Hadiths Corpus

	Pray	Fasting
when	17	14
Where	11	0
Not related to when or where	54	36
Total	82	50

Thus, to assure the validity of system testing, the proposed system was conducted based on 12 queries which are as the following:

a. Q1: When are the five times of prayers for Muslims?
b. Q2: Where was the first Friday prayer?
c. Q3: When is the fasting month of Muslims?
d. Q4: When do Muslims pray for Eid?
e. Q5: When do you pray Maghrib?
f. Q6: Where was the first Qibla of Muslims?
g. Q7: When does fasting begin?
h. Q8: When can Muslims do the Taraweeh prayer?
i. Q9: When should a traveler shorten the prayer?
j. Q10: When does fasting end?
k. Q11: When do Muslims pray to God?
l. Q12: When is the time of Al-Fajr prayer?

All above queries were selected based on the proposed questions about pray and fasting subjects that provided by 15 students from UKM universities according to discussion of the proposed system objectives. The proposed queries of when type for pray subject are six, where type for pray are three queries, when type for fating are three queries, and there no any query of where type for fasting. This considered as logic distribution of proposed queries. The fasting subject cannot be related to where question (there is no specific places for fasting); The fasting is usually related to time. Also, it is logic to proposed queries of when type for pray subject more than other queries; the pray controlled by many specifications of time such as hours, week, and Eids.

In the experiments, the author uses about (61%) of each hadith class as training set while the rest (39%) of each class is used as testing set for the classification system. Of course, for each class the (39%) hadith in the testing set are not included in the training set of hadith and the training phase calculation. The training set is used to learn the system how to analyse and classify the hadiths queries and documents while the testing is set to evaluate the accuracy of the proposed methods. Actually, the testing set is the hadiths that contained features of where, when, pray, and fasting

4. RESULTS AND DISCUSSION

In order to clarify the experimental results effectively, the four tests were accomplished based on the experimental data in previous section. Firstly, the similarity result of only Cosine Similarity (CS) was conducted. Secondly, the similarity result of only Longest Common Subsequence (LCS) was conducted. Thirdly, the similarity result based on the combination between CS and LCS was conducted. Lastly, the similarity result based on the combination between CS, LCS, and SVM was conducted. The first test is conducted through using only the cosine similarity technique to measure the similarity between user query and documents. the average of answers accuracy of all tested queries record 67%. The most accurate F_score (0.74) is belonging to third query (When is the fasting month of Muslims?) while the lowest F_score (0.55) is belong to 8th query (When can Muslims do Taraweeh prayer?). In this test the system evaluation is approximately same the external expert evaluation. The second test is conducted through using only the LCS technique to measure the similarity between user query and documents. the average of answers accuracy of all tested queries record 66%. The most accurate F_score (0.76) is belonging to 2nd query (where was the first Friday prayers?) while the lowest F_score (0.55) is belonging to 6th query (Where was the first Qibla of Muslims?), and 9th query (When should the traveller shorten the prayer?). In this test the system evaluation is approximately same the external expert evaluation. The third test is based on the combination between CS and LCS techniques. The combination between CS and LCS techniques was conducted through testing each proposed query using these two techniques and selects the better F-score of CS and LCS. For example, if F_score of first query using CS technique is higher than F_score of first query

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

using LCS then the system will select F-Score of CS. The measurement of F-score and results selection was accomplished using proposed QA system. Thus, the combination results represent the best possible answers using similarities techniques. the average of F_score using combination of CS and LCS was recorded 70%. The most accurate F_score (0.76) is belonging to 2nd query (where was the first Friday prayers?) while the lowest F_score (0.60) is belonging to 9th query (When Should the Traveller Shorten the Prayer?). In this test the system evaluation is approximately same the external expert evaluation.

The fourth test is based on CS and LCS techniques with SVM method. Table 4 summarizes the accuracy results of proposed question-answering system based on the combination of CS, LCS with SVM techniques. The combination of these techniques represents the using of SVM to classify Hadiths quires and documents according to Hadiths subjects (Pray and Fasting) and questions types (Where and When) before calculate F-score of CS and LCS techniques. As noticed from Table 3, the average of F_score using combination of CS, LCS, and SVM was recorded 80%. The most accurate F_score (0.86) is belonging to 2nd query (where was the first Friday prayers?) while the lowest F_score (0.73) is belonging to 8th and 12th queries.

Table 4: Accuracy Measurements based on CS, LCS, and SVM (at the end of this paper)

Whereas Table 5 summarizes the accuracy results of proposed question-answering system based on CS and LCS techniques with SVM according the internal evaluation by system. It can be noticed that the system evaluation (0.73) is less than the external expert evaluation (0.80). This indicates that the system falls in some lack in the evaluation of retrieved documents due to complexity of evaluation matrix (matching of query question type and subject with the documents question type and subject). However, the system evaluation presents the accuracy of CS and LCS with SVM is more than CS or LCS without SVM.

Here, the results are discussed according to external evaluation by expert rather than system evaluation. The expert evaluation method considered as reliable more than system evaluation. According to findings of experimental results (Figure 3), the combination of CS, LCS, and SVM techniques record the highest accurate records of answers (80%) followed by the combination of CS and LCS techniques (70%), then CS technique (67%), and finally LCS technique (66%). This is results considered logical; the CS technique is effective to measure the similarity based on words in small documents contents (few sentences) while the LCS is effective to measure the similarity of sequence of words in large document contents. However, LCS is also working effectively in small documents. The contents of hadiths documents are small (few contents).

Thus, there are no major differences between the results of CS and LCS. The main significance of the result is when the documents classification is conducted using SVM. SVM work on classify the hadiths documents as classes and search on the most suitable class according to query purpose. Here, the number of retrieved and irrelevant documents is reduced which lead to balance between recall and precision and consequently increase the records of F-score.

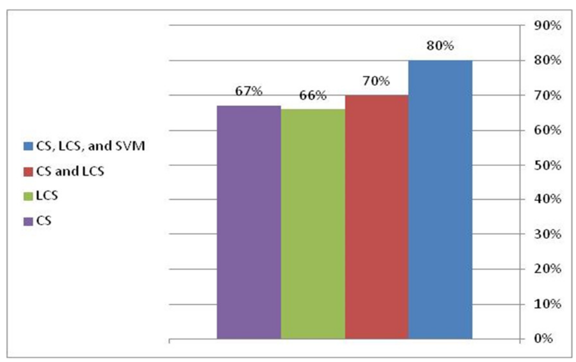

Figure 3: Accuracy Records

Figure 4 illustrates the accuracy records of each proposed query according to various methods (CS, LCS, CS and LCS, and CS, LCS, and SVM).

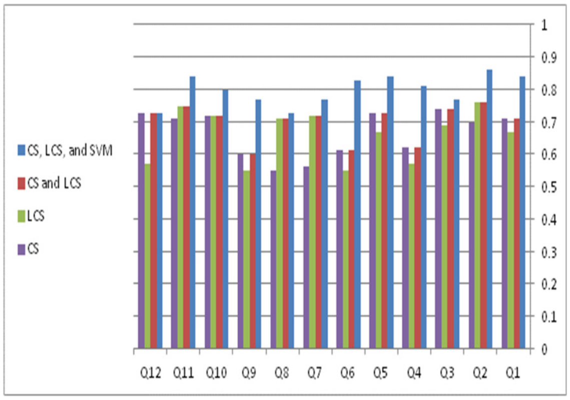

Figure 4: Accuracy of Each Query According to Various Methods

From the two figures above, SVM technique plays an important role in improving the accuracy results of the proposed question-answering system. The average accuracy results of all queries were improved by 10% when apply SVM with other techniques. On the other hand, most individual results of queries record accurate answers based on SVM with other techniques. SVM reduce the searching space of Hadiths documents through classify the Hadiths depend on proposed question types and documents subjects. The reducing of searching space increases the opportunities of retrieving true answers that match with users' queries. This finding can be justified clearly through compare the results of SVM of pray and fasting subjects. The pray subject can be classified as when and where question types documents but the fasting subject can be classified as when question type documents. the results of pray queries based on SVM is more accurate than Fasting queries due to possibility of minimize the searching space of pray subject more than the fasting subject.

5. CONCLUSIONS

The large number of answers corpus and the lack of typing standard query concepts are from the main challenges of question answering systems. This work enhances the accuracy of question answering system through three main stages: 1) Pre-processing methods like tokenization and stop-word removal is used to identify the main concepts of users' query. 2) Answering processing methods and techniques like N-gram, WordNet, CS, and LCS are used to update and enrich the extracted concepts of users' query based on the formal representation of hadiths answers or documents. 3) Support Vector Machine (SVM) and Name Entity Recognition (NER) methods are conducted to classify documents based on relevant subjects and questions types in order to reduce the searching scope of answers documents.

Documents in Hadith corpus are classified according to proposed question types, and related subjects as four main classes which are: when for pray, where for pray, when for fasting, and where for fasting. The results of experimental tests show that the proposed methods are effective to improve the accuracy of question-answering system for Hadith domain. Significantly, SVM technique reduces the searching space of answers which improve the accuracy of provided answers.

6. ACKNOWLEDGEMENT

This research has been supported by the UKM Research Grant GUP-2015-003.

REFERENCES

[1] Xu-Dong Lin, Hong Peng & Bo Liu, "Support Vector Machines for Text Categorization in Chinese Question Classification", In, *Proceeding International Conference on Web Intelligence* (WI 2006 Main Conference Proceedings) College of Computer Science and Engineering, South China University of Technology.

[2] Marcin Skowron, Kenji Araki, "Evaluation of the New Feature Types for Question Classification with Support Vector Machines", In, *International Symposium on Communication and Information Technology* (ISCIT), 2004.

[3] Dell Zhang & Wee Sun Lee, "Question Classification using Support Vector Machines", National University of Singapore. Paper work, Singapore-MIT Alliance, Toronto, Canada, 28 August 2003.

[4] Gharehchopogh, F., & Lotfi, Y., "Machine Learning based Question Classification Methods in the Question Answering Systems", *International Journal of Innovation and Applied Studies* 4(2), 2013, 264-273.

[5] Gillet, P, Trojahn, C. & Haemmerl´e, C., "Complex Correspondences for Query Patterns Rewriting", IRIT and Universit´e de Toulouse 2, Toulouse, France, 2013.

[6] Harb.A, Michel Beigbeder, Jean-Jacques, "Evaluation of Question Classification Systems Using Differing Features", Report, Institute of Electrical and Electronics Engineers, 2009.

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

[7] Tan, W., Jianrong Cao, Hongyan Li, Algorithm of Shot Detection based on SVM with Modified Kernel Function. International Conference on Artificial Intelligence and Computational Intelligence, Shan Dong Jianzhu University, Jinan 250101, China, 2009.

[8] Srihari, R. & Li, W, "Information extraction supported question answering", In, *Proceedings 8th Text Retrieval Conference* (TREC-8), NIST Special Publication, 2000, pp. 500-246.

[9] Zhang, D. & Lee, W, "Question classification using support vector machines". In, *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2003, pp. 26-32.

[10] Chinchor. N., "Overview of MUC-7/Met-2". In. *proceeding of the seventh Message Understanding Conference* (MUC-7). Fairfax, VA, 1997.

[11] Bhaskar, P., Pakray, P., Banerjee, S. & Banerjee, S. "Question Answering System for QA4MRE", Report, Department of Computer Science and Engineering, Jadavpur University, India, 2012.

[12] Bdour, N., & Gharaibeh, K., "Development of Yes/No Arabic Question Answering System", *arXiv preprint arXiv*, 1302(5675), 2013.

[13] Motik, Boris, Sergio Antonio Berná Niñerola, Pablo Castellanos Garcia & Carlos González-Cadenas, "Natural Language Question Answering System and Method Based On Deep Semantics", U.S. Patent Application 13/171,391, filed June 28, 2011.

[14] Jian-fang, S., Zong-tian, L., & Jian-feng, F., "Event-network clustering using similarity", In, *Natural Computation* (ICNC), 2010 *Sixth International Conference*, Vol. 8, 2010, pp. 3970-3973.

[15] Brants, T., & Franz, A. "Web IT 5-gram version 1", www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006T13. 2006.

[16] Poibeau, T. & Kosseim, L., "Proper Name Extraction from Non-Journalistic Texts", In, *Proceeding Computational Linguistics in the Netherlands*, 2001, pp. 144–157.

[17] Biggins, S., Mohammed, S., Oakley, S., Stringer, L., Stevenson, M. & Priess, J., "University_of_Sheffield: Two Approaches to Semantic Text Similarity", *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2. Proceedings of the Sixth International Workshop on Semantic Evaluation*, 2012, pp. 655-661.

[18] Miller, G.A., "WordNet: A lexical Database for English", *Communications ACM. USA*, 38(11), 1995, 39.

[19] Mihalcea, R., Corley, C. & Strapparava, C., "Corpus-based and Knowledge-based Measures of Text Semantic Similarity", In, *Proceedings of AAAI'06 the 21st national conference on Artificial intelligence*, vol. 1, 2006, pp. 775-780.

[20] Mohler, M. & Mihalcea, R., "Text-to-Text Semantic Similarity for Automatic Short Answer Grading", In, *Proc. 12th Conf. European Chap*. ACL, Athens, Greece, 2009, pp. 567-575.

[21] Madylova, A. & Oguducu, S., "A Taxonomy Based Semantic Similarity of Documents Using the Cosine Measure", *In, Proceeding 24th International Symposium Computer and Information Sciences* (ISCIS 2009), pp. 129-134.

[22] Hassan, S., & Mihalcea, R., "Semantic Relatedness Using Salient Semantic Analysis", In, *AAAI*, August 2011.

[23] Gunopulos, D. & Das, G., "Time series similarity measures (tutorial PM-2)", In, *Tutorial notes of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2000, pp. 243-307.

[24] Lin, C. & Och, F., "Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics", In, *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, 2004, p. 605.

[25] Day. M, & Chorng-Shyong Ong, "Question Classification in English-Chinese Cross-Language Question Answering: An Integrated Genetic Algorithm and Machine Learning Approach", Report, Institute of Information Science, Academia Sinica, Taiwan, Department of Information Management, National Taiwan University, Taiwan, IEEE, 2007.

[26] Molina-González, M.D., Martínez-Cámara, E., Martín-Valdivia, M.T. & Perea-Ortega, J.M., "Semantic orientation for polarity classification in Spanish reviews", *Expert Systems with Applications*, 2013, 40(18): 7250-7257.

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

[27] Turney, P.D., “Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews”, In, *Proceedings of the 40th annual meeting on association for computational linguistics*, 2002.

[28] Kim, Y., “Convolutional neural networks for sentence classification”, arXiv preprint arXiv:1408.5882, 2014.

[29] Mao, Y., Wei, C. H., & Lu, Z., “NCBI at the 2014 BioASQ Challenge Task: Large-scale Biomedical Semantic Indexing and Question Answering”, In CLEF (Working Notes) (pp. 1319-1327), (2014, September).

[30] Joachims, T., “*Making large scale SVM learning practical*. n.p”, Universität Dortmund, 1999.

[31] Cortes, C., & Vapnik, V., “Support-vector networks. *Machine learning*, 1995, 20(3): 273-297.

[32] Liljenback, M.E., “ContextQA: Experiments in Interactive Restricted-Domain Question Answering”, Doctoral dissertation, San Diego State University, 2007.

[33] Chung, H., Song, Y.I., Han, K.S., Yoon, D.S., Lee, J.Y., Rim, H.C., & Kim, S.H., “A practical QA system in restricted domains”, In, *Proceedings of the Workshop Question Answering in Restricted Domains*, within ACL, 2004.

[34] Shiyan, M. & Orasan, C., A”n ontology-based question answering method with the use of textual entailment”, In, *Proceeding International Conference on Natural Language Processing and Knowledge Engineering*, 2009. (NLP-KE 2009).

[35] Tian, Y., Li, H., Cai, Q. & Zhao, S., “Measuring the similarity of short texts by word similarity and tree kernels”, In, *IEEE Youth Conference on Digital Object Identifier Information Computing and Telecommunications* (YC-ICT), 2010, pp. 363-366.

[36] Panicker, A.D., Athira, U. & Venkitakrishnan, S., “Question classification using machine learning approaches”, *International Journal of Computer Applications*, 2012.

[37] Eka, K., Azhari S.N., “Review of Ontology-Based Question Answering System”, Report, Computer Science and Electronics Department. Faculty of Mathematics and Natural Sciences, Gadjah Mada University, Yogyakarta, 2001.

[38] Jbara, K., “Knowledge discovery in AlHadith using text classification algorithm”, *Journal of American Science*, 2010, 6(11): 409-419.

[39] Bdair, A.H., “Question answering system for hadiths domain concepts using CS and LCS similarity techniques”, 2014.

[40] Siddiqui, M.A., Saleh, M.E.S. & Bagais, A.A., “Extraction and Visualization of the Chain of Narrators from Hadiths using Named Entity Recognition and Classification”, 2014.

[41] Faisal, M., “U.S. Patent No. 6,094,652 Washington, DC: U.S. Patent and Trademark Office”, 2000.

Journal of Theoretical and Applied Information Technology
15th April 2017. Vol.95. No 7

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

Figure 2: Proposed Research Method

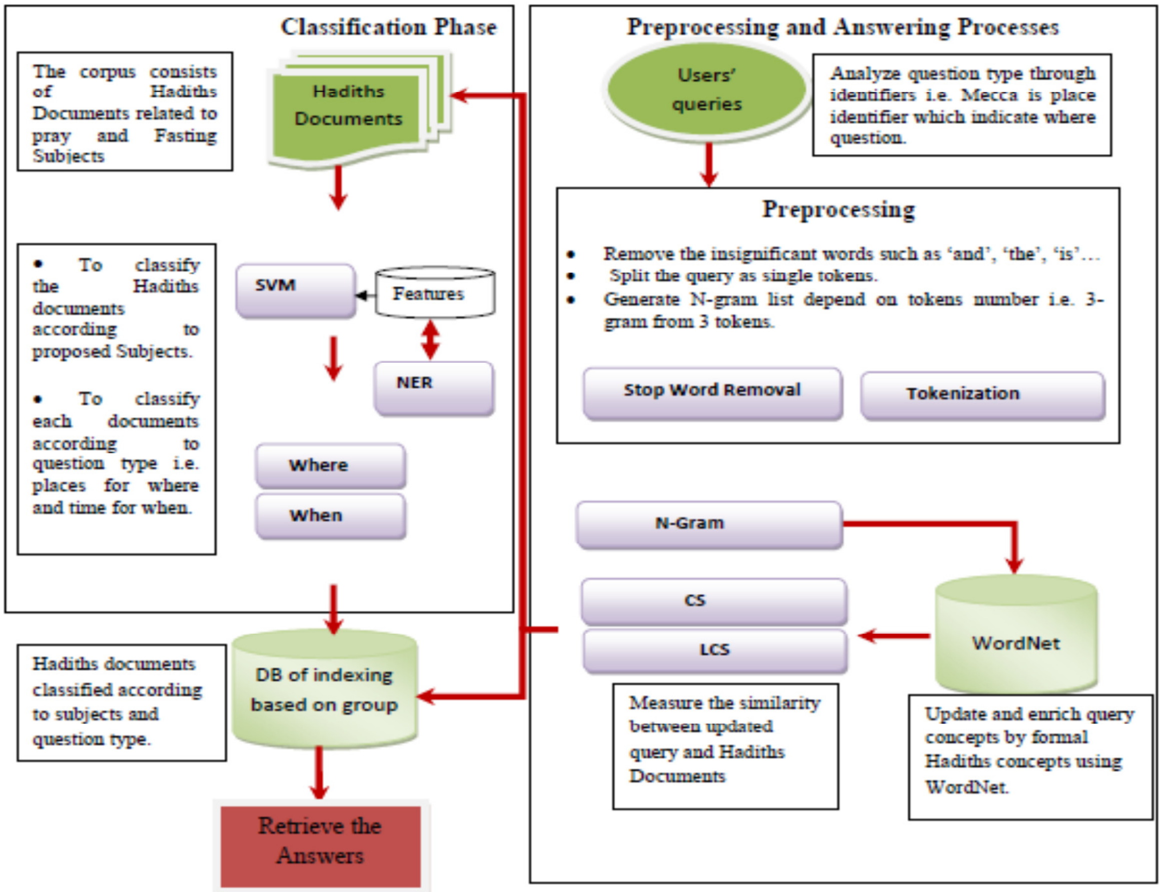

Table 4: Accuracy Measurements based on CS, LCS, and SVM

Query	Hadith	Output system N.H	T	F	Precision	Recall	F_Score
Q1	17	14	13	1	0.93	0.76	0.84
Q2	11	10	9	1	0.90	0.82	0.86
Q3	14	12	10	2	0.83	0.71	0.77
Q4	17	15	13	2	0.87	0.76	0.81
Q5	17	14	13	1	0.93	0.76	0.84
Q6	11	11	10	1	0.91	0.77	0.83
Q7	14	12	10	2	0.83	0.71	0.77
Q8	17	13	11	2	0.85	0.65	0.73
Q9	17	14	12	2	0.86	0.71	0.77
Q10	14	11	10	1	0.91	0.71	0.80
Q11	17	14	13	1	0.93	0.76	0.84
Q12	17	13	11	2	0.85	0.65	0.73
Av							0.80

Table 5: System Evaluation based on CS and LCS Techniques with SVM

Query	Hadith	Output system N.H	Relevant	irrelevant	Precision	Recall	F_Score
Q1	17	14	12	2	0.71	0.77	0.71
Q2	11	10	8	2	0.80	0.76	0.80
Q3	14	12	9	3	0.75	0.69	0.75
Q4	17	15	12	3	0.80	0.75	0.80
Q5	17	14	11	3	0.79	0.71	0.79
Q6	11	11	9	2	0.82	0.82	0.82
Q7	14	12	8	3	0.67	0.62	0.67
Q8	17	13	11	2	0.85	0.73	0.85
Q9	17	14	10	3	0.71	0.65	0.71
Q10	14	11	11	0	1.00	0.88	1.00
Q11	17	14	12	2	0.86	0.77	0.86
Q12	17	13	9	4	0.69	0.60	0.69
Av							0.73