BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Stemming adalah suatu proses pencarian bentuk dasar dari suatu *term*. Yang dimaksud dengan *term* itu adalah tiap kata yang berada pada suatu dokumen teks. *Stemming* diperlukan selain untuk memperkecil jumlah indeks yang berbeda dari suatu dokumen, juga untuk melakukan pengelompokan kata-kata lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk atau *form* yang berbeda karena mendapatkan imbuhan yang berbeda, metode *stemming* adalah salah satu cara yang digunakan untuk mengubah kata untuk menemukan akar kata dengan menerapkan aturan morfologi bahasa yang baik dan benar. Proses *stemming* dilakukan dengan menghilangkan semua imbuhan (*affiks*) baik yang terdiri dari awalan (*prefiks*) sisipan (*infiks*) maupun akhiran (*suffiks*) dan kombinasi awalan dan akhiran (*konfiks*)[1,6,7,9,10].

Beberapa algoritma yang termasuk kedalam metode *stemming* yaitu Algoritma Nazief Adriani,Algoritma Porter,Algoritma Arifin Setiono,Algoritma *Confix Stripping* (CS), Algoritma Vega, Algoritma *Enhanded Confix Stripping* (ECS),Algoritma *Connected Component.*Algoritma Porter adalah algoritma stemming yang digunakan untuk stemming dokumen teks berbahasa Inggris namun pada penelitian ini Algoritma Porter digunakan pada dokumen text berbahasa Indonesia, dari penelitian dalam lingkup analisis performansi algoritma ini maka menjadi dasar penelitian ini.

Terdapat penelitian sebelumnya mengenai perbandingan Algoritma Porter yang berjudul "Perbandingan Algoritma Stemming Porter dan Algoritma Stemming Adriani Nazief Untuk Stemming Dokumen Teks Bahasa Indonesia" oleh Ledy Agusta dalam jurnalnya yang menganalisis perbandingan pada dokumen berbahasa Indonesia. Berdasarkan hasil penelitian tersebut, disimpulkan bahwa algoritma porter lebih baik dalam hal kecepatan waktu namun memiliki kelemahan dalam hal keakuratan[1].

Penelitian lainnya dari algoritma Arifin Setiono yang berjudul “Pengaruh Stemming Kata Dalam Peningkatan Untuk Kerja Dokumen Clustering Untuk Dokumen Teks Berbahasa Indonesia”[4]. Algoritma Arifin Setiono digunakan karena memiliki kelebihan dalam hal mengatasi *Overstemming* yaitu jika kata tidak ditemukan setelah penghapusan maka algoritma ini kemudian mencoba untuk mengembalikan semua kombinasi yang dihapus untuk mendapatkan kata yang valid [5].

Dalam penelitian ini akan dilakukan analisis performansi pada dokumen dengan menggunakan metode *stemming.* Berdasarkan hal tersebut maka akan dilakukan analisis performansi metode *stemming* dengan menggabungkan dari kedua algoritma Arifin setiono dan Porter yang nantinya akan diterapkan pada dokumen berbahasa Indonesia agar menghasilkan hasil yang lebih baik. Parameter yang akan diuji yaitu waktu proses,akurasi yang berpengaruh pada presentasi algoritma yang di implementasikan.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang diuraikan, maka dapat dirumuskan masalah dalam penelitian ini yaitu Bagaimana performansi penggabungan Algoritma Arifin Setiono dan Algoritma Porter.

1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah untuk mengimplementasikan dengan menggabungkan Algoritma Arifin Setiono dan Algoritma Porter dalam mentransformasi kata-kata dalam sebuah dokumen teks berbahasa indonesia.

Sedangkan tujuan yang akan dicapai dalam penelitian adalah untuk mengetahui performansi berupa kecepatan dan akurasi dari algoritma Arifin Setiono dan Algoritma Porter dengan metode Stemming yang diimplementasikan.

1.4 Batasan Masalah

Batasan masalah dari penelitian ini yaitu:

1. Penerapan algoritma pada sistem hanya memproses dokumen berbahasa Indonesia.
2. Parameter yang akan di hasilkan pada aplikasi ini adalah kecepatan,ketepatan, dan jumlah langkah dari kedua algoritma.
3. Dokumen yang akan digunakan adalah dokumen jurnal berbahasa Indonesia berextensi *.txt.
4. Keluaran sistem yang diharapkan berupa kata dasar yang sesuai dengan kamus bahasa Indonesia.
5. Kamus sebagai pembanding kata yang di *stemming*.

1.5 Metodologi Penelitian

Metode yang digunakan dalam penelitian ini adalah metode deskriptif. Tujuan dari penelitian deskriptif ini adalah untuk membuat deskripsi, gambaran atau lukisan secara sistematis, faktual dan akurat mengenai fakta-fakta, sifat-sifat serta hubungan antarfenomena yang diselidiki.

1.5.1 Tahap Pengumpulan Data

Pengumpulan data adalah cara atau prosedur yang sistematis dan standar untuk memperoleh data yang diperlukan. Adapun metode pengumpulan data pada penelitian ini yaitu studi pustaka dengan cara mengkaji buku, media, atau hasil penelitian orang lain yang memiliki keterkaitan dengan penelitian ini.

1.5.2 Metode Pembangunan Perangkat Lunak

Teknik analisis data dalam pembuatan perangkat lunak menggunakan paradigma perangkat lunak secara *prototype,* yang meliputi beberapa proses diantaranya:

1. *Communication*

   Komunikasi antara *developer* dan *customer* mengenai tujuan pembuatan dari *software*, mengindetifikasi apakah kebutuhan diketahui.

2. *Quick Plan*

   Perancangan cepat setelah terjalin komunikasi antara *developer* dan *customer*.

3. *Modeling, Quick design*

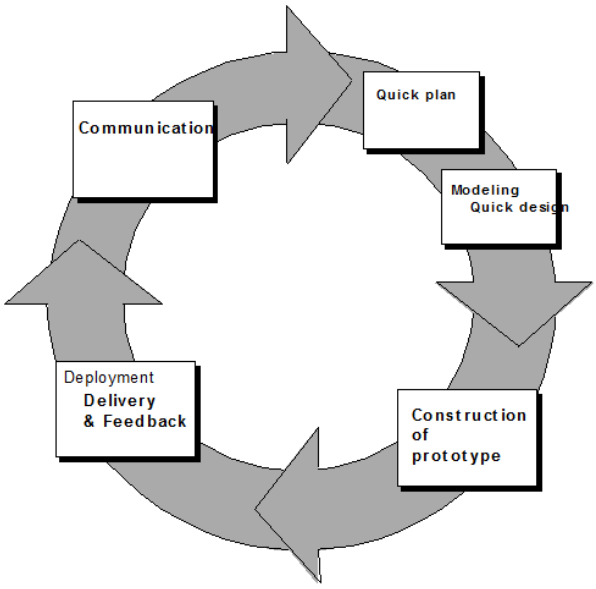
   Segera membuat model, dan *quick design* fokus pada gambaran dari segi *software* apakah *visible* menurut customer.

4. *Construction of Prototype*

   Quick design menuntun pada pembuatan dari *prototype*

5. *Deployment, Delivery & Feedback*

   Prototype yang dikirimkan kemudian dievaluasi oleh *customer*, *feedback* digunakan untuk menyaring kebutuhan untuk software.



Gambar 1.1 Metode perancangan perangkat lunak *prototype*[14]

1.6 Sistematika Penulisan

Sistematika penulisan proposal penelitian ini disusu untuk memberikan gambaran umum tentang penelitia yang dijalankan. Sistematika penulisan tugas akhir ini adalah sebagai berikut :

BAB 1 PENDAHULUAN

Bab ini menjelaskan secara singkat mengenai latar belakang masalah, identifikasi masalah, maksud dan tujuan, metodologi penelitian, batasan masalah, serta sistematika penulisan.

BAB 2 LANDASAN TEORI

Pada bab ini berisi teori – teori yang melatarbelakangi penulisan tugas akhir ini, yaitu teori tentang stemming,dan khususnya Algoritma Arifin Setiono dan Algoritma Porter.

BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Bab ini menerangkan analisis kecepatan,keakuratan,dan jumlah langkah kedua algoritma pada aplikasi stemming dengan pendekatan analisis UML. Selain itu terdapat juga perancangan untuk aplikasi yang akan dibangun sesuai dengan analisis yang sedang dibuat.

BAB 4 IMPLEMENTASI SISTEM DAN PENGUJIAN

Pada bab ini berisi tentang analisis kebutuhan dalam membangun aplikasi ini yang sesuai dengan metode pembangunan perangkat lunak yang digunakan. Selain itu terdapat juga perancangan antarmuka dan pengujian untuk aplikasi yang akan dibangun sesuai dengan hasil analisis.

BAB 5 KESIMPULAN DAN SARAN

Bab ini terdiri dari kesimpulan dan saran yang berisikan hal-hal terpenting yang penulis bahas dan kemudian dibuat kesimpulan. Bab ini juga berisi saran-saran yang penulis berikan untuk pengembangan selanjutnya.