BAB IV

GAMBARAN UMUM

4.1 Gambaran Umum Program

Penelitian ini bertujuan untuk membandingkan algoritma *stemming* antara algortima Nazief & Adriani dan algortima Porter. Dengan mengembangkan ide mengubah atau memisahkan sebuah dokumen menjadi sebuah kumpulan (*corpus*) kata. Dari *corpus* ini kemudian dilakukan *stemming* uantuk mendapatkan kata dasar (*root word*) dari masing-masing kata tersebut, dengan terlebih dahulu menghilangkan *stopwords list*.

Proses *stemming* diawali dengan mengubah seluruh kata menjadi huruf kecil (*lower case*), dilanjutkan dengan menghilangkan seluruh tanda baca dan angka, kemudian dilanjutkan dengan menghilangkan kata yang termasuk dalam *stopwords list*. Setiap kata yang muncul lebih dari satu kali dalam setiap dokumen, proses stemming akan dilakukan hanya pada salah satu kata.

Setelah mendapatkan *root word*, kemudian dilakukan pembandingan hasil dari kedua algoritma yang digunakan. Dari hasil ini akan di ketehui, algortima mana yang lebih baik di terapkan pada Sistem Temu Kembali Informasi (*Information Retrieval*) untuk dokumen berbahasa Indonesia.

4.2 Analisis Algortima Nazief & Adriani

Algortima Nazief & Adriani merupakan salah satu algortima *stemming* untuk bahasa Indonesia dengan pendekatan *confix stripping* dengan menggunakan

kamus untuk mencari kada dasar atau akar kata. Algoritma ini dikembangkan oleh Bobby Nazief dan Mirna Adriani dalam penelitiannya yang tidak dipublikasikan untuk umum dari Universitas Indonesia (Asian, 2007).

Dalam penelitiannya, Asian menjelaskan bahwa algoritma Nazief & Adriani terdiri dari beberapa langkah, yaitu:

1. Setiap kata akan dicek di kamus kata dasar. Jika kata tersebut ditemukan, maka algoritma dihentikan dan kata tersebut dianggap sebagai kata dasar, misalnya kata : “makan”, ”pergi”, ”marah”, dll. Jika tidak, maka di lanjutkan ke langkah berikutnya, misalnya : “memakan”, ”kepergiannya”, ”menemukannya”, dll.
2. Hapus *inflection suffixes*, terdiri dari:
   a. Hapus *particle* (“lah”,”kah”,”tah”,”pun”).
   b. Berikutnya adalah hapus *inflection possesive pronouns* atau kata ganti milik (“ku”,”mu”,”nya”).

   Kata yang diperoleh tersebut akan diperiksa dalam kamus, jika ditemukan maka kata tersebut dianggap sebagai kata dasar dan proses dihentikan, jika tidak lanjut ke langkah berikutnya, misalnya: “sepedamukah”, sehingga dihasilkan kata “sepedamu”. Kata tersebut dicari dalam kamus, karena tidak ditemukan, kemudian di hilangkan kata ganti milik, sehingga di peroleh kata “sepeda”. Kata tersebut ada dalam kamus, sehingga proses *stemming* dihentikan.
3. Hapus *Derivation Suffix* (“-i” atau ”-an”), kemudian cek kata dalam kamus, jika ditemukan, stemming berhenti, jika tidak, lanjut ke langkah berikut:

a. Jika setelah akhiran “-an” dihapus dan huruf terakhir adalah “k”, maka huruf “k” tersebut juga di hilangkah, misalnya: “sembunyikan”, setelah akhiran “-an” di hapus, diperoleh kata “sembunyik” dan kata tersebut tidak ada dalam kamus serta huruf terakhirnya adalah “k”, sehingga diperolah kata “sembunyi”. Kemudian langsung dilanjutkan ke langkah 4.

b. Jika kata tetap tidak di temukan dalam database, maka semua akhiran (“-i”,”-an”, dan “-kan” yang telah dihapus, dikembalikan lagi.

4. Hapus *Derivation Prefix* (“be-”,”di-”,”ke-”,”me-”,”pe-“,”se-” dan “te-“). Tahap ini akan dihentikan jika:

a. Ditemukan kombinasi awalan dan akhiran yang tidak diijinkan berdasarkan Table 2.1 pada BAB II, misalnya kata “begini”, karena awalannya adalah “be-” dan akhiranya adalah “-i”, sementara kombinasi awalan dan akhiran ini tidak diperbolehkan, maka kata tersebut tidak mengalami stemming, sehingga kata tersebut dianggap sebagai kata dasar dan algoritma dihentikan.

b. Ditemukan awalan yang sama dengan awalan yang telah dihapus sebelumnya, maka proses stemming di hentikan, misalnya ditemukan kata “didiskusikan”. Setelah akhiran “kan” dihilangkan, kemudian awalan “di-“ telah dihapus, sehingga diperolah kata “diskusi” dan ternyata 2 huruf awalnya sama dengan awalan yang telah dihapus sebelumnya, maka algoritma dihentikan dan diperoleh kata dasar “diskusi”.

c. Penghapusan awalan diulang sebanyak tiga kali, dan jika dalam setiap perulangannya ada awalan yang dihapus yang artinya ada tiga awalan yang dihapus, maka proses berhenti. Dalam kenyataannya sangat jarang ditemukan kata yang memiliki tiga awalan dalam kata bentukannya, kebanyakan satu atau dua awalan.

Proses penghapusan awalan, terdiri dari dua kelompok, yaitu:

a. Kata dengan awalan “di-“,”ke-“,”se-“, maka proses *stripping* dilakukan langsung dengan menghilangkan awalan tersebut secara langsung, misalnya “disimpan” sehingga diperoleh kata “simpan”, kata “keluar” sehingga diperolah kata “luar”, kata “serupa” sehingga diperolah kata “rupa”.

b. Kata dengan awalan “te-“,”be-“,”me-“ dan “pe-“, maka proses stripping dilakukan dengan mangacu pada Tabel 2.2 dan Tabel 2.3 pada BAB II, misalnya kata “terlambat”. Setelah awalan “te-“ dihilangkan, diperolah kata “rlambat”. Huruf pertama kata yang diperolah adalah “r” kemudian diikuti huruf “l”. Berdasarkan aturan nomor 8, maka pemenggalan kata yang tepat adalah “ter-lambat”, sehingga diperolah akar kata “lambat”. Kata tersebut ada dalam database, maka proses dihentikan.

   Contoh lain adalah kata “mempertinggi”. Berdasarkan aturan nomor 12, dari kata “mempertinggi” sehingga diperoleh kata “pertinggi”, kata ini tidak ditemukan dalam kamus. Kemudian tahap 4 ini diulangi kembali (recursive process), sehingga didapatkan “tinggi” setelah menghilangkan awalan “per”, dan kata ini ada dalam kamus.

Jika suatu kata tetap tidak ditemukan dalam database setelah proses perulangan penghapusan awalan (*recursive process*) dan tiga kondisi yang menyebabkan proses stemming dihentikan diatas tidak terpenuhi, maka proses dilanjutkan dengan melakukan recoding. Sebagai contoh kata “menangkap”, berdasarkan aturan nomor 15 untuk tipe awalan “me” (awalan “men” dikuti huruf “a”). Setelah menghilangkan awalan “men”, diperolah kata “angkap” dan kata ini bukan kata dasar. Penghapusan ulang awalan (recursive process) tidak bisa dilakukan karena tidak ada awalan yang dihapus.

Perubahan aturan sebagaimana dituangkan pada Tabel 2.4 diterapkan untuk mengatasi kata-kata yang tidak didapatkan akar kata yang benar jika hanya menggunakan aturan Tabel 2.2, misalnya kata “mempidakan”, tanpa menggunakan aturan tambahan, maka akan didapat hasil “mempidanakan” dan ini adalah salah. Dengan menerapkan aturan yang baru, sehingga didapat *root word* yang benar yaitu “pidana”. Begitu juga pada kata “mengkomunikasikan”, akan didapat kata dasar yang benar yaitu “komunikasi”.

c. Jika dua huruf pertama sebuah kata bukan “di”,”ke”,”se”,”te”,”be”,”me” atau “pe”, maka algoritma dihentikan dan kata tersebut dianggap sebagai kata dasar.

5. Jika semua awalan sudah dihilangkan dan tetap tidak ditemukan kata dasar yang tepat, maka langkah selanjutnya adalah melakukan pengkodean ulang (*recoding*) awalan dengan mangacu pada Tabel 2.2 pada kolom terakhir yang menunjukan variasi awalan dan karakter untuk melakukan *recoding*.

Sebagai contoh, kata “menangkap” diatas, ada dua kemungkinan untuk *recoding* awalan beradasarkan aturan nomor 15 yaitu “n” (untuk “men-nV...”) dan “t” (untuk “men-tV...”). Penambahan awalan “n” pada kata “angkap”, sehingga menjadi “nangkap”. Kata “nangkap” bukan merupakan kata dasar, sehingga dilakukan recoding dengan memberikan awalan “t” sehingga diperolah kata “tangkap”, yang merupakan kata dasar yang benar, sehingga proses dihentikan.

6. Setelah semua proses dilalui dan tetap tidak ditemukan kata dasar yang benar, maka algoritma dihentikan dan kata tersebut dianggap sebagai kata dasar.

Pada setiap tahapannya, algoritma ini selalu melakukan pengecekan kata dalam database.

Pada kondisi tertentu, beberapa perubahan susunan proses algoritma dilakukan untuk mendapatkan *root word* yang benar, yaitu:

a. Untuk kata dengan awalan “be” dan akhiran “lah”, maka *stripping* akan diawali dengan menghilangkan awalan terlebih dahulu, baru kemudian menghilangkan akhiran. Kata “bersebelah”, jika menghilangkan akhiran terlebih dahulu yaitu “-lah” sehingga didapat kata “bersebe”, kemudian menghilangkan awalan “ber”, didapat kata “sebe”, dan ini adalah hasil yang salah. Dengan menghilangkan awalan “ber” telebih dahulu, sehingga didapat kata “sebelah”. Kata tersebut di cek di database dan ditemukan, sehingga kata tersebut dianggap sebagai kata dasar, dan ini adalah hasil yang benar.

Aturan ini berlaku untuk kata dasar yang berakhiran “-lah”, misalnya “sekolah”, “masalah”, “sebelah” dan lain sebagainya.

b. Untuk kata dengan awalan “be” dan akhiran “-an”, proses *stripping* dilakukan dengan menghilangkan awalan terlebih dahulu, baru kemudian menghilangkan akhiran. Pada kata “berteman”, jika proses akhiran yang dihilangkan lebih dulu, maka akan dihasilkan kata “tem”. Dengan menghilangkan awalan ”ber” telebih dahulu, sehingga menghasilkan kata “teman”, dan ini adalah akar kata yang benar, maka proses dihentikan.

c. Untuk kata yang berawalan “me” dan berakhiran “-i”, proses *stripping* dilakukan denga menghilangkan awalan terlebih dahulu, kemudian menghilangkan akhiran. Contoh pada kata “menuai”, dengan menghilangkan awalan “me” dan mengubah “n” menjadi “t” sehingga didapatkan kata “tuai”. Kata ini ditemukan dalam database, sehingga ini adalah kata dasar yang benar.

d. Untuk kata dengan awalan “di” dan akhiran “-i”, proses stripping diawali dengan menghilangkan awalan terlebih dahulu, kemudian baru menghilangkan akhiran. Pada kata “dipakaikan”, dengan menghilangkan awalan “di”, sehingga didapat kata “pakaikan”, ini bukan kata dasar yang benar. Selanjutnya dengan menghilangkan akhiran “-kan”, sehingga didapat kata “pakai” dan ini adalah kata dasar yang benar.

e. Untuk kata dengan awalan “pe” dan akhiran “-i”, proses stripping diawali dengan menghilangkan awalan, kemudian akhiran. Pada kata “penelitian”, proses *stripping* diawali dengan menghilangkan awalan “pe” dan mengubah huruf “n” menjadi “t” sehingga didapat kata “telitian”.

Kemudian dilanjutkan menghilangkan ahiran “an”, sehingga didapat kata “teliti”. Kata ini ditemukan dalam database, sehingga ini kata ini merupakan kata dasar yang benar.

f. Untuk kata dengan awalan “ter” dan akhiran “-i”, proses *stripping* diawali dengan menghilangkan awalan terlebih dahulu, kemudian menghilangkan akhirannya. Contoh pada kata “tercapai”, dengan menghilangkan awalan “ter”, sehingga didapat kata “capai”. Kata ini ditemukan dalam database, sehingga ini dianggap sebagai kata dasar yang benar.

4.3 Analisis Algoritma Porter

Algoritma Porter pertama kali dikenalkan oleh M.F. Porter (1980) dalam penelitiannya yang berjudul “An algorithm for suffix stripping”. Algoritma Porter dikembangkan untuk Bahasa Inggris.

Perbedaan morfologi antara Bahasa Indonesia dan Bahasa Inggris, sehingga dikembangkanlah algoritma Porter untuk Bahasa Indonesia. Penelitian tentang algoritma Porter untuk bahasa Indonesia dilakukan oleh Fadillah Z Tala (2003). Berbeda dengan algoritma Nazief & Adriani, algoritma ini tidak menggunakan kamus kata dasar dalam proses pencarian akar kata.

Berikut adalah tahapan dalam algoritma Porter untuk bahasa Indonesia yang dikembangkan oleh Tala (2003):

1. Menghilangkan partikel (“-lah”,”-kah”,”-tah”,”-pun”), misalnya kata “terdengarlah”, sehingga diperoleh kata “terdengar”.

2. Menghilangkan kata sandang dan kepunyaan (“-ku”,”-mu”,”-nya”), misalnya kata “sepedaku”, sehingga diperolah kata “sepeda”, kata “kalimatnya”, sehingga diperolah kata “kalimat”.
3. Menghilangkan awalan pertama berdasarkan Tabel 2.5 pada Bab II. Sebagai contoh kata “menyisir”, terdiri dari awalan “meny” dan kata “isir”. Huruf pertama setelah awalan “meny” dihilangkan adalah “i”. Sesuai aturan nomor 2, maka awalan “meny” diubah menjadi “s”, sehingga didapatkan kata “sisir”.
4. Langkah ini terdiri dari 2 sub proses dengan syarat kondisi, yaitu:
   a. Hapus awalan kedua berdasarkan aturan Tabel 2.6 pada Bab II jika pada langkah 3 tidak ada awalan yang dihapus. Sebagai contoh kata “berpegangan”, karena pada langkah 3 tidak ada awalan yang dihilangkan, maka pada awalan “ber” dihilangkan, sehingga didapat kata “pegangan”. Selanjutnya langsung ke langkah 5a.
   b. Hapus akhiran jika pada langkah 3 ditemukan awalan yang dihilangkan, misalnya kata “membacakan”. Pada langkah 3, ada awalan “mem” yang merupakan kelompok awalan pertama yang dihapus, sehingga didapat kata “bacakan”. Karena ada awalan yang dihapus, maka pada langkah ini dilakukan penghapusan akhiran “kan”, sehingga didapat kata akhir “baca”. Kemudian lanjut ke langkah 5b. Jika pada langkah ini tidak ada akhiran yang dihilangkan, maka kata tersebut dianggap sebagai kata dasar (*root word*) dan proses dihentikan, misalnya kata “menyisir” pada langkah 3 yang kemudian didapat kata “sisir”, karena pada langkah 4 tidak ada akhiran yang

dihapus, maka kata “sisir” merupakan kata dasar (*root word*) dan algoritma dihentikan.

5. Seperti halnya langkah 4, langkah 5 ini terdiri dari 2 sub proses dengan syarat kondisi:
   a. Hapus akhiran dan kata terakhir dianggap sebagai kata dasar (*root word*). Sebagai contoh kata “berpegangan” pada langkah 4a yang menghasilkan kata “pegangan”. Pada langkah ini akhiran “an” dihilangkan, sehingga dari kata “pegangan” didapatkan kata “pegang”. Kata ini dianggap sebagai kata dasar (*root word*) dan algoritma dihentikan.
   b. Hapus awalan kedua dan kata terakhir yang didapat dianggap sebagai kata dasar. Sebagai contoh kata “dipekerjakan”, hasil dari langkah 1 sampai 3 didapatkan kata “pekerjakan” setelah awalan “di” dihilangkan. Karena pada langkah 3 ada awalan yang dihilangkan, maka dilanjutkan ke langkah 3b dengan menghilangkan akhiran “kan”, sehingga didapatkan kata “pekerja”. Pada tahap ini dihilangkan awalan kedua “pe” sehingga didapat kata “kerja” dan kata ini dianggap sebagai kata dasar (*root word*). Algoritma berhenti.