



Penerapan Model Information Retrieval Untuk Pencarian Konten Pada Perpustakaan Digital

Aang Alim Murtopo^a, Didi Haryadi^b, Nurul Fadilah^c

^a Teknik Informatika, aang.alim@gmail.com, STMIK Tegal

^b Teknik Informatika, didihar84@gmail.com, STMIK Tegal

^c Teknik Informatika, nurulfadilah18@gmail.com, STMIK Tegal

ABSTRACT

The growth of the library will not be separated from the progress of science and information technology. In this case the library is closely related to information technology and science. Each helping the other. Libraries have a great responsibility to improve the reading habits of their users. Therefore, this digital library has a very positive influence on users' reading motivation. A digital library is a library that uses a collection of books that are in digital format and can be accessed via computers and smartphones. The digital library has implemented a repository. Repository is a container for storing various documents or information. The use of a repository greatly facilitates digital library users in finding the information they need. In this study, the method used to search for content information uses (information retrieval), in processing the search using the N-Gram Algorithm and Cosine Similarity to measure the similarity between the documents sought. Using the above method can speed up the process of searching for content in digital libraries and to find the relevance between search results and keywords.

Keywords: *digital library, information retrieval, n-gram algorithm, cosine similarity.*

ABSTRAK

Pertumbuhan pada perpustakaan tidak akan lepas oleh kemajuan ilmu pengetahuan dan teknologi informasi. Dalam hal ini perpustakaan sangat berkaitan erat dengan dan teknologi informasi dan ilmu pengetahuan. Masing-masing saling membantu satu sama lain. Perpustakaan memiliki tanggung jawab yang besar untuk meningkatkan kebiasaan membaca para penggunanya. Oleh karena itu, perpustakaan digital ini memiliki pengaruh yang sangat positif terhadap motivasi membaca pengguna. Perpustakaan digital yakni perpustakaan menggunakan koleksi buku yang format digital dan dapat diakses melalui komputer dan *smartphone*. Perpustakaan digital sudah menerapkan repository. *Repository* merupakan suatu wadah untuk menyimpan berbagai dokumen atau informasi. Penggunaan *repository* sangat memudahkan pengguna perpustakaan digital dalam mencari informasi yang dibutuhkan. Pada penelitian ini metode yang digunakan untuk mencari informasi konten menggunakan (*information Retrieval*), dalam pengolahan pencariannya menggunakan Algoritma *N-Gram* serta *Cosine Similarity* untuk pengukuran kemiripan antara dokumen yang dicari. Dengan penggunaan metode diatas dapat mempercepat proses pencarian konten pada perpustakaan digital serta untuk menemukan relevansi antara hasil pencarian dan kata kunci.

Kata Kunci: *perpustakaan digital, information retrieval, algoritma n-gram, cosine similarity.*

1. PENDAHULUAN

Pertumbuhan pada perpustakaan tidak akan lepas oleh kemajuan ilmu pengetahuan dan teknologi informasi. Dalam hal ini perpustakaan sangat berkaitan erat dengan dan teknologi informasi dan ilmu pengetahuan. Masing-masing saling membantu satu sama lain. Perpustakaan ikut andil dalam kemajuan ilmu pengetahuan dengan menyimpan banyak sekali berita/informasi dan sejarah dari kemajuan ilmu pengetahuan, sementara itu teknologi informasi berkontribusi dalam fasilitas akses pada sebuah sistem

perpustakaan digital. Seiring dengan peningkatan ketiganya, saat ini dikenal dengan nama perpustakaan digital (*digital library*)[1].

Perpustakaan memiliki tanggung jawab yang besar untuk meningkatkan kebiasaan membaca para penggunanya. Oleh karena itu, perpustakaan digital ini memiliki pengaruh yang sangat positif terhadap motivasi membaca pengguna. Perpustakaan digital yakni perpustakaan menggunakan koleksi buku yang format digital dan dapat diakses melalui komputer dan *smartphone*.

Perpustakaan digital sudah menerapkan repository. *Repository* merupakan suatu wadah untuk menyimpan berbagai dokumen atau informasi. Penggunaan *repository* sangat memudahkan pengguna perpustakaan digital dalam mencari informasi yang dibutuhkan[2]. Segala macam informasi dan sumber informasi yang dimiliki perpustakaan akan terus meningkat seiring dengan perkembangan informasi. Perpustakaan digital membutuhkan sistem *Information Retrieval*. Algoritma untuk memperoleh informasi menggunakan *N-Gram* dan teknik koreksi ejaan diperlukan untuk mengoreksi konten dalam database[3].

Information retrieval adalah sistem yang secara otomatis mengambil informasi dari kumpulan informasi sesuai dengan kebutuhan pengguna[4]. *Information retrieval* mempunyai keunggulan dalam mendapatkan informasi yang paling akurat sehingga keperluan pengguna dapat tersalurkan secara otomatis. Sistem temu kembali informasi (*information retrieval*) menjadi media layanan buat mendapat informasi yang diperlukan bagi pengguna. Sistem temu kembali informasi (*information retrieval*) bertindak sebagai penghubung antara keperluan informasi pengguna dan sumber daya yang tersedia.

Algoritma *n-gram* merupakan contoh probabilistik dibuat dan dikembangkan oleh matematikawan Rusia pada abad ke-20 algoritma ini disempurnakan untuk memperkirakan elemen berikutnya dari sekumpulan elemen. Elemen tersebut bisa berupa kata, huruf, karakter, atau yang lainnya. Algoritma *n-gram* yang berbasis kata digunakan untuk memperkirakan kata berikutnya dalam urutan kata tertentu. Sebuah *n-gram* dapat diberi nilai secara sembarang berdasarkan preferensi pengguna. Bobot pada *N* akan digunakan untuk menguraikan kalimat menjadi bagian yang lebih kecil sesuai dengan bobot pada *N*. Algoritma *N-gram* adalah urutan potongan(karakter atau huruf) sejauh nilai *n*. Dalam *N-Gram* memiliki beberapa macam. Antara lain *uni-gram*, *bi-gram*, *tri-gram*, *quad-gram*, *5-gram* dan sejumlah nilai *n* dalam *n-gram*[5].

Pada penelitian ini mengimplementasikan algoritma *N-Gram* dipakai untuk memeriksa bahasa dan metode *Cosine Similarity* untuk memperkirakan kesesuaian antara *query* dan dokumen yang diwakili oleh *Vector Space Model (VSM)*. Kumpulan dokumen yang dipakai merupakan kumpulan artikel jurnal atau buku *pdf* yang nantinya akan dicari kemiripan dengan *query* yang dimasukkan dalam proses pencarian *repository* untuk dijadikan data acuan untuk proses pencarian. Penerapan *Cosine Similarity* dan *N-Gram* diinginkan bisa memajukan kemampuan dari pencarian pada perpustakaan digital.

Oleh sebab itu, tujuan ini mengacu untuk pencarian konten di perpustakaan digital dengan menggunakan algoritma *n-gram* dan *cosine similirity* yang berguna untuk memudahkan proses pencarian informasi dan menyampaikan informasi yang terbaik untuk memenuhi kebutuhan pencarian konten pengguna. Pengguna dapat mengakses dan mengunduh konten yang mereka butuhkan.

2. TINJAUAN PUSTAKA

2.1. Perpustakaan Digital

Perpustakaan digital/*digital library* merupakan layanan distributor informasi yang berlandaskan TIK yang menyuplai layanan dan koleksi yang berbentuk digital, dan menggunakan internet untuk mengaksesnya[6].

Pokok dari perpustakaan digital adalah sebuah sistem yang digunakan untuk menyimpan dan menyampaikan informasi/konten berbasis digital kepada pengguna.

2.2. *N-Gram*

N-Gram sendiri sering dipakai untuk menyelesaikan masalah seperti perkiraan kata, perbaikan tulisan, perbaikan kalimat terjemahan dan pencarian kata[7].

Inti mengenai *N-Gram* dapat mempermudah dalam proses pencarian informasi yang terbaik untuk memuaskan keinginan pengguna perpustakaan digital.

2.3. *Information Retrieval*

Information Retrieval ialah metode yang bermaksud untuk mendapatkan kembali informasi yang terdokumentasi pada sumber(*resources*) yang terbaik[8]. Fungsi dari *Information Retrieval* untuk melakukan pencarian terbaik yang dibutuhkan oleh pengguna.

3. METODOLOGI PENELITIAN

Dalam penelitian ini, metode yang digunakan dalam pengumpulan data-data memakai metode studi pustaka menggunakan pencarian berdasarkan kumpulan referensi, data yang diambil berhubungan dengan data perpustakaan, seperti judul buku, nama penulis, penerbit, dan tahun terbit[9]. Pengumpulan data, memeriksa dan mengkaji data dari referensi tersebut dengan tepat didasarkan dari data yang berkaitan pada penelitian ini. Tujuan penelitian ini ialah menerapkan *model information retrieval* untuk pencarian konten pada perpustakaan digital yang dikombinasikan dengan algoritma pengolahan *predictive text N-Gram*.

3.1. Analisa Bahasa

Algoritma *N-gram* bermanfaat untuk menganalisis bahasa berdasarkan perkiraan pola alokasi *N-gram*, dalam sebuah bahasa terdapat hal-hal unik berdasarkan aliran pasangan dan penggunaan huruf. Bahasa yang sering kali digunakan berbeda dengan bahasa lain. Untuk mendapatkan tanda baca berupa huruf, *padding* digunakan untuk membubuhkan tanda “_” pada awalan dan akhiran kosakata. Contoh, kalimat “BUKU” bisa dibagi dengan *N-gram* sebagai berikut (“_” berarti spasi),

Tabel 1. Contoh Pecahan N-gram

N-Gram	Pemecahan Karakter
Uni-Gram	B, U, K, U
Bi-Gram	_B, BU, UU, KU, U_
Tri-Gram	_BU, BUU, KUU, KU_
Quad-Gram	_BUU, BUKU, UUK_
Quin-Gram	_BUKU, BUKU_

Dibawah ini adalah penjelasan dalam pemenggalan kalimat *N-gram* pada gambar diatas, masukan berupa “senang meneliti dan mengembangkan algoritma *N-gram*”.

- 1) *Uni-Gram Char*: ('s','e','n','a','n','g','m','e','n','e','l','i','t','i','N','g','r','a','m')
- 2) *Bi-Gram Char*: ('se','na','ng','me','ne','li','ti','N-g','ram')
- 3) *6-Gram Char*: ('senan','nang','men','neliti','N-gr','gram')
- 4) *Uni-Gram word*: ('senang', 'senang meneliti', 'meneliti dan', 'mengembangkan', 'algoritma N-gram', 'N-gram')

Alur proses pendeteksian bahasa menggunakan N-Gram ditunjukkan pada gambar 1



Gambar 1. Alur deteksi bahasa dengan *N-gram*

Algoritma n-gram digunakan untuk mengekstrak bagian karakter dari sebuah kata atau kalimat, seperti jumlah karakter dalam sebuah kata. Salah satu keuntungan n-gram adalah n-gram kurang rentan terhadap kesalahan ejaan[10]. N-gram memiliki keunikan tersendiri[11], antara lain:

- 1) Dapat bekerja dengan baik meskipun ada kesalahan dalam teks
- 2) Berjalan dengan baik, membutuhkan perawatan yang mudah
- 3) Waktu pemrosesan cepat

3.2. Pencarian Informasi

Pencarian informasi memakai kata kunci yang dimasukkan oleh pengunjung ada di database buku, maka pencarian dilakukan berdasarkan dari pencarian yang selalu dicari akan ditemukan secara otomatis[12].

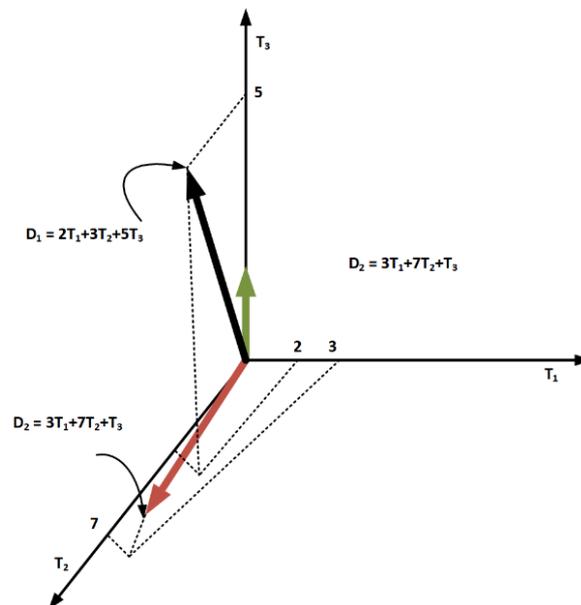
3.3. Data Preprocessing

Pada Tahapan selanjutnya dilakukan sesudah pendeteksian bahasa merupakan data *preprocessing*. Data *Preprocessing* dengan tujuan membersihkan data per *noise*, mempunyai dimensi yang kecil, & lebih terstruktur sebagai akibatnya bisa diolah lebih lanjut. Berikut merupakan tahapan preprocessing[13],

- 1) *Case folding* digunakan untuk mengganti seluruh huruf ke dalam sebuah kalimat sebagai huruf kecil dan huruf a-z yang akan terdeteksi[14].
- 2) *Stopword removal* merupakan tahapan untuk menghilangkan istilah-istilah yg tidak diperlukan oleh sebuah teks. *Stopwords* adalah istilah yg tidak mempunyai makna apabila berdiri sendiri & wajib dibersihkan pada pendekatan *bag-of-words*[15].
- 3) *Stemming* merupakan sebuah proses pemetaan istilah dalam sebuah kalimat yang berguna sebagai istilah asli (tanpa awalan, akhiran, penyisipan, kombinasi) yg dihukum prosedur pemecahan tertentu[16].
- 4) Tokensasi merupakan suatu proses memisah paragraf atau kalimat yang berupa teks pada dokumen sebagai poin-poin khusus[17].

3.4. Vector Space Model(VSM)

VSM adalah *Model Information Retrieval* yang merepresentasikan *query* dan dokumen sebagai *vektor* diruang multidimensi. Kesesuaian pada *query* dan dokumen bisa dinilai dari vektor *query* dan vektor dokumen. Pada *VSM* Model bobot dijumlah dari tiap *term*(istilah) yang ditemukan disetiap *query* dan dokumen yang diajukan oleh pengguna. *Term*(istilah) adalah kumpulan kata yang mengungkapkan suatu pengertian. Penetapan relevansi suatu dokumen terhadap suatu *query* dilakukan sebagai ukuran kemiripan antara vektor *query* dan dokumen. Gambar 2 menunjukkan contoh representasi relevansi antara *query* dengan dokumen[18].



Gambar 2. Vektor Space Model

3.5. Pembobotan *Term*

Pada fase ini, *query* pencarian dan dataset konten dibobot berdasarkan kata atau frasa, dan *term frequency* dari setiap kata dalam konten dataset dihitung. TF-IDF adalah algoritma pembobotan setiap term/frase dengan menghitung jumlah kemunculan term dalam setiap dokumen saat mencari informasi. Algoritma ini juga dikenal karena efisiensi, kesederhanaan dan hasil yang akurat. pembobotan *term weighting* digunakan rumus TF-IDF adalah sebagai berikut:

$$w = tf \times idf \quad (1)$$

3.6. *Cosine Similarity*

Cosine Similarity adalah metode pengukuran kemiripan antara dua dokumen atau teks dalam pencarian dokumen. Saat mengukur kecenderungan *Cosine Similarity* mengukur nilai berdasarkan sudut yang didapatkan. Nilai yang didapatkan berdasarkan sudut *vector* berkisar dari 0-1, semakin mendekati 1 maka tingkat kemiripan antara dokumen dan *query* semakin tinggi dan semakin mendekati 0 maka semakin rendah kemiripannya[4]. *Cosine Similarity* bisa dihitung menggunakan 2 persamaan sebagai berikut:

$$\text{CosSim}(di, qi) = \frac{di, qi}{|di, qi|} \quad (2)$$

$$\frac{\sum_j^t 1(qij \cdot dij)}{\sqrt{\sum_j^t 1(qij \cdot)^2 + \sum_j^t 1(dij \cdot)^2}}$$

3.7. Alur Pencarian Dokumen

Tahap pertama adalah mengolah data masukan untuk kontrol, Tahap kedua mengolah data masukan kalimat dalam dokumen membentuk kelompok kata, Tahap ketiga untuk diproses memeriksa kata-kata data. Sesudah di periksa tahap selanjutnya akan dibandingkan dan akan mendapatkan solusi perbaikan. Pada tahap akhir proses ini menghasilkan perbedaan dari data yang dimasukkan.

3.8. Keunggulan Algoritma

Keunggulan algoritma N-Gram dalam pencarian konten ialah adanya pengkoreksian ejaan per kata. Maka kata-kata yang dicari harus sesuai dengan konten yang ada di *database*[19] dan N-gram tidak terlalu sensitif terhadap kesalahan ejaan yang ditemukan dalam dokumentasi.

4. HASIL DAN PEMBAHASAN

Hasil implementasi dari pencarian konten perpustakaan digital dengan menggunakan model Information Retrieval dikombinasikan dengan Predictive *N-Gram* sebagai pengolahannya, pengenalan bahasa menggunakan *N-Gram*, lalu dilakukan pra-pemrosesan dokumen menggunakan *algoritma stemming sastrawi* untuk bahasa Indonesia, dan Pembobotan *term* dilakukan dalam *TF-IDF* dan diwakili oleh model ruang vektor (VSM). Pada langkah terakhir, kesamaan dokumen dihitung dengan persamaan *cosinus*. Data yang digunakan berupa data buku, yaitu judul. Berikut sample data buku pada tabel 2.

Tabel 2. Data Buku

No	Data Buku
D1	ASP.NET Core 5 dan Angular
D2	ASP.NET Core 3 dan Angular 9, Edisi Ketiga
D5	Memulai Django API dengan React Build Django 4 Web API dengan React Full Stack Applications

4.1. *Pre-Processing*

Berikutnya data diatas diolah menjadi lebih kecil dan membersihkan noise per data. Berikut hasil data setelah *preprocessing* pada tabel 3.

Tabel 3. Hasil Data Setelah Preprocessing

No	Data setelah preprocessing
D1	'asp', 'net', 'core', 'angular'
D2	'asp', 'net', 'core', 'angular', 'edisi', 'tiga'
D5	'mulai', 'django', 'api', 'react', 'build', 'django', 'web', 'api', 'react', 'full', 'stack', 'applications'

4.2. Pembobotan Term dan TFIDF

Data diatas kemudian dicari frekuensi kemunculan(term) setiap kata pada konten di dataset. Berikut hasil pencarian term pada konten pada tabel 4.

Tabel 4. Hasil Pencarian Term dan TFIDF

TERM	TF			TF Normalisasi			DF	IDF	TF IDF		
	d1	d2	d3	d1	d2	d3			d1	d2	d3
asp	1	1	0	0,25	0,167	0	2	0,176	0,176	0,176	0
net	1	1	0	0,25	0,167	0	2	0,176	0,176	0,176	0
core	1	1	0	0,25	0,167	0	2	0,176	0,176	0,176	0
angular	1	1	0	0,25	0,167	0	2	0,176	0,176	0,176	0
edisi	0	1	0	0	0,167	0	1	0,477	0	0,477	0
tiga	0	1	0	0	0,167	0	1	0,477	0	0,477	0
mulai	0	0	1	0	0	0,083	1	0,477	0	0	0,477
django	0	0	2	0	0	0,167	1	0,477	0	0	0,954
api	0	0	2	0	0	0,167	1	0,477	0	0	0,954
react	0	0	2	0	0	0,167	1	0,477	0	0	0,954
build	0	0	1	0	0	0,083	1	0,477	0	0	0,477
web	0	0	1	0	0	0,083	1	0,477	0	0	0,477
full	0	0	1	0	0	0,083	1	0,477	0	0	0,477
stack	0	0	1	0	0	0,083	1	0,477	0	0	0,477
applications	0	0	1	0	0	0,083	1	0,477	0	0	0,477
Panjang Dokumen	4	6	12								

4.3. Consine Similarity

Data tersebut kemudian dilakukan perhitungan untuk pengukuran kemiripan antara dua dokumen atau teks dalam pencarian dokumen/konten. Berikut hasil perhitungan *consine similarity* pada konten/dokumen dan perangkingan *consine similarity* pada tabel 5.

Tabel 5. Hasil Perhitungan Consine Similarity dan Perangkingan

No	Consine Similarity				Perangkingan	
	d1	d2	d3	q(pemanding)	Rank	Jumlah
d1	1,000	0,756	0,000	0,551	d1	0,551
d2	0,756	1,000	0,000	0,417	d2	0,417
d3	0,000	0,000	1,000	0,000	d3	0,000
q	0,551	0,417	0,000	1,000		

4.4. Implementasi

Penerapan(Implementasi) *information retrieval* dilakukan pada sistem perpustakaan digital dalam proses pencarian konten(buku) . adanya data buku, pencarian buku, data peminjaman.

Pada gambar 3 merupakan tampilan proses pencarian data buku, pada proses ini pengguna harus memasukan kata kunci data buku seperti dari judul buku, penulis atau penerbit yang ingin di cari.

No	Sampul	ISBN	Title	Penerbit	Tahun Buku	Stok Buku	Dipinjam	Tanggal Masuk	Aksi
4		2898-2282-9872	Kecerdasan Buatan untuk boneka	: John Wiley & Sons, Inc	2018	1	0	2022-07-31 17:37:11	Detail
5		978-1-78961-216-5	ASP.NET Core 3 dan Angular 9, Edisi Ketiga	BIRMINGHAM - MUMBAI	2020	1	1	2022-07-31 17:39:06	Detail
6		978-1-80056-033-8	ASP.NET Core 5 dan Angular 8	BIRMINGHAM - MUMBAI	2021	1	0	2022-07-31 17:40:25	Detail

Gambar 3. Pencarian Buku

Pada gambar 4 merupakan tampilan data peminjaman buku, pada tahap ini pengguna telah melakukan proses peminjaman buku pada perpustakaan.

Data Transaksi		Pinjam Buku									
No Peminjaman	: PJ0012	Status	: Dipinjam								
Tgl Peminjaman	: 2022-07-31	Tgl Kembali	: belum dikembalikan								
Tgl pengembalian	: 2022-08-07	Denda	: Tidak Ada Denda								
ID Anggota	: AG003	Kode Buku	: 1, BK0011								
Biodata	: Nama : Arya Kamandanu Anggota Telepon : 085322599812 E-mail : arya1356@gmail.com Alamat : Pemalang Level : Anggota	Data Buku	: <table border="1"> <thead> <tr> <th>No</th> <th>Title</th> <th>Penerbit</th> <th>Tahun</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>ASP.NET Core 3 dan Angular 9, Edisi Ketiga</td> <td>BIRMINGHAM - MUMBAI</td> <td>2020</td> </tr> </tbody> </table>	No	Title	Penerbit	Tahun	1	ASP.NET Core 3 dan Angular 9, Edisi Ketiga	BIRMINGHAM - MUMBAI	2020
No	Title	Penerbit	Tahun								
1	ASP.NET Core 3 dan Angular 9, Edisi Ketiga	BIRMINGHAM - MUMBAI	2020								
Lama	: 7 Hari										

Gambar 4. Peminjaman Buku

Pada gambar 5 merupakan tampilan data buku, pada data buku ini terdapat judul buku, nama penulis atau penerbit, dan tahun terbit.

No	Sampul	ISBN	Title	Penerbit	Tahun Buku	Stok Buku	Dipinjam	Tanggal Masuk	Aksi
1			Memulai Flutter dengan Dart	Leanpub book	2021	1	0	2022-07-31 17:13:13	
2		978-1-491-90352-0	Belajar Arduino	O'Reilly Media, Inc	2020	1	0	2022-07-31 17:14:36	
3		3548-2948-9834	Memulai Django API dengan React Build Django 4 Web API dengan React Full	Greg Lim	2022	1	0	2022-07-31 17:18:40	

Gambar 5. Data Buku

4.5. Prosedur pencarian konten pada perpustakaan digital

Pada pencarian konten ini memanfaatkan indeks/kata kunci. pencarian dilakukan dengan memasukan kata kunci yang dilakukan oleh pengguna. mesin pencari menjalankan proses pencarian konten berdasarkan indeks, memungkinkan konten untuk ditemukan. Dari proses ini didapatkan rekomendasi apakah pencarian dapat menampilkan dokumen yang sesuai dengan kata kunci. Dokumen hasil pencarian adalah dokumen yang disimpan di database.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil dan pembahasan diatas bisa disimpulkan, bahwa *N-gram* dan *Information retrieval* dapat mempercepat proses pencarian konten pada perpustakaan digital. Sebab adanya pengkoreksian pada konten yang dicari. Pencarian dokumen teks berbasis konten menggunakan mesin pencari memudahkan untuk menemukan relevansi antara hasil pencarian dan kata kunci dan untuk pemeringkatan hasil pencarian konten yang sering dicari oleh pengguna perpustakaan digital menggunakan *cosine similarity*.

DAFTAR PUSTAKA

- [1] A. M. Aminullah *et al.*, "Meningkatkan Kesadaran Generasi Muda Terhadap Pengembangan Koleksi Digital Dalam Membangun Perpustakaan Digital di Perpustakaan UIN Alauddin Makassar," *Juournal Educ. Psychol. Couns.*, vol. 3, no. 1, pp. 88–94, 2021.
- [2] J. Setiyono and M. M., "Persepsi Pemustaka Terhadap Pengembangan Institutional Repository Di Perpustakaan Isi Surakarta," *Publ. Libr. Inf. Sci.*, vol. 3, no. 1, p. 20, 2019, doi: 10.24269/pls.v3i1.1576.
- [3] 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.
- [4] I. G. Anugrah, "Penerapan Metode N-Gram dan Cosine Similarity Dalam Pencarian Pada Repositori Artikel Jurnal Publikasi," *Build. Informatics, Technol. Sci.*, vol. 3, no. 3, pp. 275–284, 2021, doi: 10.47065/bits.v3i3.1058.
- [5] Z. Pratama, E. Utami, and M. R. Arief, "Analisa Perbandingan Jenis N-GRAM Dalam Penentuan Similarity Pada Deteksi Plagiat," *Creat. Inf. Technol. J.*, vol. 4, no. 4, p. 254, 2019, doi: 10.24076/citec.2017v4i4.118.
- [6] A. S. Wahyuni and A. O. . Dewi, "Persepsi Pemustaka Terhadap Desain Antarmuka Pengguna (User Interface) Aplikasi Perpustakaan Digital "Jogja" Berbasis Android," *J. Ilmu Perpust.*, vol. VII, no. 1, pp. 21–30, 2018, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/jip/article/view/22810>.
- [7] A. I. Fahma, I. Cholissodin, and R. S. Perdana, "Identifikasi Kesalahan Penulisan Kata (Typographical Error) Pada Dokumen Berbahasa Indonesia Menggunakan Metode N-Gram Dan

- Levenshtein Distance,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 1, pp. 53–62, 2018.
- [8] J. Pardede, M. Gustiana Husada, and R. Riansyah, “Implementasi Dan Perbandingan Metode Okapi BM25 Dan PLSA Pada Aplikasi Information Retrieval,” *Itenas Repos.*, pp. 1–10, 2018.
- [9] A. A. Muin and F. Fitriani, “Implementasi Sistem Informasi Pengelolaan Data Perpustakaan Berbasis Web (Studi Kasus: Perpustakaan Daerah Kabupaten Sinjai),” *J. INSYPRO (Information Syst. Process.*, vol. 5, no. 2, pp. 1–8, 2021, doi: 10.24252/insypro.v5i2.19398.
- [10] A. Nurfalah and A. A. Suryani, “Analisis Sentimen Berbahasa Indonesia dengan Pendekatan Lexicon-Based pada Media Sosial,” *J. Masy. Inform. Indones.*, vol. 2, no. 1, pp. 1–8, 2017.
- [11] R. Sadida, M. R. Azkia, B. P. Candra, N. Rezeki, and M. O. C. Rendy, “Perancangan Sistem Analisis Sentimen Masyarakat pada Sosial Media dan Portal Berita,” *Semin. Nas. Teknol. Inf. dan Multimed. 2017*, no. 2015, pp. 2–7, 2017.
- [12] T. Zebua and N. Silalahi, “Aplikasi Saran Buku Bacaan Bagi Pengunjung Perpustakaan AMIK STIEKOM Sumatera Utara Berdasarkan Algoritma Brute Force,” *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 3, no. 3, p. 66, 2018, doi: 10.30645/jurasik.v3i0.67.
- [13] M. Z. Naf’an, A. Burhanuddin, and A. Riyani, “Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen,” *J. Linguist. Komputasional*, vol. 2, no. 1, pp. 23–27, 2019.
- [14] D. S. Indraloka and B. Santosa, “Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia,” *J. Sains dan Seni ITS*, vol. 6, no. 2, pp. 6–11, 2017, doi: 10.12962/j23373520.v6i2.24419.
- [15] M. S. Anwar, I. M. I. Subroto, and S. Mulyono, “Sistem Pencarian E-Journal Menggunakan Metode Stopword Removal Dan Stemming,” *Pros. Konf. Ilm. Mhs. UNISSULA 2*, pp. 58–70, 2019, [Online]. Available: <http://lppm-unissula.com/jurnal.unissula.ac.id/index.php/kimueng/article/viewFile/8420/3887>.
- [16] P. F. Ariyani, A. Rahmala, and N. Juliasari, “Implementasi Metode Stemming Tala Dan Fungsi Jaccard Pada Aplikasi Katalog Perpustakaan,” *Semin. Nas. Inov. dan Apl. Teknol. di Ind. 2019*, pp. 128–133, 2019.
- [17] I. Made Suwija Putra, N. Putu Ayu Widiari, and I. Wayan Gunaya, “Implementasi Generalized Vector Space Model (GVSM) dalam Pencarian Buku di Perpustakaan,” *J. Ilm. Merpati (Menara Penelit. Akad. Teknol. Informasi)*, p. 86, 2019, doi: 10.24843/jim.2019.v07.i01.p10.
- [18] S. Fauziah, D. N. Sulistyowati, and T. Asra, “Optimasi Algoritma Vector Space Model Dengan Algoritma K-Nearest Neighbour Pada Pencarian Judul Artikel Jurnal,” *J. Pilar Nusa Mandiri*, vol. 15, no. 1, pp. 21–26, 2019, doi: 10.33480/pilar.v15i1.27.
- [19] D. A. N. Lempel, Z. I. V Welch, D. A. N. Lempel, Z. I. V Welch, and L. Z. W. Yang, “METODE N-GRAM UNTUK ALGORITMA KOMPRES HUFFMAN,” no. 1102090.

NOMENKLATUR

Tf arti dari variabel yang muncul pada tiap dokumen
 Idf arti dari *Invers* dokumen frekuensi
 q_{ij} arti dari *term* ke i untuk dokumen ke j dari q
 d_{ij} arti dari *term* ke i untuk kueri ke j (*keyword term*)
 t arti dari Jumlah istilah j pada d atau q