

INTISARI

PENGENALAN ENTITAS BERNAMA PADA ARTIKEL BERITA BAHASA INDONESIA MENGGUNAKAN METODE BERBASIS ATURAN

ISTHOFANY IRFANA AZKIYA

11/316772/PA/13899

Pengenalan entitas bernama merupakan subpekerjaan dari ekstraksi informasi. Pengenalan entitas bernama bertujuan untuk mengidentifikasi entitas di dalam teks. Pengenalan entitas bernama sering kali digunakan untuk mengidentifikasi hubungan dalam ekstraksi informasi.

Salah satu pendekatan pengenalan entitas bernama adalah menggunakan metode *rule-based*, yaitu sekumpulan aturan untuk mengidentifikasi entitas dalam suatu dokumen teks. Pada penelitian ini dikembangkan sistem dengan metode *rule-based* dimana fitur yang diperhatikan adalah morfologi kata, kontekstual, dan kelas kata. Kelas kata tidak ditentukan secara manual melainkan menggunakan POSTagger. Pada penelitian ini pengenalan entitas bernama lebih difokuskan untuk artikel berita berbahasa Indonesia.

Berdasarkan pengujian sistem yang telah dilakukan, pengenalan entitas bernama menggunakan metode *rule-based* untuk artikel berita berbahasa Indonesia ini mendapatkan *f-measure* 75,98% dengan nilai *recall* 77,36% dan *precision* 74,64%.

Kata kunci: ekstraksi informasi, pengenalan entitas bernama, rule-based, kelas kata, morfologi kata

ABSTRACT

NAMED ENTITY RECOGNITION IN INDONESIAN NEWS ARTICLES BASED ON RULE-BASED APPROACH

ISTHOFANY IRFANA AZKIYA

11/316772/PA/13899

Named entity recognition is a subtask of information extraction. Named entity recognition objective to identify entities in the text. Named entity recognition are often used to identify relations between information extraction.

One of approaches that can be used in named entity recognition is rule-based method, which is a set of rules to identify entities in a text document. In this study, a system developed with rule-based method which is concerned on the morphological features, contextual and word class. Word class is not specified manually, but using POSagger. News articles in Indonesian language is used in this study.

Based on system testing that has been done, named entity recognition using rule-based method for the Indonesian language news articles get 75,98% for f-measure with 77,36% of recall value and 74,64% of precision value.

Keywords: information extraction, named entity recognition, rule-based, part-of-speech, NER

