

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Stemming merupakan suatu proses yang terdapat dalam sistem IR untuk mentransformasi kalimat atau dokumen ke kata-kata akarnya (*root word*) dengan menggunakan aturan-aturan tertentu. Sebagai contoh, kata bersama, kebersamaan, menyamai, akan distem ke root wordnya yaitu “sama”. Proses *stemming* pada teks berbahasa Indonesia berbeda dengan *stemming* pada teks berbahasa Inggris. Pada teks berbahasa Inggris, proses yang diperlukan hanya proses menghilangkan *sufiks*. Sedangkan pada teks berbahasa Indonesia, selain *sufiks*, *prefiks*, dan *konfiks* juga dihilangkan.

Porter adalah salah satu pencipta algoritma *stemming* bahasa indonesia yang diberi nama porter *stemming for* bahasa indonesia. Sampai saat ini algoritma porter sering digunakan untuk melakukan *stemming* bahasa indonesia, dikarenakan waktu yang dibutuhkan sangat cepat.

Namun algoritma porter *stemming* masih memiliki kelemahan dalam melakukan proses *stemming*. Mungkin dikarenakan porter berkewarganegaraan Belanda. Sehingga masih banyak kata yang hasilnya tidak sesuai dengan kata dasar yang ada. Oleh karena itu dibutuhkan sebuah metode untuk memperbaiki hasil dari algoritma porter *stemming* ini.

Knuth-Morris-Pratt(KMP) adalah algoritma yang digunakan untuk pencarian *string*. Algoritma KMP akan digunakan untuk memperbaiki masalah dari hasil algoritma porter *stemming*.

Hal inilah yang melatarbelakangi untuk membuat sebuah penelitian mengenai **“Implementasi Porter *Stemming* menggunakan Algoritma Knuth-Morris-Pratt (KMP) Untuk Menentukan Kata Dasar dalam Bahasa Indonesia**“ dimana dalam penerapannya nanti, aplikasi ini diharapkan dapat membantu dalam pencarian kata dasar bahasa indonesia menjadi lebih meningkat presisi ketepatannya.

1.2 Perumusan Masalah

Dari latar belakang di atas, maka perumusan masalah yang dapat disimpulkan adalah hasil dari algoritma porter *stemming* memiliki masalah dalam menentukan kata dasar. Sehingga dibutuhkan suatu metode untuk memperbaikinya kekurangan algoritma porter *stemming*.

1.3 Tujuan

Tujuan dari penelitian ini adalah untuk memperbaiki masalah dari hasil porter *stemming* dengan cara memodifikasi hasil porter *stemming* untuk lebih mencocokkan kata dasar dengan menggunakan metode knuth-morris-pratt. Serta untuk meningkatkan presisi ketepatan pencarian kata dasar dari hasil algoritma porter *stemming*.

1.4 Batasan Masalah

Dalam penelitian ini dilakukan batasan masalah agar penelitian lebih terpusat dalam melakukan penelitian. Batasan masalah dalam penelitian ini hanya memperbaiki kalimat yang tidak utuh dari hasil algoritma porter *stemming*.

1.5 The State of The Art

The State of The Art adalah analisa dari penelitian yang sudah dilakukan sebelumnya yang sudah ada,. Dimaksudkan untuk menganalisa penelitian sebelumnya yang pernah ada, sejalan dan mempunyai konsep yang hampir sama dengan penelitian sebelumnya dengan tujuan untuk dijadikan acuan dalam pengembangan yang akan datang.

Utomo, dkk(2008), membahas tentang *stemming* bahasa indonesia menggunakan Algoritma Tala. Dimana algoritma ini mengadopsi dari Algoritma Porter yang tidak menggunakan kamus untuk melakukan stemm pada kalimat berimbuhan. Aplikasi ini dibuat dengan berbasis web yang berfungsi untuk melakukan pencarian kata berimbuhan menjadi kata dasar dalam bahasa indonesia.[13]

Selanjutnya, Agusta(2009), membahas tentang perbandingan antara algoritma Porter *stemming* dengan Algoritma Nazieb andriani di dalam sebuah dokumen teks berbahasa indonesia. Hasil dari perbandingan ini adalah presisi ketepatan kata dasar algoritma nazieb andriani sedikit lebih unggul daripada algoritma Porter *stemming*. Namun masalah presisi waktu, algoritma Porter *stemming* jauh lebih unggul dibandingkan Algoritma Nazieb Andriani.[1]

Selanjutnya, Utomo, dkk(2013) yang membahas tentang pengukuran panjang pola yang mempengaruhi waktu pencarian pola di dalam teks. Algoritma yang digunakan adalah algoritma Brute Force(BF), Knuth Morris Pratt(KMP),Booyer Moore(BM) dan Karp Rabin(KR). Teks yang digunakan adalah teks alkitab. Hasil dari penelitian adalah menggunakan KMP dan KR waktu pencarian teks lebih meningkat dibandingkan yang lain. Namun secara signifikan BM lebih unggul dibandingkan yang lainnya.[12]

Selanjutnya, Wibowo Thio, dkk(2012). Membahas tentang Pembuatan Aplikasi untuk Mendeteksi Kebenaran Perintah SQL Query Menggunakan Metode Knuth-Morris-

Pratt(KMP). Dalam penelitian ini dilakukan pengujian dan analisa tentang algoritma Knuth Morris Pratt yang telah diterapkan pada perangkat lunak. Hasil dari penelitian ini adalah bahwa pencocokan ataupun pendeteksian kebenaran perintah SQL Query pada mata kuliah praktikum basis data lanjut dapat dilakukan. Kemudian, berdasarkan pengujian dan analisa metode KMP dapat ditarik kesimpulan bahwa penerapan metode KMP pada pencocokan string dapat diimplementasikan.[14]

Selanjutnya, Mandala, dkk(2004). membahas tentang sistem *stemming* otomatis untuk kata dalam bahasa indonesia. penelitian yang dikerjakan pada makalah ini adalah algoritma yang diusulkan dapat diaplikasikan dalam bahasa Indonesia dengan memodifikasi aturan imbuhan dan kamus yang dipakai sesuai dengan morfologi bahasa Indonesia. Kemudian fektifitas penentuan kata dasar dipengaruhi oleh jumlah aturan yang dikenali sistem dan jumlah kata dasar yang ada di dalam kamus. Semakin banyak aturan imbuhan yang dikenali sistem dan semakin banyak kata dasar yang dimasukkan ke dalam kamus akan semakin baik hasil yang didapat. Dalam penelitian ini dilakukan penerapan 4 algoritma dan dari empat algoritma yang diterapkan, algoritma Othman_Versil memberikan hasil yang terbaik karena menghasilkan kata yang tidak benar paling sedikit dibandingkan dengan ketiga algoritma lainnya.[5]

Tabel 1.1 Tabel Perbandingan Studi Literatur

No	Penelitian	Teknologi	Metode
1	Utomo, dkk(2008)	Web	Algoritma Tala
2	Agusta(2009)	-	Algoritma Porter dan Algoritma Nazien Andriani
3	Utomo, dkk(2013)	Bahasa C	Brute Force, Knuth-Morris-Pratt, Booyer Moore dan Karp Rabin
4	Wibowo Thio, dkk (2012)	MySQL	Algoritma Knuth-Morris-Pratt dan Algoritma Boyer-Moore
5	Mandala, dkk(2004)	-	Algoritma Othman
6	Mubarokah Syahrana(2014)	Java	Algoritma Porter dan Algoritma Knuth Morris Pratt

1.6 Graphic User Interface(GUI)

Perancangan Graphic User Interface sangat diperlukan sebagai acuan program yang akan dibuat. Gambar 1.1 adalah GUI dari aplikasi porter *stemming* bahasa indonesia.

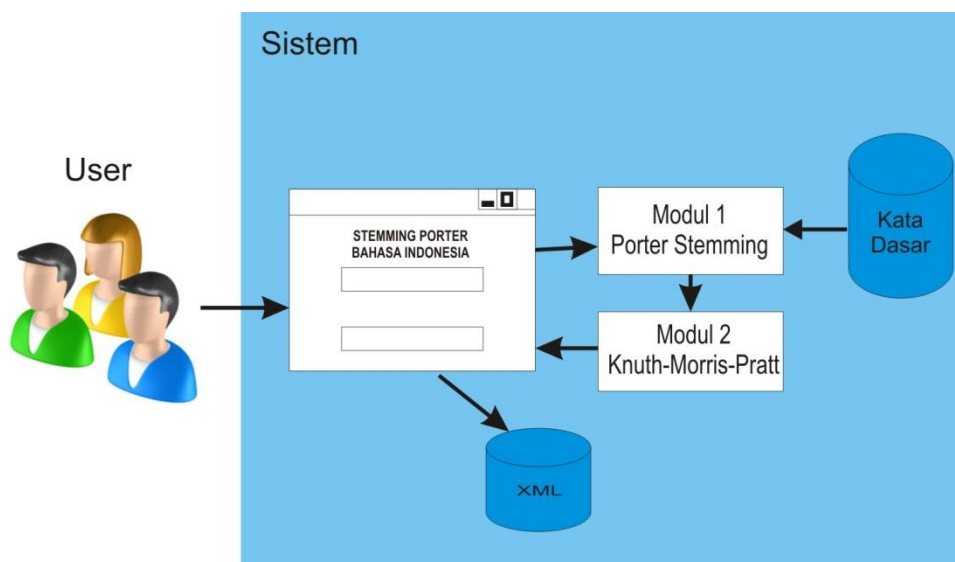
The screenshot shows a graphical user interface for a Porter stemming application. The main window is titled "STEMMING PORTER". It contains the following elements:

- Input:** A text box labeled "Masukan Kata" for entering the word to be stemmed.
- Action:** A "PROSES" button to initiate the stemming process.
- Output:** Two text boxes for displaying results: "Hasil Porter Stemming" and "Hasil Implementasi Algoritma Knuth-Morris-Pratt".
- Export:** "Save to:" options for "SQL" and "XML" to save the results.
- File Management:** A "Direktori file" section with a "Browse" button and another "PROSES" button.
- Status:** A label "Jumlah Kata : 0" indicating the current word count.
- Table:** A table with four columns: "NO", "Kata Berimbuhan", "Hasil Porter Stemming", and "Hasil Knuth-Morris-Pratt".
- Additional:** A "toTXT" button at the bottom right.

Gambar 1.1 Rancangan Aplikasi

1.7 Arsitektur program

Dalam aplikasi *stemming* dengan penerapan algoritma Knuth Morris Pratt pada Algoritma Porter for Bahasa Indonesia ini dapat digambarkan alur program yang berjalan pada gambar 1.2.

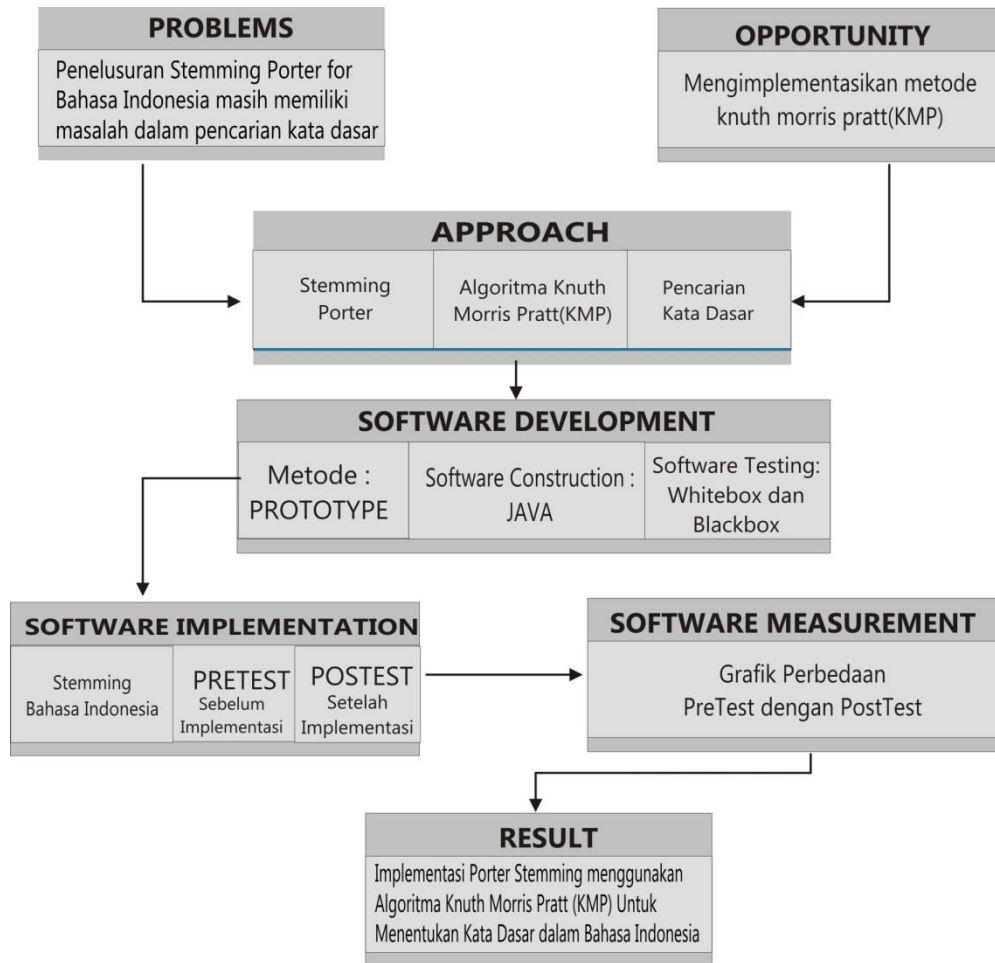


Gambar 1.2 Arsitektur program

Gambar 1.2 mendeskripsikan arsitektur antara pengguna dan sistem. User digambarkan sebagai pengguna aplikasi. Kemudian sistem yang terdiri dari aplikasi porter *stemming* untuk mencari kata dasar, serta *database* untuk menyimpan data kata dasar dalam bahasa Indonesia.

1.8 Kerangka pemikiran

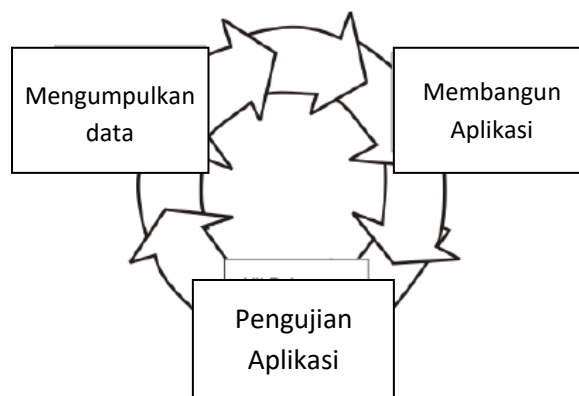
Kerangka pemikiran ini dibuat untuk mempermudah pembaca agar lebih mudah memahami dalam mempelajari alur dari penelitian. Adapun kerangka pemikiran dari “Implementasi Porter *Stemming* menggunakan Algoritma Knuth-Morris-Pratt (KMP) Untuk Menentukan Kata Dasar dalam Bahasa Indonesia” dapat dilihat pada gambar 1.3.



Gambar 1.3 Kerangka Pemikiran

1.9 Metode Pengembangan Perangkat Lunak

Adapun metode pengembangan yang digunakan dalam pembuatan aplikasi ini dapat dilihat pada gambar 1.4.



Gambar 1.4 Prototipe Paradigma [11]

Proses-proses tersebut dapat dijelaskan

a. Pengumpulan kebutuhan

Mencari semua artikel, jurnal dan seminar yang berhubungan dengan Porter *stemming* dan Algoritma Knuth Morris Pratt. Serta kebutuhan yang diketahui dan gambaran bagian-bagian yang akan dibutuhkan berikutnya;

b. Perancangan

Perancangan dilakukan cepat dan perancangan aspek software yang diketahui, dan rancangan ini menjadi dasar pembuatan *prototype*;

c. Evaluasi *Prototype*

Klien mengevaluasi *prototype* yang dibuat untuk memperjelas kebutuhan *software*.

1.10 Sistematika Penulisan

Untuk memberikan gambaran yang jelas mengenai penulisan skripsi, maka ditetapkan sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Dalam bab ini dijelaskan deskripsi umum dari isi skripsi yang meliputi latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi penelitian dan sistematika penyusunan.

BAB II LANDASAN TEORI

Dalam bab ini dijelaskan mengenai teori-teori yang berhubungan dengan masalah yang dikemukakan pada laporan ini.

BAB III ANALISIS DAN PERANCANGAN SISTEM

Dalam bab ini akan dijelaskan semua analisis dan perancangan yang berhubungan dengan Analisis dan Perancangan Aplikasi.

BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM

Dalam bab ini berisikan cara-cara penyajian teknik implementasi serta pengujian sistem, termasuk *preview* dari hasil akhir pada analisis dan perancangan aplikasi.

BAB V PENUTUP

Dalam bab ini dijelaskan mengenai hasil kesimpulan dari pembahasan yang telah dijelaskan dalam skripsi ini dan berbagai saran yang dapat berguna serta membantu untuk pembuatan laporan ke arah yang lebih baik lagi.

