

Pengembangan morphological analyzer bahasa Indonesia informal menggunakan Finite-State Transducer (FST) = Development of informal Indonesian morphological analyzer using Finite-State Transducer (FST)

Muhammad Salman Al-Farisi, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920526898&lokasi=lokal>

Abstrak

Penelitian ini bertujuan untuk meningkatkan kemampuan morphological analyzer pada Aksara agar dapat melakukan pemrosesan bahasa Indonesia informal. Metode yang digunakan pada Aksara adalah rule-based menggunakan finite-state transducer dengan compiler bernama Foma. Adapun komponen yang ditingkatkan adalah komponen tokenizer, lemmatizer, dan POS tagger. Untuk menguji penelitian ini, dibuatlah sebuah gold standard yang terdiri dari 102 kalimat dengan 1434 token. Hasil pengujian memperlihatkan bahwa penelitian ini berhasil memiliki peningkatan akurasi tokenisasi sebesar 4.6% dari Aksara v1.1. Untuk tahapan lematisasi pada kasus case sensitive terjadi peningkatan akurasi sebesar 11.82%. Evaluasi POS tagging juga berhasil mengalami peningkatan pada nilai F1-Score sebesar 14% dibandingkan dengan Aksara v1.1.

.....This study aims to improve the ability of the morphological analyzer in Aksara in order to be able to process the informal Indonesian. The method used in Aksara is rule-based, using a finite-state transducer with a compiler named Foma. The components that are being improved are tokenizer, lemmatizer, and POS tagger components. To test this research, a gold standard was created; It consists of 102 sentences with 1434 tokens. The test results show that this study has an increase in tokenization accuracy of 4.6% compared to Aksara v1.1. For the lematization stage in the case of case-sensitive word, there is an increase in accuracy of 11.82%. The POS tagging evaluation also increased its F1-Score value by 14% compared to Aksara v1.1.