# **BAB 1 Pendahuluan**

## **Latar Belakang**

Di dalam kehidupan sosial manusia disebut juga sebagai sebuah entitas, oleh karena itu manusia dianggap sebagai makhluk sosial. Manusia menggunakan bahasa atau bisa juga menggunakan dokumen sebagai media untuk berinteraksi maupun bersosialisasi. Bahasa adalah media yang digunakan untuk saling berkomunikasi maupun berinteraksi oleh sesama manusia, dan bahasa tidak akan pernah bisa lepas maupun hilang dari kehidupan manusia. Bahasa merupakan sebuah sistem lambang bunyi yang arbitrer, yang digunakan oleh masyarakat untuk bekerja sama, berinteraksi, dan mengidentifikasikan diri (Harimurti Kridalaksana, 2001:21).

Kata dokumen menurut Bahasa Inggris berasal dari kata *document* yang memiliki arti suatu yang tertulis atau tercetak dan segala benda yang mempunyai keterangan-keterangan dipilih untuk dikumpulkan, disusun, di sediakan atau untuk disebarkan. Kata dokumen ini menurut Louis Gottschalk (1986; 38) sering digunakan oleh para ahli sebagai:

* 1. Sebagai sumber yang tertulis bagi informasi sejarah sebagai kebalikan dari kesaksian lisan, artefak, peninggalan-peninggalan terlukis, dan petilasan-petilasan arkeologis.
  2. Sebagai sesuatu yang diperuntukan seperti surat-surat resmi dan surat-surat negara seperti surat perjanjian, undang-undang, hibah, konsesi, dan lainnya.

Pengertian dokumen dalam arti yang lebih luas merupakan proses yang didasarkan atas jenis sumber apapun, baik yang bersifat tulisan, lisan, gambaran, atau arkeologis. G.J. Renier (1997; 104) menjelaskan ada tiga pengertian yaitu:

* + - 1. Dalam arti yang luas, yaitu semua sumber tertulis maupun sumber lisan.
      2. Dalam arti yang sempit, yaitu semua sumber tertulis saja.
      3. Dalam arti yang spesifik, yaitu hanya surat-surat resmi dan surat-surat negara, seperti surat perjanjian, undang-undang, konsesi, hibah dan sebagainya.

Moleong (2007;216-217) menjelaskan bahwa dokumen adalah suatu *record* yang digunakan untuk membedakan suatu dokumen. Adapun definisi record adalah pernyataan tertulis yang disusun suatu lembaga / seseorang yang digunakan untuk pengujian suatu peristiwa atau menyajikan akunting.

Kesalahan penulisan pada dokumen adalah hal yang sering terjadi dalam lingkup mahasiswa, seperti kesalahan penulisan kata baku pada sebuah dokumen mahasiswa yang masih belum sesuai dengan kamus besar Bahasa Indonesia (KBBI). Contoh dokumen seperti laporan KP (Kerja Praktek), laporan MTP (Manajemen Tugas Proyek), dan laporan Metopen (Metodologi Penelitian), dari beberapa contoh tersebut masih sering terjadinya kesalahan-kesalahan ejaan khususnya bahasa tulis yang cenderung dipengaruhi oleh faktor dari penulisan yang kurang paham mengenai ejaan, kurang terbiasa menggunakan ejaan, maupun factor Lingkungan penulisan yang terbiasa menggunakan bahasa yang tidak baku (Gunawan dan Retnawati, 2017). Berikut beberapa contoh kesalahan penulisan kata baku seperti:

**Tabel 1.1** Contoh Penulisan Kata Baku

|  |  |
| --- | --- |
| **Tidak Baku** | **Baku** |
| selebritis | selebritas |
| ijasah | ijazah |
| cuman | cuma |
| pratikum | praktikum |

Pada dasarnya bahasa memegang peran yang sangat penting yang digunakan manusia untuk alat komunikasi di dalam kehidupan manusia. Bahasa bisa digunakan manusia untuk mengekspresikan pikiran dan perasaan, dan hampir di setiap kegiatan maupun aktifitas manusia selalu berhubungan dengan bahasa. Adapun pengertian bahasa adalah suatu kebutuhan yang sangat penting bagi manusia untuk mengirim pesan atau untuk menyampaikan amanat kepada orang lain. Bahasa sangat diperlukan untuk mencari sebuah informasi yang relevan dari sekumpulan informasi-informasi yang sudah tersedia atau biasa dikenal dengan *Information Retrieval* (sistem temu kembali informasi). Salah satu aplikasinya adalah mesin pencarian (*search-engine*) yang ada di *browser*. Mesin pencarian ini dapat digunakan oleh pengguna untuk mencari halaman web yang dibutuhkannya.

*Stemming* merupakan suatu cara yang digunakan untuk meningkatkan sebuah peforma dari *Information Retrieval* yaitu dengan cara mengubah kata atau kalimat yang ada di dalam dokumen ke kata dasarnya. S*temming* adalah sebuah proses untuk menghilangkan semua imbuhan yang ada pada kata maupun kalimat (*afiks*) baik dari awalan (*prefiks*), dari sisipan (*infiks*), maupun dari akhiran (*sufiks*), dan ada juga kombinasi dari awalan dan akhiran (*konfiks*). *Stemming* digunakan untuk merubah bentuk dari kata atau kalimat menjadi kata dasar yang baik dan benar sesuai dengan struktur morfologi bahasa. Metode s*temming* untuk setiap bahasa berbeda, sebagai contoh: Bahasa Inggris dan Bahasa Indonesia memiliki morfologi yang berbeda, sehingga metode *stemming* yang digunakan juga berbeda. Pada proses *stemming* untuk teks berbahasa Indonesia lebih rumit dan lebih kompleks karena terdapat berbagai variasi imbuhan yang harus dibuang untuk mendapatkan kata dasar dari sebuah kata atau kalimat. Untuk meminimalisir adanya kesalah pahamanan karena komunikasi yang kurang baik maka diperlukan penerjemah atau kamus besar Bahasa Indonesia (KBBI), karena kamus besar Bahasa Indonesia (KBBI) merupakan salah satu sarana dalam mempelajari Bahasa Indonesia yang baik dan benar. Ada beberapa hal yang perlu diketahui yaitu pada kamus besar Bahasa Indonesia (KBBI) tidak ada kata berimbuhan sehingga untuk mempelajari bahasa yang memiliki kata yang berimbuhan tersebut diperlukan adanya algoritma *stemming* yang digunakan untuk mencari kata dasar pada kata yang berimbuhan. Karena kata berimbuhan tidak bisa diterjemahkan secara langsung oleh kamus maka perlu perlakuan khusus yaitu salah satu alternatifnya adalah menggunakan algoritma *stemming*.

Algoritma yang sering digunakan untuk melakukan proses *stemming,* khususnya digunakan untuk *stemming* pada Bahasa Indonesia adalah seperti pada penelitian yang dilakukan Wahyu dkk. (2017) menerapkan tentang implementasi dan analisis algoritma *stemming* Nazief & Adriani dan Porter pada dokumen berbahasa Indonesia, pada penelitian ini menggunakan 2.132 kata yang diuji dan 2.031 *root word* yang benar dihasilkan dari Algortima Nazief & Adriani, sedangkan Algoritma Porter menghasilkan *root* *word* yang benar sebanyak 1.687 kata. Tingkat kesalahan *stemmer* pada Nazief & Adriani hanya 5%, sedangkan 21%pada Porter. Tetapi waktu proses yang diperlukan pada Nazief & Adriani hampir 2 kali lebih lama jika dibandingkan Algoritma Porter. Maka dapat disimpulkan bahwa hasil yang lebih baik untuk *stemming* pada dokumen berbahasa Indonesia adalah Algoritma Nazief & Adriani. Namun Algoritma Nazief & Adriani tidak lebih baik untuk efisiensi waktu proses dibanding Algoritma Porter.

Penelitian lainnya dilakukan oleh Novitasari. (2016) memaparkan penelitian tentang perbandingan *stemming* Algoritma Porter dengan Arifin Setiono untuk menentukan tingkat ketepatan kata dasar. Pada penelitian ini dilakukan uji coba menggunakan 40 sampel dokumen teks kata berimbuhan Bahasa Indonesia, tingkat katagori hasil *stemmer* *Exact* *Match* menghasilkan sebesar 90% untuk stemmer Porter dan 95% untuk *stemmer* Arifin Setiono, *Unchange* sebesar 2,5% untuk *stemmer* Porter dan 2,5% untuk *stemmer* Arifin Setiono, *Spelling* *Exception* sebesar 7,5% untuk stemmer Porter dan 0% untuk *stemmer* Arifin Setiono, dan *Overstemming* sebesar 0% untuk *stemmer* Porter dan 2,5% untuk *stemmer* Arifin Setiono. Dari proses *stemming* didapatkan hasil yang menunjukkan bahwa prosentase keakuratan menggunakan Algoritma Arifin & Setiono lebih tinggi dibandingkan menggunakan Algoritma Porter pada dokumen teks berbahasa Indonesia.

Berdasarkan latar belakang masalah di atas maka penelitian ini akan membuat sebuah aplikasi untuk mencari kata dasar dan kata baku yang sesuai dalam KBBI (Kamus Besar Bahasa Indonesia) pada dokumen teks berbahasa Indonesia menggunakan metode Nazief & Adriani. Metode Nazief & Adriani dipilih karena presentase keakuratan (presisi) lebih besar untuk dokumen teks berbahasa Indonesia dibandingkan dengan *stemming* menggunakan Algoritma *stemming* yang lain. Sebagai upaya untuk membantu pemahaman dan mengurangi kesalahan dalam penulisan kata baku, maka pada penelitian ini diambil judul “Aplikasi pencarian kata dasar dan kata baku pada dokumen berbahasa Indonesia menggunakan metode Nazief & Adriani”.

## **Identifikasi Masalah**

Berdasarkan permasaahan yang telah diuraikan pada latar belakang, maka identifikasi masalahnya adalah sebagai berikut:

* 1. Kesalahan penulisan kata baku yang dibuat mahasiswa pada sebuah dokumen berbahasa Indonesia.

## **Batasan Masalah**

* 1. Masukan atau inputan ke sistem berupa dokumen teks dengan huruf latin dalam Bahasa Indonesia.
  2. Sistem mengabaikan semua tanda baca seperti (koma, petik, tanda tanya, tanda seru, titik, dsb).
  3. Hanya melakukan lingkup *stemming* pada dokumen berbahasa Indonesia.

## **Rumusan Masalah**

* 1. Bagaimanakah penerapan *stemming* pada aplikasi pencarian kata dasar dan kata baku pada dokumen berbahasa Indonesia menggunakan Algoritma Nazief & Adriani.
  2. Bagaimanakah mengukur tingkat presisi pada aplikasi pencarian kata dasar dan kata baku pada dokumen berbahasa Indonesia menggunakan Algoritma Nazief & Adriani.

## **Tujuan Penelitian**

* 1. Penerapan *stemming* pada apikasi pencarian kata dasar dan kata baku menggunakan Algoritma Nazief & Adriani.
  2. Mengukur tingkat presisi pada aplikasi pencarian kata dasar dan kata baku pada dokumen berbahasa Indonesia menggunakan Algoritma Nazief & Adriani.

## **Manfaat Penelitian**

* 1. Sebagai tambahan informasi mahasiswa untuk media pembelajaran pada penulisan kata baku di sebuah dokumen berbahasa Indonesia.
  2. Mempermudah mahasiswa dalam pencarian kata dasar dan kata baku pada sebuah dokumen berbahasa Indonesia.
  3. Menambah referensi untuk mahasiswa tentang kata baku pada Bahasa Indonesia.