

Abstrak

Penelitian ini bertujuan untuk mengatasi kekurangan situs *website* komprehensif yang menyediakan daftar rinci dari *Isim Ma'rifat* dalam Al-Quran dan Hadis Sahih Muslim. Ketidakhadiran sumber daya komprehensif tersebut menghambat kemampuan untuk mempelajari dan membandingkan *Isim Ma'rifat* antara dua teks Islam penting ini. Untuk mengatasi masalah ini, penelitian ini mengembangkan pendekatan pemrosesan bahasa alami dengan menggunakan *Tokenizer* berbasis Java yang terintegrasi dengan *database MySQL* yang berisi teks Hadis Sahih Muslim dan Al-Quran. Program ini mengidentifikasi kemunculan prefix alif lam, diikuti dengan *Diacritics removal* untuk memudahkan perbandingan ayat yang akurat antara dua kitab tersebut. Penelitian ini fokus pada identifikasi *Isim Ma'rifat* yang diawali dengan alif lam yang ada dalam Al-Quran, Hadis Sahih Muslim, dan persamaan di antara keduanya. Analisis ini menghasilkan pemahaman komprehensif tentang perbedaan dan persamaan *Isim Ma'rifat* yang diawali dengan alif lam antara Al-Quran dan Sahih Muslim. Temuan ini memberikan masukan berharga untuk proyek Quranpedia, berkontribusi pada pengembangan sumber daya yang komprehensif dan mudah diakses untuk studi Islam. Diharapkan bahwa penelitian ini akan meningkatkan pemahaman tentang *Isim Ma'rifat* dalam konteks agama dan linguistik, serta memberikan kontribusi yang signifikan pada bidang pemrosesan bahasa alami, terutama dalam bahasa Arab.

Kata kunci : Diacritics, Prefix, Sahih Muslim, Tokenizer Al-Quran

