Abstract

The religion of Islam is based on a sacred text called Qur’an, a divine speech expressed in Arabic language. Qur’an constitutes the main root of Islam jurisprudence which has a second source of inspiration known as Hadiths. As the Muslim’s life is governed by those holy texts, need of their authenticity is required. Using VSM (Vector Space Model), we can represent Hadiths as a vector of words. The Term Weighting obtained by multiplying term frequency by the inverse document frequency does not take into account the word order, however, order of narrators is critical to classify Hadith. In this paper we propose a new method considering the words order (in our case the narrator’s order), to classify Hadiths into four categories: Sahih, Hasan, Da’if and Maudu’. We use in this purpose LVQ (Learning Vector Quantization). We got good results for classifying Sahih and Maudu’ categories.


Index Terms

Computer Science

Networks

Keywords

Arabic Natural Language Processing, Learning vector quantization, Term Weighting, Text
categorization, Vector Space Model.