

APLIKASI PENCARIAN KATA DASAR DOKUMEN BERBAHASA INDONESIA DENGAN METODE *STEMMING* PORTER MENGGUNAKAN PHP & MYSQL

*Dini Nopiyanti*¹
*Kemal Ade Sekarwati*²

^{1,2}*Jurusan Sistem Informasi, FIKTI, Universitas Gunadarma*
¹*dini.nopiyanti@gmail.com,* ²*ade@staff.gunadarma.ac.id*

Abstrak

Pencarian informasi yang relevan dari suatu kumpulan informasi memerlukan proses stemming. Stemming adalah proses menggabungkan atau memecahkan setiap varian-varian morfologi suatu kata menjadi kata dasar. Berdasarkan struktur dasar morfologi kata, stemming Porter terlihat cocok untuk melakukan pencarian kata dasar pada dokumen berbahasa Indonesia. Untuk kebutuhan tersebut dibuatlah aplikasi untuk mencari kumpulan kata dasar yang sesuai dalam Kamus Besar Bahasa Indonesia (KBBI) dari dokumen berbahasa Indonesia dengan menggunakan metode stemming Porter. Implementasi aplikasi dilakukan menggunakan bahasa pemrograman Php (Hypertext Preprocessor) dan basis data server MySql (My Structure Query Language). Dalam program terdapat 5 tahapan stemming Porter untuk menghapus imbuhan pada kata. Pengujian dilakukan pada 20 artikel dokumen Prosiding Kommit 2012 untuk menilai keberhasilan aplikasi dalam mencari kumpulan kata dasar. Hasil yang didapat dari pengujian tersebut adalah aplikasi dapat beroperasi dengan baik dalam mencari kumpulan kata dasar bahasa Indonesia.

Kata Kunci : Bahasa Indonesia, Dokumen, Kata Dasar, Porter Stemmer, Stemming Porter.

PENDAHULUAN

Melakukan pencarian informasi yang relevan dari suatu kumpulan informasi yang tersedia dikenal dengan Sistem Temu Kembali Informasi (*Information Retrieval*). Salah satu aplikasi umum dari sistem temu kembali informasi adalah *search-engine* atau mesin pencarian yang terdapat pada jaringan internet. Pengguna dapat mencari halaman-halaman web yang dibutuhkannya melalui mesin tersebut.

Stemming adalah salah satu cara yang digunakan untuk meningkatkan performa *Information Retrieval* dengan cara mentransformasi kata-kata dalam sebuah dokumen teks ke kata dasarnya. Proses *stemming* dilakukan dengan menghilangkan semua imbuhan (afiks) baik yang terdiri dari awalan (prefiks) sisipan (infiks) maupun akhiran (sufiks) dan kombinasi dari awalan dan akhiran (konfiks). *Stemming* ini digunakan untuk mengganti bentuk dari suatu kata menjadi kata dasar sesuai dengan

struktur morfologi bahasa yang baik dan benar.

Metode *Stemming* untuk bahasa yang satu berbeda dengan metode *stemming* untuk bahasa lainnya. Sebagai contoh bahasa Inggris memiliki morfologi yang berbeda dengan bahasa Indonesia sehingga metode *stemming* untuk kedua bahasa tersebut juga berbeda. Proses *stemming* pada teks berbahasa Indonesia lebih rumit atau kompleks karena terdapat variasi imbuhan yang harus dibuang untuk mendapatkan kata dasar (*root word*) dari sebuah kata.

Berdasarkan permasalahan yang telah diuraikan diatas maka penulis membuat aplikasi *stemming* untuk mencari kumpulan kata dasar yang sesuai dalam Kamus Besar Bahasa Indonesia (KBBI) dari dokumen berbahasa Indonesia dengan menggunakan metode *stemming* Porter. Metode Porter dipilih berdasarkan pertimbangan bahwa ide dasar metode *stemming* Porter terlihat cocok untuk struktur dasar morfologi kata dalam bahasa Indonesia. Selain itu, salah satu metode penyelesaian masalah *stemming* yang terbukti empiris untuk bahasa morfologis (bahasa Inggris) adalah *stemming* yang dibuat oleh Porter atau dikenal dengan nama *Porter stemmer*.

METODE PENELITIAN

Metode penelitian yang digunakan pada penulisan ini adalah metode SDLC (*System Development Life Cycle*). Tahapan perencanaan, penulis mendefinisikan masalah yang ada, yaitu mencari kata dasar dari kata-kata berimbuhan pada suatu dokumen berbahasa Indonesia. Tahapan analisa, penulis mengumpulkan berbagai informasi dalam proses pembuatan aplikasi. Seperti mempelajari berbagai sumber pustaka dan menganalisa kebutuhan yang diperlukan. Tahapan

perancangan, penulis membuat rancangan aplikasi yang terdiri dari perancangan struktur navigasi, perancangan tampilan aplikasi, perancangan basis data dan tahapan terakhir perancangan diagram *use case*. Tahapan implementasi, penulis melakukan implementasi aplikasi menggunakan bahasa pemrograman Php dan basis data MySQL. Pada tahapan uji coba, dilakukan pengujian terhadap 20 artikel dokumen Prosiding Kommit menggunakan *server local* pada komputer pribadi penulis.

HASIL & PEMBAHASAN

Perancangan Aplikasi

Proses perancangan aplikasi terdiri dari beberapa tahapan, yaitu perancangan struktur navigasi, perancangan tampilan aplikasi, perancangan basis data dan tahapan terakhir perancangan diagram *use case*.

Aplikasi pencarian kata dasar dokumen berbahasa Indonesia menggunakan struktur navigasi linear. Pada struktur navigasi linear ini pengguna dapat berpindah antar halaman secara *sequential* atau berurutan.

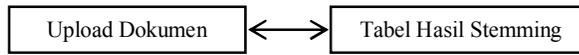
Aplikasi terdiri dari dua tampilan perancangan, yaitu perancangan tampilan halaman *upload* dan perancangan tampilan halaman hasil *stemming*. Halaman *upload* dokumen merupakan halaman yang pertama kali muncul ketika website dibuka. Pada halaman ini pengguna dapat memilih satu atau beberapa dokumen yang akan diunggah untuk dilakukan proses *stemming*. Halaman hasil *stemming* merupakan halaman yang menampilkan hasil *stemming* dalam bentuk tabel. Tabel terdiri dari nomor, nama file dan hasil *stemming*.

Basis data yang digunakan dalam aplikasi adalah basis data kamus. Kamus yang dimaksud merupakan kamus

bahasa Indonesia yang berisi tabel *tb_katadasar* yang memuat kata dasar sesuai dengan Kamus Besar Bahasa Indonesia (KBBI).

Diagram *use case* menggambarkan interaksi antara pengguna (*user*) dengan

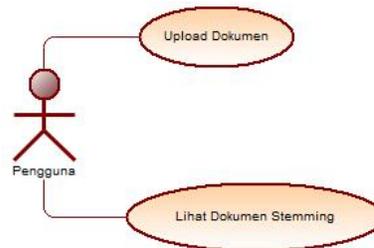
sistem aplikasi. Pada diagram *use case* aplikasi dapat dilihat bahwa pengguna dapat melakukan *upload* dokumen dan memilih link untuk menampilkan dokumen hasil *stemming*.



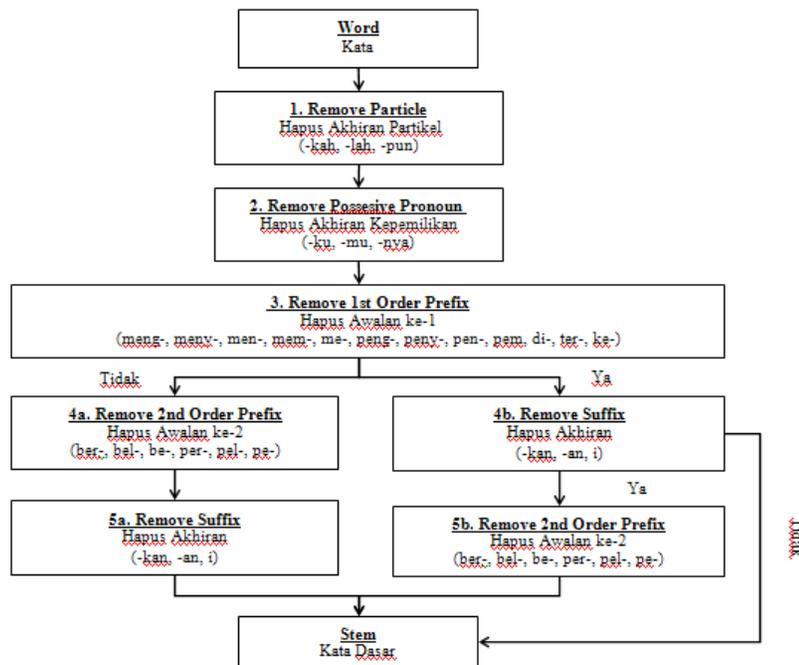
Gambar 1. Struktur Navigasi Aplikasi

Tabel 1. Struktur Tabel *tb_katadasar*

Field	Tipe	Not Null
<i>id_katadasar</i>	Integer(10)	√
<i>Katadasar</i>	Varchar(20)	√
<i>tipe_katadasar</i>	Varchar(20)	√



Gambar 2. Diagram *Use Case* Aplikasi



Gambar 3. Stemming Porter Bahasa Indonesia
Sumber : Tala (2013)

Tahapan-tahapan yang dilakukan dalam proses *stemming* Porter bahasa Indonesia yaitu :

- 1) Hapus akhiran partikel
- 2) Hapus akhiran kepemilikan.
- 3) Hapus awalan pertama. Jika tidak ada lanjutkan ke langkah 4a, jika ada cari maka lanjutkan ke langkah 4b.
- 4) a. Hapus awalan kedua, lanjutkan ke langkah 5a.
b. Hapus akhiran, jika tidak ditemukan maka kata tersebut diasumsikan sebagai kata dasar.

Jika ditemukan maka lanjutkan ke langkah 5b.

- 5) a. Hapus akhiran. Kemudian kata akhir diasumsikan sebagai kata dasar.
b. Hapus awalan kedua. Kemudian kata akhir diasumsikan sebagai kata dasar.

Hasil Pengujian

Aplikasi terdiri dari dua tampilan yaitu halaman *upload* dan halaman hasil *stemming*.



Gambar 4. Halaman *Upload*



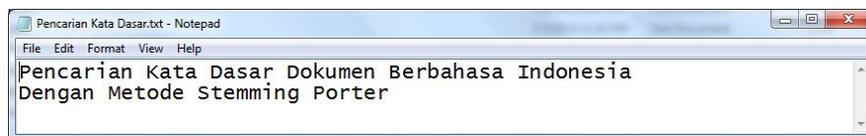
Gambar 5. Halaman Hasil *Stemming*

Pada halaman hasil *stemming* terdapat tabel hasil pencarian kata dasar dari dokumen yang diunggah. Kolom kata berisi asal kata sebelum kata dilakukan proses *stemming*, kolom kata dasar berisi kata dasar hasil proses *stemming*, dan kolom proses *stemming* berisi akhir proses *stemming* untuk

mendapatkan kata dasar. Dalam tabel terdapat pula keterangan **Link File**. Jika link tersebut dipilih maka halaman akan menampilkan dokumen hasil *stemming* yang berisi kumpulan kata dasar dari dokumen yang dipilih. Tampilannya yaitu sebagai berikut :



Gambar 6 Contoh Dokumen Hasil *Stemming*



Gambar 7 Contoh Dokumen Sebelum Diunggah dan Dilakukan Proses *Stemming*

Dari gambar 6, didapatkan kumpulan kata dasar bahasa Indonesia dari

dokumen 'Pencarian Kata Dasar.txt' yang telah diunggah. Apabila kata yang

dihasilkan dari proses *stemming* Porter bukan kata dasar bahasa Indonesia, maka kata akan diapit oleh tanda kurung (). Kata *stemming* bukanlah kata dasar bahasa Indonesia karena kata tersebut berasal dari kamus bahasa Inggris. Sedangkan kata Porter bukanlah kata dasar bahasa Indonesia karena kata tersebut merupakan nama orang.

Untuk menilai keberhasilan pencarian kata dasar dokumen berbahasa Indonesia dengan metode *stemming* Porter, dilakukan pengujian pada 20 artikel dokumen Prosiding Kommit 2012. Hasil pengujian dari 20 artikel tersebut terdapat pada tabel 2.

Berdasarkan pengujian pada tabel 2 hasil pencarian kata dasar yang ditemukan dari 20 artikel dokumen Prosiding Kommit adalah 16.025 kata dasar dari 19.212 total kata. Persentase terbesar kata dasar yang presisi ditemukan pada atikel ‘99. Penentuan

Keputusan untuk Membantu Program Genre Bagi Pasangan Muda.txt’ dengan persentase 92.98%. Sedangkan persentase terendah ditemukan pada artikel ‘88. Aplikasi Multimedia Audio Video Player dengan Menggunakan Visual Basic Net 2008’ dengan persentase 72.14%.

Aplikasi Pencarian Kata Dasar Dokumen Berbahasa Indonesia Dengan Metode *Stemming* Porter Menggunakan Php & MySql telah berhasil dibuat. Pada halaman hasil *stemming* aplikasi terdapat link yang dapat menampilkan dokumen hasil *stemming* yang berisi kumpulan kata dasar bahasa Indonesia.

Kata yang mengalami kesalahan proses *stemming*, tidak terdapat pada basis data kamus, maupun merupakan bagian dari kamus bahasa asing atau kamus lain diapit oleh tanda kurung () pada dokumen hasil *stemming*.

Tabel 2. Hasil Pengujian Pencarian Kata Dasar Dengan Metode *Stemming* Porter

No	Nama File	Total Kata	Kata Dasar	Bukan Kata Dasar	Presisi
1	76. Sistem Penunjang Keputusan Memilih Kriteria Lagu Pop Indonesia yang Baik.txt	655	600	55	91.60%
2	77. Evaluasi Algoritma Prim dan Kruskal Terhadap Pemasangan Kabel Telepon di DKI Jakarta.txt	749	617	132	82.38%
3	78. Aplikasi Pemetaan Pusat Perbelanjaan Kota Bekasi Menggunakan Android.txt	539	470	69	87.20%

4	79. Penerapan Algoritma Greedy pada Studi Kasus Pencarian Rumah Sakit Terdekat di Jakarta Selatan.txt	620	544	76	87.74%
5	80. Implementasi Algoritma Dijkstra Guna Optimasi Jalur Pendistribusian Produk Seluler.txt	434	348	86	80.18%
6	86. Aplikasi Pendidikan Bagi Anak di Bawah Umur 7 Tahun.txt	2,004	1,701	303	84.88%
7	87. Sistem Pencarian Fasilitas Umum Terdekat Menggunakan Augmented Reality dengan Minimum Spanning Tree.txt	865	689	176	79.65%
8	88. Aplikasi Multimedia Audio Video Player dengan Menggunakan Visual Basic Net 2008.txt	1,260	909	351	72.14%
9	89. Aplikasi Energy Usage Calculator untuk Menghitung Penggunaan dan Biaya Energi Listrik Berbasis Python Versi 323.txt	1,008	898	110	89.09%
10	90. Implementasi Algoritma Kruskal untuk Optimasi Pengangkutan Sampah.txt	733	623	110	84.99%
11	91. Pemilihan Aplikasi Permainan untuk Perkembangan Motorik dan Simbolik Anak Usia 1 - 7 Tahun.txt	767	700	67	91.26%
12	92. Sistem Informasi Geografis SMA di Bogor.txt	1,982	1,658	324	83.65%
13	93. Pembuatan Website World Watch Shop Menggunakan Magento Commerce.txt	634	535	99	84.38%
14	94. Pembuatan Aplikasi Pemetaan Tempat Usaha di Sekitar Kampus Depok Gunadarma Menggunakan Android 2.1.txt	922	771	151	83.62%
15	95. Aplikasi Penjualan Makanan Khas Garut pada Toko Aneka Sari dengan Menggunakan Visual Basic Net.txt	1,704	1458	246	85.56%
16	96. Pengembangan E-Government pada Layanan Informasi Publik Pemerintahan Daerah Sulawesi Barat Menuju Good Governance.txt	926	735	191	79.37%
17	97. Perlunya Penerapan Teknologi Web Semantik pada Situs Pencarian Lowongan Pekerjaan di DKI Jakarta.txt	565	484	81	85.66%
18	98. Program Aplikasi Enkripsi dan Dekripsi SMS pada Ponsel Berbasis Android dengan Algoritma DES.txt	718	591	127	82.31%
19	99. Penentuan Keputusan untuk Membantu Program Genre Bagi Pasangan Muda.txt	655	609	46	92.98%
20	100. Pembuatan Aplikasi Pembelajaran Bahasa Jerman Berbasis Mobile Android.txt	1,472	1085	387	73.71%
Total : 20 artikel		19,212	16,025	3,187	

SIMPULAN DAN SARAN

Simpulan

Berdasarkan hasil uji coba yang telah dilakukan, dapat disimpulkan bahwa Aplikasi Pencarian Kata Dasar Dokumen Berbahasa Indonesia Dengan Metode *Stemming* Porter Menggunakan Php & MySql telah berhasil dibuat. Hasil pencarian kata dasar yang ditemukan dari 20 artikel dokumen Prosiding Kommit adalah 16.025 kata dasar dari 19.212 total kata. Dari 20 artikel dokumen Prosiding Kommit didapatkan hasil 92.98% sebagai

persentase terbesar kata dasar yang presisi pada dokumen. Serta didapatkan hasil 72.14% sebagai persentase terendah kata dasar yang presisi pada dokumen. Perbedaan hasil persentasi didapat dari jumlah kata dasar yang berhasil ditemukan oleh aplikasi. Kata yang terdapat pada kolom bukan kata dasar pada halaman hasil *stemming* merupakan kata yang tidak berhasil menjadi kata dasar bahasa Indonesia akibat kesalahan proses *stemming*, tidak terdapat pada basis data kamus, maupun merupakan bagian dari kamus bahasa asing atau kamus lain. Pada dokumen

hasil *stemming*, kata yang tidak berhasil menjadi kata dasar bahasa Indonesia diapit oleh tanda kurung ().

Saran

Aplikasi pencarian kata dasar dokumen berbahasa Indonesia ini masih memerlukan pengembangan dan penyempurnaan. Salah satunya menambah kelengkapan basis data kamus serta memperbaiki kata yang tidak berhasil menjadi kata dasar akibat kesalahan proses *stemming*. Penulis berharap pembaca dapat mengembangkan aplikasi ini menjadi lebih baik di masa mendatang.

DAFTAR PUSTAKA

- Anhar. 2010. *Panduan Menguasai PHP & MySQL secara Otodidak*. Jakarta : Mediakita.
- E. Zaenal Arifin dan Junaiyah H. M. 2009. *Morfologi : Bentuk, Makna dan Fungsi* (Edisi Kedua). Jakarta : Gramedia Widiasarana Indonesia.
- Henky Prihatna. 2005. *Struktur Navigasi*. Jakarta : Elex Media Komputindo.
- Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schutze. 2009. *An Introduction to Information Retrieval* (Online Edition). Cambridge : Cambridge University Press (<http://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf> , diakses 14 Juli 2014).
- Rossa A. S dan M. Shalahuddin. 2013. *Rekayasa Perangkat Lunak*. Bandung : Penerbit Informatika.
- Stendy B. Sakur. 2010. *PHP 5 Pemrograman Berorientasi Objek - Konsep & Implementasi*. Yogyakarta : Penerbit Andi.
- Tala, F Z. 2003. Tesis “A Study of Stemming Effect on Information Retrieval in Bahasa Indonesia” (Online). Netherland : Universiteit van Amsterdam. (<http://dare.uva.nl/document/443347> , diakses 18 Juni 2014).
- URL
(http://id.wikipedia.org/wiki/Sistem_temu_balik_informasi , diakses 3 Juni 2014).