

Penerapan Algoritma Pencocokan String dalam Mengidentifikasi Struktur Kalimat

Ridwan Faturrahman

Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13517150@std.stei.itb.ac.id

Abstract— Setiap bahasa memiliki tata bahasanya masing-masing termasuk bagian bagaimana struktur kalimat terbentuk. Membentuk kalimat yang baik artinya membentuk kalimat berdasarkan tata bahasa itu sendiri. Dengan struktur kalimat yang baik maka orang-orang dapat menerima pesan yang hendak disampaikan lewat kalimat dengan baik. Salah satu metode untuk menguji struktur suatu kalimat adalah menentukan jenis setiap kata (dapat berupa kata benda, kata kerja, kata depan, dan seterusnya) dan mencocokkan semua itu dengan struktur kalimat yang mungkin. Algoritma pencocokan string dapat berperan dalam melakukan identifikasi terhadap setiap kalimat dan dapat memperbaharui tingkat akurasi seiring dengan banyaknya kalimat yang digunakan.

Keywords—Boyer-Moore, Struktur, Kalimat, Kata.

I. PENDAHULUAN

Tata bahasa merupakan ilmu yang mempelajari aturan-aturan yang mengatur penggunaan bahasa. Aturan-aturan tersebut merupakan acuan setiap ahli bahasa dalam mempelajari atau membuat sebuah tulisan dan percakapan untuk menarik atau meletakkan inti sari dalam kesatuan kalimat yang terbentuk di dalamnya. Bukan hanya ahli bahasa saja, melainkan juga setiap masyarakat yang menggunakan bahasa dalam kegiatan sehari-harinya menggunakan tata bahasa dalam membuat tulisan ataupun melakukan percakapan.

Tentunya tidak semua aturan dalam tata bahasa yang salah satunya dalam membentuk struktur kalimat diikuti oleh masyarakat pada umumnya. Hal tersebut merupakan hal yang manusiawi dikarenakan pada dasarnya manusia bukanlah mesin yang mengikuti setiap algoritma dan data yang melekat pada dirinya secara konsisten. Dan bahkan mungkin penulis sendiri melanggar aturan tata bahasa dalam menulis makalah ini.

Dalam perkembangan teknologi saat ini, sangat dimungkinkan sebuah mesin dapat mengoreksi tata bahasa seorang manusia. Hal tersebut merupakan latar belakang pembuatan makalah ini, yaitu mengidentifikasi struktur kalimat yang dapat mengeluarkan hasil berupa ketepatan dalam pembentukan kalimat dan memperbaharui tingkat akurasi algoritma pengidentifikasian tersebut.

Peran algoritma pencocokan string dalam kasus ini adalah penentuan jenis kata dan struktur kalimat yang cocok. Sebelumnya telah dibuat daftar kata berdasarkan jenisnya dan daftar struktur kalimat yang umum. Sehingga kalimat yang ingin diidentifikasi akan ditentukan jenis setiap katanya sehingga selanjutnya dapat dianalisis bentuk struktur kalimatnya. Jenis kata yang didapat berasal dari pencocokan kata dengan daftar kata yang telah dibuat menggunakan algoritma *Boyer-Moore*, karena algoritma ini lebih efektif dalam pencocokan string dengan karakter yang bervariasi.

Hasil dari pengidentifikasian ini dapat menilai seberapa tepat penulisan kalimat yang diuji dan memperbaharui tingkat keakurasiannya dengan menambah beberapa kata atau bentuk kalimat yang baru ke dalam daftar kata atau daftar struktur kalimat.

Di dalam makalah ini hanya membahas ketepatan suatu kalimat dari segi tata kalimatnya saja, sehingga memungkinkan untuk mendapatkan suatu kalimat yang benar secara tata kalimat tapi tidak masuk akal atau tidak lazim dituliskan atau diucapkan. Karena hanya melihat tata kalimatnya saja sehingga lebih mudah dianalisis rumusan tata kalimat itu.

Pada makalah ini, akan dibahas mengenai penggunaan data berupa kata atau bentuk kalimat baru yang didapat dari beberapa kalimat yang diuji menggunakan algoritma pencocokan string, mengolah dan menganalisis data yang telah didapat. Makalah ini diharapkan untuk memberikan

manfaat dalam penggunaan bahasa di kehidupan sehari-hari.

II. LANDASAN TEORI

Persoalan yang sering dijumpai dalam bidang informatika salah satunya adalah pencarian string. Permasalahan ini memunculkan banyaknya algoritma pencarian atau pencocokan string. Pada dasarnya, sebuah algoritma pencocokan string seperti menggunakan algoritma brute force yaitu mencocokkan teks dengan mencocokkan setiap hurufnya dengan kata kunci.

Algoritma brute force tersebut menyelesaikan permasalahan pencocokan string, akan tetapi tingkat keefektifannya sangat tidak baik. Lalu dikenal algoritma pencocokan string yang menjawab permasalahan keefektifan algoritma brute force tersebut yaitu algoritma Knuth-Morris-Pratt, algoritma Boyer-Moore, dan pencocokan string dengan regular expression. Pada makalah ini, algoritma yang dibahas hanyalah algoritma Boyer-Moore sebagai algoritma pencocokan string dalam pencarian setiap jenis kata dalam beberapa kalimat yang diuji.

A. Algoritma Boyer-Moore

Algoritma Boyer-Moore merupakan algoritma pencocokan string merupakan algoritma pencocokan string yang paling efisien dibandingkan algoritma pencocokan string lainnya. Algoritma Boyer-Moore ditemukan oleh J. Strother Moore dan Robert S. Boyer pada tahun 1977. Algoritma ini menggunakan beberapa kasus pengecekan teks dengan kata kunci.

Algoritma pencarian string ini mencocokkan dengan cara membandingkan sebuah huruf dengan huruf yang ada di kata kunci yang dicari, menggeser kata kunci tersebut hingga posisinya sama dengan posisi teks yang dicari.

Proses penggeseran ini merupakan salah satu teknik yang menjadi salah satu metode dasar pada Boyer Moore yaitu Character-Jump. Algoritma ini berbeda dari algoritma pencocokan string lainnya, dikarenakan proses pencocokan kata dimulai dari karakter terakhir pada kata kunci dan lalu sampai karakter pertamanya. Apabila terjadi perbedaan antara karakter terakhir pada kata kunci dengan karakter pada teks, maka akan dilakukan pencocokan antara karakter – karakter kata kunci satu per satu.

Pencocokan ini dilakukan untuk mengecek adanya karakter dalam teks yang sama dengan karakter pada kata kunci untuk selanjutnya melakukan proses Character-Jump. Apabila terdapat kesamaan, maka kata kunci akan digeser

sehingga posisi karakter yang sama terletak sejajar, dan kemudian dilakukan kembali pencocokan karakter terakhir dari kata kunci. Sebaliknya jika tidak terdapat kesamaan karakter, maka seluruh karakter kata kunci akan bergeser ke kanan sebanyak karakter dari kata kunci. Proses ini akan dibantu array dengan panjangnya adalah banyak karakter yang muncul pada sebuah teks. Array ini digunakan sebagai pencatatan kemunculan terakhir dari kata kunci yang dicari pada sebuah teks.

B. Struktur Kalimat

Struktur kalimat adalah pola kalimat yang terdiri dari partikel-partikel kalimat yang unik yang dimiliki oleh setiap bahasa. Partikel-partikel unik itu dapat berupa sebuah kata atau lebih sehingga memunculkan aturan baru yaitu struktur partikel. Sebagai contoh dalam bahasa Indonesia, kita bisa menyebutkan partikel-partikel yang membentuk kalimat itu berupa subjek, predikat, objek dan keterangan. Lebih dalam lagi, subjek dapat berupa kata benda yang dapat terdiri dari satu kata atau lebih sehingga memunculkan aturan baru dalam penulisan kata benda, misalnya kata benda harus ditulis dalam rumusan “kata benda” + “kata sifat”, “kata benda” + “kata depan” + “kata benda”, “kata benda” + “kata konjungsi” + “kata benda”, dan seterusnya. Sehingga selain struktur kalimat yang akan dibuat, dibuat juga daftar struktur partikel atau struktur frase sebagai tempat menampung aturan pembentukan sebuah frase dalam kalimat.

III. METODOLOGI

Langkah -- langkah yang dilakukan untuk menghasilkan analisis struktur kalimat menggunakan pencocokan string adalah sebagai berikut :

1. Menentukan bahasa yang ingin dianalisis struktur kalimatnya

Struktur kalimat bergantung pada tata bahasanya sehingga harus ditentukan terlebih dahulu bahasa apa yang hendak dianalisis struktur kalimatnya. Bahasa yang dimaksud haruslah bahasa yang jelas rumusan tata kalimatnya sehingga tidak menimbulkan banyak pengecualian di dalamnya.

2. Membuat daftar jenis kata

Tahap ini hanya menggali sejumlah kata dalam bahasa yang telah kita pilih sebagai acuan untuk mengetahui jenis dari kata yang lain. Dalam penentuan jenis kata ini dimungkinkan suatu kata dapat dikelompokkan ke dalam dua jenis kelompok kata atau lebih sehingga memunculkan permasalahan baru. Pada kenyataannya bisa jadi kata dalam suatu bahasa yang memiliki padanan kata

dalam bahasa lain berbeda jenis katanya .

3. Membuat daftar bentuk kalimat

Bentuk kalimat atau pola kalimat merupakan kumpulan partikel kalimat yang tersusun berdasarkan tata kalimat dalam bahasa itu sendiri . Daftar ini dibuat untuk selanjutnya dicocokkan dengan bentuk kalimat yang diuji dengan melakukan pencocokan setiap partikelnya.

4. Membuat daftar bentuk partikel kalimat

Partikel kalimat merupakan satu kata atau lebih yang menempati satu posisi dalam kalimat yang disusun berdasarkan tata kata dalam bahasa itu sendiri . Bentuk partikel kalimat dalam setiap bahasa bersifat unik sama halnya dengan bentuk kalimat . Misalnya dalam suatu bahasa partikel subjek dapat didirikan dengan kata benda dulu lalu disandingkan dengan kata sifat , sedangkan dalam bahasa lain partikel subjek didirikan dengan kata sifat dulu baru disandingkan dengan kata benda .

Lebih dalam lagi masing-masing bahasa memiliki keunikannya sendiri dalam bentuk partikel yang salah satunya membagi jenis kata ke dalam beberapa jenis kata lagi sehingga urutan dalam pembentukan partikel diperhatikan . Salah satunya terdapat dalam bahasa Inggris adalah bagaimana partikel kata sifat atau *adjective* harus memenuhi urutan OSASCOMP (*Opinion , Size , Age , Shape Color , Origin , Material , Purpose*) . Dan terdapat keunikan bentuk partikel yang lain .

5. Mengambil data kalimat yang akan dianalisis

Daftar kalimat yang akan dianalisis sebisa mungkin merupakan kalimat-kalimat yang telah sesuai dengan kaidah-kaidah bahasanya . Hal ini dimaksudkan agar supaya pembaharuan kepada algoritma ini mendekati kaidah-kaidah yang benar . Sehingga pada akhirnya , setelah data sudah cukup lengkap maka dapat mengukur tingkat ketepatan penulisan kalimat dengan lebih akurat .

6. Melakukan pengujian kalimat

Pengujian kalimat dilakukan dalam tiga tahap yaitu:

- Mengkonversi setiap kata menjadi simbol yang merepresentasikan jenis dari kata tersebut . Pengkonversian ini bisa memunculkan lebih dari satu versi dikarenakan kemungkinan suatu kata dikelompokkan lebih dari satu jenis kata .
- Mengkonversi setiap simbol yang telah

terbentuk tadi menjadi simbol partikel kalimat . Pengkonversian ini dilakukan dengan mencocokkan setiap bentuk partikel pada daftar kalimat dengan susunan simbol hasil konversi jenis kata . Pengkonversian ini pun dapat memunculkan lebih dari satu versi konversi .

- Mencocokkan susunan partikel kalimat dengan susunan pada daftar bentuk kalimat . Apabila ditemukan bentuk kalimat yang cocok maka dapat disimpulkan sementara bahwa kalimat yang diuji memiliki tata kalimat yang benar.

7. Pembaharuan daftar

Pembaharuan daftar dilakukan apabila dalam satu alur terdapat bagian yang tidak dapat dikonversi atau dicocokkan baik itu ke dalam jenis kata , partikel kalimat , ataupun dengan bentuk kalimat yang dikarenakan terbatasnya data pada daftar . Sehingga dengan melakukan penyeleksian terhadap penentuan jenis kata , partikel kalimat , atau bentuk kalimat , dapat ditambahkan hal yang tidak ditemukan tadi ke dalam daftar .

IV. PENCOCOKAN STRING

A. Pengujian Kalimat

Pada proses ini , kita akan melakukan hal-hal yang tertulis pada metodologi tadi menggunakan bahasa abstrak . Pertama-tama kita membuat daftar kata yang terdapat informasi mengenai jenis katanya .

Contoh Daftar Kata yang Dibuat dalam Bahasa Abstrak :

Kata	Jenis Kata
Aaa	A
Ababab	A
Bbb	B
Ccc	C
Ddd	D
Dedede	D

Kemudian kita membuat daftar bentuk partikel kalimat . Bentuk partikel kalimat pada permulaan akan lebih baik jika

berisi susunan jenis-jenis kata yang terdapat pada daftar jenis kata tadi .

Contoh Daftar Bentuk Partikel Kalimat yang Dibuat Dalam Bahasa Abstrak :

Bentuk Partikel Kalimat	Posisi Pada Kalimat
AA	X
BA	Y
BD	Y
CA	Z
BC	W

Kemudian kita membuat daftar bentuk kalimat . Bentuk kalimat pada permulaan akan lebih baik jika berisi susunan partikel kalimat yang terdapat pada daftar bentuk kalimat tadi .

Contoh Daftar Bentuk Kalimat yang Dibuat Dalam Bahasa Abstrak :

Bentuk Kalimat
XY
XYW
XYZ

Setelah membentuk daftar permulaan ini , kita dapat menganalisis atau menguji ketepatan struktur suatu kalimat . Kalimat yang diuji pada permulaan ini akan lebih baik memiliki kata-kata, bentuk partikel , atau bentuk kalimat yang terdapat pada daftar .

Misalnya pengujian terhadap kalimat berikut :

Aaa ababab bbb aaa ccc ababab .

Langkah selanjutnya dalam metodologi di atas adalah mengkonversi kalimat yang hendak di uji berdasarkan daftar di atas .

- Konveri Jenis Kata

Akan dilakukan pengkonversian jenis kata terhadap kalimat tersebut sehingga didapat :

A -- A -- B -- A -- C -- A

- Konversi Bentuk Partikel

Apabila disusun kembali susunan di atas maka susunan terdekat yang mudah dikonversi adalah :

AA -- BA -- CA

Yang setelah dikonversi menjadi :

X -- Y -- Z

- Uji Struktur Kalimat

Di dalam daftar bentuk kalimat terdapat bentuk 'XYZ' sehingga pengujian mendapatkan hasil bahwa kalimat tersebut tepat secara tata kalimat .

B. Pembaharuan Daftar

Selanjutnya akan dibahas bagaimana pembaharuan daftar terjadi . Pembaharuan terjadi akibat tidak ditemukannya salah satu dari jenis kata , bentuk partikel kalimat atau bentuk kalimat yang menyebabkan pengkonversian terhambat . Hal ini juga menimbulkan permasalahan baru sesungguhnya bagaimana agar supaya pembaharuan yang terjadi mendekatkan algoritma pengidentifikasian kalimat ini sesuai dengan kaidah-kaidah bahasa dari kalimat yang diuji.

Misalnya dengan menggunakan bahasa abstrak tadi dan daftarnya , akan dianalisis kalimat sebagai berikut :

Aaaa ababab bbb ccc dedede.

Pada pengkonversian pertama akan didapat :

A -- A -- B -- C -- D

Lalu pengkonversian kedua mengalami hambatan karena

A -- A dapat menjadi X

Tetapi B -- C -- D tidak dapat diubah karena tidak terdapat pada daftar

Pendekatan yang bisa dilakukan adalah :

- Menjadikan B -- C -- D sebagai bentuk partikel kalimat baru
- Menjadikan seluruh kombinasi B -- C -- D sebagai bentuk partikel kalimat baru
- Menjadikan sebagian kombinasi B -- C -- D sebagai bentuk partikel kalimat baru
- Menunda proses ini sampai ditemukan dalam kalimat yang lain bentuk partikel kalimat yang sama dalam ketentuan tertentu
- Dan pendekatan yang lain .

Misalnya pendekatan yang dilakukan adalah menjadikan B -- C -- D sebagai bentuk partikel kalimat baru . Sehingga daftar bentuk partikel kalimat menjadi sebagai berikut .

Bentuk Partikel Kalimat	Posisi Pada Kalimat
AA	X
BA	Y
BD	Y
CA	Z
BC	W
BCD	V

Kemudian kita dapat mengkonversi susunan bentuk partikel kalimat itu menjadi :

X -- V

Bentuk X -- V tidak ditemukan dalam daftar bentuk kalimat sehingga karena ini dalam proses pembaharuan maka kita dapat menambahkan bentuk tersebut ke dalam daftar bentuk kalimat sehingga daftar bentuk kalimat menjadi sebagai berikut .

Bentuk Kalimat
XY
XYW
XYZ
XV

C. Ambiguitas dalam Konversi

Seperti yang sebelumnya dibahas bahwa dalam mengkonversi dapat memiliki lebih dari satu versi sehingga perlu ditentukan pendekatan dalam memecahkan masalah ini . Opsi pendekatan ini salah satunya adalah menguji semua versi konversi ke tahap selanjutnya sampai ditemukan versi yang tidak terhambat oleh keterbatasan pada daftar .

Misalnya menggunakan daftar dari bahasa abstrak sebagai berikut :

- Daftar Kata

Kata	Jenis Kata
------	------------

Af	A
Bf	B
Cf	C
Df	D
Ef	E

- Daftar Bentuk Partikel Kalimat

Bentuk Partikel Kalimat	Posisi Pada Kalimat
AB	X
CDE	Y
ABC	Z
DE	W

- Daftar Bentuk Kalimat

Bentuk Kalimat
XY
XYZ
XYW

Maka dengan daftar tersebut dapat diidentifikasi bahwa kalimat di bawah ini ambigu dalam konversi bentuk partikel kalimat.

Af bf cf df ef .

Yang jika dikonversi dalam jenis kata menghasilkan :

A--B--C--D--E

Kalimat tersebut ambigu karena dapat menghasilkan 2 versi konversi ke dalam bentuk kalimat yaitu :

- 1) X -- Y
- 2) Z -- W

Maka dari itu salah satu versi bisa diambil dengan mengunggurkan versi yang lain . Versi 1 yang kita pilih karena

terdapat dalam daftar bentuk kalimat sedangkan versi 2 tidak . Sehingga ambiguitas yang terjadi bisa diselesaikan dengan melanjutkan proses masing-masing versi

D. Hasil Akhir

Hasil yang diharapkan dari semua proses ini adalah daftar jenis kata , bentuk partikel kalimat serta bentuk kalimat dalam suatu bahasa yang menjadikan algoritma pengidentifikasi kalimat ini berjalan dengan baik . Pencocokan string dilakukan di setiap konversi dan pencocokan yang terjadi di dalam proses tersebut . Dengan input dan pendekatan pembaharuan serta penyelesaian terhadap ambiguitas yang baik maka tingkat akurasi algoritma ini dalam hal-hal mengikuti kaidah-kaidah yang ada dalam bahasa yang diuji menjadi lebih baik.

V. KESIMPULAN

Penggunaan bahasa dalam kehidupan sehari-hari merupakan hal yang tidak bisa dihindarkan oleh manusia sebagai makhluk sosial . Dengan salah satu algoritma pencocokan string kita dapat melakukan analisis terhadap setiap tata kalimat yang kita gunakan dalam kehidupan sehari-hari . Algoritma *Boyer- Moore* merupakan algoritma pencocokan string yang cocok untuk mengemban tugas itu karena dapat memaksimalkan permasalahan yang memerlukan pola terhadap suatu teks , seperti pencarian jenis kata dalam daftar untuk menentukan jenis kata dari kata yang berada dalam kalimat yang diuji.

VI. UCAPAN TERIMA KASIH

Puji syukur selalu penulis panjatkan kepada Tuhan Yang Maha Esa atas rahmat dan serta – Nya sehingga penulis dapat diberi kesempatan dan kekuatan untuk menyelesaikan makalah ini. Penulis juga mengucapkan banyak terima kasih kepada orang tua dan teman – teman selama proses penyelesaian makalah ini. Penulis mengucapkan terima kasih kepada seluruh dosen IF2211 Strategi Algoritma, Ibu Masayu Leylia Khodra, S.T., M.T., Ibu Dr. Nur Ulfa Maulidevi, S.T., M.Sc., dan Bapak Dr. Ir. Rinaldi Munir, M.T., khususnya kepada Bapak Rinaldi selaku dosen pengampu mata kuliah Strategi Algoritma di kelas atas bimbingan dan materi yang telah diajarkan kepada saya.

REFERENSI

- [1] Munir, Rinaldi. *Pencocokan String*.
[http://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2017-2018/Pencocokan-String-\(2018\).pdf](http://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2017-2018/Pencocokan-String-(2018).pdf)
- [2] <https://typoonline.com/blog/10-jenis-jenis-kata-yang-perlu-kamu-tahu/>
- [3] <http://mahir-msoffice.blogspot.com/2015/12/adjective-order-urutan-kata-sifat-dalam.html>
- [4] <https://idschool.net/smp/contoh-struktur-kalimat-yang-benar/>

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 26 April 2019



Ridwan Faturrahman
13517150