Abstract

Hadith has several levels of authenticity, among which are weak (dhaif), and fabricated (maudhu) hadith that may not originate from the prophet Muhammad PBUH, and thus should not be considered in concluding an Islamic law (sharia). However, many such hadiths have been commonly confused as authentic hadiths among ordinary Muslims. To easily distinguish such hadiths, this paper proposes a method to check the authenticity of a hadith by comparing them with a collection of fabricated hadiths in Indonesian. The proposed method applies the vector space model and also performs spelling correction using symspell to check whether the use of spelling check can improve the accuracy of hadith retrieval, because it has never been done in previous works and typos are common on Indonesian-translated hadiths on the Web and social media raw text. The experiment result shows that the use of spell checking improves the mean average precision and recall to become 81% (from 73%) and 89% (from 80%), respectively. Therefore, the improvement in accuracy by implementing spelling correction make the hadith retrieval system more feasible and encouraged to be implemented in future works because it can correct typos that are common in the raw text on the Internet.

Keywords: hadith, vector space model, symspell, spelling correction, document retrieval