

Entity Recogniton untuk Alquran Terjemahan Bahasa Inggris dengan Pendekatan Supervised Learning

1. Pendahuluan

The Quran, according to the *Kamus Besar Bahasa Indonesia* (Indonesian Dictionary) or KBBI is a Muslim holy book that contains the word of God revealed to the Prophet Muhammad. Through the angel Gabriel to be read, understood, and practiced as a guide or a way of life for humankind. The Quran is a holy book from the beginning until the time of its contents has not changed. This Muslim religious scripture consists of 6236 verses divided into 144 chapters. A chapter contains a different narrative between the other chapters and sometimes has no connection with one other chapter. In each of the chapters, there are also many entities scattered in each verse, which can make it difficult for someone to search for a particular entity. Entities in the Quran can refer to the name of people or groups. A classification system can be developed to make it easier for someone to find a particular entity.

Therefore, the researcher offers a technique for finding entities using Entity Recognition through the Supervised Learning approach to facilitate finding specific entities. Unlike the NER (Named-Entity Recognition) which classifies a named-entity, Entity Recognition limited to entities. The system overview can be seen in Fig. ???. The use of a dataset needed by following the Supervised Learning approach where the program will first be trained to achieve optimum accuracy where previously the dataset has been preprocessed. The primary input is the verses of the Saheeh International Quran version. The researcher chose The Saheeh International because it provided convenience to students, and brought readers closer to their original nuances ?. After the *preprocessing* process, the entity that has been identified by the system will be classified into the entity class by taking the multiple-choice approach. To be more clearly given an example of an input of a verse in chapter As-Saff verse 6 "*when Jesus, the son of Mary, said, " O children of Israel, indeed I am the messenger of God to you confirming what came before me of the Torah and bringing good tidings of a messenger to come after me, whose name is Ahmad. "But when he came up with clear evidences, they said," This is obvious magic. "*", then detected "Jesus the son of Mary", "Ahmad" and "Children of Israel" which classify as entities. After that, the system builds a model that can classify entities in the Quran and make it easier for someone to understand the Quran. In conducting the research, researchers only use the Quran with Saheeh English translation taken through the website *tanzil.net*. The data obtained by us is still in the form of textit raw and unstructured. Therefore, preprocessing needs to be done first.