

ANALISIS PERBANDINGAN METODE VECTOR SPACE MODEL DAN LEVENSHTTEIN DISTANCE DALAM SISTEM TEMU KEMBALI INFORMASI PADA PERPUSTAKAAN DIGITAL STMIK PRIMAKARA (PRIMAKARA LIBRARY)

I Nyoman Wiyana¹⁾ I Nyoman Purnama²⁾ Ida Bagus Kresna Sudiatmika³⁾

Program Studi Teknik Informatika¹⁾

Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Primakara, Denpasar, Bali^{1) 2) 3)}
inyomanwiyana10@gmail.com¹⁾ purnama@primakara.ac.id²⁾ kresna@primakara.ac.id³⁾

ABSTRACT

Digital Library is an application of Information technology that is used for storage, retrieving, and distributing information of all kinds of knowledge in digital form. STMIK Primakara currently has a Digital Library under the name PRILI (Primakara Library). The number of books in PRILI is still relatively small, but as time goes by, it will definitely continue to grow. The more books there are, search engines will definitely need to make it easier to find the desired book, but the search engine is only able to find the title. To be able to do a search according to user needs, the author will analyze the search engine using the Information Retrieval System. In the Information Retrieval System, a specific method is needed to be applied to the system. The author will analyze two methods, namely the Vector Space Model (VSM) and Levenshtein Distance. The two methods will compare the search results, in the form of relevance values. This value will determine which method the results are most in accordance with the wishes of the user. The test uses data of 298 book titles. The data will then be tested using 10 keywords, where each keyword uses 1-4 terms which are totaled into 40 keywords. In the values of Precision, Recall, accuracy, and F-measure in each term test in the Vector Space Model method, where the fewer terms used, the greater the relevance value, while the time required is faster, while in the Levenshtein Distance method where The more terms used, the greater the relevance value, while for the longer time it takes. The average obtained if all terms are calculated, in the Vector Space Model method with an average relevance of Precision 58.1%, Recall 91.1%, accuracy 94.5%, F-measure 63% and processing time needed 1.43 seconds; while the Levenshtein Distance method has an average relevance of 9.6% Precision, 24.4% Recall, 75.4% accuracy, F-measure 11.1% and the required processing time is 0.87 seconds. The relevance of the Vector Space Model method is superior to the Levenshtein Distance method, but the average time required for the Vector Space Model method is longer than the Levenshtein Distance method.

Keywords: *Information Retrieval, Vector Space Model, Levenshtein Distance, Library.*

ABSTRAK

Perpustakaan Digital adalah suatu penerapan teknologi Informasi yang digunakan untuk sarana penyimpanan, mengambil, dan mendistribusikan informasi segala jenis ilmu pengetahuan dalam bentuk digital. STMIK Primakara saat ini sudah memiliki Perpustakaan Digital dengan nama PRILI (Primakara Library). Buku yang ada pada PRILI masih terbilang sedikit namun seiring berjalannya waktu pasti akan terus bertambah. Semakin banyaknya buku pasti dibutuhkan mesin pencari untuk mempermudah dalam pencarian buku yang diinginkan namun mesin pencari hanya mampu mencari judulnya saja. Untuk dapat melakukan pencarian sesuai kebutuhan user, penulis akan menganalisis mesin pencari menggunakan Sistem Temu Kembali Informasi (Information Retrieval). Dalam Sistem Temu Kembali Informasi diperlukan sebuah metode untuk diterapkan pada sistem, Penulis akan melakukan analisis pada dua metode yaitu Vektor Space Model (VSM) dan Levenshtein Distance. Dua metode tersebut akan dibandingkan hasil pencariannya, berupa nilai relevansinya. Nilai tersebut akan menentukan metode mana yang hasilnya paling sesuai dengan keinginan user. Pengujian menggunakan data sejumlah 298 judul buku. Data tersebut selanjutnya akan dilakukan pengujian dengan menggunakan 10 kata kunci, dimana tiap kata kuncinya menggunakan 1 – 4 term yang ditotalkan menjadi 40 kata kunci. Pada nilai dari Precision, Recall, accuracy, dan F-measure pada tiap pengujian term-nya dalam metode Vector Space Model, dimana makin sedikit term yang

digunakan maka nilai relevansinya makin besar, sementara waktu yang dibutuhkan lebih cepat, sedangkan pada Metode Levenshtein Distance dimana makin banyak term yang digunakan maka nilai relevansinya makin besar, sementara untuk waktu yang dibutuhkan lebih lama. Rata-rata yang didapat jika semua term dihitung, pada metode Vector Space Model dengan relevansi rata-rata pada Precision 58.1%, Recall 91.1%, accuracy 94.5%, F-measure 63% dan waktu proses yang dibutuhkan 1,43 detik; sedangkan metode Levenshtein Distance relevansi rata-rata pada Precision 9.6%, Recall 24.4%, accuracy 75.4%, F-measure 11.1% dan waktu proses yang dibutuhkan 0,87 detik. Relevansi metode Vector Space Model lebih unggul dibandingkan dengan metode Levenshtein Distance, namun rata-rata waktu yang dibutuhkan pada metode Vector Space Model lebih lama dari pada Levenshtein Distance.

Kata Kunci: Sistem Temu Kembali Informasi, Vektor Space Model, *Levenshtein Distance*, Perpustakaan.

PENDAHULUAN

Perpustakaan Digital adalah suatu penerapan teknologi Informasi yang digunakan untuk sarana penyimpanan, mengambil, dan mendistribusikan informasi segala jenis ilmu pengetahuan dalam bentuk digital. Perpustakaan digital bukan merupakan jenis perpustakaan yang berdiri sendiri namun lebih pengembangan dalam sistem layanan perpustakaan. Seperti perpustakaan di perguruan tinggi yang menyediakan sumber informasi atau koleksi bahan pustaka bagi mahasiswa, lalu bentuk informasinya diubah menjadi bentuk digital[1].

STMIK Primakara saat ini sudah memiliki Perpustakaan Digital dengan nama PRILI (Primakara Library)[2]. Buku yang ada pada PRILI masih terbilang sedikit namun seiring berjalannya waktu pasti akan terus bertambah. Semakin banyaknya buku pasti dibutuhkan mesin pencari untuk mempermudah dalam pencarian buku yang diinginkan namun mesin pencari hanya mampu mencari judulnya saja yang sesuai berdasarkan kata kunci atau query yang dimasukan. Misalnya, query yang dimasukan adalah “panduan aplikasi android” maka yang ditampilkan semua dokumen yang mengandung kata kunci tersebut. Dokumen yang ditampilkan bisa tidak relevan atau tidak penting. Bahkan dokumen yang tidak penting ini bisa saja muncul paling atas sedangkan yang relevan bisa ada paling dibawah, maka user harus menyaringnya sendiri agar bisa sesuai dengan kebutuhan informasi yang diinginkan.[3]

Pada kasus mesin pencari di Primakara Library penulis melakukan pengujian dengan memasukan beberapa query.

Hasilnya, Jika memasukan query lebih dari satu kata misalnya “panduan aplikasi android” maka hasilnya tidak ditemukan, hal ini karena kata dalam query tersebut tidak dipisahkan (tokenisasi) dan dokumen atau buku yang dicari harus memiliki judul yang memiliki penyisipan kata dari query yang dimasukan. Namun, Jika menggunakan satu kata saja hasilnya bisa ditemukan akan tetapi jika suatu saat dokumen atau buku yang dimiliki akan terus bertambah dan hasil yang dicari menjadi banyak serta banyak yang tidak relevan maka user harus menyaringnya sendiri sesuai kebutuhan informasi yang diinginkan.

Perbedaan antara temu kembali informasi dan pencarian data adalah, sistem temu kembali informasi terutama berhubungan dengan pencarian informasi yang isinya tidak memiliki struktur, yang kemudian ditemukan hasil suatu informasi yang dibutuhkan atau relevan dengan kata kunci. Sifat pencarian sistem temu-kembali informasi berbeda dengan sistem temu-kembali data (misalnya dalam sistem manajemen basis data) dalam beberapa segi, antara lain spesifikasi query yang tidak lengkap, dan tingkat ketanggapan kesalahan yang tidak peka[4].

Sistem pencarian langsung dari database tujuan utamanya untuk menentukan dokumen yang memiliki kata kunci yang sesuai dengan query yang diberikan oleh pengguna di dalam sekumpulan dokumen, sehingga saat melakukan pencarian data jika data tidak sama persis maka informasi yang relevan dengan kata kunci tidak berhasil ditemukan. Hal ini belum dapat memecahkan masalah user akan kebutuhan informasi[4].

Untuk dapat melakukan pencarian sesuai kebutuhan user, penulis akan menganalisis mesin pencari menggunakan Sistem Temu Kembali Informasi (Information Retrieval). Sistem Temu Kembali Informasi menggunakan bertujuan untuk mengambil informasi serta menampilkan dokumen dari koleksi dokumen. User akan memasukan (input) berupa kata kunci (query) dari informasi yang akan dicari. Daftar dokumen yang ditampilkan akan relevan dan sesuai dengan kebutuhan informasi dari pengguna[5].

Dalam Temu Kembali Informasi diperlukan sebuah metode untuk diterapkan pada sistem, salah satunya adalah Vektor Space Model (VSM). VSM menggunakan vektor untuk merepresentasikan dokumen dan query. Vektornya dibentuk dari seluruh term atau kata-kata yang menyusun dokumen. Pada vektor dokumen dan query, tiap elemen merepresentasikan bobot kata-kata yang berhubungan dalam dokumen dan query-nya[6]. Pada VSM menggunakan rumus untuk mencari nilai cosinus sudut antara dua vector pada tiap bobot dokumen dan juga kata kunci. Rumus untuk menghitung kedekatan nilai antar dua dokumen atau kecocokan ini dikenal dengan metode Cosine Similarity.[7]

Selain VSM, metode lain dalam Temu Kembali Informasi dapat menggunakan Metode Levenshtein Distance, dimana metode ini digunakan untuk menentukan jarak levenshtein berdasarkan nilai yang paling terkecil atau paling sedikit jumlah modifikasinya. Oleh karena itu, string atau kata yang nilainya paling sedikit berubah saat dibandingkan dengan string atau kata lain dianggap paling cocok atau paling mendekati[8].

Berdasarkan pemaparan masalah diatas, maka penulis ingin melakukan analisa perbandingan Metode Vector Space Model Dan Levenshtein Distance dalam menentukan hasil akurasi sistem temu kembali informasi pada Perpustakaan Digital Stmik Primakara.

METODOLOGI PENELITIAN

Penulis akan melakukan analisa perbandingan metode dalam sistem temu kembali informasi yaitu Vektor Space Model dan metode Levenshtein Distance. Analisa perbandingan metode ini dilakukan untuk

mengetahui perbandingan hasil akurasi dan relevansi dari metode Vektor Space Model dan metode Levenshtein Distance pada saat melakukan proses pengambilan informasi sesuai kebutuhan informasi yang diinginkan oleh user.

Jenis data yang penulis gunakan dalam penelitian adalah data kuantitatif dan untuk pengujian keakuratan hasil pencarian akan dievaluasi nilai dari precision, recall dan accuracy. Precision mengevaluasi kemampuan sistem untuk menemukan peringkat yang paling relevan, dan didefinisikan sebagai presentase dokumen yang di-retrieve dan benar-benar relevan terhadap query. Recall mengevaluasi kemampuan sistem untuk menemukan semua item yang relevan dari koleksi dokumen dan didefinisikan sebagai presentase dokumen yang relevan terhadap query. Nilai precision, recall, dan accuracy dapat dihitung dengan menggunakan confusion matrix [31].

Tabel 1
Confusion Matrix

	<i>Relevant</i>	<i>Non relevant</i>
<i>Retrieved</i>	<i>True positive (tp)</i>	<i>False positive (fp)</i>
<i>Not retrieved</i>	<i>False negative (fn)</i>	<i>True negative (tn)</i>

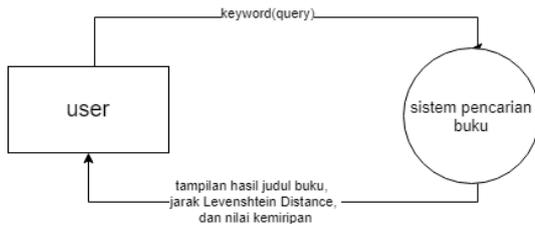
Precision = $(tp) / ((tp + fp))$ (III.1)

Recall = $(tp) / ((tp + fn))$ (III.2)

Accuracy = $((tp + tn) / (tp + fp + tn + fn))$ (III.3)

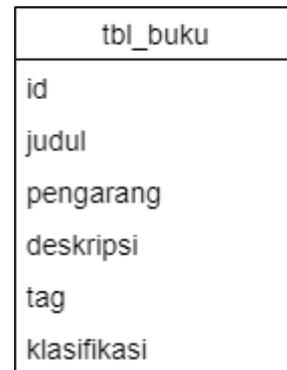
Selain penilaian precision, recall dan accuracy yang dilakukan. pada penelitian ini juga dilakukan penilaian rata-rata F-Measure. F-Measure adalah harmonic mean atau nilai rata-rata dari precision dan recall [23] . Rumus F-Measure adalah sebagai berikut :

$$F - Measure = 2 * \frac{Precision * Recall}{Precision + Recall}$$



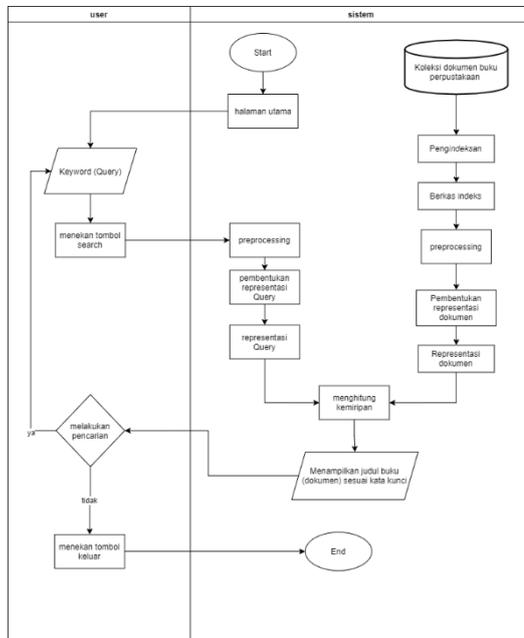
Gambar 1 Diagram Konteks

Pada gambar diagram diatas dapat dilihat proses secara umum yang terjadi didalam sistem, proses ini dimulai dari user memasukan kata kunci kemudian output yang didapatkan oleh pengguna berupa tampilan judul buku, jarak Levenshtein Distance dan nilai similarity.



Gambar 3 Desain Logikal

Tabel 2
Desain Fisikal Database

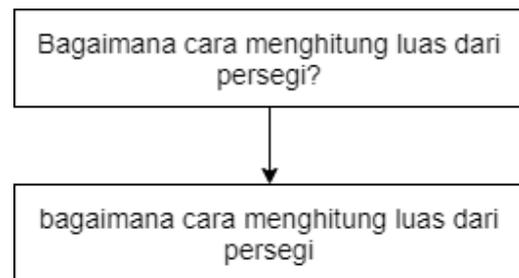


Gambar 2 Diagram Aktifitas

Pada gambar diatas menggambarkan aktivitas yang dilakukan oleh user dan sistem yang dilakukan pengguna adalah memasukan keyword kemudian menekan tombol search, selanjutnya pada sistem melakukan proses pencarian dan menampilkan hasil yang dicari sesuai masukan user.

HASIL DAN PEMBAHASAN

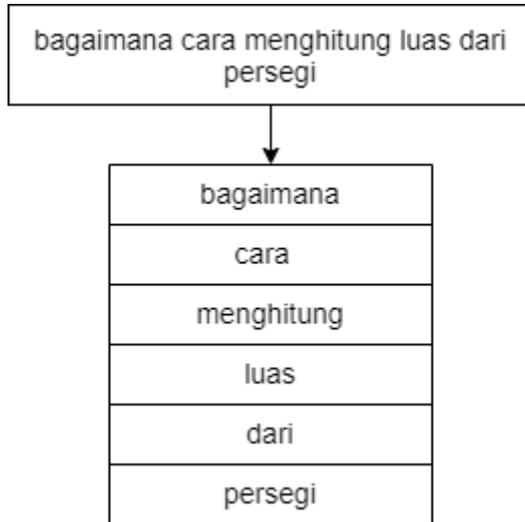
Pengujian ini dilakukan untuk memeriksa apakah sistem dapat melakukan pencarian menggunakan metode Vector Space Model dan Levenshtein Distance dengan kata kunci yang dimasukan oleh user. Selanjutnya sistem akan menghitung nilai kedekatan antara kata kunci dan semua kata yang ada pada database. Kata yang memiliki nilai kedekatan lebih besar akan ditampilkan sebagai hasil dari pencarian. Berikut proses Preprocessing dan pengujian dari metode Vector Space Model dan Levenshtein Distance:



Gambar 4 Case Folding

No.	Nama Field	Type	Size	Keterangan
1	id*	Int	5	Id tabel buku
2	Judul	Text		Judul buku
3	deskripsi	Text		Deskripsi buku
4	Tag	Varchar	100	Tag buku
5	Klasifikasi	Varchar	100	Klasifikasi buku

Case Folding yaitu mengubah semua karakter dalam teks menjadi huruf kecil dan menghapus karakter selain huruf [12].



Gambar 5 Tokenisasi

Tokenisasi adalah tahap pemrosesan pemisahan kata yang ada pada kalimat [13]. *Filtering* adalah proses menghilangkan kata-kata yang dianggap sebagai kata yang tidak penting dan tidak mempengaruhi makna kata tersebut [14]. *Stemming* merupakan suatu proses mengubah kata-kata yang terdapat dalam suatu dokumen ke kata-kata aslinya atau dasarnya (root word) [15].

Pada pengujian ini sebelum diimplementasikan pada sistem, penulis melakukan pengujian perhitungan manual. Perhitungan ini menggunakan satu Query (Q) dan tiga dokumen (D).

- Q = pengetahuan logistik
- D1 = manajemen transaksi logistik
- D2 = pengetahuan antar individu
- D3 = dalam manajemen pengetahuan terdapat transfer pengetahuan logistik

Token	tf				df	D/df	idf= log(D/df)
	Q	D1	D2	D3			
pengetahuan	1	0	1	1	2	1.5	0.17609125
logistik	1	1	0	1	2	1.5	0.17609125
manajemen	0	1	0	1	2	1.5	0.17609125
transaksi	0	1	0	0	1	3	0.47712125
individu	0	0	1	0	1	3	0.47712125
transfer	0	0	0	1	1	3	0.47712125

Gambar 6 Hasil Hitung Metode Vector Space Model (VSM) Bagian ke-1

Q	W			Q*2	D1*2	D2*2	D3*2	Q*D1	Q*D2	Q*D3	
	D1	D2	D3								
0.176091	0	0.176091	0.176091	0.031008	0	0.031008	0.0310081	0	0.000962	0.000962	
0.176091	0.176091	0	0.176091	0.031008	0.031008	0	0.0310081	0.000962	0	0.000962	
0	0.176091	0	0.176091	0	0.031008	0	0.0310081	0	0	0	
0	0.477121	0	0	0	0.227645	0	0	0	0	0	
0	0	0.477121	0	0	0	0.227645	0	0	0	0	
0	0	0	0.477121	0	0	0	0.2276447	0	0	0	
				Q	0.062016						
				sqrt(Q)	0.249031						
						sqrt(D1)				SUM(Q*D1)	
					0.538202	0.508579	0.5662765	0.000962	0.000962	0.001923	
								Rank Score	0.007174	0.007592	0.015856

Gambar 7 Hasil Hitung Metode Vector Space Model (VSM) Bagian ke-2

Dari pengujian perhitungan ini, hasil yang didapatkan adalah berupa rank score, dimana pada dokumen ke-3 (D3) mendapatkan nilai tertinggi alias mendekati dengan query dengan nilai 0.013636, dilanjutkan dokumen ke-2 (D2) dengan nilai 0.007592 dan terakhir dokumen ke-1 (D1) dengan nilai 0.007174.

Hasil Perhitungan Metode Levenshtein Distance. Pada pengujian ini sebelum diimplementasikan pada sistem, penulis melakukan pengujian perhitungan manual. Perhitungan ini melakukan membandingkan dua string.

Perhitungan 1

String 1 = barang

String 2 = barang

S1/S2		b	a	r	a	n	g	
		0	1	2	3	4	5	6
b		1	0	0	0	0	0	0
a		2	1	0	0	0	0	0
r		3	2	1	0	0	0	0
a		4	3	2	1	0	0	0
n		5	4	3	2	1	0	0
g		6	5	4	3	2	1	0

Gambar 8 Hasil Perhitungan ke-1 pada Metode Levenshtein Distance

Setelah mendapatkan jarak edit distance atau jarak Levenshtein Distance maka dilakukan perhitungan similarity :

Jarak = 0

String terpanjang = 6

Maka $\left(1 - \frac{0}{6}\right) * 100 = 100$

Perhitungan 2

String 1 = barang

String 2 = burung

S1/S2	b	u	r	u	n	g
	0	1	2	3	4	5
b	1	0	0	0	0	0
a	2	1	1	1	1	1
r	3	2	2	1	1	1
a	4	3	3	2	2	2
n	5	4	4	3	3	2
g	6	5	5	4	4	3

Gambar 9 Hasil Perhitungan ke-2 pada Metode Levenshtein Distance

Setelah mendapatkan jarak edit distance atau jarak Levenshtein Distance maka dilakukan perhitungan similarity :

Jarak = 2

String terpanjang = 6

Maka $(1 - \frac{2}{6}) * 100 = 66.7$

Pengujian Relevansi Hasil Pencarian Judul dengan Confusion Matrix. Confusion Matrix digunakan untuk menghitung Nilai precision, recall, dan accuracy. Berikut hasil Confusion Matrix menggunakan kata kunci “ilmu komputer”.

Tabel 3 Confusion Matrix

	Relevant	Non relevant
Retrieved	26	23
Not retrieved	1	248

	precision	recall	accuracy	F-measure	time
4 term	30.7%	86.2%	90.5%	39.0%	00:00:01.50
3 term	36.7%	89.3%	92.1%	45.9%	00:00:01.40
2 term	64.9%	90.5%	95.6%	71.3%	00:00:01.42
1 term	100.0%	98.3%	99.7%	99.1%	00:00:01.40

$$Precision = \frac{26}{(26 + 23)} = 0.5306122$$

$$Recall = \frac{26}{(26+1)} = 0.962963$$

$$Accuracy = \frac{(26 + 248)}{26 + 23 + 248 + 1} = 0.919463$$

$$F - Measure = 2 * \frac{0.5306122 * 0.962963}{0.5306122 + 0.962963} = 0.684211$$

Tabel Error! No text of specified style in document. Confusion Matrix Levenshtein Distance

	Relevant	Non relevant
Retrieved	5	28
Not retrieved	22	243

$$Precision = \frac{5}{(5 + 28)} = 0.151515152$$

$$Recall = \frac{5}{(5+22)} = 0.185185185$$

$$Accuracy = \frac{(5 + 243)}{5 + 28 + 243 + 22} = 0.832214765$$

$$F - Measure = 2 * \frac{0.151515152 * 0.185185185}{0.151515152 + 0.185185185} = 0.166666667$$

Data dari penelitian ini diambil dari website Primakara Library (Prili) dengan cara web scraping. Data yang berhasil didapatkan sejumlah 298 judul buku. Data tersebut selanjutnya akan dilakukan pengujian dengan menggunakan 20 kata kunci dan 5 kata kunci salah eja (typo). Hasil dari pencarian tersebut akan dipakai untuk pengujian Precision, Recall, accuracy, F-measure, dan waktu pencarian pada sistem yang telah dibangun.

Pengujian ini dilakukan untuk mengetahui relevansi antara kata kunci dengan data buku. Berikut ini hasil dari pengujian Precision, Recall, accuracy, F-measure, dan waktu pencarian:

Tabel 5 Rata-rata hasil pengujian tiap term Metode Vector Space Model

Dari pengujian tersebut Metode Vector Space Model memiliki relevansi rata-rata tiap term

berbeda-beda, dimana makin sedikit term yang digunakan maka nilai relevansinya makin besar, hasil ini didapat dikarenakan makin spesifik query tersebut maka dokumen yang ditampilkan akan sesuai dengan kata pada query tersebut. Sementara waktu yang dibutuhkan jika term yang digunakan lebih sedikit maka prosesnya akan lebih cepat, hasil ini didapat karena makin sedikit term yang didapat dari query dan juga dari dokumen maka prosesnya akan membutuhkan waktu yang lebih cepat. Untuk rata-rata jika dihitung pada semua term hasilnya adalah Precision 58.1%, Recall 91.1%, accuracy 94.5%, F-measure 63% dan waktu proses yang dibutuhkan 1,43 detik.

	precision	recall	accuracy	F-measure	time
4 term	9.9%	29.2%	75.7%	12.1%	00:00:01.34
3 term	9.8%	27.8%	76.4%	12.0%	00:00:00.98
2 term	9.7%	26.1%	78.1%	11.8%	00:00:00.74
1 term	8.9%	14.4%	71.1%	8.4%	00:00:00.41

Tabel 1 Rata-rata hasil pengujian tiap term Metode Levenshtein Distance

Dari pengujian tersebut Metode Levenshtein Distance memiliki relevansi rata-rata tiap term berbeda-beda, dimana makin banyak term yang digunakan maka nilai relevansinya makin besar, hasil ini diperoleh dikarenakan makin panjang term atau string tersebut, maka nilai kedekatannya akan makin kecil. Sementara untuk waktu proses pada saat menggunakan banyak term, hasil yang dibutuhkan lebih lama. Hasil ini didapat karena makin banyak string yang digunakan dari query dan juga dari dokumen maka prosesnya akan membutuhkan waktu yang lebih lama. Untuk rata-rata jika dihitung pada semua term hasilnya adalah Precision 9.6%, Recall 24.4%, accuracy 75.4%, F-measure 11.1% dan waktu proses yang dibutuhkan 0,87 detik.

SIMPULAN

Berdasarkan hasil penerapan Information Retrieval dalam proses perbandingan

metode Vector Space Model dan Levenshtein Distance dengan menggunakan data buku pada website Primakara Library (Prili), maka dapat disimpulkan:

- Keluaran akhir dari program ini berupa rekomendasi daftar judul buku yang memiliki kemiripan dengan query yang diinputkan oleh user.
- Waktu dalam proses pencarian sebuah informasi tergantung dari jumlah dokumen yang ditampilkan dan jumlah kata kunci yang digunakan sebagai query pencarian.
- Pada nilai dari Precision, Recall, accuracy, dan F-measure pada tiap term-nya dalam metode Vector Space Model, dimana makin sedikit term yang digunakan maka nilai relevansinya makin besar, sementara waktu yang dibutuhkan lebih cepat, sedangkan pada Metode Levenshtein Distance dimana makin banyak term yang digunakan maka nilai relevansinya makin besar, sementara untuk waktu yang dibutuhkan lebih lama.
- Rata-rata yang didapat jika semua term dihitung, pada metode Vector Space Model dengan relevansi rata-rata pada Precision 58.1%, Recall 91.1%, accuracy 94.5%, F-measure 63% dan waktu proses yang dibutuhkan 1,43 detik; sedangkan metode Levenshtein Distance relevansi rata-rata pada Precision 9.6%, Recall 24.4%, accuracy 75.4%, F-measure 11.1% dan waktu proses yang dibutuhkan 0,87 detik. Relevansi metode Vector Space Model lebih unggul dibandingkan dengan metode Levenshtein Distance, namun rata-rata waktu yang dibutuhkan pada metode Vector Space Model lebih lama dari pada Levenshtein Distance.
- Rata-rata dari nilai precision yang kecil dikarenakan kata kunci dan judul buku pada database memiliki nilai kemiripan yang hampir mendekati ketika dalam pemrosesannya, meski judul buku tersebut tidak sesuai dengan yang diinginkan oleh user.

DAFTAR PUSTAKA

- U. N. Medan, "Perpustakaan Digital sebagai Wujud Penerapan Teknologi Informasi di Perguruan Tinggi," *Pustaka*, vol. 2, no. 1, pp. 32–36, 2006.

- [2] D. YASA, “IMPLEMENTASI PERPUSTAKAAN DIGITAL MENGGUNAKAN FRAMEWORK CODEIGNITER DENGAN METODE PERSONAL EXTREME PROGRAMMING PADA STMIK PRIMAKARA,” Denpasar, 2016.
- [3] A. N. Langville and C. D. Meyer, “Information Retrieval and Web Search,” 2011. doi: 10.1007/978-3-642-19460-3.
- [4] K. D. Putung, A. Lumenta, and A. Jacobus, “PENERAPAN SISTEM TEMU KEMBALI INFORMASI PADA KUMPULAN DOKUMEN SKRIPSI,” *E-journal Tek. Inform.*, vol. 8, no. 1, pp. 18–23, 2016.
- [5] E. Fitriani, R. E. Indrajit, and R. Aryanti, “Penerapan Model Information Retrieval Untuk Pencarian Konten Pada Perpustakaan Digital,” *J. Perspekt.*, vol. 15, no. 2, pp. 170–176, 2017, doi: 10.31294/JP.V15I2.2350.
- [6] I. M. Ayudita, Indriati, and P. P. Adikara, “Sistem Pencarian Jurnal Ilmiah Cross Language dengan Metode Vector Space Model (VSM),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 12, pp. 6837–6841, 2018.
- [7] H. W. A. Kesuma, “Penerapan Metode TF-IDF dan Cosine Similarity dalam Aplikasi Kitab Undang-Undang Hukum Dagang,” 2016.
- [8] W. Widiatry, V. H. Pranatawijaya, N. N. K. Sari, and P. B. A. A. Putra, “Penerapan Algoritma Levenshteindistance untuk Pencarian pada Sistem Informasi Perpustakaan Fakultas Kedokteran Universitas Palangka Raya,” *J. Saintekom*, vol. 9, no. 1, pp. 66–82, 2019.
- [9] *UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 43 TAHUN 2007 TENTANG PERPUSTAKAAN*. Jakarta: Sekretariat Negara, 2007.
- [10] C. D. Manning, P. Raghavan, and H. Schütze, *An Introduction to Information Retrieval*. Cambridge, 2009.
- [11] B. Mitra and N. Craswell, “An introduction to neural information retrieval,” *Found. Trends Inf. Retr.*, vol. 13, no. 1, pp. 1–129, 2018, doi: 10.1561/15000000061.
- [12] A. Setiawan, I. F. Astuti, and A. H. Kridalaksana, “Klasifikasi Dan Pencarian Buku Referensi Akademik Menggunakan Metode Naïve Bayes Classifier (NBC) (Studi Kasus: Perpustakaan Daerah Provinsi Kalimantan Timur),” *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 10, no. 1, p. 1, 2016, doi: 10.30872/jim.v10i1.17.
- [13] F. Amin and Purwatiningtyas, “Rancang Bangun Information Retrieval System (IRS) Bahasa Jawa Ngoko pada Palintangan Penjebar Semangad dengan Metode Vector Space Model (VSM),” *J. Teknol. Inf. Din.*, vol. 20, no. 1, pp. 25–35, 2015.
- [14] T. Hariguna and F. S. Utomo, “Implementation Of Information Retrieval Indonesian Text Documents Using The Vector Space Model,” no. May 2012, 2016.
- [15] L. Agusta, “Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia,” *Konf. Nas. Sist. dan Inform. 2009*, no. KNS&I09-036, pp. 196–201, 2009.
- [16] D. Muliadi, “PERBANDINGAN ALGORITMA STEMMING UNTUK BAHASA INDONESIA DENGAN PARAMETER AKURASI DAN WAKTU PROSES,” pp. 7–37, 2015.
- [17] A. Hendrawan, T. Winarti, and H. Indriyawati, “Pengembangan stemming untuk artikel berbahasa indonesia,” 2020.
- [18] R. Mandala, E. Koryanti, and R. Munir, “Sistem Stemming Otomatis untuk Kata dalam Bahasa Indonesia,” *Seminar*, pp. 29–36, 2004.
- [19] F. Amin, “IMPLEMENTASI SEARCH ENGINE (MESIN PENCARI) MENGGUNAKAN METODE VECTOR SPACE MODEL,” *Din. Tek.*, vol. V, no. 1,

- pp. 45–58, 2011.
- [20] B. P. Pratama and S. A. Pamungkas, “Analisis Kinerja Algoritma Levenshtein Distance dalam Mendeteksi Kemiripan Dokumen Teks,” *J. “LOG!K@,”* vol. 6, no. 2, pp. 131–143, 2016.
- [21] N. H. Ariyani, Sutardi, and R. Ramadhan, “Aplikasi Pendeteksi Kemiripan Isi Teks Dokumen Menggunakan Metode Levenshtein Distance,” *semanTIK*, vol. Vol 2, no. 1, pp. 279–286, 2016, [Online]. Available: <http://ojs.uho.ac.id/index.php/semantik/article/download/1030/661>.
- [22] D. Rosmala and Z. M. Risyad, “Algoritma Levenshtein Distance dalam Aplikasi Pencarian isu di Kota Bandung pada Twitter,” *MIND J.*, vol. 2, no. 2, pp. 1–12, 2018, doi: 10.26760/mindjournal.v2i2.1-12.
- [23] M. A. Kusuma, M. Kamayani, and A. Avorizano, “Pencarian Full Text Pada Koleksi Skripsi Fakultas Teknik UHAMKA Menggunakan Metode Vector Space Model,” *Semin. Nas. Teknoka*, vol. 2, no. 2502, pp. 96–102, 2017.
- [24] D. L. Khuseri Andesa, “Implementasi Vector Space Model Untuk Meningkatkan KUALITAS PADA SISTEM PENCARIAN BUKU PERPUSTAKAAN,” *Semin. Nas. Inform. 2015*, no. May 2013, pp. 8–15, 2015.
- [25] A. Fauzi and G. Ginabila, “Information Retrieval System Pada File Pencarian Dokumen Tesis Berbasis Text Menggunakan Metode Vector