

**PENARIKAN KESIMPULAN PADA *WEBSITE* BERITA *ONLINE*  
DENGAN METODE LSA (*LATENT SEMANTIC ANALYSIS*) DAN  
KMP (*KNUTH MORRIS PRATT*)**

**SKRIPSI**

**Oleh:**

**Salvatores Pitison 1620250019**

**Program Studi Teknik Informatika**



**STMIK  MDP**

**STMIK Global Informatika MDP  
Palembang  
2020**

## STMIK Global Informatika MDP

---

Program Studi Teknik Informatika  
Skripsi Sarjana Komputer  
Semester Gasal Tahun 2019/2020

### **PENARIKAN KESIMPULAN PADA WEBSITE BERITA ONLINE DENGAN METODE LSA (LATENT SEMANTIC ANALYSIS) DAN KMP (KNUTH MORRIS PRATT)**

Salvatores Pitison 1620250019

#### **Abstrak**

Membaca teks terlalu panjang tentunya menjadi tantangan sendiri bagi para pembaca berita. Sajian berita yang terlalu panjang tentunya membuat bosan. Salah satu solusi yang ditemukan adalah penarikan kesimpulan atau peringkasan teks pada berita sehingga isi berita menjadi lebih mudah dipahami. Peringkasan teks terbukti dalam menarik kesimpulan dari sebuah teks yang panjang. Metode yang biasa digunakan adalah dengan LSA atau kNN. Namun LSA banyak dipakai karena metode ini memberikan hasil yang lebih baik karena adanya *preprocessing* pada teks. *Preprocessing* pada LSA umumnya memakai proses *stemming* sebagai salah satu bagiannya. Namun dengan *stemming* ada beberapa makna kata yang menjadi hilang. Pencarian String dengan KMP dapat menggantikan tahap *stemming*. Pada penelitian kali ini, dilakukanlah pengujian algoritma KMP dan metode LSA dalam penarikan kesimpulan teks berita *online*. Berdasarkan hasil yang didapat, penggunaan algoritma KMP dan metode LSA dapat menarik kesimpulan dengan nilai *error* lebih kecil dari 30%.

**Kata kunci:** Penarikan kesimpulan, *LSA*, *KMP*, berita *online*



# STMIK MDP

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER  
GLOBAL INFORMATIKA MDP

## **BAB 1**

### **PENDAHULUAN**

Bab ini menjelaskan dasar pemikiran dan pertimbangan pelaksanaan penelitian yang berfungsi sebagai panduan pelaksanaan penelitian yang sistematis dan terarah. Penjelasan pertama adalah mengenai latar belakang penelitian (subbab 1.1), dilanjutkan dengan penjelasan rumusan masalah berdasarkan latar belakang (subbab 1.2), pembatasan ruang lingkup penelitian (subbab 1.3), penjelasan tujuan dan manfaat yang diharapkan dari penelitian (subbab 1.4), metodologi penelitian (subbab 1.5), dan sistematika penulisan (subbab 1.6).

#### **1.1 Latar Belakang**

Informasi merupakan hal yang sangat dibutuhkan oleh setiap orang. Semakin banyak informasi yang tersedia dalam suatu dokumen maka akan semakin panjang pula sebuah dokumen tersebut (Hariharan & Srinivasan, 2008). Dengan banyak dan kompleksnya informasi yang ada tentunya kita ingin untuk mendapatkan kesimpulan intisari dari informasi tersebut. Untuk mendapatkan inti dari informasi tersebut sendiri, dapat dilakukan dengan berbagai cara. Salah satunya adalah membaca keseluruhan informasi itu sendiri sehingga kita paham akan isi dari informasi itu dan mendapatkan suatu inti atau pokok bahasan dari informasi itu. Inti atau pokok bahasan tadi dapat menjadi esensi dari

keseluruhan informasi itu sendiri. Cara demikian tentunya akan memakan waktu yang lumayan panjang apalagi teks yang akan diringkas merupakan teks yang panjang.

Dengan kelemahan dari cara manual di atas, tentu kita perlu untuk mengatasi masalah tersebut. Sebuah peringkasan teks otomatis yang membantu kita tanpa harus membaca keseluruhan teks namun tetap mendapatkan esensi dari informasi itu sendiri tentulah menjadi keinginan kita. Peringkasan teks secara otomatis atau lebih dikenal sebagai Automatic Text Summarization merupakan solusi yang dapat digunakan dalam hal peringkasan teks dokumen dan tetap mempertahankan kualitas dari hasil ringkasan tersebut (Dokun & Celebi, 2015). Peringkasan teks otomatis ini tentulah menjadi hal yang dapat meningkatkan keefisienan kita sebagai pembaca informasi misalnya pada informasi berita.

Penelitian mengenai peringkasan teks sendiri telah banyak dilakukan oleh para peneliti sebelumnya. Peringkasan teks dengan menggunakan metode *Fuzzy Genetic Algorithm and Genetic Programming* yang dimana hasil akhir dari penelitian tersebut didapatkan hasil peringkasan sebesar 73% menyerupai hasil peringkasan manual (Kiani & Akbarzadeh-T, 2006). Peringkasan teks menggunakan metode *Fuzzy Logic* yang dikombinasikan dengan metode *Latent Semantic Analysis* mendapatkan hasil yang lebih baik dibandingkan tanpa menggunakan *Latent Semantic Analysis* (Babar & Patil, 2015). Penelitian terkait lainnya mengenai metode *Knuth Morris Pratt* yang dipadukan dengan

metode Trimmer. Hasil akhir dari penelitian tersebut didapat hasil peringkasan sebesar 70% (Widi Madiyanto,dkk, 2003).

*Latent Semantic Analysis* sendiri merupakan metode statistik aljabar yang digunakan untuk mencari kesamaan unsur semantik pada tiap kalimat dan kata, dengan menggunakan metode *Singular Value Decomposition* (SVD) yang dapat mengurangi kesalahan serta meningkatkan akurasi. Penelitian mengenai peringkasan teks dengan membandingkan metode *Latent Semantic Analysis* dan metode *Maximal Marginal Relevance* didapatkan hasil bahwa metode *Latent Semantic Analysis* lebih baik (Xiong & Luo, 2015). Peringkasan teks menggunakan metode *Latent Semantic Analysis* pada dokumen berbahasa Kanada yang menghasilkan ringkasan dengan pengujian *Accuracy* sebesar 94% dan *Precision* sebesar 80% (Geetha J K & Deepamala N, 2015). Penelitian lainnya menerapkan metode *Latent Semantic Analysis* pada proses Clustering dokumen menghasilkan tingkat akurasi yang lebih baik (Luthfiarta,dkk, 2013).

Sementara itu, *Knuth Morris Pratt* sendiri adalah salah satu metode pencarian String yang banyak dipakai dibandingkan dengan metode algoritma pencarian String lainnya dan sesuai dengan susunan karakter (Rosario M, Susilo B, Ernawati, 2015).

Berdasarkan uraian yang telah dipaparkan sebelumnya, dapat diperhatikan bahwa hasil dari peringkasan teks dengan metode LSA atau KMP menghasilkan hasil yang optimal (tingkat akurasi 70% sampai 90% dan tingkat presisi 80%), namun belum ada penelitian yang menggabungkan kedua metode

tersebut sehingga tingkat *error* masih belum diketahui dalam melakukan hal yang sama yaitu peringkasan teks, maka penelitian ini penting untuk dilakukan.

Melalui skripsi ini, dilakukan penelitian dengan menggabungkan metode LSA dan algoritma KMP dalam meringkas teks berita *online* di beberapa kategori melalui perhitungan tingkat kesalahan (*error rate*).

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, dapat dirumuskan masalah yang ada, yaitu bagaimana tingkat *error* (*error rate*) metode LSA dan KMP untuk penarikan kesimpulan dari *website* berita *Online*.

## 1.3 Ruang Lingkup

Dalam penelitian ini akan ditentukan ruang lingkup atau batasan masalah sebagai berikut :

1. Metode yang diterapkan adalah metode LSA dan KMP
2. Aplikasi ini dibuat memakai program NetBeans
3. Bahasa pemrograman yang digunakan adalah Java
4. Bahasa yang dipakai sebagai dokumennya adalah bahasa Indonesia
5. Parameter pengujian performa metode adalah *confusion matrix* yang meliputi *error rate* .
6. *Website* yang dipakai adalah *website* berita *online* yaitu, *website* detik.com dengan kategori berita yang berbeda beda

7. Kategori berita yang adalah kesehatan, ekonomi dan olahraga dengan berita untuk *dataset* sebanyak 20 berita dan berita yang dijadikan data uji sebanyak 5 berita
8. Nilai kompres (*compression rate*) yang dipakai adalah 30%
9. Pakar yang digunakan adalah mahasiswa yang menempuh pendidikan bahasa Indonesia

#### 1.4 Tujuan dan Manfaat

Tujuan penyusunan Skripsi ini adalah mengukur tingkat *error* metode LSA dan KMP dalam melakukan penarikan kesimpulan pada teks berita *Online*.

Manfaat dari penyusunan Skripsi ini yaitu :

1. Menciptakan suatu sistem yang dapat digunakan sebagai solusi bagi para pembaca agar lebih cepat menemukan suatu kesimpulan dari *website* berita *online*
2. Memahami alur LSA dan KMP dalam melakukan peringkasan teks
3. Memberikan pembelajaran berupa tingkat keakurasian peringkasan dengan metode LSA dan KMP.
4. Membuat aplikasi yang dapat melakukan *text mining* dan *text processing* berupa meringkas isi *website* berita *online*.



## 1.5 Metodologi Penelitian

Penelitian dilaksanakan secara bertahap hingga mencapai suatu hasil yang diinginkan. Penelitian ini dilakukan dengan tahapan sebagai berikut:

Penelitian dilaksanakan secara bertahap hingga mencapai suatu hasil yang diinginkan. Penelitian ini dilakukan dengan tahapan sebagai berikut :

### 1. *Identifikasi Masalah*

Penelitian dimulai dengan mencari topik-topik mengenai penelitian terkait pemrosesan teks. Ruang lingkup dikhusus pada metode yang digunakan yaitu metode LSA dan KMP untuk menilai performanya melalui *confusion matrix*. Pada tahapan ini, dilakukan analisis perkembangan penggunaan metode terkait, pendekatan yang telah diuji, dan mencari ruang yang mampu dioptimalkan.

### 2. *Studi Literatur*

Dilakukan pencarian jurnal, buku dan hasil penelitian lain yang berhubungan dengan pemanfaatan metode LSA dan KMP. Fokus dari literatur yang dicari adalah penggunaan metode LSA dan KMP serta mendapatkan teks dari URL berita *online*. Tahapan ini ditujukan untuk mengumpulkan referensi yang relevan sehingga penelitian ini dapat memberikan kontribusi penelitian yang baru dan memiliki dasar yang kuat.

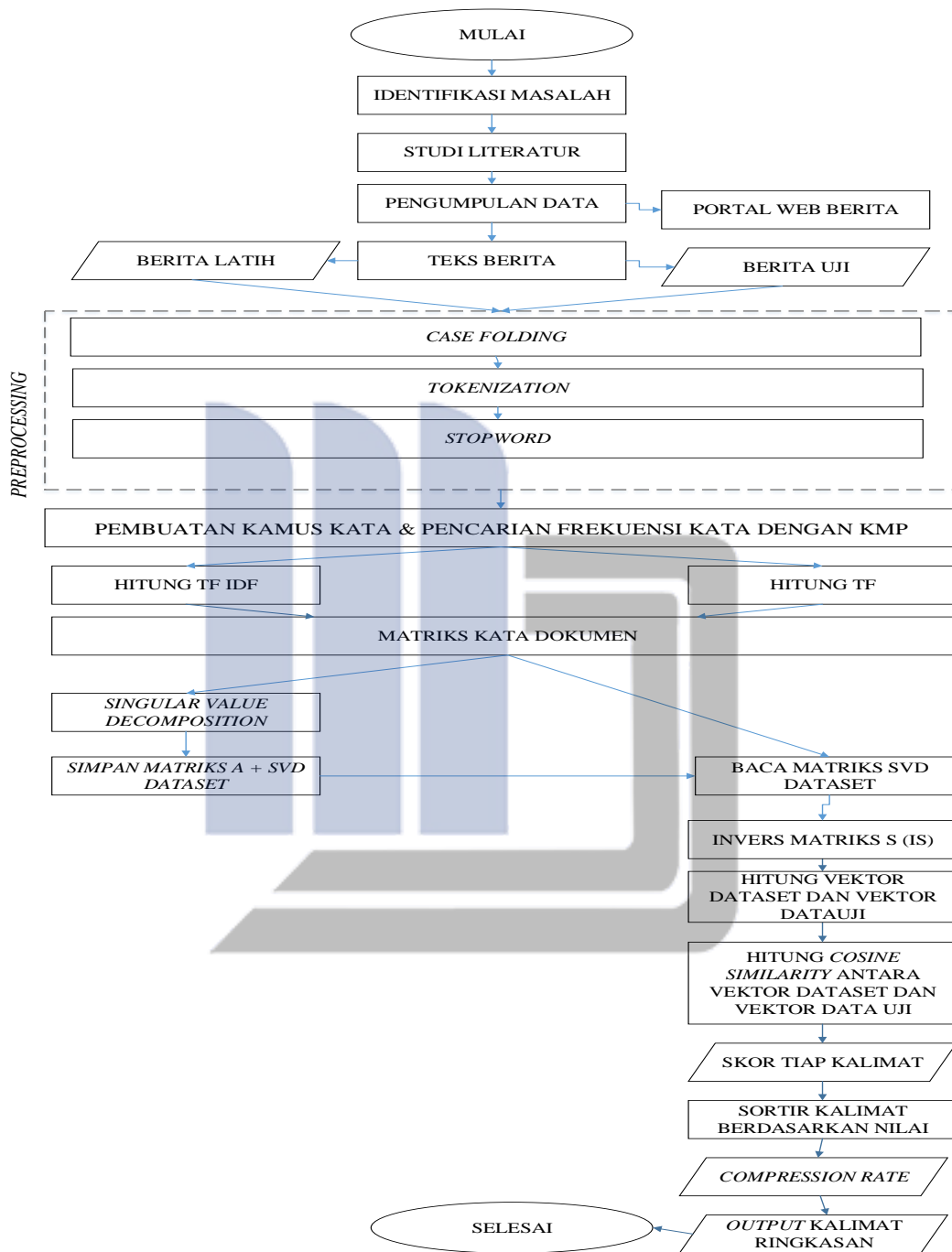
### 3. *Pembuatan Prototype*

Pembuatan *prototype* dilakukan dengan merancang sistem berupa penarikan teks melalui *website*. Setelah melalui tahap perancangan, tahap

yang dilakukan berupa penerjemahan teks sehingga bisa didapatkan kesimpulan seperti yang diharapkan. Implementasi dilakukan dengan program NetBeans dan Bahasa pemrograman Java. Tujuan tahap ini adalah untuk memperjelas kebutuhan penelitian yang disimpulkan dari tahap studi literatur pada sebuah *prototype* agar dapat diimplementasikan menjadi program yang utuh.

Berikut ini bagan metodologi pembuatan *dataset*:





**Gambar 1.1 Pseudocode KMP Pencarian**

#### 4. Pengujian

Pengujian program dilakukan setelah tahap sebelumnya dipenuhi. Tujuan tahap ini adalah untuk menguji keluaran yang dihasilkan. Hasil dari pengujian akan di ditampilkan dalam bentuk *confusion matrix* dengan parameter *error rate*. Cara pengujiannya adalah dengan memakai beberapa artikel yang telah diringkas secara manual oleh pakar (mahasiswa jurusan pendidikan Bahasa Indonesia) lalu dibandingkan dengan hasil dari peringkasan dari ringkasan sistem yang telah dibuat sebelumnya terhadap berita keseluruhan. Kalimat hasil ringkasan akan dicari nilai TP (True Positif), FP (false positif), FN (False Negatif) dan TN (True Negatif). Adapun rumus dari *error rate* sebagai berikut :

$$\frac{(FP+FN)}{(TP+TN+FP+FN)} \times 100\% \dots\dots\dots(I.1)$$

#### 5. Analisis Performa

Hasil dari pengujian program akan dianalisis untuk membentuk sebuah kesimpulan dan rekomedasi penelitian.

## **1.6 Sistematika Penulisan**

Penulisan Skripsi ini dikelompokkan ke dalam 5 bab, yaitu pendahuluan, landasan teori, rancangan pengujian algoritma, pengujian dan analisis program, dan penutup.

### **BAB 1 PENDAHULUAN**

Bab ini berisi pendahuluan yang berupa penjelasan latar belakang masalah alasan pentingnya penelitian ini dilakukan, permasalahan, ruang lingkup, tujuan dan manfaat, metodologi pengembangan aplikasi serta sistematika penulisan.

### **BAB 2 LANDASAN TEORI**

Bab ini membahas teori pendukung pengembangan yang berhubungan dengan penelitian baik dari pemrosesan teks, algoritma *KMP*, *LSA*, dan penelitian yang terkait.

### **BAB 3 RANCANGAN PENGUJIAN ALGORITMA**

Bab ini membahas tentang desain dan sistem yang akan dibuat meliputi lingkungan pengembangan program, rancangan algoritma, kerangka kerja penelitian, rancangan alur dan teks berita, spesifikasi kebutuhan aplikasi.

### **BAB 4 ANALISIS DAN PEMBAHASAN**

Bab ini berisi hasil dari pengujian algoritma yang ditulis dalam bentuk tabel dan dicantumkan tampilan masing-masing pengujian, serta analisis dari hasil pengujian.

## **BAB 5 PENUTUP**

Bab ini berisi kesimpulan dari hasil pengujian yang dilakukan serta saran untuk pengembangan aplikasi selanjutnya.





# STMIK MDP

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER  
GLOBAL INFORMATIKA MDP

## DAFTAR PUSTAKA

- B. Raharjo.2011.“*Presisi Dan Akurasi,*” Beni Raharjo – Nature, Environment, Remote Sensing, GIS, IT and Myself, 17-Mar-2011. Tersedia pada : <http://www.raharjo.org/math/presisi-dan-akurasi.html>. (diakses 20 Oktober 2019)
- Babar & Patil. 2015.”Improving Performanca Text Summarization”. *Research Gate*. 354:363  
*Berorientasi Objek Edisi Revisi*, Bandung:Infomatika
- Celebi & Dokun. 2015. “A comparative study of efficient initialization method Singular Value Decomposion”.*Expert System with Application*. 40(1).
- Gaikwad, D. K., & Mahender, C. N. 2016. “A Review Paper on Text Summarization”, 5(3), 154–160. <https://doi.org/10.17148/IJARCCE.2016.5340>
- Geetha J. K. & Deepamala N. 2015. “Kanada Text Summarization using Latent Semantic Analysis”. *Jurnal Teknologi*. 2:10
- Imbar, dkk. 2014. “Implementasi Cosine Similarity dan Algoritma Smith Waterman untuk Mendeteksi Kemiripan Teks. *Jurnal Informatika*”. Volume 10.Nomor 1
- Informatikalogi. 2016. “Praprocessing Teks”. Praprocessing pada teks. Tersedia pada : <https://informatikalogi.com/text-preprocessing/> (diakses 5 September 2019)
- Informatikalogi. 2016. “Tf Idf”. Pembobotan. Tersedia pada: <https://informatikalogi.com/term-weighting-tf-idf/> (diakses 5 September 2019)
- Luthfiarta, dkk. 2013.*Algoritma LSA pada peringkas dokumen otomatis untuk proses clustering dokumen*. Seminar nasional Teknologi Informasi & Komunikasi Terapan : Bandung
- R. Peter, S. G, D. G, & S. Kp, 2012. "Evaluation of SVD and NMF Methods for Latent Semantic Analysis," *InternationalJournal Engineering*. Vol 1. Pp 308-310



Rossaria M, Susilo B, Ernawati. 2015. *Implementasi Algoritma Pencocokan String Knuth-Morris-Pratt Dalam Aplikasi Pencarian Dokumen Digital Berbasis Android*. Jurnal Komputasi 2 : 183-195

Torres-Moreno, J.-M. 2014. *Automatic Text Summarization (Cognitive Science and Knowledge Management) (1st ed.)*. Wiley-ISTE.

Tus Sa'diah, Halimah.2017. "Implementasi Algoritma Knuth-Morris-Pratt Pada Fungsi Pencarian Judul Tugas Akhir Repository". Jurnal Komputasi.Vol 14. No 1. Pp115 -124.ISSN :1693-7554

