

PENERAPAN ANALISIS MORFOLOGI UNTUK PENGGUNAAN KATA BERIMBUHAN PADA POS TAGGER BAHASA INDONESIA BERBASIS STATISTIK

UMRIYA AFINI

(Pembimbing : Muljono, SSi, M.Kom)

Teknik Informatika - S1, FIK, Universitas Dian Nuswantoro

www.dinus.ac.id

Email : 111201207305@mhs.dinus.ac.id

ABSTRAK

POS tagging (pelabelan kelas kata) merupakan salah satu proses penting pada aplikasi-aplikasi NLP. POS tagging otomatis dibutuhkan karena POS tagging manual membutuhkan waktu yang lama dan biaya mahal. Masalah utama POS tagging otomatis adalah kata ambigu dan kata Out-of-Vocabulary (OOV). Salah satu pendekatan untuk mengatasi masalah kata ambigu adalah POS tagging menggunakan pendekatan statistik dengan Hidden Markov Model (HMM). Sistem POS tagger Bahasa Indonesia yang menerapkan HMM adalah IPOSTagger. Sedangkan penanganan kata OOV dalam penelitian ini digunakan metode pengalisis morfologi dengan sistem yang diterapkan adalah Morphind. Selain untuk penanganan kata OOV, Morphind juga digunakan untuk pemotongan klitik pada kata imbuhan berbentuk frasa. Dengan menggabungkan kedua sistem dihasilkan beberapa model POS tagging Bahasa Indonesia. Korpus yang digunakan yaitu korpus latih terdiri dari 10.000 kata yang telah diberi 31 label POS dan 3 jenis korpus uji masing-masing berisi 3000 kata. Keakuratan tertinggi sebesar 95.683~% dihasilkan oleh model HMM trigram + MA pada korpus uji 1 dengan tingkat kata OOV 10%, pada korpus 2 dengan tingkat kata OOV 20% dihasilkan oleh model HMM trigram + MA dan HMM bigram + MA dengan keakuratan yang sama yaitu 92.809~%, sedangkan pada korpus 3 dengan tingkat kata OOV 30% dihasilkan oleh model HMM bigram + MA yaitu sebesar 88.3279~%.

Kata Kunci : Part of Speech Tagger, Hidden Markov Model, Analisis Morfologi

IMPLEMENTING MORPHOLOGICAL ANALYZER FOR AFFIXES HANDLING IN STATISTIC BASED INDONESIAN POS Tagger

UMRIYA AFINI

(Lecturer : Muljono, SSi, M.Kom)

*Bachelor of Informatics Engineering - S1, Faculty of Computer
Science, DINUS University*

www.dinus.ac.id

Email : 111201207305@mhs.dinus.ac.id

ABSTRACT

POS tagging (word class tagging) is one of essential processes in many NLP applications. Automatic POS tagging is needed because manually POS tagging is time-consuming and costly. The main problem of automatic POS tagging is word ambiguity and Out-of-Vocabulary (OOV) word. One of approach for handling word ambiguity problem is using Hidden Markov Model (HMM) with IPOSTagger. Another POS tagging problem is OOV words handling, in this research, is used morphology analyzer method with system applied is MorphInd. In addition to handling OOV word, system MorphInd is used for clitic segmentation in phrase form. Combining both systems produced several models for POS tagging in Bahasa Indonesia. Used corpus is training corpus which consists of 1.0000 words given 31 POS tag and 3 kinds test corpus consist of 3000 words. Highest accuracy in test corpus 1 with 10% OOV is 95.683~% resulted by HMM trigram + MA model, in test corpus 2 with 20% OOV is 92.809~% resulted by HMM trigram/bigram + MA model, and test corpus 3 with 30% OOV is 88.3279~% resulted by HMM bigram + MA model.

Keyword : Part of Speech Tagger, Hidden Markov Model , Morphology Analysis