

IMPLEMENTASI KINERJA ALGORITMA ROCCHIO DAN SEQUENTIAL SEARCH DALAM PROSES TEMU KEMBALI INFORMASI PADA KOLEKSI DIGITAL PERPUSTAKAAN SMA NEGERI 1 KUALA

Dio Febrian Surbakti¹, Safrizal², Nurainun³, Aulia Rahma⁴, Annisa Purwanti⁵,
Aginga Sri Malem Br Sitepu⁶

Sistem Informasi, ^{1,3,4,5}STMIK Kaputama Binjai, ²Universitas Muhammadiyah Asahan, Sumatera Utara

E-mail: [*diofebrian495@gmail.com](mailto:diofebrian495@gmail.com)¹, rizalsyl75@gmail.com², ainunn0702@gmail.com³,
auliarhmaa131@gmail.com⁴, annisapurwanti666@gmail.com⁵, agintasitepu12@gmail.com⁶

ABSTRAK

Perkembangan teknologi informasi menyoroti perlunya sekolah menyediakan layanan informasi yang tepat waktu dan relevan. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja Algoritma Rocchio dan Algoritma Sequential Search dalam proses pengambilan informasi dari koleksi digital SMA Negeri 1 Kuala. Algoritma Rocchio digunakan dengan pendekan yang relevan umpan balik yang dapat memperbaiki kueri pada dokumen bertebaran relevan dan tidak relevan, Sedangkan Sequential Search menggunakan metodekan linear yang menelusuri dokumen secara berurutan. Studi ini menggunakan parameter akurasi, presisi, recall, dan pencarian waktu untuk menilai kuantitas dokumen digital. Temuan studi ini menunjukkan bahwa Algoritma Rocchio memberikan hasil pencarian yang lebih relevan karena dapat melakukan "bobot kata penyesuai" pada kueri. Sementara itu, Pencarian Sekuensial Algoritma unggul pada memproses sederhana namun kurang efektif ketika jumlah dokumen semakin besar. Dengan demikian, implementasi Algoritma Rocchio lebih direkomendasi untuk meningkatkan efektivitas temu kembali informasi pada perpustakaan digital disekolah SMA Negeri 1 Kuala.

Kata kunci

Sistem Temu Kembali, Algoritma Rocchio, Sequential Search, SMA Negeria 1 Kuala

ABSTRACT

The development of information technology highlights the need for schools to provide timely and relevant information services. This study aims to analyze and compare the performance of the Rocchio Algorithm and the Sequential Search Algorithm in the process of retrieving information from the digital collection of SMA Negeri 1 Kuala. The Rocchio Algorithm is used with a relevant feedback approach that can improve query searching for relevant and irrelevant scattered documents, while Sequential Search uses a linear method that searches documents sequentially. This study uses parameters of accuracy, precision, recall, and search time to assess the quantity of digital documents. The findings of this study indicate that the Rocchio Algorithm provides more relevant search results because it can perform "adjustment word weights" on queries. Meanwhile, the Sequential Search Algorithm excels in simple processing but is less effective when the number of documents increases. Thus, the implementation of the Rocchio Algorithm is more recommended to increase the effectiveness of information retrieval in digital universities and digital libraries.

Keywords

Information Retrieval System, Rocchio Algorithm, Sequential Search, SMA Negeri 1 Kuala

1. PENDAHULUAN

Kemajuan teknologi informasi telah menghasilkan perubahan yang signifikan dalam pengolahan informasi, terutama di bidang perpustakan. Perpustakan tidak hanya menyediakan koleksi nyata tetapi juga koleksi digital seperti e-book, jurnal elektronik, dan dokumen digital lainnya. Menurut (Pendit, 2005), perpustakan digital adalah sistem yang menyediakan akses terorganisir ke koleksi digital untuk memenuhi kebutuhan informasi pengguna dengan cepat dan efektif. Dengan munculnya teknologi digital, diharapkan pengguna dapat memperoleh informasi dengan lebih mudah tanpa dibatasi oleh waktu atau ruang.

Seiring dengan bertambahnya jumlah koleksi digital, tantangan dalam proses pencarian dan pengambilan informasi menjadi lebih kompleks. Pengguna sering mengalami kesulitan menemukan informasi terkait dengan kunci yang disediakan. Menurut (Baeza-Yates & Ribeiro, 2011), tujuan sistem ini adalah untuk menemukan dokumen yang relevan dari kumpulan data besar berdasarkan kebutuhan informasi pengguna. Jika informasi sistem tidak diproses oleh algoritma yang sesuai, hasil analisis akan kurang relevan dan kurang ideal.

Algoritma adalah komponen penting dari sistem informasi. Salah satu algoritma yang sering digunakan adalah algoritme Rocchio, yang merupakan komponen dari model vektor dan menggunakan umpan balik relevansi untuk meningkatkan kualitas hasil pencarian. Menurut (SALTON, 1983), algoritme Rocchio dapat meningkatkan kueri pencarian dengan meningkatkan kueri dalam dokumen yang relevan dan menguranginya dalam dokumen yang tidak relevan. Selain itu, algoritme Pencarian Sequential adalah metode sederhana pencarian yang bekerja dengan menganalisis data satu per satu secara berurutan. Menurut Knuth (1998), pencarian sekuensial dapat diterapkan pada data yang cukup kecil untuk mudah diimplementasikan, meskipun memiliki kerugian dalam hal efisiensi waktu.

Perpustakan SMA Negeri 1 Kuala memiliki koleksi digital yang mendukung kegiatan belajar mengajar. Namun, sistem pencarian yang digunakan masih memiliki kemampuan untuk memberikan hasil yang tepat waktu dan relevan. Oleh karena itu, penelitian tentang pengembangan algoritme pencarian yang sesuai diperlukan. Masalah utama dalam penelitian ini adalah bagaimana menerapkan algoritma Rocchio dan Sequential Search dalam proses pengambilan informasi dari koleksi digital SMA Negeri 1 Kuala, serta bagaimana algoritma ini bekerja sama untuk menghasilkan hasil pencarian yang relevan dan efisien.

Penelitian ini bertujuan untuk menerapkan algoritme Pencarian Sequential dan Rocchio pada koleksi informasi digital SMA Negeri 1 Kuala dan menganalisis kinerja kedua algoritme berdasarkan tingkat relevansi dan efisiensi pencarian. Diharapkan hasil kajian ini berkontribusi secara teoritis terhadap pengembangan pengetahuan berbasis informasi serta praktis kepada pengurus sekolah dalam meningkatkan kualitas layanan pengumpulan digital.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan tujuan menganalisis dan membandingkan efisiensi algoritma Rocchio dan Sequential Search dalam proses pengambilan informasi dari koleksi digital SMA Negeri 1 Kuala. Pendekatan ini dipilih karena memungkinkan kinerja algoritma disesuaikan secara objektif berdasarkan parameter yang dievaluasi. Data teoretis dikumpulkan dari berbagai jurnal

ilmiah dan literatur akademis yang relevan (seperti (Rahmanto et al., n.d.) ; (Suparman & Azizah, 2022)

Tujuan penelitian ini adalah sistem pengumpulan informasi digital SMA Negeri 1 Kuala. Dataset yang digunakan terdiri dari kumpulan dokumen digital, seperti e-book, modul pendidikan, dan dokumen akademik lainnya. Setiap dokumen diproses melalui langkah preproses yang mencakup tokenisasi, penghapus kata penghapus, dan pemrosesan kata menggunakan metode TF-IDF.

2.1 Sistem Temu Kembali

Pengambilan informasi adalah proses terkomputerisasi yang menghasilkan dokumen yang relevan berdasarkan masukan pengguna dengan membandingkan kueri pengguna dengan indeks yang dihasilkan dari dokumen yang ada dalam proses pengambilan informasi (Yugianus et al., 2013). Setiap kata diwakili oleh kata kunci, yang juga dikenal sebagai istilah indeks. Menurut Yugianus (2013), istilah indeks adalah kata yang membantu pengguna memahami gagasan utama dokumen.

Teks, suara, gambar, atau data, antara lain, adalah contoh dokumen yang menggunakan pengambilan informasi. Menurut Fachruddin (2011), data koleksi yang bisa menjadi pencari sumber juga bisa berupa pesan teks, seperti email, fax, dan dokumen tertulis, termasuk dokumen yang tersedia secara online. Kumpulan dokumen di tugas akhir ini terdiri dari artikel jurnal abstrak yang berasal dari materi lokal dari Perpustakan Universitas Brawijaya.

2.2 Metode Rocchio

Algoritme Rocchio (Jr., 1971) adalah algoritma klasik untuk mengimplementasikan umpan balik relevansi. Sejak tahun 1970, sistem SMART Salton's telah mengembangkan dan mempopulerkan mekanisme umpan balik yang relevan menggunakan algoritma Rocchio (UP, 2007). Dalam konteks kueri, sistem mengambil informasi dari kueri pengguna dan menghasilkan dokumen yang relevan dan tidak relevan, seperti yang ditunjukkan pada Gambar 2.1 di bawah ini (UP, 2007):

$$\vec{q}_m = \alpha \vec{q}_0 + \beta \frac{1}{|D_r|} \sum_{\vec{d}_j \in D_r} \vec{d}_j - \gamma \frac{1}{|D_{nr}|} \sum_{\vec{d}_j \in D_{nr}} \vec{d}_j$$

q_m = new query vector

q_0 = initial query vector

d_j = Relevant document vector

d_k = Non-relevant document vector

D_r = relevant document vector

D_{nr} = non-relevant document vector

α = Original query weight

β = Relevant document weight

γ = Non-relevant document weight

3. HASIL DAN PEMBAHASAN

3.1 Uji Proses Pencarian data

Lakukan pencarian data terhadap data koleksi perpustakaan. Langkah-langkah penerapan algoritma rocchio pada data data di atas adalah sebagai berikut :

D1: Pembahasan mengenai pemanasan global dan dampak teknologi.

D2: Dasar sistem informasi dan peran teknologi memiliki dampak yang besar.

D3: Asal usul perkembangan internet secara global.

D4: Fungsi ekosistem mangrove bagi lingkungan.

D5: Konsep algoritma pencarian data.

Query awal (Q): teknologi pemanasan global

Langkah 1 – Tentukan Vobulary

Dari seluruh dokumen + query, diperoleh 13

Tabel 1. Kata Dokumen Unik

No	Istilah
1	Pemanasan
2	Global
3	Dampak
4	Dasar
5	Sistem
6	Informasi
7	Teknologi
8	Internet
9	ekosistem
10	Mangrove
11	lingkungan
12	algoritma
13	data

Langkah 2 – Representasikan Dokumen dalam Bentuk Vektor Biner

Tabel 2. Dokumen Vektor Biner

Term/Doc	D1	D2	D3	D4	D5
Pemanasan	1	0	0	0	0
Global	1	0	1	0	0
Dampak	1	1	0	0	0
Dasar	0	1	0	0	0
Sistem	0	1	0	0	0
Informasi	0	1	0	0	0
Teknologi	1	0	0	0	0
Internet	0	0	1	0	0
ekosistem	0	0	0	1	0
Mangrove	0	0	0	1	0
lingkungan	0	0	0	1	0
algoritma	0	0	0	0	1
data	0	0	0	0	1

LANGKAH 3 — Representasi Query Awal

Query awal: teknologi pemanasan global

Tabel 3. Representasi Query Awal

Term/Doc	Ada	Bobot
Pemanasan	1	1
Global	1	1
Teknologi	1	1
lainnya		

Vektor Q:

$$Q = [1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$$

LANGKAH 4 — Tentukan Dokumen Relevan & Tidak Relevan

Tabel 4. Dokumen Relavan dan Tidak Relavan

Kategori	Dokumen	Jumlah
Relevan (Dr)	D1, D2, D3	
Tidak Relevan (Dnr)	D4, D5	

Parameter Rocchio:

$$\alpha = 1$$

$$\beta = 0.75$$

$$\gamma = 0.25$$

LANGKAH 5 — Hitung Σ Dokumen Relevan

Penjumlahan vektor D1 + D2 + D4:

Tabel 5. Dokumen Relevan

Term/Doc	D1	D2	D3	$\Sigma(Dr)$
Pemanasan	1	0	0	1
Global	1	0	1	2
Dampak	1	1	0	2
Dasar	0	1	0	1
Sistem	0	1	0	1
Informasi	0	1	0	1
Teknologi	1	0	0	1
Internet	0	0	1	1
ekosistem	0	0	0	0
Mangrove	0	0	0	0
lingkungan	0	0	0	0
algoritma	0	0	0	0
data	0	0	0	0

$$\Sigma(Dr) = [1, 2, 2, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0]$$

LANGKAH 6 — Hitung Rata-rata Dokumen Relevan

$$\text{avg}(Dr) = \frac{\sum Dr}{3}$$

Tabel 6. Rata-rata Dokumen Relevan

Term/Doc	$\Sigma(Dr)$	avg
Pemanasan	1	0,3333
Global	2	0,6666
Dampak	2	0,6666
Dasar	1	0,3333
Sistem	1	0,3333
Informasi	1	0,3333
Teknologi	1	0,3333
Internet	1	0,3333
ekosistem	0	0
Mangrove	0	0
lingkungan	0	0
algoritma	0	0
data	0	0

LANGKAH 7 — Kalikan dengan $\beta = 0.75$

$$\beta \cdot \text{avg}(Dr)$$

Hasil: [0.25, 0.50, 0.50, 0.25, 0.25, 0.25, 0.25, 0.25, 0, 0, 0, 0, 0]

LANGKAH 8 — Hitung Σ Dokumen Tidak Relevan

Tabel 7. Dokumen Tidak relevan

Term/Doc	D4	D5	$\Sigma(Dnr)$
Pemanasan	0	0	0
Global	0	0	0
Dampak	0	0	0
Dasar	0	0	0
Sistem	0	0	0
Informasi	0	0	0
Teknologi	0	0	0
Internet	0	0	0
ekosistem	1	0	1
Mangrove	1	0	1
lingkungan	1	0	1
algoritma	0	1	1
data	0	1	1

$$\Sigma Dnr = [0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1]$$

LANGKAH 9 — Hitung Rata-rata Dnr

$$\text{avg}(Dnr) = \frac{\sum Dnr}{2}$$

Hasil: [0, 0, 0, 0, 0, 0, 0, 0, 0.5, 0.5, 0.5, 0.5, 0.5]

LANGKAH 10 — Kalikan dengan $\gamma = 0.25$

$$\gamma \cdot \text{avg}(Dnr)$$

Hasil: [0, 0, 0, 0, 0, 0, 0, 0.125, 0.125, 0.125, 0.125, 0.125, 0.125]

LANGKAH 11 — Hitung Komponen αQ

$$\alpha Q = Q = [1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$$

LANGKAH 12 — Hitung Query Baru Q_new

Formula Rocchio: $Q_{new} = \alpha Q + (\beta \cdot avg(Dr)) - (\gamma \cdot avg(Dnr))$

Hitung per elemen:

Tabel 8. Hitung Query Baru

Term/Doc	αQ	$\beta \cdot avg(Dr)$	$\gamma \cdot avg(Dnr)$	Q new
Pemanasan	1	0,25	0	1,25
Global	1	0,50	0	1,5
Dampak	1	0,50	0	1,5
Dasar	0	0,25	0	-0,25
Sistem	0	0,25	0	-0,25
Informasi	0	0,25	0	-0,25
Teknologi	0	0,25	0	-0,25
Internet	0	0,25	0	-0,25
ekosistem	0	0	0,125	-0,125
Mangrove	0	0	0,125	-0,125
lingkungan	0	0	0,125	-0,125
algoritma	0	0	0,125	-0,125
data	0	0	0,125	-0,125

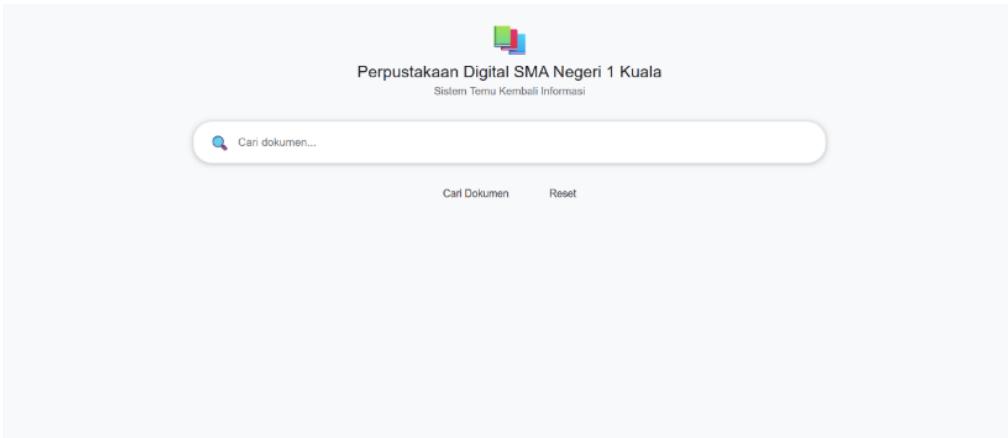
HASIL AKHIR — Vektor Query Baru

$$Q_{new} 1.25, 1.5, 1.5, -0.25, -0.25, -0.25, -0.25, -0.25, -0.125, -0.125, -0.125, -0.125, -0.125$$

3.2 Implementasi Sistem

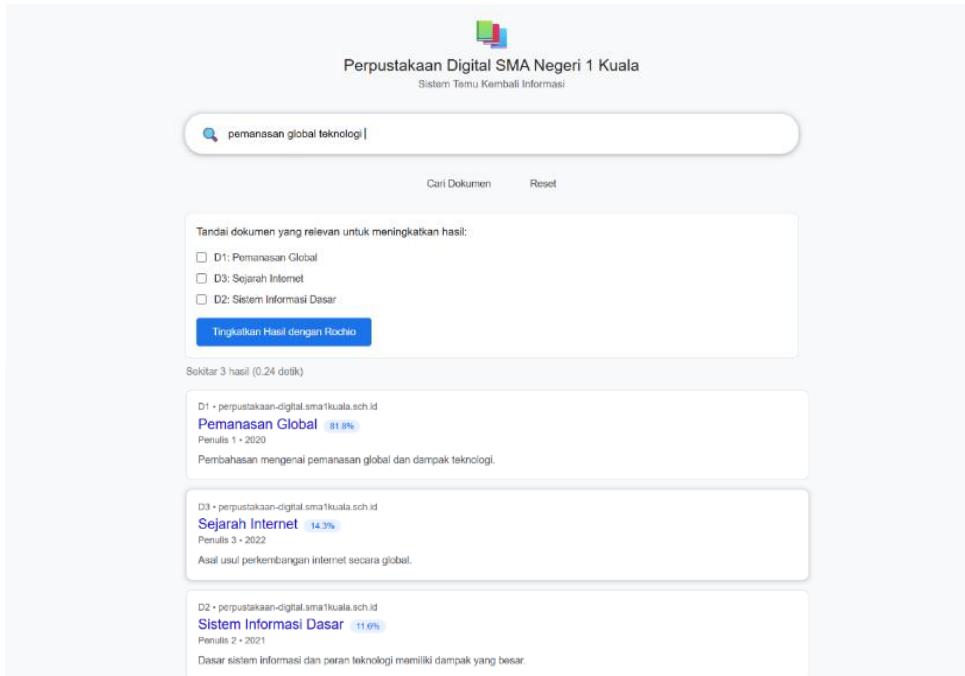
Implementasi algoritma Rocchio dan Sequential Search pada penelitian ini diterapkan dalam sistem temu kembali informasi untuk koleksi digital perpustakaan SMA Negeri 1 Kuala dengan tujuan meningkatkan relevansi hasil pencarian dokumen. Sistem dibangun menggunakan pendekatan information retrieval berbasis vector space model, di mana dokumen dan kueri direpresentasikan dalam bentuk vektor melalui tahapan praprosesan teks, meliputi tokenisasi, penghapusan stopword, dan pembobotan istilah menggunakan TF-IDF.

Algoritma Rocchio digunakan untuk melakukan penyesuaian kueri berdasarkan umpan balik relevansi pengguna dengan memperkuat bobot istilah dari dokumen relevan dan mengurangi pengaruh dokumen tidak relevan, sedangkan Sequential Search diterapkan sebagai metode pencarian linear dengan membandingkan kueri terhadap dokumen secara berurutan. Evaluasi kinerja sistem dilakukan menggunakan parameter akurasi, presisi, recall, dan waktu pencarian. Hasil penelitian menunjukkan bahwa algoritma Rocchio menghasilkan tingkat relevansi pencarian yang lebih baik dibandingkan Sequential Search, meskipun Sequential Search lebih sederhana dalam implementasi namun kurang efisien pada jumlah dokumen yang besar.



Gambar 1. Tampilan Awal Antarmuka

Tahap selanjutnya adalah proses pencarian dokumen pada sistem temu kembali informasi Perpustakaan Digital SMA Negeri 1 Kuala. Pengguna memasukkan kueri pencarian melalui kolom yang tersedia, kemudian sistem memproses kueri tersebut untuk menampilkan daftar dokumen yang memiliki tingkat kecocokan tertinggi. Selanjutnya, pengguna dapat menandai dokumen yang dianggap relevan sebagai umpan balik untuk meningkatkan kualitas hasil pencarian. Sistem kemudian menerapkan metode Rocchio untuk melakukan penyesuaian kueri berdasarkan dokumen relevan yang dipilih.



Gambar 2. Proses Pencarian Data Menggunakan Algoritma Sequential Search

Berdasarkan hasil implementasi sistem temu kembali informasi pada koleksi digital Perpustakaan SMA Negeri 1 Kuala, dapat disimpulkan bahwa algoritma Sequential Search dan Rocchio telah berhasil diterapkan sesuai dengan fungsi dan tujuan masing-masing. Sequential Search digunakan sebagai metode pencarian awal dengan cara membandingkan kata kunci terhadap setiap data secara berurutan, serta mampu menampilkan proses pencarian secara step-by-step, jumlah langkah, dan hasil pencarian ketika data ditemukan. Selanjutnya, algoritma Rocchio diterapkan untuk meningkatkan kualitas pencarian melalui penyesuaian kueri berdasarkan umpan balik relevansi

pengguna, sehingga sistem mampu menampilkan dokumen dengan tingkat kecocokan yang lebih baik.

Hasil pengujian menunjukkan bahwa Rocchio memberikan relevansi pencarian yang lebih tinggi dibandingkan Sequential Search, sementara Sequential Search unggul dari sisi kesederhanaan implementasi namun kurang efisien pada jumlah data yang besar. Dengan demikian, kombinasi kedua algoritma ini efektif digunakan dalam mendukung proses pencarian dokumen pada sistem perpustakaan digital.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa sistem temu kembali informasi pada koleksi digital Perpustakaan SMA Negeri 1 Kuala berhasil diimplementasikan menggunakan algoritma Sequential Search dan Rocchio dengan pendekatan information retrieval berbasis vector space model dan pembobotan TF-IDF. Sequential Search berfungsi sebagai metode pencarian dasar yang sederhana dan mudah diterapkan, namun kurang efisien ketika jumlah dokumen meningkat. Sementara itu, algoritma Rocchio mampu meningkatkan relevansi hasil pencarian melalui penyesuaian kueri berbasis relevance feedback. Berdasarkan evaluasi menggunakan parameter akurasi, presisi, recall, dan waktu pencarian, algoritma Rocchio terbukti lebih efektif dan direkomendasikan untuk meningkatkan ketepatan hasil pencarian pada sistem perpustakaan digital.

5. DAFTAR PUSTAKA

- Ardi Widodo, K., Adi Wibowo, S., & Vendyansyah, N. (2021). PENERAPAN SEQUENTIAL SEARCH UNTUK PENGELOLAAN DATA BARANG. *Antivirus: Jurnal Ilmiah Teknik Informatika*, 15(1), 86–97. <https://doi.org/10.35457/antivirus.v15i1.1385>
- Baeza-Yates, R. (Ricardo), & Ribeiro, B. de A. N. (2011). *Modern information retrieval : the concepts and technology behind search* (2nd ed). Pearson. <https://cir.nii.ac.jp/crid/1970867909879259840.bib?lang=ja>
- Febryanto, A. (2022). *Penerapan Algoritma Sequential Search untuk Mencari Data Siswa Pada Sekolah Menengah Kejuruan Negeri 3 Bengkalis*. 2(1), 51–59.
- Jr., R. J. J. (1971). Relevance feedback in information retrieval. *The SMART Retrieval System : Experiments in Automatic Document Processing*. <https://cir.nii.ac.jp/crid/1570291225586316544.bib?lang=ja>
- Pendit, P. L. (2005). *Perpustakaan digital: Perspektif perpustakaan perguruan tinggi Indonesia*. 1.
- Rahmanto, Y., Alfian, J., & Indra Borman, R. (n.d.). *Penerapan Algoritma Sequential Search pada Aplikasi Kamus Bahasa Ilmiah Tumbuhan 21*.
- SALTON, G. (1983). *Modern Information Retrieval*. <https://cir.nii.ac.jp/crid/1573668923832757248.bib?lang=ja>
- Setyawati Wahyuni, W., Andryana, S., Rahman, B., Teknologi Komunikasi dan Informatika, F., Nasional Ps Minggu, U., Jakarta Selatan, K., & Khusus Ibukota Jakarta, D. (n.d.). *PENGGUNAAN ALGORITMA SEQUENTIAL SEARCHING PADA APLIKASI PERPUSTAKAAN BERBASIS WEB*.
- Suparman, P., & Azizah, A. L. (2022). PENERAPAN ALGORITMA SEQUENTIAL SEARCH PADA DATA NILAI SISWA DI SMK BINA NASIONAL INFORMATIKA MENGGUNAKAN PHP DAN MYSQL DI SMK BINA NASIONAL INFORMATIKA. *Jurnal Informatika SIMANTIK*, 7(1). <https://www.simantik.fst-panca-sakti.ac.id/>

Yugianus, P., Dachlan, H. S., & Hasanah, R. N. (2013). Pengembangan sistem penelusuran katalog perpustakaan dengan metode Rocchio Relevance Feedback. *Jurnal EECCIS (Electrics, Electronics, Communications, Controls, Informatics, Systems)*, 7(1), 47–52.