

Kode Makalah M-7

FINITE STATE AUTOMATA UNTUK PARSING (PEMENGALAN) SUKU KATA DALAM BAHASA INDONESIA

Suprihatin, Program Studi Ilmu Komputer FMIPA UAD

Abstrak

Komputasi mempunyai peranan penting dalam beberapa bidang, termasuk bidang bahasa. Pembacaan suatu kalimat memerlukan pemenggalan suku kata-suku kata. Tulisan ini membahas bagaimana memenggal kata atau kalimat menjadi suku kata-suku kata sehingga pembacaannya benar. Struktur suku kata disesuaikan dengan ejaan yang disempurnakan (EYD). Finite State Automata (FSA) sebagai mesin abstrak akan dipergunakan untuk membantu merancang dalam parsing (pemenggalan) suku kata. Hasil tulisan ini dapat dikembangkan atau sebagai langkah awal program untuk mengubah suatu teks menjadi ucapan yang dapat dimengerti.

Kata kunci: suku kata, parsing, Finite State Automata

PENDAHULUAN

Komputer sebagai alat bantu menjadi bertambah penting, seiring dengan perkembangan kehidupan manusia. Komputer dapat dipergunakan untuk mengubah teks bahasa Inggris menjadi ucapan. Selama ini belum ada program komputer untuk mengubah teks bahasa Indonesia menjadi ucapan. Untuk itu akan dibuat tahap awal program mengubah teks bahasa Indonesia menjadi ucapan.

Bahasa Indonesia mengalami perkembangan cukup baik. Bahasa Indonesia digunakan untuk alat komunikasi baik formal maupun nonformal, untuk itu dibentuklah buku pedoman ejaan bahasa Indonesia, yang diberi nama Ejaan yang Disempurnakan (EYD).

EYD memberi ketentuan mengenai susunan kalimat yang baku, cara penulisan yang berkenaan dengan huruf yang digunakan dan pemenggalan suku kata dalam mengeja suatu kata.

Bahasa Indonesia sebagai alat komunikasi mempergunakan lisan dan tulisan sebagai pengantaranya. Bahasa lisan sering disebut dengan *fonologi* yaitu lafal kata yang

diucapkan sama dengan hurufnya. Bahasa formal adalah bahasa tulisan yang sering berbeda dengan bahasa lisannya.

Suku kata dalam bahasa Indonesia adalah bahasa lisan, kadang-kadang pemenggalan lisan ataupun tulisan berbeda. Pemenggalan suku kata ditujukan pada bahasa lisan tentu saja ada beberapa hal yang tidak dapat didekati dengan teori bahasa formal.

Komputer sebagai alat akan digunakan untuk membantu mengeja kata untuk pengucapan suku kata, penelitian ini dimaksudkan untuk membuat program komputer untuk pemenggalan kata menjadi suku kata-suku kata yang sesuai dengan EYD. Penelitian dapat dilanjutkan untuk mengubah teks bahasa Indonesia menjadi ucapan. Finite state Automata (FSA) digunakan untuk menggambarkan struktur vokal, konsonan, ataupun struktur suku kata bahasa Indonesia yang sesuai dengan EYD.

DASAR TEORI

Secara formal Finite State Automata (FSA) didefinisikan sebagai sebuah 5 tupel $(Q, \Sigma, \delta, q_0, F)$, dimana Q : himpunan berhingga status, Σ : himpunan berhingga simbol input (Alpabet), q_0 dalam Q adalah status awal, $F \subseteq Q$ adalah himpunan status akhir (finish) dan δ : fungsi transisi yang memetakan $Q \times \Sigma$ ke Q (Hopcroft, 1979).

Sebuah FSA dapat digambarkan sebagai graff berarah yang titik-titikya merupakan status-statusnya. Jika sebuah transisi dari status q ke status p dalam input a , maka sebuah garis berlabel a akan menghubungkan status q ke status p dalam graff tersebut.

Bahasa Indonesia mengenal bahasa tulisan dan bahasa lisan. Bahasa lisan mengenal adanya fonem dan diftong yang merupakan kesatuan bahasa terkecil sebagai simbol bunyi dan dapat membedakan arti. Fonem bahasa tulisan dilambangkan dengan huruf dan gabungan dua huruf mati, sedang diftong adalah gabungan dua vokal (huruf hidup). Fonem digolongkan menjadi 5 vokal (a, i, u, e, o) dan 25 konsonan yaitu: b, c, d, f, g, h, j, k, kh, l, m, n, ng, ny, p, q, r, s, sy, t, v, w, x, y, z serta 3 diftong yaitu: ai, au, oi, dalam hal ini diftong digolongkan sebagai vokal.

Suku kata adalah struktur yang terjadi dari suatu urutan fonem yang merupakan bagian dari kata. Berikut adalah pola suku kata dalam bahasa Indonesia dengan V

sebagai simbol vokal dan K konsonan. Pola tersebut terdiri dari 11 yaitu: V, VK, VKK, KV, KVK, KVKK, KKV, KKVK, KKVKK, KKKV, KKKVK. Berikut ini contoh pola struktur suku kata:

1. V :**i**-bu, **a**-suh, **a**-nak
2. VK :**an**-da, **in**-dah
3. VKK :**eks**, **ons**
4. KV :**ka**-mi, **ma**-kan
5. KVK :**bu**-kan, mu-**lut**
6. KVKK :**teks**-til, **pers**
7. KKV :**pra**-bu, **tro**-pis
8. KKVK :**prak**-tis, **frak**-si
9. KKVKK : kom-**pleks**
10. KKKV :**stra**-tegi
11. KKKVK :**struk**-tur

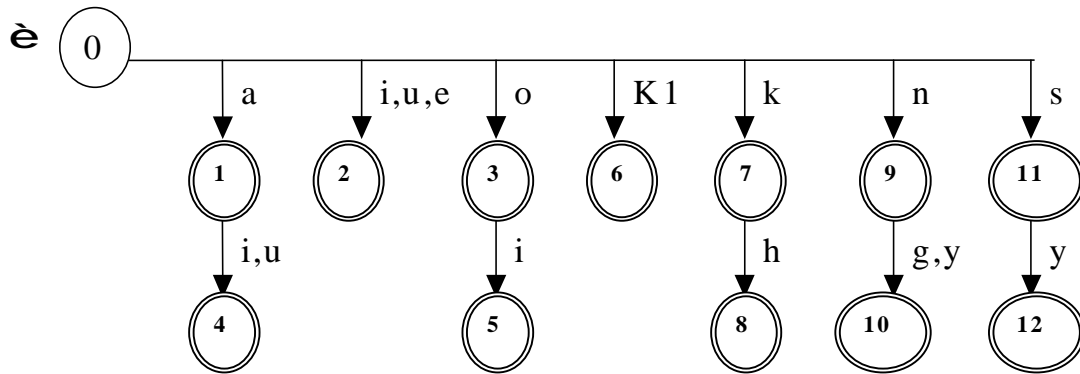
Pola V, VK, KV, KVK, merupakan pola suku kata bahasa Indonesia asli, sedangkan lainnya merupakan pola suku kata dari bahasa asing dan bahasa daerah yang sudah diadopsi dalam bahasa Indonesia. Berikut adalah pedoman untuk memenggal suku kata:

1. Jika ditengah kata terdapat dua vokal berurutan selain diftong maka pemenggalannya diantara dua vokal tersebut.
2. Jika ditengah terdapat konsonan diantara dua vokal, maka pemenggalan dilakukan sebelum konsonan tersebut.
3. Jika ditengah kata terdapat dua atau lebih konsonan, maka pemenggalan setelah konsonan pertama.
4. Imbuhan atau partikel yang biasanya ditulis serangkai dengan kata dasar pada penyukuan dipisahkan.

RANCANGAN PROGRAM

Fonem digolongkan menjadi 5 vokal (a, i, u, e, o) dan 25 konsonan yaitu: b, c, d, f, g, h, j, k, kh, l, m, n, ng, ny, p, q, r, s, sy, t, v, w, x, y, z serta 3 diftong yaitu: ai, au, oi.

Jika diftong digolongkan dalam vokal maka terdapat 25 konsonan dan 8 vokal. FSA dapat mengenali fonem, seperti pada gambar 1 berikut:

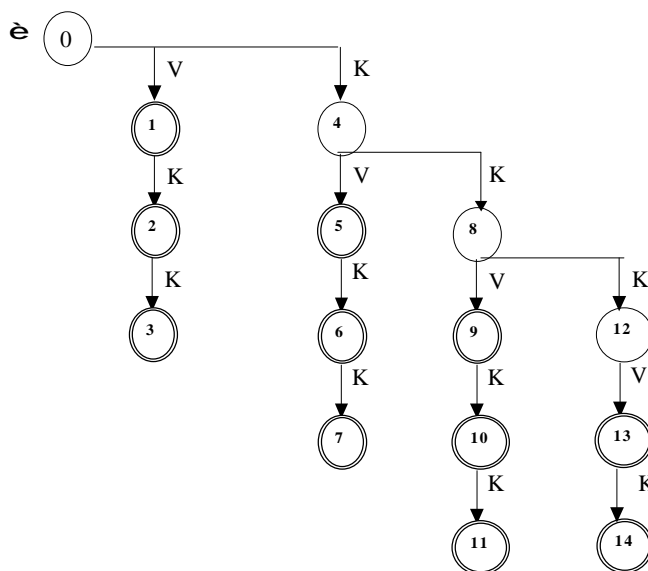


$K1 = \{ b, c, d, f, g, h, j, l, m, p, q, r, t, v, w, x, y, z \}$

Gambar 1 FSA Pengenalan Fonem

Lingkaran dua menunjukkan status akhir pengenalan fonem. FSA gambar 1 di atas ada 12 status akhir yang menunjukkan struktur fonem bahasa Indonesia. Status akhir 1, 2, 3, 4, 5 menunjukkan fonem vokal, status 6 sampai 12 menunjukkan konsonan

Pola struktur suku kata terdiri dari 11 yaitu: V, VK, VKK, KV, KVK, KVKK, KKV, KKVK, KKVKK, KKKV, KKKVK yang jika digambarkan dengan FSA akan terlihat seperti pada gambar 2 di bawah ini:



Gambar 2 FSA Struktur Suku Kata