

STEMMING BAHASA INDONESIA SEBAGAI MEDIA BELAJAR SISWA SEKOLAH MENGGUNAKAN ALGORITMA PORTER

Dony Yudha Lestiyanto, A12.2009.03537
Sistem Informasi – S1
Universitas Dian Nuswantoro Semarang 2014

Abstrak

Stemming adalah salah satu cara yang digunakan untuk meningkatkan performa IR dengan cara mentransformasi kata- kata dalam sebuah dokumen teks ke kata dasarnya. Algoritma *Stemming* untuk bahasa yang satu berbeda dengan algoritma *stemming* untuk bahasa lainnya. Sebagai contoh Bahasa Inggris memiliki morfologi yang berbeda dengan Bahasa Indonesia sehingga algoritma *stemming* untuk kedua bahasa tersebut juga berbeda. Proses *stemming* pada teks berbahasa Indonesia lebih rumit/kompleks karena terdapat variasi imbuhan yang harus dibuang untuk mendapatkan root word dari sebuah kata. Beberapa algoritma *stemming* Bahasa Indonesia telah dikembangkan sebelumnya. Penggunaan algoritma *stemming* yang sesuai mempengaruhi performa sistem IR. Dalam penelitian ini penulis akan membahas mengenai algoritma *stemming* yaitu algoritma *Porter*.

Kata Kunci : Stemming, Algoritma, Porter, Bahasa Indonesia

1. PENDAHULUAN

1.1 Latar Belakang

Pencarian informasi berupa dokumen teks atau yang dikenal dengan istilah Information Retrieval (IR) merupakan proses pemisahan dokumen-dokumen yang dianggap relevan dari sekumpulan dokumen yang tersedia. Bertambahnya jumlah dokumen teks yang dapat diakses di internet diikuti dengan meningkatnya kebutuhan pengguna akan perangkat pencarian informasi yang efektif dan efisien. Efektif berarti user mendapatkan dokumen yang relevan dengan query yang diinputkan. Efisien berarti waktu pencarian yang sesingkat-singkatnya.

Stemming adalah salah satu cara yang digunakan untuk meningkatkan performa IR dengan cara mentransformasi kata- kata dalam sebuah dokumen teks ke kata dasarnya. Algoritma *Stemming* untuk bahasa yang satu berbeda dengan algoritma *stemming* untuk bahasa lainnya. Sebagai contoh Bahasa Inggris memiliki morfologi yang berbeda dengan Bahasa Indonesia sehingga algoritma *stemming* untuk kedua bahasa tersebut juga berbeda. Proses *stemming* pada teks berbahasa Indonesia lebih rumit/kompleks karena terdapat variasi

imbuhan yang harus dibuang untuk mendapatkan root word dari sebuah kata. Beberapa algoritma *stemming* Bahasa Indonesia telah dikembangkan sebelumnya. Penggunaan algoritma *stemming* yang sesuai mempengaruhi performa sistem IR. Dalam penelitian ini penulis akan membahas mengenai algoritma *stemming* yaitu algoritma *Porter*.

Algoritma *stemming* memiliki kelebihan dan kekurangannya. Efektifitas algoritma *stemming* dapat diukur berdasarkan beberapa parameter, seperti kecepatan proses, keakuratan, dan kesalahan. Dalam tulisan ini, penulis akan membahas efektifitas algoritma *Porter* untuk proses *stemming* pada teks berbahasa Indonesia, sehingga akhirnya akan diketahui mengenai tingkat kecepatan, keakuratan dan kesalahan *stemming*. Tujuan penelitian ini adalah untuk mengetahui kemampuan dan ketepatan algoritma *Porter* untuk proses *stemming* pada teks agar dapat membantu siswa sekolah dalam proses belajar bahasa Indonesia.

SMP Kesatrian 2 Semarang merupakan sekolah menengah tingkat pertama yang memiliki berbagai prestasi akademik baik di kota Semarang maupun tingkat provinsi dan nasional. Khusus dalam proses belajar

mengajar mata pelajaran Bahasa Indonesia mengenai kata dasar terkadang memiliki kendala, karena pengajaran dilakukan secara manual oleh guru dengan menuliskan materi pelajaran di papan tulis dan mencari kata dasar satu per satu melalui kamus besar bahasa Indonesia. Salah satu kendalanya adalah siswa akan kesulitan dan membutuhkan waktu yang lama dalam hal mencari kata dasar dari suatu kata dalam kamus besar bahasa Indonesia satu per satu, padahal jumlah kata dalam kamus jumlahnya mencapai puluhan ribu kata.

Untuk mengatasi masalah pada SMP Kesatrian 2 Semarang tersebut dibutuhkan sebuah aplikasi stemming Bahasa Indonesia yang dapat membantu siswa dan guru dalam hal proses belajar mengajar mencari sebuah kata dasar dari suatu kata tertentu secara cepat dan efisien.

Oleh karena itu dalam penelitian ini penulis akan mengambil judul "STEMMING BAHASA INDONESIA SEBAGAI MEDIA BELAJAR SISWA SEKOLAH MENGGUNAKAN ALGORITMA PORTER". Sistem ini direncanakan akan dibuat menggunakan bahasa pemrograman PHP, algoritma porter dan database MySQL.

1.2 Perumusan Masalah

Bagaimana merancang aplikasi *stemming porter* Bahasa Indonesia supaya efektif dan mampu menangani proses *stemming* (mencari kata dasar dari suatu kata) secara tepat dan akurat ?.

1.3 Tujuan Penelitian

Tujuan penulisan dalam penelitian adalah membuat aplikasi *stemming porter* Bahasa Indonesia supaya efektif dan mampu menangani proses *stemming* (mencari kata dasar dari suatu kata) secara tepat dan akurat sehingga didapatkan hasil *stemming* yang sesuai dan mampu di dalam memproses pencarian data, sehingga dapat membantu siswa sekolah dalam belajar Bahasa Indonesia dengan baik dan benar.

1.4 Manfaat Penelitian

Sebagai masukan dan bahan untuk perbaikan sistem pembelajaran mata

pelajaran Bahasa Indonesia yang berjalan dengan memanfaatkan kemajuan teknologi internet mengembangkan *stemming porter* Bahasa Indonesia secara *online*.

2. ANALISA DAN PERANCANGAN

2.1 Identifikasi Masalah

Identifikasi Masalah

Pada umumnya orang Indonesia itu sering menggunakan bahasa Indonesia tanpa menghiraukan susunan kata dan bagaimana penulisannya. Penulis mencoba menerapkan teknik *stemming* dalam bahasa Indonesia yang bisa digunakan oleh orang awam sekalipun. Sehingga bisa membantu mempermudah dalam pembelajaran bahasa Indonesia.

2.2 Alternatif Yang Diusulkan

Dengan ditemukan permasalahan yang ada, maka alternatif yang diusulkan adalah *stemming* bahasa Indonesia. Teknik *stemming* terdiri dari berbagai macam metode. Metode pertama yakni *stemming* dengan acuan tabel pemenggalan imbuhan. Proses *stemming* suatu term dengan metode ini dilakukan dengan cara menghilangkan imbuhan dari term tersebut sesuai dengan tabel acuan pemenggalan imbuhan yang digunakan. Metode kedua merupakan pengembangan dari metode pertama. Metode kedua ini selain menggunakan tabel acuan pemenggalan imbuhan, juga menggunakan suatu kamus kata dasar. Kamus kata dasar ini digunakan sebagai acuan hasil *stemming* saat proses pemenggalan imbuhan selesai dilakukan. Hasil dari proses *stemming* dengan metode ini harus ada pada kamus kata dasar, jika tidak maka term yang diinputkan dianggap sebagai bentuk dasar.

2.3 Perancangan Sistem

Untuk merancang sistem baru ini, penulis menggunakan metode pengembangan sistem *prototyping*. Sedangkan untuk melakukan proses analisa, penulis menggunakan metode analisa *abbot*. Analisa *abbot* adalah suatu cara (metode) untuk mengabstraksikan benda (objek) atau biasa disebut analisis sistem berorientasi objek. Terakhir dalam melakukan perancangan sistem berorientasi objek, penulis

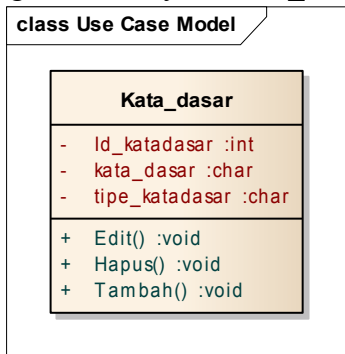
menggunakan alat bantu perancangan berupa diagram UML yang terdiri dari diagram *class*, diagram *use case* dan diagram *sequence*.

2.4 Narasi Sistem Baru

Narasi sistem baru dimulai dari pengguna memasukkan kata yang akan distemming, lalu untuk mengetahui kata dasarnya pengguna harus mengklik tombol Stemming pada web. Setelah itu akan muncul tabel hasil stemming dengan algoritma Porter. Untuk menambah kata dasar dapat mengklik tombol "Tambah" dan untuk menghapus kata dapat mengklik tombol "Hapus".

2.5 Diagram Class

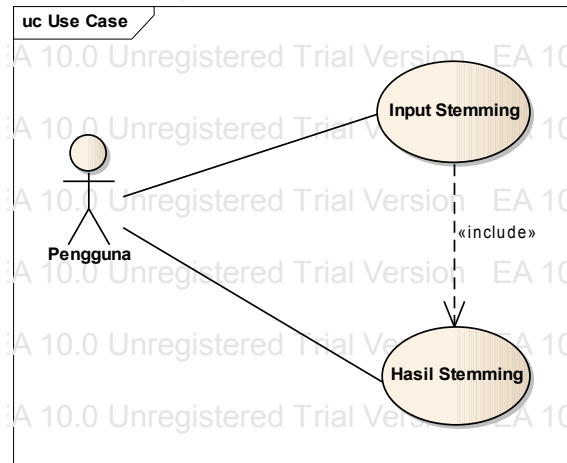
Dari gambar *Class Diagram* dibawah ini dapat dijelaskan bahwa *class diagram* mendeskripsikan jenis-jenis objek dalam sistem yang saling berelasi, dan nantinya digunakan untuk merancang database sistem. *Class* yang terbentuk yaitu kata_dasar.



Gambar 1 Diagram Class

2.6 Diagram Use Case

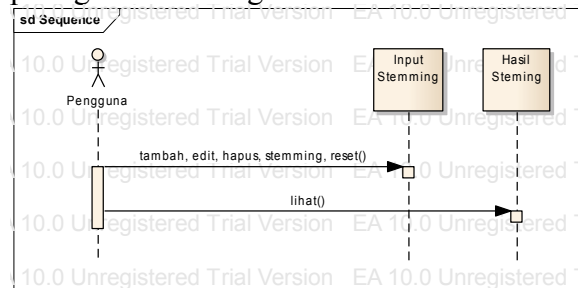
Gambar dibawah ini menjelaskan *graphical* dari beberapa atau semua actor, use case dan interaksi diantara komponen-komponen tersebut yang memperkenalkan suatu sistem yang akan dibangun.



Gambar 2 Diagram Use Case

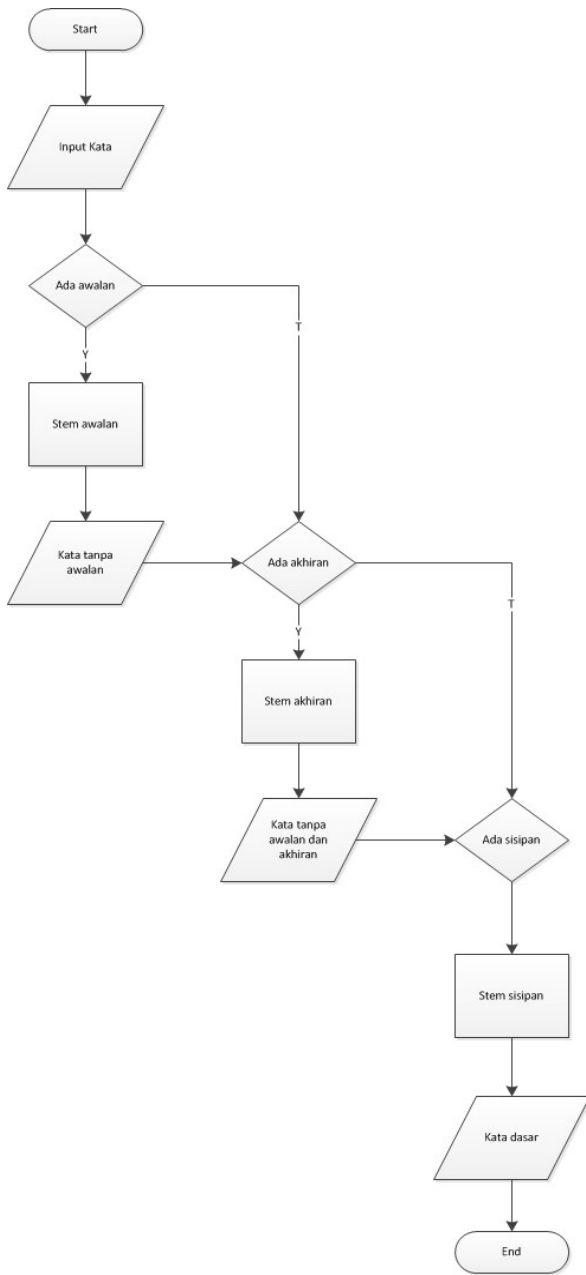
2.7 Diagram Sequence

Sequence diagram dibawah ini menunjukkan serangkaian pesan yang dipertukarkan oleh obyek-obyek yang melakukan suatu tugas atau aksi tertentu. Obyek-obyek tersebut kemudian diurutkan dari kiri ke kanan dan aktor di letakkan paling kiri dari diagram..



Gambar 3 Diagram Sequence

3. Flowchart Stemming Porter Bahasa Indonesia



Gambar 4 Flowchart Stemming Porter Bahasa Indonesia

4. Perancangan Antar Muka

Rancangan desain antar muka web ini dibuat untuk merancang tata letak konten web supaya terlihat menarik. Rancangan desain antar muka web tersebut adalah sebagai berikut :

a. Halaman Input Kata

Dalam rancangan halaman input kata ini terdiri dari beberapa konten yaitu *header* yang berisi *title bar*. Lalu dibawah *header*

terdapat konten form yaitu untuk mengisi kata yang akan distemming.

Gambar 5 Halaman Input Kata

b. Halaman Hasil Stemming

Halaman ini berfungsi menampilkan hasil stemming yang diinputkan oleh pengguna.

No	Inputan	Hasil Partikel	Hasil Pasifive Proses	Hasil Analisa 1	Hasil Analisa 2	Hasil Akhiran	Kata Dasar
1	menyagar	menyagar	menyagar	agar	agar	agar	agar
2	menemang	menemang	menemang	emang	emang	emang	emang
3	menemukul	menemukul	menemukul	emukul	emukul	emukul	emukul
4	menemukul	menemukul	menemukul	emukul	emukul	emukul	emukul
5	demas	demas	demas	emas	emas	emas	emas

Gambar 6 Halaman Output Laporan Surat Masuk

5. Implementasi

a. Halaman Input Kata



Gambar 7 Halaman Input Kata

Dalam halaman input kata ini terdiri dari beberapa konten yaitu *header* yang berisi *title bar*. Lalu dibawah *header* terdapat konten form yaitu untuk mengisi kata yang akan distemming.

b. Halaman Hasil Stemming

No	Inputan	Hasil Partikel	Hasil Pasifive Proses	Hasil Analisa 1	Hasil Analisa 2	Hasil Akhiran	Kata Dasar
1	menyagar	menyagar	menyagar	agar	agar	agar	agar
2	menemang	menemang	menemang	emang	emang	emang	emang
3	menemukul	menemukul	menemukul	emukul	emukul	emukul	emukul
4	demas	demas	demas	emas	emas	emas	emas

Gambar 8 Halaman Hasil Stemming

Halaman ini berfungsi menampilkan hasil stemming yang diinputkan oleh pengguna.

6. KESIMPULAN

Setelah penulis merancang dan membangun suatu stemming Bahasa Indonesia dengan menggunakan bahasa pemrograman PHP, algoritma Porter dan database MySQL, maka dapat diambil kesimpulan dari hasil penulisan skripsi ini seperti berikut :

1. Sistem ini dapat membantu siswa dan guru untuk belajar Bahasa Indonesia terutama mengenai pencarian kata dasar dari suatu kata dengan lebih cepat dan efisien.
2. Sistem ini dibangun dengan menggunakan bahasa pemrograman PHP, algoritma Porter dan database MySQL.
3. Sampel kata yang terdapat dalam kamus database aplikasi mencapai lebih dari 28000 kata..
4. Akurasi tingkat kebenaran stemming Bahasa Indonesia ini maksimal mencapai 86% keberhasilan dan sisanya 14% gagal.
5. Pencarian kata dasar manual tetap harus dilakukan khususnya pada kata-kata yang gagal dilakukan stemming.

7. SARAN

Saran-saran yang dapat penulis buat sesuai dengan berjalannya sistem baru yang telah diterapkan adalah :

1. Agar penerapan aplikasi stemming Bahasa Indonesia dapat berjalan dengan baik, maka dibutuhkan sumber daya manusia yang minimal menguasai komputer serta melakukan training atau pelatihan tentang cara penggunaan sistem ini.
2. Pemeliharaan dan penyempurnaan sistem harus terus dilakukan sesuai perkembangan dan kurikulum pelajaran agar sistem ini dapat berfungsi sesuai dengan apa yang diinginkan.
3. Diperlukan dukungan teknis yaitu beberapa perangkat keras (hardware), perangkat lunak (software), dan pengguna tersebut, serta dukungan non

teknis yaitu dukungan kedisiplinan user yang menangani sistem baru ini.

4. Data-data yang sudah lama sebaiknya perlu di backup guna menghindari kehilangan data bila terjadi kerusakan pada sistem atau pada perangkat keras.

DAFTAR PUSTAKA

- Abdul Kadir & Terra CH. Triwahyuni. (2005) *Pengenalan Teknologi Informasi*, Yogyakarta, Penerbit Andi.
- Bambang Hariyanto, Ir., MT. (2004) *Rekayasa Sistem Berorientasi Objek*, Bandung, Penerbit Informatika
- Edhy Sutanta, (2003) *Sistem Informasi Manajemen*, Yogyakarta, Penerbit Graha Ilmu.
- Edward, (1973), *Techniques of Attitude Scale Constructi-ons*. New York: Appleton-Century-Crofts, Inc.,
- Raymond Mcleod, Jr. (2001) *Sistem Informasi Manajemen Edisi ketujuh*, Jakarta, Penerbit PT. Prenhallindo