

Aplikasi Algoritma *Stringmatching* pada Analisa Teks (Text Analysis) untuk *Decision Support System*

Hanif Lyonuais

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia

hanif.lyonnais@gmail.com

Abstract— Sistem teknologi informasi semakin berkembang setiap harinya. Penggunaannya pun bisa dikatakan sudah mencakup semua bidang kehidupan. Kebutuhan akan sistem teknologi informasi menjadi kebutuhan semua industri saat ini. Pada intinya penggunaan teknologi informasi ini untuk memudahkan kita dalam mengelola data untuk menghasilkan informasi. Salah satu peran yang essential dari sistem informasi ini yaitu untuk membantu pengambilan keputusan (*decision making*). Sistem informasi diharapkan dapat mengasihkan informasi – informasi penting dari banyaknya data yang tersedia. Salah satu metode yang dapat digunakan yaitu analisa teks (*text analysis*). Analisa teks ini adalah teknik untuk mengekstraksi informasi dari sumber data berupa teks. Hasil ekstraksi ini kemudian dapat digunakan untuk menjadi pertimbangan dalam pengambilan keputusan

Index Terms—*decision support system, string matching, analisa teks, knuth-morris-pratt.*

I. PENDAHULUAN

Pada era globalisasi seperti sekarang,, industri berkembang pesat. Hal ini dipicu oleh keinginan pasar yang semakin meningkat, contoh mudahnya yaitu penjualan gadget, sekitar 5-10 tahun yang lalu bisnis gadget bukanlah bisnis yang menggiurkan, akan tetapi saat ini tuntutan pasar akan gadget – gadget terbaru memaksa para pelaku industri untuk terus melakukan inovasi dalam produk gadget mereka demi memuaskan keinginan pasar. Selain itu, perkembangan industri yang eksponensial ini juga sangat di dorong oleh perkembangan teknologi. Aspek geografis tidak lagi menjadi halangan untuk mengembangkan pasar, dengan bantuan media dan internet produsen bisa mengenalkan produknya lebih mudah kepada konsumen. Konsep “toko yang harus terletak di tempat strategis” pun perlahan mulai bergeser dengan adanya toko online. Amazon.com merupakan salah satu perusahaan yang berhasil memanfaatkan potensi toko online dan tumbuh menjadi salah satu toko terlaris saat ini.

Jumlah pelaku industri yang semakin banyak tentu saja akhirnya akan melahirkan kompetisi. Perusahaan yang mampu beradaptasi dengan cepat yang akan mampu bertahan dan bahkan merebut dominasi pasar.

Pengambilan keputusan menjadi faktor kunci yang menentukan arah perusahaan dalam menghadapi persaingan. Berikut ini contoh sederhana mengenai bagaimana pengambilan keputusan bisa sangat mempengaruhi keberjalanan perusahaan. Nokia pada awalnya merupakan produsen handphone nomor satu di dunia, dominasi ini telah dipegang teguh selama beberapa tahun. Saat itu kebanyakan handphone Nokia menggunakan sistem operasi Symbian, Symbian memang merupakan OS favorit pada masa itu, kemudian ternyata lahir OS baru yang lebih menjanjikan, yaitu Android. Pada saat itu Nokia harus mengambil keputusan apakah akan mulai menggunakan Android atau tetap berpegang teguh pada Symbian. Pada akhirnya Nokia memutuskan untuk tetap mengembangkan Symbian, dan Android digunakan oleh Samsung sebagai OS standar perangkat handphone mereka. Bisa kita lihat hasilnya saat ini, Samsung yang pada awalnya merupakan minoritas di industri handphone telah berkembang menguasai pasar, sedangkan Nokia malah tertinggal jauh. Ini ilustrasi sederhana bagaimana pengambilan keputusan bisa menjadi sangat penting.

Pengambilan keputusan sangat bergantung pada analisa data. Untuk dapat menghasilkan analisa yang baik dan akurat, tentu saja kita perlu mengolah berbagai data terlebih dahulu. Ini tentu saja tidak mudah karena data yang ada saat ini sangat banyak, misal saja kita ingin membaca trend penjualan, maka kita perlu menganalisa data penjualan 5 tahun ke belakang. Berangkat dari hal tersebut akhirnya muncul kebutuhan akan tool analisa yang powerful dan efektif. Disinilah cikal bakal *data mining*. Salah satu teknik dari *data mining* yaitu analisa teks. Analisa teks merupakan teknik untuk mengambil informasi dari data yang berbentuk teks. Menurut saya, ini merupakan salah satu teknik paling penting karena saat ini kebanyakan data masih tersimpan dalam bentuk teks biasa (belum dalam bentuk database).

Dalam paper ini dibahas lebih dalam mengenai algoritma string matching yang dapat digunakan dalam teknik analisa teks. Selain itu juga dibahas mengenai aplikasi dari analisa teks ini dalam beberapa bidang kehidupan.

II. DASAR TEORI

A. Decision Support System

Decision Support System (DSS) adalah suatu bentuk dari sistem informasi yang secara khusus dibuat untuk mendukung perencana dan stakeholder dalam pengambilan keputusan. DSS dapat mencerminkan berbagai konsep dari pengambilan keputusan dan kondisi yang berbeda-beda, dan akan sangat berguna untuk *semi-structured* atau *unstructured problems* dimana proses pengambilan keputusan ditingkatkan dengan dialog interaktif antara DSS dengan pengguna.

DSS biasanya dilengkapi dengan kemampuan analitik statistic, spreadsheet, dan grafik – grafik untuk membantu memvisualisasikan informasi. Kelebihan utama dari DSS adalah kemampuannya untuk memanfaatkan sistem komputer untuk membantu pengambil keputusan dalam mempelajari masalah dan mengambil kebijakan, dan meningkatkan pemahaman mengenai kondisi lingkungan dimana kebijakan tersebut akan diterapkan dengan mengakses data dan model yang bermanfaat untuk pengambilan keputusan tersebut. DSS berfungsi untuk mengembangkan dan mengevaluasi beragam alternatif solusi untuk memperoleh pemahaman mengenai permasalahan, *trade off* antara obyektif-obyektif yang ada, dan mendukung proses pengambilan keputusan.

. Penggunaan DSS ini akan sangat menguntungkan karena dapat menghemat banyak waktu serta tenaga dalam menganalisa data yang jumlahnya banyak.

Secara umum DSS dibangun oleh tiga komponen :

- Database berisi kumpulan dari semua data bisnis yang diperlukan perusahaan, baik yang berasal dari transaksi sehari – hari ataupun data master. Untuk keperluan DSS, diperlukan data yang relevan dengan permasalahan yang hendak dipecahkan.
- Model Base adalah suatu model yang merepresentasikan permasalahan ke dalam format kuantitatif sebagai dasar simulasi data pengambilan keputusan, termasuk di dalamnya tujuan dari permasalahan, komponen terkait, batasan yang ada, dan lain – lain.
- Software system adalah tools yang akan mengelola kedua komponen di atas.

B. Stringmatching

Algoritma stringmatching adalah algoritma yang digunakan untuk mencari suatu kata (pattern) dalam suatu kalimat atau teks (text). Untuk memudahkan pemahaman kata yang ingin kita cari kita beri nama *pattern*, sedangkan kalimat dimana kita mencari kata tersebut kita beri nama *text*.

Ada beberapa algoritma string matching, diantaranya :

- Brute force
- Knuth-Morris-Pratt
- Boyer-Moore

Contoh penggunaan algoritma :

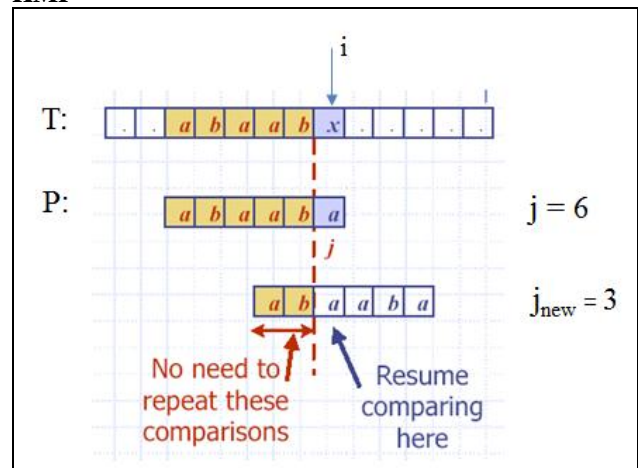
Brute force:

```

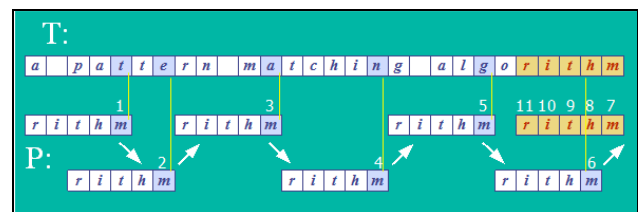
Text      :no one like her
Pattern   :like

      no one like her
s=0      like
s=1      like
s=2      like
s=3      like
s=4      like
s=5      like
s=6      like
s=7      like
    
```

KMP



Boyer-Moore

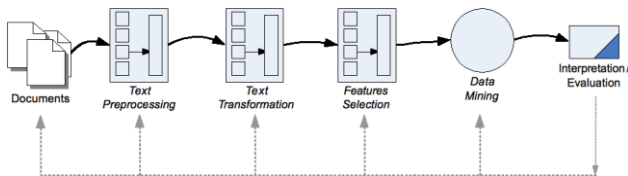


B. Text Analysis

Teks analysis juga populer dengan istilah text mining, sekarang istilah teks analysis lebih sering digunakan dalam konteks bisnis, sedangkan teks mining dalam konteks keilmuan.

Ini merupakan salah satu bidang khusus dari data mining. Text Analysis adalah proses pencarian informasi berguna dari sekumpulan data berupa dokumen teks. Kumpulan dokumen teks ini memiliki format yang tidak terstruktur atau semi-terstruktur.

Proses pada text analysis bisa dilihat sebagai berikut :



Permasalahan yang dihadapi pada text analysis sama dengan permasalahan yang terdapat pada data mining, yaitu jumlah data yang besar, dimensi yang tinggi, data dan struktur yang terus berubah, dan data noise. Perbedaan di antara keduanya adalah pada data yang digunakan. Pada data mining, data yang digunakan adalah structured data, sedangkan pada text analysis, data yang digunakan pada umumnya adalah unstructured data, atau minimal semistructured. Hal ini menyebabkan adanya tantangan tambahan pada text analysis yaitu struktur text yang complex dan tidak lengkap, arti yang tidak jelas dan tidak standard, dan bahasa yang berbeda ditambah translasi yang tidak akurat.

III. ANALISA

A. Teknik Analisa Teks

Ada beberapa teknik yang digunakan dalam analisa teks yaitu :

a. Information Extraction

Pada proses ini kita mengidentifikasi kata kunci dan keterkaitannya di dalam teks. Proses yang dilakukan yaitu dengan mencari kata-kata yang telah ditentukan sebelumnya dengan menggunakan algoritma pattern matching. Teknik ini sangat membantu, terlebih lagi jika kita digunakan dalam pencarian teks yang sangat panjang.

b. Categorization

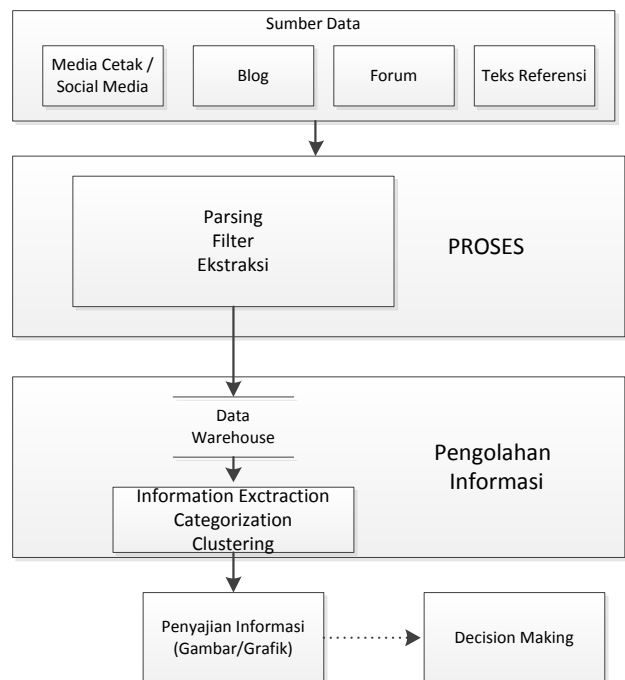
Categorization adalah proses mengidentifikasi tema utama dari suatu dokumen. Dokumen ini akan digolongkan ke beberapa topic yang telah ditentukan sebelumnya. Pada categorization kita tidak mencari informasi secara langsung seperti pada information extraction, kita hanya menghitung jumlah kemunculan term tertentu di dalam teks, setelah didapatkan jumlah katanya dokumen dipisah berdasarkan topik yang paling relevan.

Categorization ini sangat bergantung pada pencarian kata, sehingga kata yang dicari dalam teks harus merepresentasikan topik-topik yang telah ditentukan sebelumnya .

c. Clustering

Clustering adalah teknik untuk menggabungkan dokumen-dokumen yang mirip ke dalam satu grup. Pada clustering dokumen bukan dipisahkan berdasarkan topik seperti pada categorization, akan tetapi digabungkan berdasarkan kemiripan tiap dokumen tersebut.

Di bawah ini adalah diagram yang menggambarkan bagaimana proses analisa teks dalam DSS berlangsung :



Sumber data yang digunakan tidak harus seperti yang tertulis pada diagram di atas, sumber data yang digunakan disesuaikan dengan kebutuhan pengguna.

B. Penggunaan Algoritma Stringmatching

Dalam teknik analisa teks beberapa bagian membutuhkan pencarian kata atau pattern. Secara default pencarian yang digunakan yaitu brute force karena paling mudah diimplementasi.

Akan tetapi algoritma bruteforce merupakan algoritma yang tidak mangkus. Algoritma lain yang lebih mangkus yaitu Knuth-Morris-Pratt dan Boyer-Moore.

Salah satu penggunaan algoritma string matching ini yaitu pada tahap categorization, dimana kita ingin menentukan tema utama dari teks. Kita menentukan tema ini dengan mencari jumlah kemunculan term tertentu dalam teks.

Misal ada tiga kategori yaitu : Makanan, Gadget, Hiburan

Term yang kita gunakan untuk mencari yaitu : *makanan*, *handphone*, dan *TV*

Teks :

Udah lama nih ga makan di luar, jadi pengen makan di Café Halaman. Makanan disana lumayan enak dan harganya terjangkau. Porsi makanannya juga besar terus kita bisa nonton TV sambil nunggu makanan kita dimasak. Disini juga ada wifi jadi gw bisa internetan pake handphone., lumayan bisa gratisan.

Berikut ini hasil pencarian dengan kedua algoritma tersebut :

KMP

Term

Makanan : 3 kali

Handphone : 1 kali

TV : 1 kali

Boyer Moore

Makanan : 3 kali

Handphone : 1 kali

TV : 1 kali

Maka teks diatas dimasukan ke kategori *Makanan*.

Jumlah perbandingan karakter yang dilakukan oleh Boyer-Moore lebih sedikit dibandingkan dengan yang dilakukan oleh KMP

KMP : 553 kali

Boyer Moore : 324 kali

Boyer-Moore memang merupakan algoritma yang lebih cerdas sehingga lebih efektif digunakan.

Algoritma Boyer Moore membandingkan karakter dari huruf paling belakang dan jika karakter tersebut ternyata tidak ada di dalam pattern maka seluruh pattern akan bergerak maju. Ini banyak mengurangi jumlah karakter yang harus dibandingkan.

Algoritma KMP agak kurang cocok diaplikasikan pada perbandingan bahasa indonesia. Dalam KMP perbandingan mengandalkan terbentuknya pola antara prefix dan suffix sedangkan pada kenyataannya jarang sekali kata yang memiliki prefix dan suffix tersebut.

IV. APLIKASI DSS BERBASIS ANALISA TEXT

Decision Support System biasanya memiliki berbagai teknik untuk mengambil informasi dari data yang ada. Analisa teks merupakan salah satu teknik yang paling penting karena saat ini kebanyakan data ada dalam bentuk

dokumen teks. Teknik ini juga sangat cocok digunakan untuk bisnis yang B2C(Business to Consumer) dimana produk yang dihasilkan digunakan langsung oleh konsumen. Saat ini feedback dari konsumen bisa diambil langsung dari tulisan – tulisan di internet, walaupun tidak semua yang dituliskan itu benar akan tetapi feedback yang muncul cukup untuk menjadi bahan pertimbangan bagi produsen.

Beberapa bidang yang dapat memanfaatkan DSS berbasis analisa text yaitu :

- Sentiment analysis
DSS bisa membantu kita untuk mendapatkan feedback mengenai suatu produk. Sumber data yang digunakan yaitu blog, social media, artikel di internet. Informasi yang didapatkan tersebut bisa membantu pengguna untuk membuat kebijakan terkait feedback produknya.
- Security analysis
DSS juga bisa memberikan informasi mengenai adanya ancaman keamanan berdasarkan artikel, email, maupun dokumen yang beredar di internet. Informasi ini bisa menjadi early warning sebelum terjadi hal yang tidak diinginkan
- Marketing analysis
Dengan menggunakan DSS ini, pengguna bisa melihat trend pasar pada saat itu. Dengan sumber data blog, social, media, dan artikel di internet, DSS bisa melakukan analisa mengenai produk yang paling sering dibicarakan, pola belanja remaja, dan lain – lain.

V. CONCLUSION

Salah satu metode pengolahan data dalam Decision Support System yaitu analisa text. Kita dapat melakukan analisa text secara lebih efisien dengan menerapkan algoritma stringmatching pada bagian – bagian pencarian kata. Dalam aplikasinya, pencarian kata dengan algoritma Boyer-Moore lebih akurat, terutama saat kita mencari pattern yang berupa kalimat. Beberapa aplikasi DSS yang menggunakan analisa teks yaitu sentiment analysis, security analysis, marketing analysis.

REFERENCES

- [1] Richardus Eko Indrajit, "Decision Support System". 2011.
- [2] <http://itsum.wordpress.com/>
- [3] http://en.wikipedia.org/wiki/String_searching_algorithm
- [4] <http://zingar-college.blogspot.com/>
- [5] http://en.wikipedia.org/wiki/Data_mining
- [6] <http://www.informationweek.com/software/business-intelligence/text-analytics-demand-approaches-1-billi/>

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, 20 Desember 2012

A handwritten signature in black ink, appearing to read 'Hanif'.

Hanif Lyonnais