

INTISARI

MODIFIKASI ALGORITME STEMMING MENGGUNAKAN PENDEKATAN NON DETERMINISTIK UNTUK TEKS BAHASA INDONESIA

Oleh

Wafda Adita Rifai

16/403720/PPA/05237

Dalam ilmu *Artificial Intelligence* terdapat bidang ilmu yang secara spesifik berfokus pada pengolahan bahasa yaitu *Natural Language Processing* (NLP). Salah satu tahapan yang dilakukan pada NLP adalah *Preprocessing*. *Preprocessing* merupakan tahapan dalam menyiapkan data sebelum diolah. Salah satu tahap pada proses *preprocessing* adalah *Stemming*. *Stemming* merupakan proses yang melakukan pencarian kata dasar dari suatu kata. Pemilihan kata dasar yang tidak tepat dapat menjadi kesalahan informasi yang akan diperoleh. Selain itu, proses *stemming* tidak selalu menghasilkan satu kata dasar karena terdapat beberapa kata dalam Bahasa Indonesia yang memiliki dua kemungkinan, yaitu sebagai kata dasar maupun kata berimbuhan seperti pada kata “beruang”

Penelitian ini melakukan modifikasi pada salah satu metode *Stemming* yang sudah ada dengan menerapkan pendekatan non deterministik untuk meningkatkan akurasi. Penggunaan metode non deterministik dilakukan dengan menerapkan daftar kandidat kata dasar untuk kata yang memiliki kata dasar lebih dari satu. Dari daftar kandidat kata dasar tersebut kemudian dipilih salah satu kata sebagai hasil.

Modifikasi *stemming* ini telah diuji pada data sejumlah 15.934 dengan tingkat akurasi sebesar 93.15 %. Oleh karena itu modifikasi *stemming* ini dapat digunakan untuk mengidentifikasi kata yang memiliki kata dasar lebih dari satu.

Kata kunci : *stemming*, non deterministik, akurasi, ambigu

ABSTRACT

MODIFICATION OF STEMMING ALGORITHM USING A NON DETERMINISTIC APPROACH TO INDONESIAN TEXT

by

Wafda Adita Rifai

16/403720/PPA/05237

Natural Language Processing is part of Artificial Intelligence that focus on language processing. One of stage in Natural Language Processing is Preprocessing. Preprocessing is the stage to prepare data before it is processed. There are many types of process in preprocessing, one of them is stemming. Stemming is process to find the root word from regular word. Errors when determining root words can cause misinformation. In addition, stemming process does not always produce one root word because there are several words in Indonesian that have two possibilities as root word or affixes word, e.g.the word “beruang”.

To handle these problems, this study proposes a stemmer with more accurate word results by employing a non deterministic algorithm which gives more than one word candidate result. All rules are checked and the word results are kept in a candidate list. In case there are several word candidates were found, then one result will be chosen.

This stemmer has been tested to 15.934 word and results in an appropriate level of 93.15 %. Therefore the stemmer can be used to detect words with more than one root word.

Keywords : stemming, non deterministik, accuration, ambiguous