



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB III

ANALISIS DAN PERANCANGAN SISTEM

3.1 Metodologi Penelitian

Metode penelitian yang akan digunakan dalam penelitian ini antara lain:

1. Studi Literatur

Melakukan analisis teori dan konsep yang berkaitan dengan penelitian, seperti *information retrieval*, *web search engine*, algoritma *stemming*, *stop word* dan metode evaluasi Paice. Sumber yang digunakan berupa buku, penelitian terdahulu, jurnal ilmiah, artikel, dan lain – lain.

2. Perancangan Sistem

Melakukan perancangan pembuatan *stemmer* (aplikasi untuk melakukan *stemming*) menggunakan algoritma *stemming* ECS yang kemudian akan diimplementasikan ke *web search engine* situs Sindo Weekly.

3. Implementasi

Membuat aplikasi pengubah kata berimbuhan menjadi kata dasar atau disebut *stemmer* dan kemudian mengimplementasikannya ke dalam *search engine*. Setelah implementasi berhasil, kemudian mengimplemntasikan ke situs resmi Sindo Weekly.

4. Pengujian dan Evaluasi

Melakukan uji coba implementasi ke dalam *website* dan mengevaluasi hasil yang didapatkan.

3.2 Analisis Sistem

3.2.1 Stemmer

Untuk mengimplementasikan algoritma *stemming* dalam *search engine*, dibutuhkan *stemmer*. *Stemmer* merupakan aplikasi pemotongan imbuhan pada kata berimbuhan (awalan, akhiran, sisipan, kombinasi) yang dijalankan dengan algoritma tertentu (Persada, 2012). *Stemmer* ini berfungsi untuk mengubah kata berimbuhan menjadi kata dasar yang akan digunakan untuk *query* mesin pencarian. *Stemmer* dibuat menggunakan algoritma *stemming Enhanced Confix Stripping* seperti teori pada bab dua di atas. Setelah *stemmer* telah selesai dibuat, kemudian mengimplementasikannya ke dalam *search engine*.

Pada aplikasi *stemmer* ini, pengguna dapat memasukkan lebih dari satu kata yang akan dilakukan proses *stemming*. Kata tersebut akan dibagi menjadi satu persatu kata kemudian dilakukan proses *stemming* tersebut dan menampilkan hasilnya. Aplikasi ini tidak menampilkan tahapan proses pemenggalan kata hingga menjadi bentuk dasarnya, *output* hanya berupa hasil *stemming*-nya saja.

3.2.2 Sindo Search Engine

Sindo Search Engine merupakan aplikasi *web search engine* yang dibuat penulis untuk memudahkan analisis hasil implementasi algoritma. Dalam halaman situs Sindo Weekly Magazine telah terdapat fitur pencarian berupa *query* standar tanpa menggunakan algoritma tertentu. Hasil pencarian berupa edisi terbitan majalah tiap minggunya dan judul dari artikel yang dicari.

Dalam penelitian ini, penulis membuat aplikasi *web search engine* yang juga menampilkan potongan dari isi artikel dan jumlah halaman yang ditemukan. Aplikasi ini dibuat dengan tujuan untuk memudahkan dalam evaluasi hasil implementasi dengan nama *Sindo Search Engine*. Kata kunci pencarian akan dicetak tebal untuk menunjukkan kata yang dicari telah berhasil ditemukan. Setelah implementasi dari algoritma *stemming* ECS berhasil dilakukan pada *search engine* yang dibuat penulis, maka akan dilanjutkan dengan mengimplementasikannya ke dalam *website* Sindo Weekly Magazine, yaitu www.sindoweekly-magz.com.

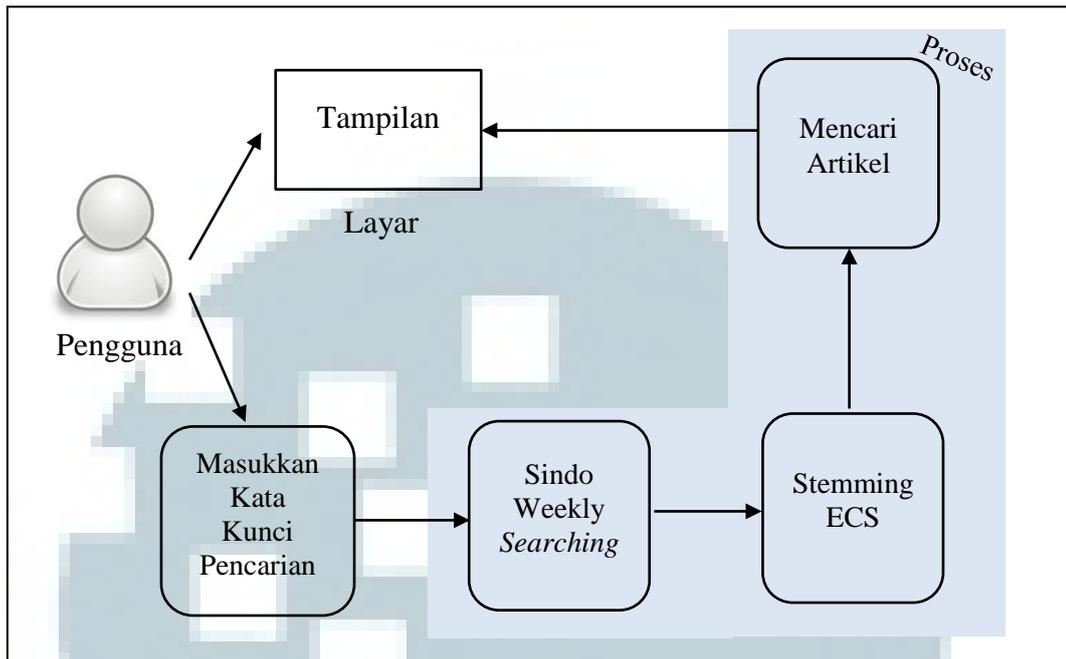
Pembuatan *stemmer* dan *search engine* menggunakan bahasa pemrograman PHP seperti pada yang digunakan pada *website* Sindo Weekly Magazine untuk memudahkan implementasi hasil *stemming* ke dalam fitur pencarian. Database yang digunakan menggunakan *mysql*.

3.3 Perancangan Sistem

Terdapat dua diagram dalam perancangan sistem ini, yaitu diagram umum, *flowchart*, struktur tabel, dan desain antarmuka. Berikut masing-masing penjelasannya.

3.3.1 Diagram Umum

Gambar 3.1 menunjukkan bagaimana alur interaksi sistem fitur pencarian dalam situs Sindo Weekly Magazine secara keseluruhan.

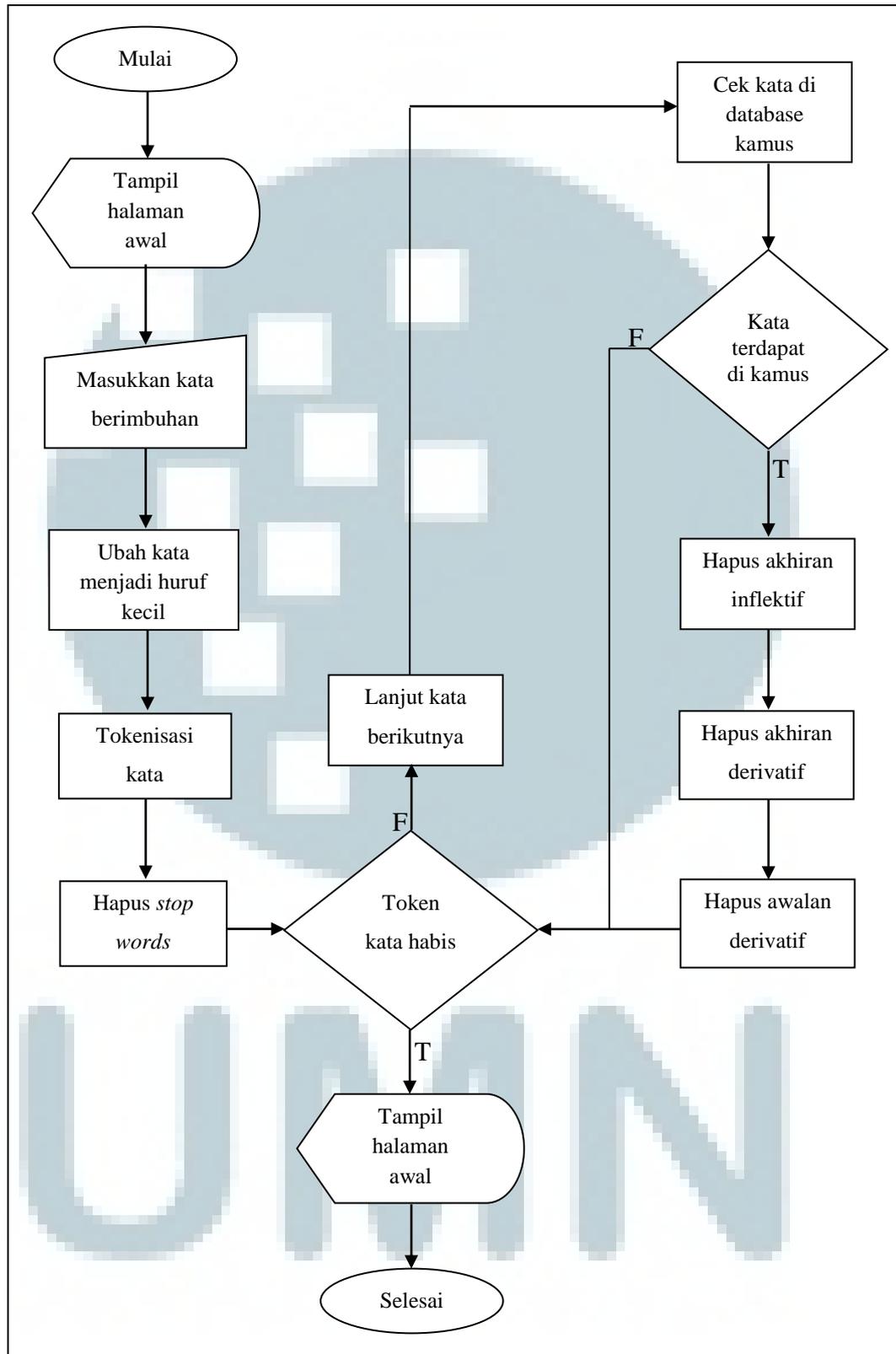


Gambar 3.1 Diagram umum sistem pencarian situs Sindo Weekly

Interaksi berawal dari pengguna menggunakan komputer mengakses situs Sindo Weekly Magazine, www.sindoweekly-magz.com. Pengguna memasukkan kata kunci di *form* pencarian. Kemudian dilakukan pemrosesan kata kunci dengan melakukan *stemming* ECS. Setelah proses *stemming* dilakukan, kemudian mencari artikel atau berita pada situs Sindo Weekly Magazine yang sesuai dengan kata kunci hasil *stemming* dan kata kunci asli yang dimasukkan pengguna. Pada akhirnya, hasil pencarian akan ditampilkan pada layar monitor komputer.

3.3.2 Flowchart

Terdapat dua diagram alir (*flowchart*) dalam penelitian ini yaitu *flowchart* ECS *stemmer* dan *flowchart* implementasi ECS *stemmer* ke dalam Sindo Search Engine.



Gambar 3.2 Flowchart ECS stemmer

Flowchart ECS stemmer pada gambar 3.2 di atas merupakan urutan kerja dari proses *stemming* aplikasi *stemmer* menggunakan algoritma *stemming* ECS.

Pertama kali aplikasi akan menampilkan form untuk memasukkan kata berimbuhan bahasa Indonesia. Kemudian setelah pengguna memasukkan kata (boleh lebih dari satu kata), kata tersebut akan diubah menjadi huruf kecil.

Setelah itu melakukan proses tokenisasi kata. Tokenisasi adalah proses pembagian kata-kata atau kalimat menjadi satu-satu kata. Misalnya "bekerja untuk kehidupan", akan dibagi menjadi "bekerja", "untuk", dan "kehidupan".

Kemudian dilanjutkan dengan menghapus *stop words*. Kata-kata yang tidak diperlukan dalam pencarian akan dihapus karena akan memperbanyak jumlah hasil pencarian tetapi tidak memiliki makna. Dalam hal ini kata "untuk" akan dihapus sehingga kata token hanya menjadi "bekerja" dan "kehidupan".

Setelah itu setiap token kata akan dilakukan proses *stemming* satu per-satu, kata "pemerintahan" akan diproses terlebih dahulu. Kemudian akan diperiksa apakah kata ada di kamus. Jika tidak ada maka akan dilanjutkan ke kata berikutnya, jika ada maka akan dilanjutkan menghapus akhiran inflektif.

Akhiran inflektif mencakup kata yang berakhiran "-lah", "-kah", "-ku", "-mu", atau "-nya". Karena kata "bekerja" tidak terdapat akhiran tersebut, maka proses akan berlanjut dengan menghapus akhiran derivatif.

Akhiran derivatif menghapus akhiran "-i", "-an", atau "-kan". Kata "bekerja" tidak memiliki akhiran tersebut sehingga tetap menjadi "bekerja". Jika terjadi proses pemotongan, maka akan dicek dalam kamus. Jika kata yang telah

dibuang akhiran tersebut terdapat dalam kamus, maka akan mengembalikan kata tersebut. Jika kata tidak terdapat dalam kamus, maka kata asal akan dikembalikan.

Proses akhir adalah menghapus awalan derivatif yaitu awalan “di-”, “ke-”, “se-”, “te-”, “be-”, “me-”, atau “pe-”. Kata "bekerja" memiliki awalan "be-" sehingga menjadi "kerja". Kemudian sistem akan memeriksa kata di *database* kamus dan kata "bekerja" ada dalam kamus. Oleh karena itu proses *stemming* untuk kata "bekerja" selesai menjadi "kerja".

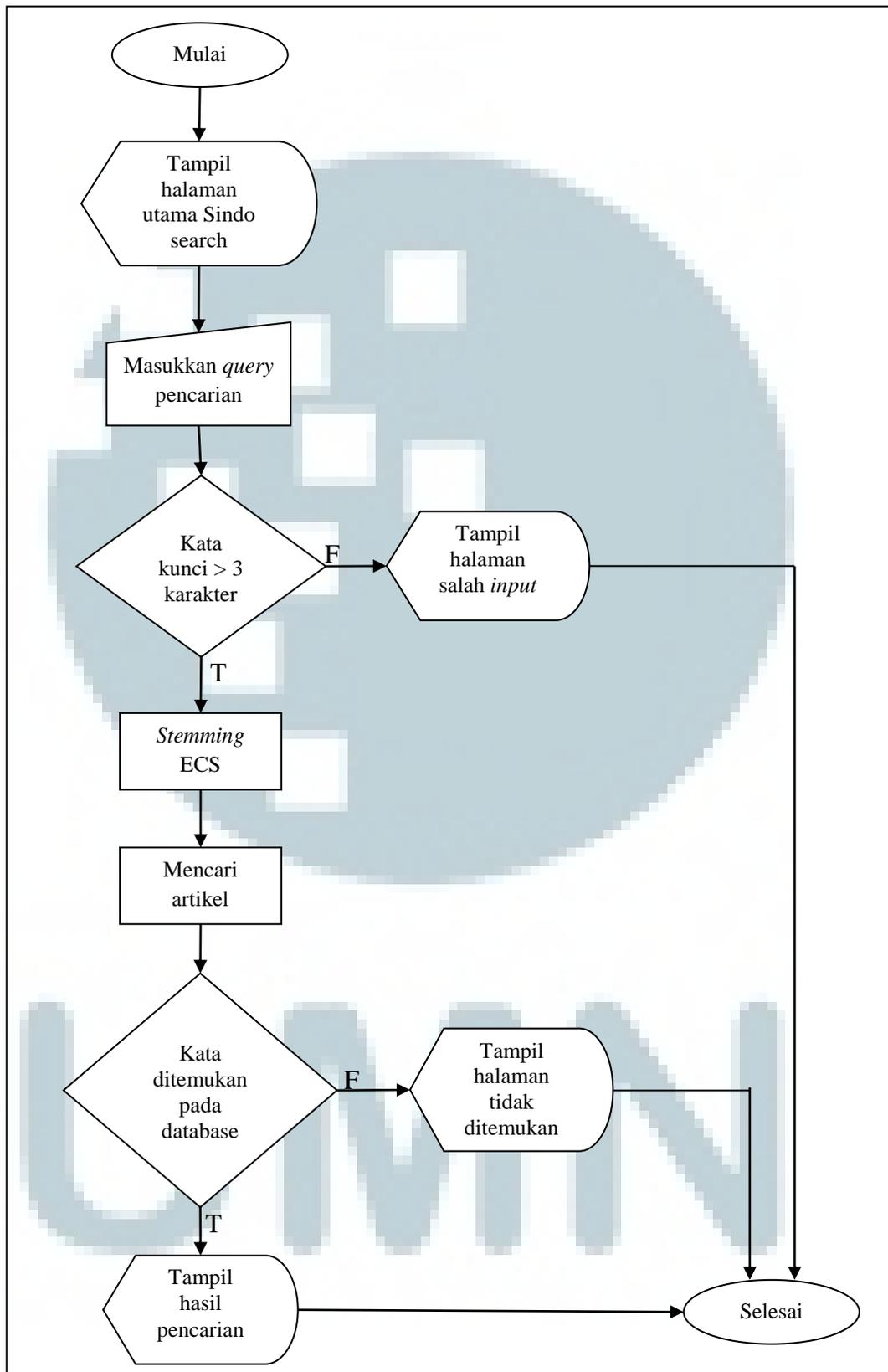
Karena token belum habis, kemudian menuju proses selanjutnya yaitu memproses kata "kehidupan" dengan urutan proses *stemming* seperti pada penjelasan di atas, sebagai berikut:

- a. Hapus akhiran inflektif ("-lah", "-kah", "-ku", "-mu", "-nya"): "kehidupan".
- b. Hapus akhiran derivatif ("-i", "-an", "-kan"): "kehidup".
- c. Hapus awalan (“di-”, “ke-”, “se-”, “te-”, “be-”, “me-”, “pe-”): "hidup".

Dengan demikian hasil proses *stemming* kata "kehidupan" menjadi "hidup".

Karena token telah habis, semua kata telah selesai diproses, maka layar akan menampilkan hasil proses *stemming* yaitu "kerja" dan "hidup".

Catatan: setiap proses pemenggalan kata mengikuti aturan pada tabel 2.1, 2.2, dan 2.3.



Gambar 3.3 Flowchart Sindo Search Engine

Flowchart Sindo Search Engine pada gambar 3.3 di atas merupakan urutan kerja dari proses mesin pencarian artikel yang dibuat dalam penelitian ini.

Pertama kali aplikasi akan menampilkan form untuk memasukkan kata kunci pencarian. Kemudian setelah pengguna memasukkan kata, akan dilanjutkan proses menghitung jumlah karakter yang dimasukkan jika kata yang dimasukkan kurang dari tiga karakter, maka akan menampilkan halaman keterangan salah masukan atau *input* kata kunci.

Jika kata yang dimasukkan lebih dari tiga karakter akan dilanjutkan proses *stemming*. Modul yang digunakan adalah modul seperti yang digunakan pada program ECS *stemmer* yang dipaparkan pada gambar 3.2.

Kata yang telah dilakukan proses *stemming* akan menjadi kata kunci pencarian. Lalu berlanjut ke proses mencari artikel. Proses pencarian artikel dilakukan dengan kata kunci yang telah di-*stemming* dan kata aslinya. Hal tersebut bertujuan untuk menghindari jika tidak artikel tidak memiliki kata yang telah di-*stemming* tetapi memiliki kata aslinya. Dalam hal ini berkaitan dengan kata berawalan "me-" dengan kata dasar berawalan "s-", "p-", "t-", "k-". Contoh kata dasar "salin" dengan awalan "me-" menjadi "menyalin" bukan "mensalin". Jika pencarian menggunakan kata "salin" tetapi dalam satu artikel hanya terdapat kata "menyalin", maka pencarian tidak ditemukan. Oleh karena itu, kata asli juga perlu dilakukan pencarian.

Jika pencarian tidak ditemukan, layar menampilkan halaman keterangan pencarian tidak ditemukan. Jika pencarian ditemukan, layar menampilkan halaman pencarian ditemukan berupa edisi, judul, dan potongan isi artikel.

3.3.3 Struktur Tabel

Dalam penelitian ini terdapat dua tabel tambahan dalam *database* untuk proses *stemming* pada situs www.sindoweekly-magz.com. Tabel lainnya adalah tabel yang telah terdapat dalam situs Sindo Weekly Magazine.

Tabel-tabel tambahan tersebut merupakan tabel *kata_dasar* dan tabel *stoplist*.

1. Tabel *kata_dasar*

Tabel *kata_dasar* berisikan *id_katadasar*, *katadasar*, dan *tipe_katadasar* dengan jumlah *records* 28.532.

Tabel 3.1 Struktur tabel *kata_dasar*

Field	Tipe Data	Panjang
<i>id_katadasar</i>	<i>integer</i>	10
<i>katadasar</i>	<i>varchar</i>	20
<i>tipe_katadasar</i>	<i>varchar</i>	20

Tabel *kata_dasar* ini berfungsi untuk menyimpan semua kata dasar dalam kamus bahasa Indonesia dan juga tipe kata dasar seperti adjektiva, adverbial, bentuk terikat, interjeksi, konjungsi, nomina, numeralia, preposisi, pronomina, dan verba.

2. Tabel *stoplist*

Tabel *stoplist* memiliki *field* *id_stoplist* dan *stoplist* ber jumlah 1147 *records*.

Tabel 3.2 Struktur tabel *stoplist*

Field	Tipe Data	Panjang
id_stoplist	<i>integer</i>	10
stoplist	<i>varchar</i>	50

Tabel *stoplist* berfungsi untuk menyimpan kata bahasa Indonesia yang tidak memiliki makna pada proses pencarian seperti kata dan, serta, ke, dari, di, dan untuk. Kata-kata tersebut tidak disertakan pada proses *query* pencarian.

3.3.4 Desain Antarmuka

Berikut adalah desain antarmuka form *stemmer*.

**Enhanced Confix Stripping Stemmer
untuk Teks Bahasa Indonesia**

Masukkan kata :

Hasil Stem dari "kata" :

==> (Kata stem)

Copyright © Universitas Multimedia Nusantara - Johan Sebastian

Gambar 3.4 Desain form *stemmer*

Form *stemmer* adalah form untuk melakukan uji coba *stemming* kata berbahasa Indonesia. Form *stemmer* ini memiliki sebuah form *input* teks untuk memasukkan kata berbahasa Indonesia. Terdapat tombol "Setem" untuk melakukan proses *stemming*. Ketika pengguna memasukkan kata dan setelah tombol "Setem" diklik, maka proses *stemming* dilakukan. Hasil *stemming* akan ditampilkan pada bagian bawa form *input* kata.

Mesin Pencarian Sindo Weekly Magazine
<http://www.sindoweekly-magz.com>

Masukkan kata kunci :

Nomor. Tahun. Tanggal	Judul Artikel ...potongan isi artikel.....potongan isi artikel.....potongan isi artikel.....potongan isi artikel... ...potongan isi artikel.....potongan isi artikel.....potongan isi artikel.....potongan isi artikel...
Nomor. Tahun. Tanggal	Judul Artikel ...potongan isi artikel.....potongan isi artikel.....potongan isi artikel.....potongan isi artikel... ...potongan isi artikel.....potongan isi artikel.....potongan isi artikel.....potongan isi artikel...
Nomor. Tahun. Tanggal	Judul Artikel ...potongan isi artikel.....potongan isi artikel.....potongan isi artikel.....potongan isi artikel... ...potongan isi artikel.....potongan isi artikel.....potongan isi artikel.....potongan isi artikel...
Nomor. Tahun. Tanggal	Judul Artikel ...potongan isi artikel.....potongan isi artikel.....potongan isi artikel.....potongan isi artikel... ...potongan isi artikel.....potongan isi artikel.....potongan isi artikel.....potongan isi artikel...
Nomor. Tahun. Tanggal	Judul Artikel ...potongan isi artikel.....potongan isi artikel.....potongan isi artikel.....potongan isi artikel... ...potongan isi artikel.....potongan isi artikel.....potongan isi artikel.....potongan isi artikel...

Copyright © Universitas Multimedia Nusantara - Johan Sebastian

Gambar 3.5 Desain form pencarian Sindo Search Engine

Gambar 3.5 di atas merupakan desain form untuk memasukkan kata kunci pencarian dan menampilkan hasil pencarian. Halaman *web* ini dibuat untuk menampilkan dan menganalisis kata hasil *stemming* yang digunakan untuk *query* pencarian karena pada fitur pencarian situs resmi Sindo tidak terdapat keterangan atau potongan isi artikel, hanya berupa edisi dan judul berita saja.

U
M
M
N