

ABSTRACT

In the Qur'an, a lot of words can be used to be researched especially in the case of Natural Language Processing (NLP), one of which is the Model semantic distribution for common words and linkages as in this research. Previous research, which uses the Qur'an dataset only performed semantic similarities to certain surah, evaluation of datasets similarities and semantic linkages but the application of the Turkish language, and application of evaluation of datasets for verbs only. The purpose of the study was to provide insight into the researcher's semantic model by evaluating the model in several attributes. Balance the pairs of words of the Quranic vocabulary dataset for a class of nouns and verbs with human frequencies to evaluate the durability of a semantic model regarding the problem of rare word occurrence. The word pair is selected to enable evaluation of a semantic model distribution with multiple Word attributes and link pairs of words such as the frequency of word occurrence, concreteness, and relationship type (e.g., synonymy, antonymy). The Dataset consists of 500 pairs of words and is assisted by 15 respondents, of which each pair has two distinct values for similarity and relevance. The method used is the Sim-Rel vector, questionnaire, and the calculation of gold standard, until the result of performance calculations using a correlation of Spearman Rank of 0.909. The Sim-Rel vector axis Gets the result with 4 areas i.e. SU = 23 pairs of words, SR = 77, DU = 192, and DR = 208.

Keywords: Natural Language Processing, Distribution Semantic Model, Sim-Rel Vector, Spearman Rank.