CHAPTER 1

INTRODUCTION

Exploring questions about Islam for Muslims and non-Muslims is currently an interesting study. Knowledge and answers about Islam all come from the Al-Quran, this is because Al-Quran can provide broad and comprehensive teaching, not only about rituals but also about all aspects of life. Today people prefer searching for religious answers from the Internet and various platforms rather than searching through classical religious books [3]. However, there are still many people who have not found accurate answers to questions about Islam. Accuracy and relevance in finding an answer to user questions is an important thing. Keywords are often used to make the search easier when users unknowingly specifically what to ask. However, searching for answers using keywords still has problems, one of which is when the answers issued are less relevant. Generally, existing search engines still have problems such as incorrect words or fetching lots of irrelevant documents, especially when the user's query is not specific enough [1][2]. Most of the search methods available via search engines cannot meet all the user's needs for precise and specific information. Although the autoresponder system is one of the most important components of a distance teaching platform, it has received attention and is an interesting one to explore [3].

Accuracy and relevance in finding an answer, as well as accuracy in searching for keywords from a question are some of the problems in the automatic answering system. Providing the main purpose of QAS is to facilitate human-machine interaction, with the right answers to quickly review the list of documents [3]. In recent years, QAS has been investigated extensively and automated QAS has become an interesting area of research and the results show a marked improvement in its performance [4]. Various techniques and approaches have been researched and undertaken by researchers over the past few years to provide a reliable Question and Answer System (QAS). However, in their research, they have not yet discussed the topic of Islam in detail, it is still looking for answers that are still too broad and not too specific on the sources used to find answers.

Based on the existing problems, this study researched to discuss the topic of questions about Islam. The QAS category used in this study is a closed domain, where the questions discussed have a specific or specific topic. In addition, this study proposes a method that focuses on accuracy in searching for keywords from a question and accuracy in choosing relevant answers. By carrying out the method proposed in this study, it is hoped that it can answer the existing problems.

This chapter includes the following subtopics, namely: (1) Rationale; (2) Theoretical Framework; (3) Conceptual Framework/Paradigm; (4) Statement of the problem; (5) Objective and Hypotheses; (6) Scope and Delimitation; and (7) Importance of the study.

1.1 Rationale

Search engine technology is part of the Information Retrieval (IR) process and has one of the next generations, namely the Question and Answers System (QAS) [5][6]. QAS summarizes the information that users need is required to comply with more accurate searching. QAS includes various types start from answering simple questions (yes/no) and complex questions, where the answers must be synthesized from various data sources. One of the functions of the QA System is to process and analyze a question posed by a user efficiently, which aims to get an accurate answer. This system is using unlimited text as the main source of knowledge [6]. Ensuring the success of QAS can be measured based on 2 factors, including; (1) analyze user requirements (Questions) efficiently using Natural Processing Language (NLP) and (2) classify and manage documents that contain candidate answers accurately based on the classification phase document. Therefore, accurate matching between user questions and answers will be found effectively [6]. QAS can be used in a linguistic approach, statistical approach, and pattern matching approach [7][8].

The question-answering framework was constructed using natural language processing and information retrieval techniques [7]. In general, this framework is divided into three modules, namely the Question Processing module, the Document Processing module, and the Answer Processing Module [9]. Several studies on QAS have been carried out based on techniques and raise interesting problems. Several issues play an important role in a question answering system such as Question Classes, Multilingual (or cross-lingual) Questions Answering, User profiling for QA,s Question Processing, Data Sources for QA, Context, Answer Extraction, Real-Time Question Answering, Answer Formulation, Interactive QA Advanced reasoning for QA, Information clustering for QA[7].

There have been many studies discussing the issues of QAS, one of which is the Question Processing issue. Question processing has the task of detecting the type of answer expected by a question. The relationship between a question and an answer can be seen from the topic or keyword of the question. With this in mind, searching for pairs of questions and answers will be easier. One of the techniques for extracting topics and keywords is using Query Expansion (QE). As has been done by several previous studies [10][11][12], the use of QE in QAS has helped in classifying and expanding the search for relevant questions and answers, based on the topic and keyword of the question. The main challenge of QE in QAS is the problem of discrepancies, which can arise because of the mismatch between the expression referred to and the text-based answers [10].

Suhaib Kh. Hamed and Mohd Juzaidin, in their study they discussed question processing and answer extraction [4]. In question processing, they use question expansion to expand the meaning of words and look for document placement in answer extraction. Question expansion is an approach that adapts from Query Expansion. Query expansion reformulates the user's original query and performs the process of adding several relevant words [1]. Apart from Suhaib Kh. Hamed and Mohd Juzaidin, other studies use

question expansion in QAS aimed at increasing the effectiveness of retrieving information on a question [1][11][10].

The breadth of the concept of reformulating a question, this study offers an expansion of the query questions that can be pursued to the main goal. The main purpose of a question can be seen from its focus and type. In addition, the results of this process will be continued by calculating the closeness of the relevant question documents. This concept is an overview that will be discussed in the proposed study.

1.2 Theoretical Framework

In previous studies, QAS has developed to efficiency in information retrieval. The expected information is an accurate answer, to achieve this the classification of the verses of the Qur'an is done, where the verse will be used as the QAS answer. To classify the document for candidate answer, this study using the Neural Network and to retrieve relevant documents based on the ranking of similarity scores in this study using N-Gram [11]. However, in their research, he still has shortcomings, namely the process of taking answers to limited documents, does not pay attention to the answers generated based on the type of question and there is no measurement of semantic closeness to see the similarity of the questions.

Based on some of the above, this study will develop QAS by transforming by using question expansion to see the type and focus of the question. In addition, you will also use the similarity of the question scores in the answer extraction process. Since the transformation of queries can lead to the expansion of the proposed domains, a Semantic Similarity Score will be used to avoid this expansion and narrow the domain search. So that the QAS that is built is expected to be more appropriate and relevant to the answers to be recommended.

1.3 Conceptual Framework/ Paradigm

In general, this study will use two main processes, namely Question Analysis and Answer Selecting. Question analysis, this study uses the concept of Query Expansion method for Question Expansion. Expansion Questions are used to see the focus of the question by expanding the query inside. Knowing the focus questions and the closeness between the relevant questions is the main focus of this study. The main focus is based on the fact that users can express the same question using different wordings and styles, these questions have representations that rarely vary in length, mostly short with small overlapping words [13]. In addition, this study is a close-domain QAS and relies on a collection of answer documents.

1.4 Statement of the Problem

The outline problem raised in this study is how to find out the headword of a question and find out the closeness between the relevant questions. This is taken from previous research, which has discussed question expansion and measures the closeness of the answer to a question.

Based on the main problems above, here are the sub-problems that are the focus of work in this study:

- How to use query expansion to find out the headword of the question?
- What if getting a focus question affect the obtaining of relevant documents?
- How to use semantic similarity to find out the closeness between documents?

1.5 Objective and Hypothesis

The objective of this study is to develop QAS using question expansion by looking at the type and the headword of the question. In the question expansion, a rule comparison is perform, to see which rule is longer in search of the type and the headword of the question. In the process of finding answers, score questions to answer relevant answers are used. Based on some of these things, this study analyzed this baseline study that the method proposed could prove the type and focus of the questions to find relevant answers.

1.6 Scope and Delimitation

The scope of this study is Close-Domain QAS. The questions asked are questions limited to the Frequently Asked Questions (FAQ) of the Quran [14]. The topic of this QAS question is about questions about Islam. The type of question tested on this QAS is a Factoid question type, and the type is 4W-1H (What, Who, Where, When and How). The series of processes carried out in the proposed study is question analysis and answer extraction.

1.7 Significance of the Study

In the consideration of previous research [4], this study aims to expand questions by using rules to find the headword of the question and measure the closeness between answer documents. In this study, the dataset is differentiate from previous research, but using the same method. In addition, this study grouped the question types into 4W-1H [15]. Based on some of the work that this study has done, it is hoped that these contributions will attract attention to further exploration.

1.8 The Contributions

There are many challenges and issues in the QAS, one of which is the exploration of methods for question analysis. The purpose of the question analysis is to find out what information is obtained from the questions asked by users. The information generated in this process must be information that can support the next process, namely the process of selecting answers. One of the issues in this study is how to get information on questions looking for similar questions to get answers. To answer the problem, motivation is raised and made contributions to the study. The contribution of this study is to add a process to question expansion by using rules to perform the headword extraction process. The addition of the question expansion process aims to prove that using headwords can make the search for relevant questions accurate. In addition, this study performs an error analysis of the rules used, to determine the lack of rules regarding the types of data used.