

INTISARI

Pengenalan Entitas Bernama pada Artikel Berita Berbahasa Indonesia Menggunakan Metode Hidden Markov Model dan Rule-Based

Oleh

ISMA MISDALENI TIARASANI
11/312728/PA/13568

Proses pengenalan entitas bernama ini dilakukan untuk teks berbahasa Indonesia. Pengenalan entitas bernama merupakan suatu proses pemberian label entitas pada token dalam kalimat. Hal yang menjadi tantangan dalam melakukan pengenalan entitas bernama yaitu sulitnya dalam melakukan analisis secara manual karena menurut Ekbal dan Bandyopadhyay (2007), hal tersebut termasuk dalam *open class of expressions* yang berarti bahwa ada banyak variasi yang tak terbatas dan ekspresi baru yang akan terus berkembang.

Metode penelitian ini dilakukan dengan menggabungkan *Hidden Markov Model* dan *Rule Based* dengan harapan mendapatkan nilai performa sistem yang baik. Sistem dikembangkan dengan memperhatikan fitur kata pada *Hidden Markov Model*, sedangkan pada *Rule Based* dengan memperhatikan fitur kelas kata, kontekstual, dan morfologi kata.

Pada proses pengujian, data yang digunakan yaitu data *corpus* penelitian Fachri (2014). Hasil evaluasi dari performa sistem dengan menggunakan metode *Hidden Markov Model* pada observasi kata tersebut menghasilkan nilai *precision* 18,3% dan akurasi 30,6%. Pada penggabungan metode *Hidden Markov Model* dan *Rule Based* dapat menghasilkan nilai *precision* 63,2% dan akurasi 88,4%. Berdasarkan hasil pengujian tersebut dapat disimpulkan bahwa penggabungan metode *Hidden Markov Model* dan *Rule Based* dapat meningkatkan nilai *precision* 44,9%, dan akurasi 57,8% dibandingkan dengan hanya menggunakan metode *Hidden Markov Model*.

Kata kunci : Pengenalan Entitas Bernama, *Hidden Markov Model*, *Rule Based*

ABSTRACT

NAMED ENTITY RECOGNITION IN INDONESIAN NEWS ARTICLES BASED ON HIDDEN MARKOV MODEL AND RULE BASED APPROACH

by

ISMA MISDALENI TIARASANI
11/312728/PA/13568

A named entity recognition process is performed for Indonesian text. The Named Entity Recognition is a process of labeling entities on token in a sentence. The challenge of Named Entity Recognition is that such expressions are hard to analysis manually because in Ekbali and Bandyopadhyay (2007), it is included in the open class of expressions which means that there are an infinite variety and new expressions are constantly being invented.

This research method is done by combining Hidden Markov Model and Rule Based in hopes of getting good system performance value. This system was developed with attention to the word feature of the Hidden Markov Model, while in Rule Based with respect to word class, contextual, and word morphology features.

In the testing process, the data used is research corpus data Fachri (2014). The result of system performance evaluation using Hidden Markov Model on word observation resulted precision value 18,3% and accuracy 30,6%. In the merger of Hidden Markov Model and Rule Based method can provide value of precision 63.2% and accuracy 88.4%. Based on the test results, it can be concluded that the merger of Hidden Markov Model and Rule Based method can increase the value of precision 44,9%, and accuracy 57,8%, when compare with only Hidden Markov Model method.

Key words: Named Entity Recognition, Hidden Markov Model, Rule Based