

**Abstract**

This research aims to address the scarcity of comprehensive websites providing detailed lists of Isim Ma'rifat in the Quran and Sahih Muslim Hadith. The absence of a comprehensive resource hinders the ability to study and compare Isim Ma'rifat between these significant Islamic texts. To overcome this issue, the study develops a natural language processing approach utilizing an integrated Java tokenizer program with a MySQL database containing the Sahih Muslim Hadith and Quranic texts. The program identifies the occurrence of the alif lam prefix, followed by diacritic removal to facilitate accurate verse comparison between the two texts. The research focuses on identifying alif lam prefixed Isim Ma'rifat exclusively present in the Quran, exclusive to Sahih Muslim Hadith, and similarities between them. The analysis yields a comprehensive understanding of the distinctions and similarities of alif lam prefixed Isim Ma'rifat between the Quran and Sahih Muslim. These findings provide valuable input for the Al-Quran project, contributing to the development of comprehensive and accessible resources for Islamic studies. It is expected that this research will enhance the understanding of Isim Ma'rifat in the religious and linguistic context, offering a significant contribution to the field of natural language processing especially in the Arabic language.

**Keywords:** Diacritics, Prefix, Sahih Muslim, Tokenizer, Quran

