

“ PENERAPAN ALGORITMA KNUTH MORRIS PRATT PADA PENCARIAN KAMUS DIGITAL TANAMAN OBAT ”

Erika candra¹⁾, Sri Setyaningsih²⁾, Arie Qur'ania³⁾

^{1, 2 & 3)}Program Studi Ilmu Komputer, FMIPA, Universitas Pakuan Bogor

¹⁾yuli.ec31@gmail.com, ²⁾@unpak.ic.id, ³⁾qurania@unpak.ac.id

Abstrak

Tanaman obat merupakan jenis tanaman yang sebagian atau seluruh isi tanaman tersebut seperti daun, batang, buah, umbi/rimpang, hingga akar digunakan sebagai obat, bahan atau ramuan obat-obatan. Saat ini perkembangan teknologi informasi dan komunikasi pada perangkat mobile sudah semakin maju. Aplikasi kamus digital tanaman obat memuat kumpulan nama tanaman obat. Penggunaan dalam aplikasi kamus digital tanaman obat sulit untuk mencari informasi yang pasti terhadap pencarian obat karena kosakata dalam penggunaannya menghasilkan beberapa arti lain dan belum tentu sama dengan apa yang dicari. Sehingga diperlukan aplikasi kamus digital tanaman obat lebih terarah dengan penggunaan kosa kata yang tepat. Tujuan dari penelitian ini ialah menerapkan algoritma Knuth Morris Pratt pada pencarian kamus digital tanaman obat. Inputan data tanaman yang akan diberikan oleh sistem kepada pengguna didapatkan melalui proses pencocokan string antara teks dan pola. Hasil dari penerapan knuth morris pratt diharapkan dapat membantu pengguna mencari nama tanaman obat sesuai kosakata penggunaannya.

Kata kunci : kamus digital, tanaman obat, algoritma knuth morris pratt.

Abstract

Medicinal plants are a type of plant in which part or all of the plant's contents, such as leaves, stems, fruit, tubers / rhizomes, and roots are used as medicine, ingredients or medicinal ingredients. Currently, the development of information and communication technology on mobile devices is increasingly advanced. The digital medicinal plant dictionary application contains a collection of medicinal plant names. It is difficult to use a digital dictionary application for medicinal plants to find definite information on drug search because the vocabulary in its use produces several other meanings and is not necessarily the same as what is being searched for. So a digital dictionary application of medicinal plants is needed to be more focused with the use of the right vocabulary. The purpose of this study is to apply the Knuth Morris Pratt algorithm to the search for a digital dictionary of medicinal plants. The plant data input that will be provided by the system to the user is obtained through a string matching process between text and patterns. The results of the application of the knuth morris pratt are expected to help users search for medicinal plant names according to the vocabulary of their use.

Key words: digital dictionary, medicinal plants, knuth morris pratt algorithm.

1. Pendahuluan

Tanaman obat merupakan jenis tanaman yang sebagian atau seluruh isi tanaman tersebut seperti daun, batang, buah, umbi/rimpang, hingga akar digunakan sebagai obat, bahan atau ramuan obat-obatan. Penggunaan obat tradisional di Indonesia sudah berlangsung sejak ribuan tahun yang lalu, sebelum obat-obatan modern ditemukan dan dipasarkan. [1]. Saat ini perkembangan teknologi informasi dan komunikasi pada perangkat mobile sudah semakin maju. Khususnya pada penggunaan aplikasi, aplikasi merupakan sebuah proses dimana keperluan pengguna dirubah kedalam bentuk paket perangkat lunak atau kedalam spesifikasi pada komputer yang berdasarkan dengan sistem informasi.

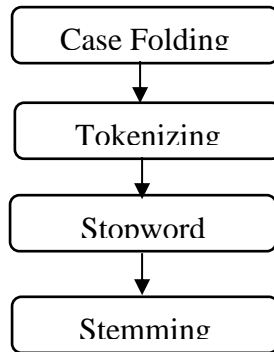
Aplikasi kamus digital tanaman obat memuat kumpulan nama tanaman obat. Penggunaan dalam aplikasi kamus digital tanaman obat sulit untuk mencari informasi yang pasti terhadap pencarian obat karena kosakata dalam penggunaannya menghasilkan beberapa arti lain dan belum tentu sama dengan apa yang dicari. Sehingga diperlukan aplikasi kamus digital tanaman obat lebih terarah dengan penggunaan kosa kata yang tepat [2]. Pencarian terdahulu dengan menggunakan penerapan algoritma *knuth morris pratt* juga dilakukan oleh Sigit Triono Haris (2015) yang berjudul Aplikasi Android Kamus Bahasa Jawa Serang- Indonesia menggunakan Algoritma *Knuth Morris Pratt* [3]. Pencarian dilakukan untuk membantu pengguna menemukan terjemahan bahasa jawa serang ke dalam bahasa Indonesia dengan mudah sehingga pengguna dapat mempelajari bahasa Jawa Serang dengan cepat.

Penelitian lain dilakukan oleh Wistiani Astuti (2017) yang berjudul Analisis *String Matching* Pada Judul Skripsi dengan Algoritma *Knuth Morris Pratt* [4]. Pada penelitian ini menggunakan algoritma *knuth morris pratt* untuk menganalisis bagaimana proses pencocokan *string* yang dihasilkan dan membandingkan sejauh mana nilai kemiripan dari beberapa judul yang sama dan serupa sehingga dapat memberikan suatu informasi yang efektif bagi mahasiswa. Penelitian selanjutnya oleh Rizal Adi Saputra (2017) dengan judul Aplikasi Kamus Tanaman Obat Menggunakan Algoritma *Boyer Moore* Berbasis Android [5]. Penelitian ini bertujuan untuk membuat aplikasi kamus tanaman obat berbasis android yang bisa membantu mahasiswa ataupun masyarakat secara umum yang ingin mencari tanaman herbal secara praktis dan efisien. Pencarian dilakukan dengan algoritma *boyer moore* dengan mencocokkan karakter dari sebelah kanan *pattern* (pola yang dicari). Penelitian selanjutnya oleh Nazar Muhammad Ikhbal (2019) dengan judul Kamus Digital Tanaman Obat Menggunakan Algoritma *Rocchio* Berbasis Mobile [6]. Penelitian ini bertujuan untuk membuat kamus digital tanaman obat berbasis mobile. Pencarian dilakukan untuk pencarian data tanaman obat dengan mengecek kedekatan *query* dalam database.

Penelitian ini bertujuan menerapkan algoritma *Knuth Morris Partt* pada pencarian kamus digital tanaman obat. Yaitu algoritma pencocokan *string* yang merupakan algoritma yang digunakan untuk melakukan pencarian sebuah *string* yang terdiri dari beberapa karakter yang biasa disebut *pattern*. Alasan penggunaan penerapan algoritma *knut morris pratt* karena dalam pencocokan *string* algoritma ini menyimpan informasi yang digunakan untuk melakukan jumlah pergeseran, algoritma *knuth morris pratt* menggunakan informasi tersebut untuk membuat pergeseran yang lebih jauh.

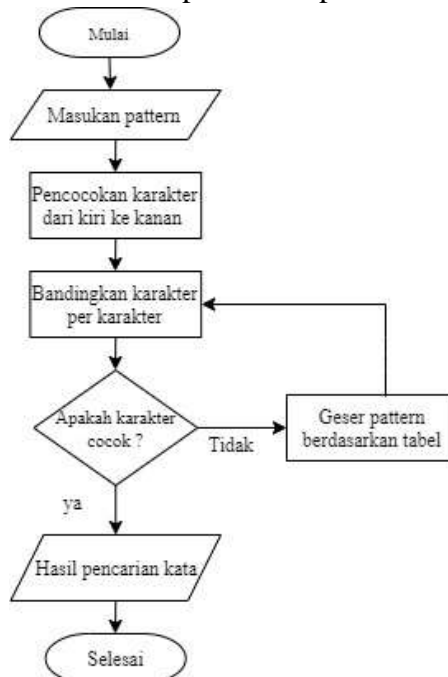
2. Metode Penelitian

Tahap awal penelitian adalah dengan pengumpulan data tentang tanaman obat sebanyak 640 data tanaman obat. Data gambar dan khasiat tanaman obat yang sudah diinput akan *filter* atribut sesuai dengan kata kunci yang biasa diinputkan untuk pencarian tanaman obat berdasarkan nama penyakit. Proses pencocokan kata kunci dan data tanaman obat akan diproses dalam tahap *preprocessing* diantaranya *Case folding*, *Tokenizing*, *Stopword*, *Stemming*. Tahapan Preprocessing dapat dilihat pada gambar 1.



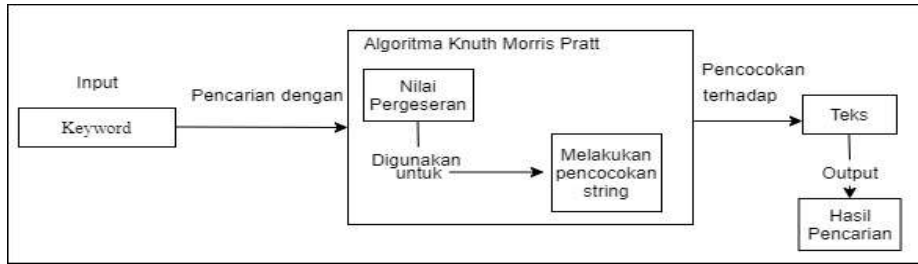
Gambar 1. Tahapan Preprocessing

Inputan data tanaman yang akan diberikan oleh sistem kepada pengguna didapatkan melalui proses pencocokan *string* antara teks dan pola. *Knuth Morris Pratt* memiliki dua tahap pada proses pencariannya, yaitu tahap *pre-KMP* yang merupakan tahap pemberian nilai pergeseran pada masing-masing karakter yang dimasukkan (*pattern*) oleh *user*. Tahap kedua yaitu *KMP Search* yang merupakan tahap pencocokkan karakter yang dimasukkan (*pattern*) oleh *user* dengan karakter yang ada pada database (*text*) berdasarkan nilai pergeseran. *Flowchart* dari pencarian menggunakan algoritma *Knuth Morris Pratt* dapat dilihat pada Gambar 2.



Gambar 2. Flowchart Algoritma *Knuth Morris Pratt*

Implementasi algoritma *Knuth-Morris-Pratt* dalam proses pencarian kamus digital tanaman obat dapat dilihat pada gambar 3.



Gambar 3. Proses Pencarian menggunakan Algoritma *Knuth Morris Pratt*

3. Hasil dan Pembahasan

Aplikasi kamus digital tanaman obat memuat kumpulan nama tanaman obat. Algoritma *knuth morris pratt* ini membagi proses pencarian dengan dua tahap yaitu proses mencari nilai pergeseran dari masing-masing karakter pada *keyword* yang dimasukkan *user*, dan proses pencocokkan karakter *keyword* yang telah memiliki nilai pergeseran pada teks yang tersedia dalam *database*.

Pembahasan 1

Inputan user : obat sakit asma

Text : Radang hati (hepatitis) akut dan kronis, pembesaran hati dan limpa (hepato splenomegali), pembesaran kelenjar limfe (limfadenopati), termasuk pembesaran kelenjar limfe pada tuberkulosis (TBC) kulit (skrofuloderma), gondongan (parotitis), sesak napas (asma bronkial), cacingan, nyeri lambung, sakit perut, kanker, kanker hati, bisul dan cacingan

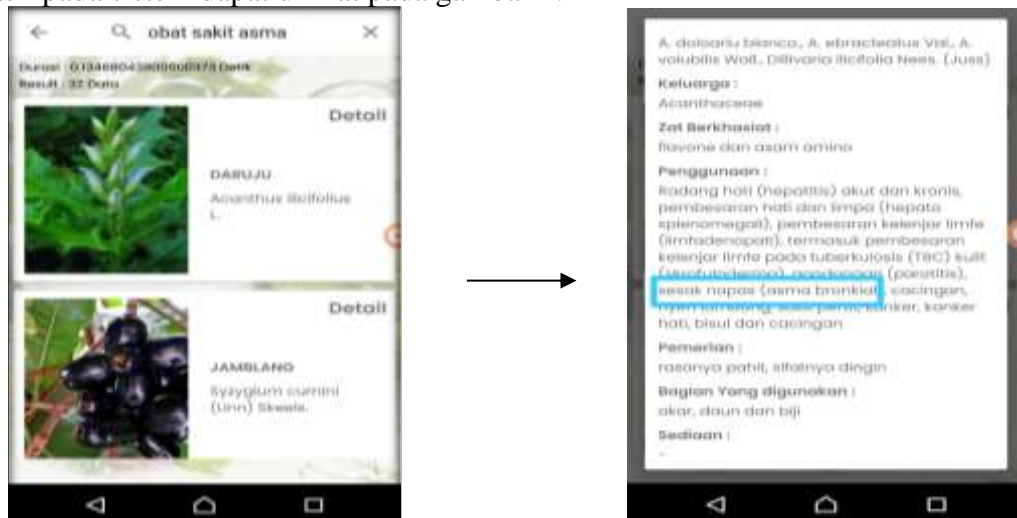
Preprocessing

: radang hati hepatitis akut kronis besar hati limpa hepato splenomegali besar kelenjar limfe limfadenopati besar kelenjar limfe tuberkulosis tbc kulit skrofuloderma gondong parotitis sesak napas asma bronkial cacing nyeri lambung perut kanker kanker hati bisul cacing.

Knuth morris pratt : asma

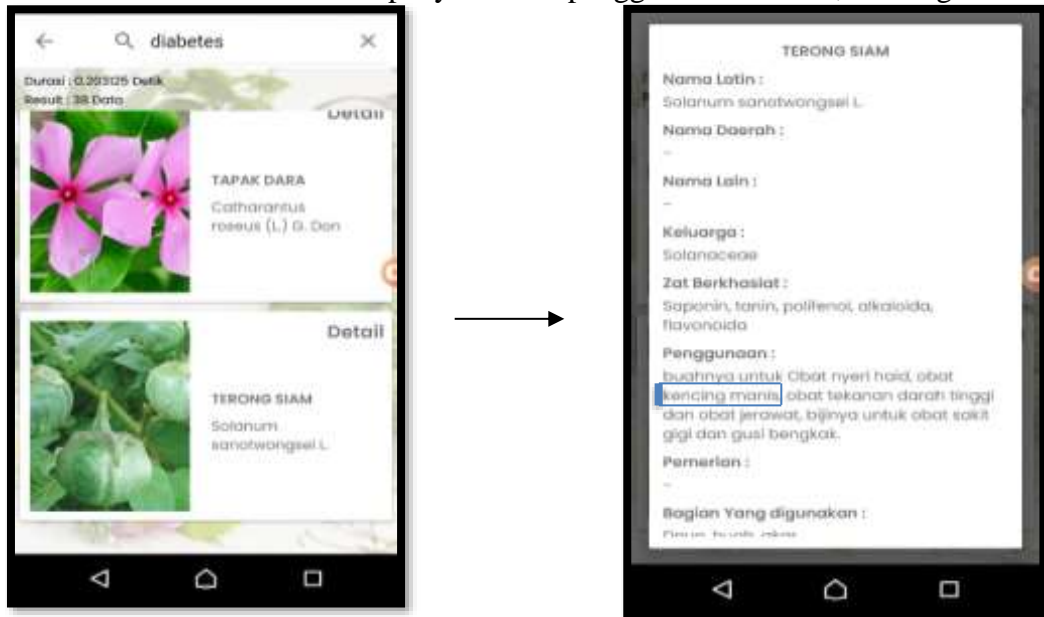
Database persamaan : asma = ampek, bengkek, manggah, sesak napas.

Hasil pada sistem dapat dilihat pada gambar 4.



Gambar 4. Pembahasan 1 hasil *knuth morris pratt*

keywords : diabetes
search : diabetes, kencing manis.
Result : Tanaman untuk penyakit atau penggunaan diabetes, kencing manis.



Gambar 5 . Pembahasan 1 hasil *knuth morris pratt* dengan data persamaan

Pembahasan 2

Data pada database terdiri dari photo tanaman, nama tanaman, nama_latin, nama_daerah, nama_lain, keluarga, zat_berkhasiat, penggunaan, pemerian, bagian yang digunakan, sediaan, waktu panen, penyimpanan dan data persamaan kemiripan penyakit. *Knuth Moris pratt* mencocokkan kata berdasarkan nama penyakit. Selain dari kata penyakit tidak dapat ditampilkan. Berikut pembahasan 2 dapat dilihat pada gambar 6.



Gambar 6 . Pembahasan 2

4. Validasi Kinerja Sistem

Penerapan algoritma *knuth morris pratt* pada sistem pencarian ini diperlukan suatu tolak ukur untuk mengukur kualitas hasil dokumen yang ditemukan kembali. Dari proses ini dapat dihasilkan 30 kata pencarian penggunaan yang berbeda – beda untuk dihitung nilai *precision* dan *recall* nya berdasarkan hasil pencarian. Hasil perhitungan ini menunjukkan evaluasi seberapa baik penerapan metode pada sistem dalam mencocokkan. Berikut hasil perhitungan *precision* dan *recall* dapat dilihat pada tabel 1.

Tabel 1. Hasil perhitungan *precision* dan *recall*

No	K	A	B	C	D	P	R	AC
1	sakit perut	76	45	31	45	59	100	100
2	diare	79	79	0	79	100	100	100
3	peluruh haid	117	105	12	105	89	100	100
4	panas	150	148	2	148	98	100	100
5	asi	51	51	0	51	100	100	100
6	diabetes	38	38	0	38	100	100	100
7	penurun panas	157	139	18	139	88	100	100
8	asma	32	32	0	32	100	100	100
9	pusing	12	12	0	12	100	100	100
10	kencing manis	95	65	30	65	68	100	100
11	sakit kepala	37	31	6	31	83	100	100
12	kanker	16	16	0	16	100	100	100
13	jantung	10	10	0	10	100	100	100
14	haus	2	2	0	2	100	100	100
15	sariawan	46	46	0	46	100	100	100
16	pegal pegal	13	13	0	13	100	100	100
17	reumatik	22	22	0	22	100	100	100
18	Luka bakar	77	77	0	77	100	100	100
19	mata minus	31	31	0	31	100	100	100
20	cacing kremi	47	47	0	47	100	100	100
21	obat sakit gigi	44	25	19	25	56	100	100
22	obat jerawat	6	6	0	6	100	100	100
23	Obat kurap	4	4	0	4	100	100	100
24	menambah nafsu makan	47	47	0	47	100	100	100
25	masuk angin	12	10	2	10	83	100	100
26	panu	5	5	0	5	100	100	100
27	mimisan	12	12	0	12	100	100	100
28	cacingan	47	47	0	47	100	100	100
29	memar	18	18	0	18	100	100	100
30	meredakan jantung berdebar	20	10	10	10	50	100	100

Rata - Rata	92.4	100	100
-------------	------	-----	-----

Keterangan :

- K : Keyword
- A : Hasil keseluruhan dokumen yang di dapatkan pada aplikasi
- B : Hit (Hasil relevan pada aplikasi)
- C : Miss (Hasil yang tidak relevan pada aplikasi)
- D : Hit Database (Hasil yang relevan pada database)
- P : Precision dalam %
- R : Recall dalam %
- AC : Acuraccy %

Hasil dari pengujian yang didapat dari 30 data yang diujikan dengan rata-rata penggunaan metode *knuth morris pratt* precision 100% recall 92.7 % acuraccy 100%. Berdasarkan hasil evaluasi kinerja sistem bahwa algoritma *knuth morris pratt* dalam pencarian pencocokan kata baik.

5. Validasi Kesalahan Eja

Uji coba validasi ini dilakukan dengan koreksi ejaan terhadap kata kunci yaitu kesalahan pengetikan penambahan/penyisipan, penghapusan, penukaran/ tidak ada perubahan karakter dengan mencari nilai jarak edit *levenshtein distance*. Berikut perhitungan jarak edit *levenshtein distance*.

Kata kunci = n

Inputan user = m

- a. Periksa setiap karakter n terhadap m untuk mencari nilai cost

Jika karakter n [i] = m[j] maka cost = 0.

Jika karakter n [i] ≠ m[j] maka cost = 1.

Hitung nilai min :

$$d[i,j] = \min((d[i-1,j]+1),(d[1,j-1]+1),(d[i-1,j-1]+cost)) \quad (1)$$

Hitung nilai max :

$$m = 4$$

$$\text{jarak_max} = \text{round}\left(\frac{40}{100}\right) \times 4 = 2 \quad (2)$$

Ukuran matriks

$$= (n+1) \times (m+1)$$

6. Kesimpulan

Kesimpulan dari penerapan algoritma *Knuth morris pratt* pada pencarian kamus digital tanaman obat dapat memudahkan mencari informasi yang pasti terhadap pencarian data tanaman obat dengan inputan data tanaman yang diberikan oleh sistem kepada pengguna didapatkan melalui proses pencocokan *string* antara teks dan pola dari pengecekan karakter yang dimasukan oleh user dengan karakter yang ada pada database. Sehingga pencarian banyak menghasilkan *match*.

Aplikasi dibuat menggunakan android yang memungkinkan pengguna bisa *install* di dalam perangkat *mobile* yang bisa digunakan secara *realtime* dengan menginputkan kata kunci sesuai dengan kebutuhan pengguna.

Referensi

- [1] Astuti, Wistiani. 2017. Analisis String Matching Pada Judul Skripsi Dengan Algoritma Knuth Morris Pratt. ILKOM. Jurnal Ilmiah Volume 9 nomor 2, Agustus 2017, I ISSN online 2548-7779.
- [2] Bonafacius Vicky Indriyono., Ema Utami., Andi Sunyoto. 2015. Pemanfaatan Algoritma Porter Stemmer Untuk Bahasa Indonesia Dalam Proses Klasifikasi Jenis Buku 301. Vol 6, No. 4. Program Pascasarjana Magister Teknik Informatika STMIK AMIKOM Yogyakarta.
- [3] Ekaputri Gahayu Handari, dan Yulie Anneria Sinaga, 2006. Aplikasi Algoritma Pencarian String *Knuth-Morris-Pratt* dalam Permainan Word Search. Departemen Teknik Informatika, Institut Teknologi Bandung, Bandung.
- [4] Harfatiani, Rina Rizky. 2007. Teknik Riset Operasi. Surabaya: Kartika.
- [5] Haris Triono Sigit. 2015. Aplikasi Android Kamus Bahasa Jawa Serang – Indonesia Menggunakan Algoritma *Knuth Morris Pratt*. Jurnal Vol. II, No. 2 September 2015: 30-34.
- [6] Harjawinata, M.B., Hardhienata, S., & Qur'ania, A. 2015. Aplikasi Pencocokan Jenis Tanaman Obat Berdasarkan Penyakit Berbasis WEB. Bogor: Universitas Pakuan.
- [7] Herny, F. 2013. Perancangan Pengindeks Kata pada Dokumen Teks Menggunakan Aplikasi Berbasis Web. Jurnal Teknologi Informasi DINAMIK. Vol.18 No. 2.
- [8] Ikhbal Muhammad Nazar. 2019. Aplikasi Farmakognosi Tumbuhan Menggunakan Metode *Rocchio*. Program Studi Ilmu Komputer, FMIPA, Universitas Pakuan Bogor.
- [9] JetBrains. “PyCharm: the Python IDE for Professional Developers by JetBrains,” : <https://www.jetbrains.com/pycharm/>. Maret. 30, 2019. [Maret, 2019]).
- [10] Jogyanto, 2010. Analisis dan Desain Sistem Informasi, Edisi IV, Andi Offset, Yogyakarta.
- [11] Kadir & Abdul . 2005. *Dasar Pemrograman Python*. Ed. Ke-1. Andi Offset, Yogyakarta
- [12] Pratiwi., Rahutomo F., Asmara R. 2017. Study Of Hoax News Detection Using Naïve Bayes Classifier In Indonesian Language. IEEE, 2017. 73 – 78.
- [13] Manikandan, P., & Ramyachitra, D. 2018. PATSIM: Prediction and analysis of Protein sequences using hybrid Knuth- Morris Pratt (KMP) and Boyer-Moore (BM) algorithm. Gene, 657, 50-59.
- [14] Nurhadi, Achmad. 2016. Implementasi Algoritma Naive Bayes Classifier Berbasis Particle Swarm Optimization (PSO) untuk Klasifikasi Konten Berita Digital Bahasa Indonesia. Journal-Speed- Sentra Penelitian Engginering dan edukasi. Volume 8 No. 3.
- [15] Sumarni., Sutardi., Rizal adi saputra. 2017. Aplikasi Kamus Tanaman Obat Menggunakan Algoritma Boyer Moore Berbasis Android. SemanTIK, Vo. 3, No. 2, Jul-Des 2017, pp. 245-250. ISSN : 2502-8928 (Online).

- [16] Sunarto, Y. K. 2018. Studi Perbandingan Algoritma Naive Method, Knuth-Morris-Pratt dan BoyerMoore-Hoorspool pada Multi Record Database (Doctoral dissertation, Program Studi Teknik Informatika FTI-UKSW).
- [17] Triawati, Chandra. 2009. Metode Pembobotan Statistical Concept Based untuk Klastering dan Kategorisasi Dokumen Berbahasa Indonesia, Institut Teknologi Telkom Bandung.