



## Question Answering Machine for Islamic Factual Questions Based on English Translation of Hadith

Ja'far, Adiwijaya<sup>2</sup>, Said Al-Faraby<sup>3</sup>

<sup>1</sup>Fakultas Informatika, Universitas Telkom, Bandung

<sup>2</sup>Divisi Digital Service PT Telekomunikasi Indonesia

<sup>1</sup>Bandung, West Java, Indonesia

[\\*jafarassagaf@students.telkomuniversity.ac.id](mailto:*jafarassagaf@students.telkomuniversity.ac.id), [adiwijaya@telkomuniversity.ac.id](mailto:adiwijaya@telkomuniversity.ac.id),  
[saidalfaraby@telkomuniversity.ac.id](mailto:saidalfaraby@telkomuniversity.ac.id)

Received, Revised, Accepted

**Abstract** — The basic idea of this project is to create a system that can answer Islamic factual question. With English translation of Hadith Sahih Bukhari book as the corpus for answers. The corpus is not a knowledge base, it is a document consisting 7095 verses of Hadith. Hence, it is inherently harder for the system to read and extract answer directly. The system able to receive an input posed in natural language such as "When was Prophert Muhammad born?" or "What are the best deed we can do to please Allah?". A keyword based search was used. First the user question went through preprocessing phase where keyword is extracted. Then, the keywords were used to search the corpus looking for matching verses. After that, the system used scoring and ranking to find the best matched verse and then return the corresponding answer for the question.

**Keywords** – system, question, answer, hadith, islam.

Copyright © 2018 JURNAL INFOTEL

All rights reserved.

### I. INTRODUCTION

A Question Answering (QA) system is an automated approach to retrieving correct answers to question written by user in natural language. A QA system is a computer program that extracts its answers from a collection of structured (database) or unstructured (text) data. The main goal if a QA system is to facilitate human-machine interactions, by getting quickly a precise answer rather than a list of documents. QA systems are classified into two categories: Closed domain and Open domain QA systems. Closed-Domain QA systems deal with questions in a specific domain such as Quran, medicine, or agriculture; closed-domain might point to a condition where only a limited type of questions are used, such as question asking about description or fact. while open domain QA systems answer questions about anything in all domains and depend on general ontology and world knowledge; moreover, open domain systems usually need much more data to extract the answer [1]. In general, the processing of a QA system is composed of three stages: question analysis, document analysis, and answer analysis. The question is analyzed and classified to

determine the question type and the major concept involved in the question; it is also reformulated to an appropriate query. The documents are analyzed to extract candidate documents that contain answers. The answer is analyzed to extract candidate answers and then rank to find the best one. Question answering systems integrate many techniques such as artificial intelligence(AI), natural language processing(NLP), machine learning, information retrieval, and information extraction. To create an accurate QA system, we need to integrate a wide range of knowledge, and use many ideas in natural language processing and artificial intelligence.

Knowledge about Islam is a mandatory subject that all of us Muslim should be aware of. But sometimes, be it young or old people, basic information about Islam such as the number of tahajud prayer rakaah, the date of isra mi'raj, or the year of which our prophet hijra to Medina are all easily forgotten. For some, it's easy to obtain/regain the information by reading the Holy Quran or a collection of hadith, but for the people that have time off the essence but wanted to get answers directly from Book of hadith, sitting down and