

Pembangunan Data Latih Menggunakan Monolingual Word Alignment Pada Persamaan Kisah Nabi Dalam Al-Quran dan Alkitab

Nadia Aruma Dita Suwandi¹, Moch. Arif Bijaksana²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung
¹nadiaaruma@students.telkomuniversity.ac.id, ²arifbijaksana@telkomuniversity.ac.id

Abstract

For centuries the Qur'an and the Bible have been present in the midst of human civilization as a guide to the lives of their followers and it turns out that there are verses that have the same meaning / have similar stories and are interrelated between the two. To conduct a research on measuring the semantic similarity between the translations of the Koran and the Bible, a training data is needed about the similarity of the stories in the Koran and the Bible itself. In developing the training data, the right method is needed to be used, one of which is to use the monolingual alignment method. Monolingual alignment is a method of word alignment that identifies similarities between words in pairs of sentences that are in the same language. In this study, the equation of the prophet story contained in the Koran and the Bible is used as a dataset for the construction of training data (corpus) that are dialigned and function as a collection of objects. In addition, several features were applied in the construction of monolingual word alignment in the equation of the AL-Quran and Indonesian-language Bibles with this MSR format, namely identical align words, align PFA and align word sequences and intrinsic evaluations as performance benchmarks from training data (corpus) and produces F1 value of 64.89%. And for the best F1 results resulted from several combinations of alignment features, namely by combining align identical words and align PFA features with F1 results of 67.62%.

Keywords: Al-Quran, Alkitab, Monolingual Alignment, MSR
