

## INTISARI

### **PENGEMBANGAN MODEL PEMBANGUNAN KORPUS TERJEMAHAN AL-QURAN BERBAHASA INDONESIA YANG MEMILIKI INFORMASI LINGUISTIK TEKS BAHASA ARAB**

oleh  
**SUWANTO RAHARJO**  
**13/351312/SPA/00467**

Kebutuhan adanya basis data teks atau sering disebut korpus saat ini semakin meningkat untuk berbagai keperluan. Mulai dari mesin penerjemah sampai dengan penggunaan dalam bidang *Internet of Thing*. Banyak negara sudah memiliki berbagai jenis korpus, seperti korpus khusus, umum atau nasional. Indonesia dalam beberapa pustaka sudah disebut akan merencanakan pembentukan korpus nasional namun sampai penelitian ini dilakukan belum didapatkan informasi yang memadai. Terjemahan Quran berbahasa Indonesia merupakan salah satu teks yang banyak dibaca oleh masyarakat dan bisa dijadikan korpus.

Korpus terjemahan Quran tersebut dapat menjadi salah satu sub korpus nasional bahasa Indonesia. Pembangunan korpus terdiri dari beberapa tahapan dan berisi berbagai metode yang digunakan dalam prosesnya. Pengembangan model pembangunan korpus yang dilakukan dalam penelitian ini adalah pada tahapan pemrosesan korpus. Penelitian ini melakukan pengembangan metode anotasi *Part of Speech* (POS) yang berbasis aturan dengan menggunakan informasi linguistik bahasa Arab dari korpus Quran yang sudah ada. Hasil penelitian menunjukkan metode anotasi POS mampu memprediksi lokasi POS tertentu dengan akurasi 97.54%. Pemberian tambahan informasi semantik semantik bersifat independen sesuai dengan karakteristik teks Quran dimana mampu untuk memiliki banyak makna juga dilakukan dalam penelitian ini.

Berdasarkan penelitian menunjukkan bahwa metode anotasi pada korpus terjemahan Quran dapat dikembangkan berdasarkan informasi linguistik bahasa Arab dari korpus Quran. Kelebihan Sistem Manajemen Basis Data Relasional dari penelitian ini dapat digunakan untuk menyimpan anotasi semantik yang bersifat independen untuk korpus terjemahan Quran.

**Kata Kunci :** *korpus, anotasi pos, metode anotasi, pemrosesan teks*

## ABSTRACT

### THE DEVELOPMENT OF CORPUS BUILDING MODEL OF INDONESIAN TRANSLATION OF THE QURAN HAVING ARABIC LINGUISTIC INFORMATION

by  
**SUWANTO RAHARJO**  
**13/351312/SPA/00467**

The growing need for text databases, often referred to as corpus, is increasing for various purposes. Starting from the translator machine to usage in the Internet of Thing fields. Many countries already have various types of corpus, such as special corpus, public or national corpus. In several literature, Indonesia has planned to build a national corpus, but until this study has been carried out adequate information has not been obtained. Indonesian Quran translations are one of the texts that are widely read by the public and can be build as a corpus.

The result from Quran translation corpus building can be part of the national sub-corpus of the Indonesian language. The build of the corpus consists of several stages and various methods are used in the processes. The development of corpus building model on this study is in corpus processing stage. This study developed a rule-based Part of Speech (POS) annotation method using Arabic linguistic information from the existing corpus of the Quran. The result of proposed method showed that the method was able to predict POS locations with 97.54% accuracy. Independent semantic annotations in accordance with the characteristics of the Quran text which makes it possible to have many meanings also carried out in this study.

Based on the study, it can be shown that corpus annotations of the Indonesian Quran translation can be developed based on Arabic linguistic information from the Quran corpus. The advantages of the Relational Database Management System from this study can be used to store independent semantic annotations for the Indonesian Quran translation.

**Key Words :** *corpus, pos annotation, annotation method, text processing*