

Similarity Detection of Page in the Al-Quran by using Rabin Karp Algorithm and Jaccard Similarity

Winda Eka Samodra¹, Moch Arif Bijaksana².

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹samoded@student.telkomuniversity.ac.id, ²arifbijaksana@telkomuniversity.ac.id,

Abstract

Plagiarism is an action to get scientific recognition without providing an original source. On the other side, plagiarism can be used to find similarity between one document to another. It's quite difficult to measure the similarity of documents only by reading or measuring them manually. But this concept of plagiarism can be used to detect similarities between verses in the Qur'an. The dataset used is a set of verses on the pages of the Qur'an. Therefore, a system is needed that can carry out similarity detection. To build the system, this final project uses the Rabin Karp method to be processed using N-gram parsing on each input document. The data will be searched for the similarity value between paragraphs one on the first page with all the verses on the second page. Next, a similarity value of 1 and 0 will be taken. The system is built using one of the similarity methods namely, Jaccard Similarity. By using hashing techniques and using the Rabin Karp Algorithm the system can produce the best precision values of 77.4% and produce f-Measure of 55.8%.

Keywords: Plagiarism, Qur'an, Rabin Karp, Jaccard Similarity