Abstract

We conducted a semantic similarity study between concepts in the Holy Book Quran. Semantic similarity looks at how much two concepts within Quran share common attributes or properties to each other. The more they share the same characteristics to each other, the bigger the semantic similarity score is. We also gathered the Quranic gold standard; a Quranic concept pairs data set along with its similarity score manually annotated by human raters.

We chose knowledge based approach to automatically measure semantic similarity, by using the property of length and depth of a synset in WordNet lexical database, and applied it to Yuhua Li's et. al. equation that has been oonsidered to be the baseline among other researchers.

As the result, our system gained Pearson's correlation r of 0.33, while the Spearman's is ρ 0.19. By considering the inter-annotator agreement ($\rho = 0.63$) of our Quranic gold standard data set as the upper bound score, our result still relatively far from perfectly mimicking Muslim's intuition to measure the degree of similarity of concepts in Quran and Islamic domain in general.

Keywords: semantic similarity, Quran, concept, WordNet, gold standard