# **BAB II Kajian Pustaka**

## **Kajian Terdahulu**

Pada bagian ini akan dipaparkan mengenai kajian penelitian terdahulu yang akan dijadikan sebagai acuan dasar dalam penelitian. Selain kajian terdahulu, akan dijelaskan tentang kajian teori yang mendukung penelitian ini. Berikut beberapa kajian terdahulu yang menjadi acuan penelitian:

Nurul Justina Mahardianing Verdaningroem dan Aries Saifudin. (2018) memaparkan penelitian terkait tentang mengurangi kesalahan *stemming* Bahasa Indonesia untuk penerapan kamus dasar pada Algoritma Porter. Pada penelitian ini bertujuan untuk membuat program sederhana yang digunakan untuk pembanding Algoritma Nazief & Adriani dengan kemampuan Algoritma Porter, perbandingan dilakukan dengan cara menggunakan 30 dokumen teks Bahasa Indonesia untuk mengetahui waktu proses, presisi dari hasil *stemming* pada dokumen tersebut. Penelitian ini melakukan eksperimen menggunakan komputer untuk melakukan proses *stemming*. Sedangkan perangkat lunak yang digunakan untuk menyimpan *database* kamus adalah MySQL.



**Gambar 2.1** Hasil Pengujian 30 Kata Yang Dipilih Secara Acak

Dari Gambar 2.1 hasil pengujian di atas menyatakan bahwa ada beberapa kata imbuhan yang di *stemming* dengan kamus menghasilkan kata dasar yang rata-rata masih memiliki makna, namun pada hasil kata dasar yang di stemming tanpa kamus menghasilkan beberapa kata dasar yang ambigu atau membingungkan seperti “gum”, ”rogram”, “gala”, “njara”, “iksa”, ”masu” dan “ail”. Sementara untuk perbandingan waktu menunjukkan bahwa waktu yang diperlukan dalam proses *stemming* dengan kamus membutuhkan waktu beberapa detik lebih lama dari pada *stemming* tanpa menggunakan kamus. Dapat dilihat pada kolom waktu *stemming*, pada *stemming* tanpa kamus menunjukkan bahwa dari 30 kata berimbuhan yang diproses stemming rata-rata menunjukkan waktu 0,00030 detik, sementara untuk *stemming* dengan menggunakan kamus membutuhkan lebih dari 0,001 detik pada Gambar 2.1 tersebut menunjukkan bahwa rata-rata waktu *stemming* dengan kamus adalah 0,50767 detik. Percobaan waktu ini dicoba pada proses *stemming* setiap 1 kata berimbuhan. Hasil dari proses *stemming* dengan memasukkan 30 kata secara langsung dan memprosesnya secara bersamaan menghasilkan waktu 0,00813 detik untuk proses *stemming* tanpa kamus dan 15,50625 detik untuk proses *stemming* dengan menggunakan kamus. Selisih waktu yang dihasilkan adalah 15,49812 detik. Hasil dari persentase kata dasar yang di *stemming* dengan kamus dapat dilihat pada perhitungan berikut: ((30 - 5)) / 30 X 100% = 83.333%. Sementara hasil kata dasar yang di *stemming* tanpa kamus dapat dilihat pada perhitungan berikut: ((30 - 9)) / 30 X 100% = 70%. Dari hasil presentase menunjukkan bahwa proses *stemming* dengan ditambahkan kamus dasar memiliki persentase keakuratan kata yang benar lebih besar 13.333% dari pada persentase keakuratan *stemming* yang hanya menggunakan aturan yang sudah ada.

Selanjutnya dalam penelitian Wahyudi dkk. (2017) memaparkan tentang implementasi dan analisis Algoritma *stemming* Nazief & Adriani dan Porter pada dokumen berbahasa Indonesia. Penelitian ini bertujuan untuk membandingkan Algoritma Nazief & Adriani dan Algoritma Porter untuk mengetahui proses *stemming* pada teks berbahasa Indonesia. Pada penelitian ini menggunakan data berupa kata dasar 28.532 kata yang di unduh dari internet. Pada proses *stemming* menggunakan *database* untuk menyimpan kata-kata yang telah di unduh. Penelitian ini menggunakan 8.168 kata yang diujikan dari 26 dokumen, *double word* yang muncul ada 1.312, kata yang termasuk dalam *stopword* ada 4.724 dan kata yang benar diujikan adalah 2.132. Hasil dari pengujian menghasilkan 2.031 *stop word* yang benaruntuk Algortima Nazief & Adriani, sedangkan 1.687 kata dihasilkan Algoritma Porter. Nazief & Adriani tingkat kesalahan *stemmer* hanya 5%, sedangkan Porter yaitu 21%. Tetapi untuk waktu proses Nazief & Adriani membutuhkan waktu lebih lama dari pada Algoritma Porter. Dari hasil berikut dapat disimpulkan bahwa hasil *stemming* pada dokumen berbahasa Indonesia yang baik adalah Algoritma Nazief & Adriani. Tetapi Agoritma Nazief & Adriani tidak lebih baik untuk efisiensi waktu proses dibanding Algoritma Porter.

Penelitian lainnya tentang pendekatan *confix-stripping* untuk *stemming* artikel berbahasa Indonesia. Tujuan penelitian ini adalah mencari kata dasar dari kata imbuhan teks bebahasa Indonesia dalam sebuah dokumen. Penelitian ini dilakukan dengan 20 dokumen teks Bahasa Indonesia dan menghasilkan nilai akurasi rata-rata sebesar 94.85%. Tahap *stemming* ini dilakukan dengan cara: (Hapsari dan Santoso, 2015).

* + - 1. Melakukan *parsing* pada dokumen untuk memecah sebuah kata-kata yang ada pada dokumen.
      2. Melakukan *stopword* atau bisa juga disebut sebagai proses menghilangan kata-kata yang tidak penting di dalam sebuah dokumen.
      3. Melakukan proses *confix-stripping* dengan cara pemenggalan atau penghapusan kata imbuhan yang ada awalan, akhiran, maupun imbuhan gabungan dari awalan dan akhiran untuk dijadikan bentuk dasar.

Berdasarkan kajian terdahulu yang telah dikaji, maka dilakukan penelitian tentang aplikasi pencarian kata dasar dan kata baku pada dokumen berbahasa Indonesia menggunakan metode Nazief & Adriani. Masalah dari penelitian ini adalah masih seringnya terjadi kesalahan-kesalahan penulisan kata baku pada sebuah dokumen berbahasa Indonesia. Tujuannya adalah untuk menerapkan stemming pada aplikasi pencarian kata dasar dan kata baku pada dokumen berbahasa Indonesia menggunakan Algoritma Nazief & Adriani dan mengukur tingkat presisi pada aplikasi pencarian kata dasar dan kata baku pada dokumen berbahasa Indonesia menggunakan Algoritma Nazief & Adriani. Adapun hasil dari penelitian ini adalah menghasilkan suatu sistem yang dapat membantu mahasiswa dalam perncarian kata dasar pada dokumen dan dapat menjadi referensi dalam pencarian kata baku pada sebuah dokumen.