

## BAB I

### PENDAHULUAN

#### 1.1 Latar Belakang

Ketersediaan informasi di internet saat ini sudah semakin banyak bahkan tidak hanya pencarian informasi, internet sudah menjadi media penghubung aktivitas manusia di seluruh dunia. Kehadiran internet memberikan dampak positif jika digunakan dengan baik dan sesuai kebutuhan. Seiring dengan berkembangnya internet dan jumlah pengguna internet semakin bertambah mengakibatkan banyaknya informasi yang tertampung dalam internet. Oleh sebab itu muncul masalah baru yaitu menemukan informasi yang tepat dan berguna menjadi hal yang sulit karena internet sangat luas. Kesulitan ini akhirnya menjadi alasan berkembangnya teknologi *search engine*, yaitu *information retrieval system* yang digunakan khusus untuk mencari informasi yang tepat dan berguna pada suatu *website* tertentu (Baeza-Yates, 2011). Dalam penggunaan *web search engine* populer seperti Google, Yahoo!, Bing, atau *web search engine* lainnya sering menemukan masalah yang dikenal dengan masalah daftar peringkat dan ambiguitas semantik dari query yang dikirimkan pengguna ke mesin pencari. Kebanyakan query yang dipakai sangat pendek, hanya terdiri dari satu atau dua istilah, sehingga menyebabkan mesin telusur hampir tidak dapat menebak apa yang sebenarnya dicari pengguna (Jansen, Spink, & Saracevic, 2000, 2009 ; Jansen & Spink, 2006).

Dashboard merupakan salah satu alat atau media yang dapat menjembatani pengguna dalam menemukan informasi. Dashboard dibuat untuk manajemen informasi seperti mengatur, menyimpan, dan menampilkan informasi penting dari berbagai sumber data dan diatur dalam satu tempat sehingga dapat diakses oleh pengguna (Few, 2006). Dalam praktiknya, dashboard banyak digunakan di berbagai kepentingan, salah satunya dalam bidang *information retrieval* dengan membangun *search engine*. Melalui dashboard tersebut pengguna dapat melakukan pencarian dan pengelolaan informasi yang dibutuhkan. Implementasi dashboard banyak digunakan untuk keperluan *business intelligence* seperti visualisasi

monitoring dan evaluasi kinerja karyawan, atau sebuah tampilan visual dari informasi terpenting yang dibutuhkan untuk mencapai tujuan tertentu, selain itu implementasi dashboard dalam sistem pencarian digunakan untuk menampilkan informasi yang sedang dicari oleh pengguna berdasarkan query yang diinputkan. Misalnya pada dashboard pencarian artikel, dimana dashboard tersebut dapat menampilkan informasi mengenai artikel yang berhubungan dengan query yang diinputkan pengguna seperti judul, nama pengarang, tahun terbit, abstrak, dan lain-lain. Sehingga dashboard menjadi penting kaitannya dengan mendapatkan informasi yang penting.

Web search engine merupakan program komputer yang berfungsi untuk membantu pengguna dalam mencari berkas – berkas yang tersimpan dalam layanan World Wide Web pada sejumlah jaringan komputer server sesuai dengan permintaan pengguna (**Hans, Rafael. 2018**). Proses pencarian dokumen biasanya dilakukan dengan cara melakukan pembobotan dokumen. Salah satu cara yang sering digunakan dalam pembobotan dokumen adalah TF-IDF. Menurut (**Nugroho dkk, 2014**) pada penelitiannya memaparkan bahwa algoritma TF-IDF menghasilkan skor uji coba yang lebih tinggi dibandingkan dengan algoritma lainnya. *Stemming* adalah salah satu bagian dari aktivitas yang terdapat di dalam sebuah *search engine*, *Stemming* digunakan untuk menyederhanakan kata ke bentuk kata dasarnya. Tentu proses *stemming* tiap bahasa berbeda dengan bahasa lainnya. *Stemming* pada bahasa indonesia sebagian besar memiliki dependensi terhadap kamus. Prosesnya pun tergolong rumit secara morfologis karena hampir setiap kata memiliki variasi imbuhan yang sangat beragam dengan metode penyederhanaan yang berbeda- beda (**Suhartono, 2014**). Pada teks berbahasa Indonesia terdapat beberapa algoritma stemming yang dapat digunakan seperti algoritma Nazief dan Adriani, algoritma Vega, algoritma Idris dan Mustapha, algoritma Arifin dan Setiono, algoritma Tala, algoritma *Confix Stripping* dan algoritma *Enhanced Confix Stripping* (**Ningrum, B. C., 2019**).

Algoritma *Enhanced Confix Stripping* merupakan salah satu algoritma stemming pada bahasa indonesia yang dikembangkan dari algoritma sebelumnya

yaitu algoritma Confix Stripping dengan menambahkan beberapa aturan untuk meminimalisir kesalahan proses stemming pada algoritma Confix Stripping. Pada penelitian sebelumnya, Perbandingan algoritma stemming untuk bahasa Indonesia dengan parameter akurasi dan waktu proses (**Ningrum, B. C., 2019**) memaparkan bahwa algoritma Enhanced Confix stripping merupakan algoritma yang cukup bagus dalam melakukan stemming kata berbahasa Indonesia dengan presisi sebesar 86% lebih baik dibandingkan algoritma lain seperti algoritma Nazief, algoritma Tala, algoritma Arifin, dan algoritma Vega. Algoritma Enhanced Confix Stripping menghasilkan kata hasil stemming yang cukup baik karena menggunakan perbandingan kamus kata dasar dalam prosesnya sehingga kecil kemungkinannya terjadi stemming yang berlebihan pada suatu kata.

Google Scholar adalah salah satu fitur yang dimiliki Google untuk memfasilitasi pencarian khusus bagi para pelajar, mahasiswa, dan akademisi untuk menemukan informasi yang bisa dipertanggungjawabkan. Mesin pencari Google pada umumnya menampilkan hasil pencarian atas suatu kata tertentu berdasarkan tingkat popularitas dan kecocokan dengan kata kunci. Tetapi Google Scholar lebih jauh menyaring hasil pencarian tersebut hanya pada sumber-sumber ilmiah yang dapat dipercaya kebenarannya, misalnya website universitas, perpustakaan, jurnal ilmiah, dan lain sebagainya. Google Scholar tidak hanya dapat digunakan untuk mencari data penelitian atau penulisan ilmiah yang telah diterbitkan sebelumnya, tetapi juga dapat digunakan oleh penulis atau peneliti untuk membuat profile secara otomatis. Sehingga mereka dapat membagikan tulisan atau jurnal penelitiannya agar dimanfaatkan oleh pelajar, mahasiswa, dan peneliti selanjutnya (**Andy, 2020**).

Berdasarkan uraian diatas, maka penulis tertarik untuk melakukan penelitian yang bertujuan untuk mendukung pengguna dalam mengoptimalkan pencarian informasi menggunakan web. Di penelitian ini akan dijelaskan mekanisme *stemming* dengan menggunakan algoritma *Enhanced Confix Stripping* yang digabungkan dengan proses pembobotan menggunakan algoritma TF-IDF (*Term Frequency Inverse Document Frequency*) dan diimplementasikan ke dalam sebuah dashboard website.

## 1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, maka permasalahan yang akan dibahas sebagai berikut :

1. Bagaimana mengimplementasikan algoritma *Enhanced Confix Stripping* pada *web search engine* untuk pemrosesan *query*?
2. Bagaimana proses membuat program pemrosesan *query* setelah melalui proses *stemming* menggunakan algoritma *Enhanced Confix Stripping* dalam bentuk dashboard?
3. Bagaimana implementasi program hasil pemrosesan *query* setelah melalui proses *stemming* menggunakan algoritma *Enhanced Confix Stripping* dengan metode pembobotan TF-IDF dalam bentuk dashboard?

## 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan penelitian ini adalah sebagai berikut :

1. Mengimplementasikan algoritma *Enhanced Confix Stripping* pada *web search engine* untuk melakukan pemrosesan *query* yang dimasukan oleh pengguna.
2. Membuat program pemrosesan *query* setelah melalui proses *stemming* menggunakan algoritma *Enhanced Confix* dalam bentuk dashboard.
3. Mengetahui implementasi program hasil pemrosesan *query* setelah melalui proses *stemming* menggunakan algoritma *Enhanced Confix Stripping* dengan metode pembobotan TF-IDF dalam bentuk dashboard.

## 1.4 Manfaat Penelitian

Manfaat dilakukannya penelitian ini yaitu:

1. Mengetahui pengimplementasian algoritma *Enhanced Confix Stripping* dalam melakukan proses *stemming* terhadap *query* yang dimasukan oleh pengguna saat mengakses *web search engine*.
2. Mengimplementasikan dashboard pencarian dengan *query* hasil pemrosesan menggunakan algoritma *Enhanced Confix Stripping* dan pembobotan TF-IDF

sehingga memudahkan pengguna untuk menemukan artikel atau dokumen yang dimaksud.

### 1.5 Batasan Masalah

Batasan masalah yang digunakan yaitu :

1. Sumber data yang digunakan pada skripsi ini adalah bersumber dari data Google Scholar dengan cakupan artikel, penelitian atau jurnal dalam bidang matematika.
2. Pencarian yang dilakukan harus berupa teks dalam bahasa Indonesia dan berhubungan dengan artikel, penelitian atau jurnal.

