INTISARI

PENERAPAN *PARSER, STEMMER,* DAN PENDETEKSI AMBIGUITAS DALAM SISTEM PENGANALISIS TATA BAHASA BAKU DALAM KALIMAT BAHASA INDONESIA

Habib Malik
10/297422/PA/12989

Bahasa Indonesia merupakan bahasa resmi negara Indonesia yang penggunaannya telah diatur dalam Undang-Undang Nomor 24 tahun 2009. Bahasa Indonesia wajib dipakai untuk, antara lain, penulisan karya ilmiah serta publikasi informasi melalui media. Namun, kesalahan penggunaan tata bahasa Indonesia masih banyak ditemukan dalam berbagai penulisan resmi.

Penelitian ini akan menggunakan tiga teknik NLP untuk membangun sebuah sistem penganalisis tata bahasa. Teknik tersebut adalah *stemming,* pendeteksi ambiguitas kelas kata (sintaksis), dan *parsing*. Teknik *stemming* dan pendeteksi ambiguitas akan digunakan untuk membangun sistem *POS-Tagging* (penentuan kelas kata) berbasis aturan dan kamus, sedangkan teknik *parsing* digunakan untuk menguji keakuratan sintaksis dari suatu kalimat setelah kelas-kelas katanya diketahui. Setelah itu, keakuratan sistem akan diuji untuk menganalisis ketepatan data teks dari media berita daring dan abstrak penelitian ilmiah.

Dari hasil pengujian, akurasi sistem secara keseluruhan mencapai 96,17%, dengan ketepatan pendeteksian kesalahan sebesar 98,38% dan ketepatan pendeteksian kalimat tepat sebesar 93,61%.

Keyword: NLP, tata bahasa baku bahasa Indonesia, bottom-up parsing, stemming, pendeteksi ambiguitas.

ABSTRACT

IMPLEMENTATION OF PARSER, STEMMER, AND PART OF SPEECH AMBIGUITY DETECTOR FOR FORMAL GRAMMAR ANALYZER SYSTEM FOR INDONESIAN SENTENCES

Habib Malik
10/297422/PA/12989

Indonesian is the official language of Indonesia, whose utilization has been regulated by the Law No.24 of 2009. The Indonesian Language must be utilized for, among others, scientific papers and media publication. However, grammatical mistakes are still often seen in various formal writings.

This research will employ three NLP techniques to develop a grammar analyzer system. The techniques are stemming, part of speech ambiguity detection, and parsing. Stemming and ambiguity detection will be employed to develop a dictionary- and rule-based POS-Tagging system, while parsing technique will be employed to determine syntactic accuracy of the sentences after the parts of speeches have been known. Subsequently, the system's accuracy will be tested by analyzing the correctness of text data from online news media and scientific paper abstracts.

Based on the test result, the system, which was developed by combining parsing, stemming, and ambiguity detection techniques, could achieve an overall accuracy of 96.17%; specifically, it could detect grammatical mistakes in sentences with an accuracy of 98.38% and parse correct sentences with an accuracy of 93.61%.

Keyword: NLP, Indonesian grammar, parsing, stemming, ambiguity detection