

SENTENCE ANALYSIS WITH ARTIFICIAL INTELLIGENCE MACHINE LEARNING USING FINITE STATE AUTOMATA

Yos Merry Raditya Putra

Program Studi Teknik Informatika, Unika Soegijapranata Semarang

truefalseboy@gmail.com

Abstract

Sentence pattern in Indonesian is very important. Incorrect pattern can cause misinterpretation and different meaning. The basic sentence pattern consists of subject, predicate, and object (SPO). To analyze the sentence, this project implements Finite State Automata (FSA) algorithm, which is useful to identify the class of words in a sentence.

The given sentence processed through parsing process to break down the sentence into array of tokens. The program will then examine the tokens to determine the class of each word. The discovered pattern will be checked against existing vocabularies on the database. If the word exists, the program will return a response in accordance to the existing one. Otherwise, the program will store the word onto a temporary file until it find the class of the word.

The Intelligence of the program depends on the number of vocabularies on the database. This program supports learning process, so the more sentences given, the more intelligent it gets.

Keywords: finite state automata algorithm, sentence structure, word class, artificial intelligence, machine learning, sentence breaker

Pendahuluan

Pemrosesan bahasa adalah salah satu bidang ilmu kecerdasan buatan yang mempelajari komunikasi antara manusia dan komputer. Di Indonesia ada beberapa struktur kalimat dasar yang menjadi pedoman berdasarkan karakteristik SPO. Pemeriksaan SPO menjadi salah satu dasar untuk menulis kalimat bahasa Indonesia dengan baik sesuai dengan aturan EYD.

Untuk menentukan struktur SPO yang benar, maka pengetahuan yang diperlukan adalah mengetahui kelas kata, sebelum masuk ke dalam struktur SPO kalimat. Untuk mengatasi ini, algoritma finite state automata berguna untuk memeriksa kata berimbuhan yang akan dikategorikan dalam kelas kata. Dengan konsep yang jelas dari kelas kata akan mendapatkan struktur kalimat yang benar.

Tujuan utama dari program ini adalah untuk menentukan struktur kalimat dengan pedoman kelas kata dan karakteristik SPO. Semakin banyak data yang diinputkan, program akan semakin cerdas dalam menentukan kelas kata karena ada proses pembelajaran di dalamnya.

Penelitian tentang struktur SPO kalimat dan kata kelas telah banyak diteliti. Banyak metode dan berbagai algoritma yang digunakan untuk memecahkan dan merumuskannya. Metode dengan penyelesaian finite state automata yang belum pernah digunakan atau dilakukan dalam penelitian sebelumnya untuk menyelesaikan dan menentukan struktur kalimat dan kata kelas dan menggunakan metode meletakkan file sementara untuk proses pembelajaran dalam penentuan program kelas kata.

Ada tiga kelas untuk mendefinisikan kata dalam program, yaitu kata benda, kata sifat dan kata kerja. Dalam tiga kelas kata-kata berikut, program ini dapat menentukan struktur kalimat subjek, frase subjek, predikat dan frase predikat dalam kalimat. Menentukan kelas kata menggunakan algoritma finite state automata dan metode pembelajaran dari file sementara, semakin program ini akan lebih canggih, dan dapat menentukan banyak kelas kata. Dengan begitu banyak kelas kata yang dapat didefinisikan oleh program, struktur kalimat dapat diketahui.

Landasan Teori

Algoritma Finite State Automata

Finite state automata adalah mesin abstrak berupa sistem model matematika dengan masukan dan keluaran diskrit yang dapat mengenali bahasa paling sederhana (bahasa reguler) dan dapat diimplementasikan secara nyata.

Finite State Automata (FSA) adalah model matematika yang dapat menerima input dan mengeluarkan output yang memiliki state yang berhingga banyaknya dan dapat berpindah dari satu state ke state lainnya berdasarkan input dan fungsi transisi. Finite state automata tidak memiliki tempat penyimpanan/memory, hanya bisa mengingat state terkini.

Finite State Automata dinyatakan oleh pasangan 5 tuple, yaitu:

$$M=(Q, \Sigma, \delta, S, F)$$

Q = himpunan state

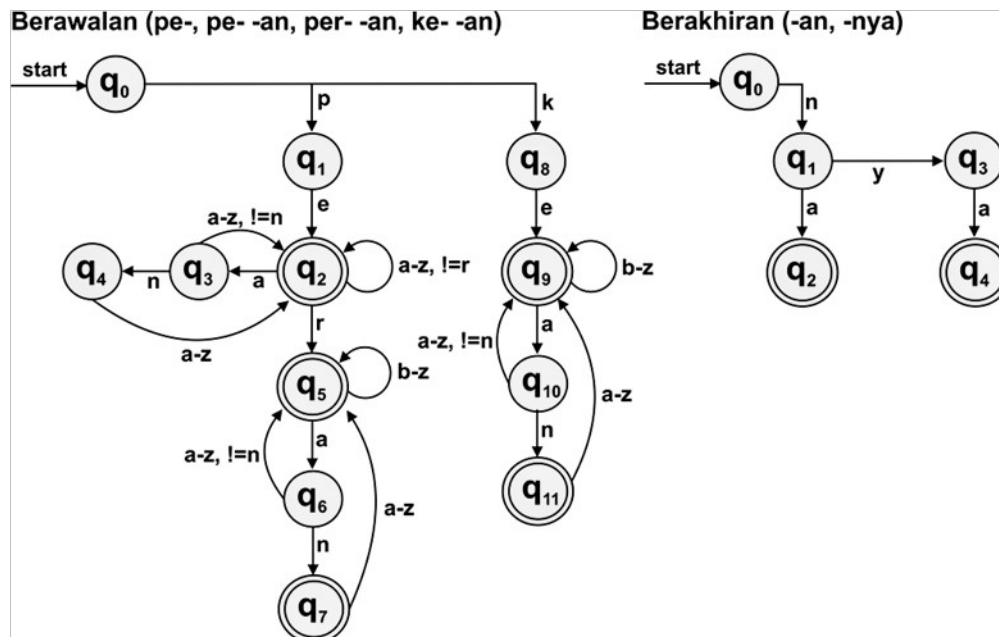
Σ = himpunan simbol input

δ = fungsi transisi $\delta : Q \times \Sigma$

S = state awal / initial state, $S \in Q$

F = state akhir, $F \subseteq Q$

Salah satu contoh diagram finite state automata yang program ini buat adalah kata benda seperti di bawah ini:



Gambar 1: Diagram FSA Kata Benda

Metodologi Penelitian

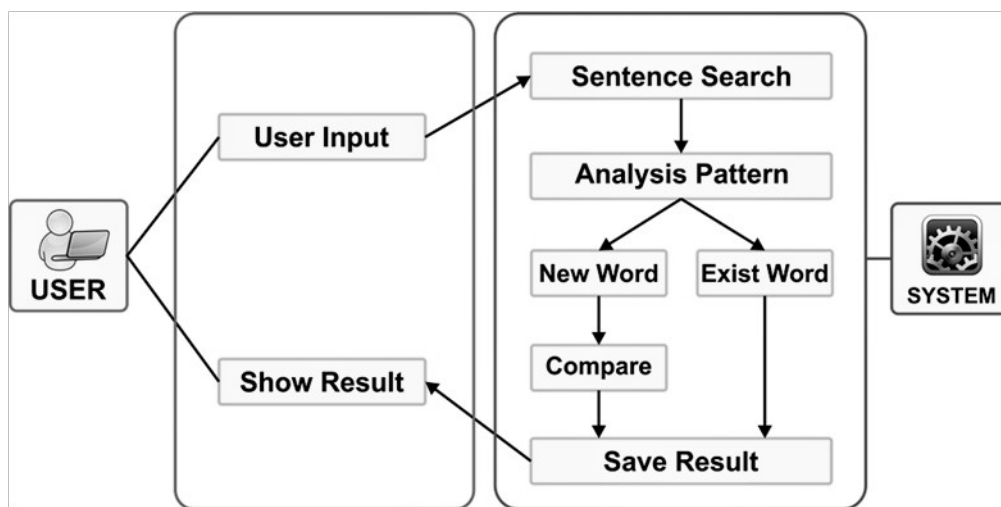
Dalam pelaksanaan program ini, ada beberapa langkah yang harus diambil. Berikut adalah beberapa langkah.

1. **Data Finding:** Mengumpulkan data referensi yang dapat digunakan dalam program ini termasuk algoritma dan struktur data yang harus digunakan. Membandingkan algoritma yang sama dan mencari keuntungan dan kerugian dari masing-masing algoritma.
2. **Analysis and learning:** Menganalisis masalah yang akan terjadi dan bagaimana mengatasinya untuk tetap menjalankan projek. Belajar tata bahasa Indonesia dan membuat percobaan yang mungkin dibuat dalam program, membuat rumus yang diperlukan.
3. **Implementing:** Menerapkan struktur data, algoritma dan user interface. Pelaksanaan proses pembelajaran di program ini berjalan dengan baik atau tidak. Analisis dari setiap kata menggunakan rasio dari brain file yang ada dan menggunakan algoritma FSA. Memperbaiki bugs selama proses program. Penggunaan pembelajaran yang tepat dan mencari celah untuk setiap rumus yang akan dibuat.

4. **Testing:** Menguji semua proses dari langkah pertama sampai terakhir. Semua metode pengujian harus berfungsi dengan baik tanpa bugs. Ketika ada bugs, akan kembali ke langkah kedua untuk memperbaiki program.

Hasil dan Pembahasan

Seperti yang terlihat pada gambar, program ini akan mendapatkan input dari user. Input harus kalimat bahasa Indonesia. Setelah kalimat dimasukkan, program akan memproses input dengan menganalisis pola yang ada. Setelah mendapatkan hasil, data akan disimpan. Selanjutnya hasil analisis akan ditampilkan kepada user.



Gambar 2: Use Case Diagram

Program akan menganalisis kalimat dan kemudian menampilkan hasilnya. Karena hasil analisis program bergantung pada kalimat yang user inputkan, maka ada beberapa tes untuk menguji bagaimana program kerja dengan memasukkan kalimat yang memiliki kelas kata berbeda. Analisis struktur kalimat yang mengandung kelas kata tertentu di beberapa bagian dalam kalimat, kata-kata tertentu yang mungkin dikenali oleh program.

Tabel 1: Daftar Percobaan Kalimat

No	Sentence	Subject/ Subject Phrase	Predicate/ Predicate Phrase	Information
1	lukisan bergerak	lukisan	bergerak	Subjek selalu nominal, Predikat verbal
2	hidupnya menyenangkan	hidupnya	menyenangkan	Subjek selalu nominal, Predikat verbal
3	gambar bagus	gambar	bagus	Subjek selalu nominal, Predikat belum terdefinisi dengan baik
4	gambar bagus	gambar, bagus	-	Subjek selalu nominal, Predikat adjektival
5	ayahnya petani	ayahnya	petani	Subjek selalu nominal, Predikat nominal
6	budi telah pergi	budi	telah pergi	Subjek selalu nominal, predikat verbal
7	hanry masih tidur lelap	hanry	masih tidur lelap	Subjek selalu nominal, predikat verbal, tidur lelap adalah frase verbal
8	kita akan pergi jauh	kita	akan pergi jauh	Subjek selalu nominal, predikat verbal, pergi jauh adalah frase verbal
9	penari itu berlari	penari itu	berlari	Subjek selalu nominal, predikat verbal
10	pesulap itu bagus	pesulap itu	bagus	Subjek selalu nominal, predikat adjektival
11	hidup ini indah	hidup ini	-	Subjek selalu nominal, predikat belum terdefinisi dengan baik

Tabel 2: Percobaan untuk Pembelajaran Program

Input	Sentence	Define	Undefined	Information
1	gambar bagus	gambar	bagus	Subjek selalu nominal, Predikat belum terdefinisi dengan baik
2	gambar bagus	Gambar, bagus	-	Subjek selalu nominal, Predikat adjektival
3	putrinya cantik	putrinya	cantik	Subjek selalu nominal, Predikat belum terdefinisi dengan baik
4	putrinya sangat cantik	Putrinya, cantik	-	Subjek selalu nominal, predikat adjektival
5	putrinya cantik	Putrinya, cantik	-	Subjek selalu nominal, Predikat adjektival

Kesimpulan

Algoritma finite state automata dapat menentukan kelas kata kerja dan kelas kata benda berdasarkan konteks kalimat. Menggunakan rumus preposisi untuk menentukan kata sifat membuat program dapat belajar kata baru yang diinputkan oleh user. Proses pembelajaran di program ini dapat membantu untuk menentukan subjek, frase subjek, predikat dan frase predikat yang masih belum diketahui program. Dengan ini, program dapat menentukan kelas kata dari setiap kata dalam kalimat inputan oleh user dan untuk mengetahui struktur kalimat. Semakin banyak kata-kata yang telah disimpan oleh program, program semakin cerdas dalam menentukan kelas kata dan struktur kalimat.

Program ini dapat menganalisis tiga kelas kata, yaitu kata sifat, kata kerja dan kata benda dan struktur kalimat hanya subjek, frase subjek, predikat dan frase predikat. Selanjutnya, program ini dapat dikembangkan dengan kelas kata yang lebih lengkap dan begitu juga dengan struktur kalimatnya.

Daftar Pustaka

- [1] Gopalakrishnan, "Computation Engineering : Applied Automata Theory and Logic", Ganesh, New York : Springer, 2006.
- [2] Gorys, "Tata Bahasa Rujukan Bahasa Indonesia", Jakarta: PT Gramedia.
- [3] Hopcroft, John E. et al, "Introduction to Automata Theory, Languages, and Computation", 2nd Edition, New York: Addison-Wesley, 2001.
- [4] Kosasih, Dr.E., "Kompetensi Ketatabahasaan dan Kesusastraan", Bandung: CV YramaWidya, 2003.
- [5] Kusno, B.S., "Tata Bahasa Indonesia", Bandung: CV Rosda. 1985.
- [6] Ramlan, M., "Sintaksis". Yogyakarta: CV Karyono. 1987.
- [7] Sudaryanto, "Predikat Objek dalam Bahasa Indonesia", Yogyakarta: Djambatan.
- [8] Sugono, Dendy, "Berbahasa Indonesia dengan Benar". Jakarta: Puspa Swara. 1994.
- [9] Tarigan, Henry Guntur, "Prinsip-prinsip Dasar Sintaksis", Bandung: Angkasa. 1983.
- [10] Heru Cahya Rustamadji, "Teori Bahasa dan Otomata", Fakultas Teknik Industri Universitas Pembangunan Nasional, 2004.