



Search Engine Menggunakan Metode Information Retrieval

Rudiansyah¹, Rafly Ariyansyah², Rian Nanda³, Ovi Wiranda⁴
^{1,2,3,4}Universitas Abdurrah

e-mail: 1rudiansyah20@student.univrab.ac.id, 2rafly.ariyansyah20@student.univrab.ac.id,
3rian.nanda20@student.univrab.ac.id, 4ovi.wiranda20@student.univrab.ac.id

Abstrak

Pencarian data pada database yang biasa dilakukan oleh kebanyakan orang hanya mampu mencari judul artikel yang sesuai berdasarkan kata kunci yang diinputkan, Contoh, jika kata kunci yang dimasukkan adalah Kecerdasan buatan maka akan yang ditampilkan semua dokumen yang mengandung kata Kecerdasan buatan namun sistem tidak bisa mengukur mana dokumen yang paling mirip. Oleh karena itu, untuk dapat memenuhi kebutuhan pengguna dalam menemukan *keyword* yang cocok tetapi dapat mewakili data tersebut, kemudian sistem menghitung kemiripan antara *keyword* dengan daftar dokumen yang diwakili oleh kata perkata di dalam form. Artikel akan di tampilakan berdasarkan kata yang paling mirip. Dari hasil yang di dapat Ketika kita masukan *keyword* "android" maka akan tampil empat dokumen yang diurutkan sesuai tingkat kemiripannya, yaitu docId 3 dengan tingkat kemiripan 0.9512, docId 4 dengan tingkat kemiripan 0.5020, docId 2 dengan tingkat kemiripan 0.2671, docId 8 dengan tingkat kemiripan 0.1522.

Kata Kunci: Mesin Pencari, Pencarian Informasi, Kata Kunci

Abstract

Searching data in the database that is usually done by most people is only able to find the appropriate article title based on the keywords entered, for example, if the keyword entered is "Artificial Intelligence" then all documents containing the word "Artificial Intelligence" will be displayed but the system can not measure which documents are most similar. Therefore, to be able to meet user needs in finding keywords that match but can represent the data, then the system calculates the similarity between the keywords and the list of documents represented by words in the form. Articles will be displayed based on the most similar words. From the results obtained, when we enter the keyword "android" it will display four documents sorted according to the level of similarity, namely docId 3 with a similarity level of 0.9512, docId 4 with a similarity level of 0.5020, docId 2 with a similarity level of 0.2671, docId 8 with a similarity level 0.1522.

Keywords: Search Engine, Information Retrieval, Keyword

1. Pendahuluan

Perkembangan teknologi informasi yang sangat pesat telah membawa perubahan besar dalam semua aspek kehidupan manusia. Kemajuan dalam bidang teknologi informasi ini juga telah berdampak terhadap perilaku seseorang dalam menyimpan dan menelusuri informasi[1]. Perilaku dalam mendapatkan informasi ini juga diungkapkan oleh Jonner Hasugian bahwa, “perilaku pencarian informasi yang muncul dimaksudkan untuk memperoleh informasi yang benar-benar sesuai dengan kebutuhan terutama di lingkungan akademik dan peneliti melalui penelusuran secara Online”[2].

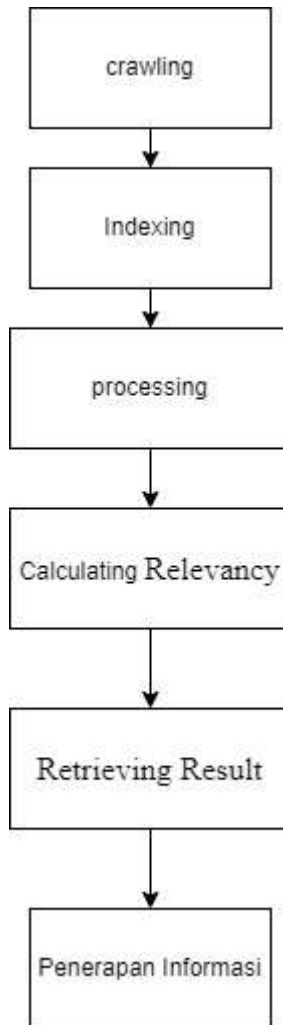
Penerapan teknologi digital dan jaringan komputer telah menyebabkan terjadinya “ledakan” informasi yang berkembang eksponensial. Hal ini menyebabkan Sistem temu kembali informasi (information retrieval =IR) mengalami kesulitan [3]. Information Retrieval (IR) merupakan bagian dari computer science yang berhubungan dengan pengambilan informasi dari dokumendokumen yang didasarkan pada isi dan konteks dari dokumen-dokumen itu sendiri. Informasi yang diinginkan pengguna direpresentasikan dalam bentuk query dan mengandung satu atau lebih term yang akan digunakan dalam pencarian[4][5][6].

Search engine atau mesin pencari merupakan teknik dari temu-kembali dalam menemukan dokumen dan sekaligus mengeksekusi algoritma peringkat dalam menampilkan dokumen. Pengguna dapat mencari halaman web yang dibutuhkan melalui search engine[7][8][9]. Search engine tidak lain sebuah mesin pencari yang ulet dan teliti, yang melakukan eksplorasi atas informasi-informasi yang di-request tanpa memandang kapan, di mana dan oleh siapa itu dilakukan[10]. Mesin pencari menggunakan indeks (yang sudah dibuat dan disusun secara teratur) untuk mencari file setelah pengguna memasukkan kriteria pencarian. Indexing atau pengindeksan merupakan proses membangun basis data indeks dari koleksi dokumen. Indexing dilakukan terhadap dokumen sebelum pencarian dilakukan [11][12].

Dari latar belakang diatas, maka penulis ingin mengembangkan sistem “*Search Engine Menggunakan Metode Information Retrieval*” jadi sistem akan secara otomatis melakukan pencarian secara (*retrival*).Proses retrieval dimulai dengan mengambil query dari pengguna, menerapkan stop word removal sehingga dihasilkan keyword yang cocok tetapi dapat mewakili query tersebut, kemudian sistem menghitung kemiripan antara keyword dengan daftar dokumen yang diwakili oleh kata-perkata di dalam index[13][14][15]. Dokumen akan ditampilkan diurutkan berdasarkan dokumen yang paling mirip.

2. Metode Penelitian

Adapun metode Penelitian Pada Gambar 1, Search Engine Menggunakan Metode Information Retrieval sebagai berikut:



Gambar 1. Kerangka Metode

1. Crawling atau perayapan merupakan suatu proses pengumpulan informasi dari semua halaman website yang terhubung pada situs website tersebut. Jadi dari halaman satu ke halaman lainnya, dari link satu ke link lainnya. Proses ini adalah tahapan awal kerja search engine optimasi.
2. Indexing
Setelah itu proses selanjutnya adalah indexing. Pada dasarnya proses indexing adalah identifikasi dari kata-kata dan tulisan yang menjadikan halaman tersebut masuk ke dalam database search engine dengan keyword tertentu. Jadi setelah proses crawling maka web crawler akan menyimpan informasinya ke dalam database atau pada entry list: index. Maka dari itu Crawling dan Indexing merupakan dua hal yang tidak dapat dipisahkan pada proses SEO(Search Engine Optimization)pada search engine/mesin pencarian.
3. Processing
Pada processing ini dilakukan setelah search engine menerima permintaan dari user contohnya dengan mengetikkan keyword “Artikel” pada mesin pencarian, maka pada search engine atau mesin pencarian tersebut melakukan proses dengan membandingkan keyword “Artikel” ke dalam permintaan pencarian dengan halaman website lainnya dari berbagai situs, yang telah ter-index di database mesin pencari atau search engine.

4. Calculating Relevancy

Pada satu halaman website pasti memiliki lebih jumlah kata kunci yang diinputkan ke dalam mesin pencarian. Contoh saja pada kata kunci atau keyword “Apa itu Artikel”, pada mesin pencari akan melakukan proses dengan menghitung relevansi atau keterkaitan dari setiap halaman website yang telah ter-index untuk setiap kata kunci yang diterima.

5. Retrieving Result

Langkah terakhir adalah mengambil result yang paling relevan dari kata kunci atau keyword yang telah diproses sebelumnya. Kemudian ditampilkan ke dalam daftar hasil mesin pencarian atau search engine yang ada di browser. Nah Hasil nya akan diperingkatkan dan ditampilkan secara berurutan, berdasarkan pada pencarian yang paling sesuai atau paling relevan.

6. Penerapan Information

Penerapan Information mulai dari tahapan awal hingga tahap akhir. Adapun tahap dari penggunaan Information Retrieval dapat di sajikan menggunakan flowchart sebagai berikut:

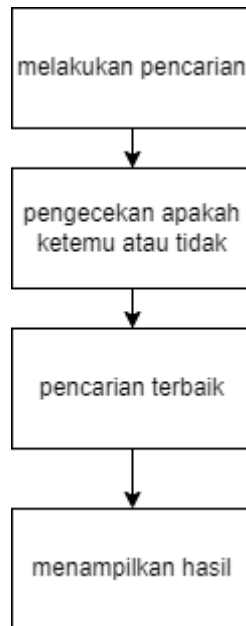


Gambar 2. Flowchart IR

Dari gambar 2, terlihat bahwa di dalam flowchart terdapat proses operasi dalam sistem IR. Proses pertama dimulai dari koleksi dokumen dan proses kedua dimulai dari data pengguna. Proses pertama yaitu pemrosesan terhadap koleksi dokumen menjadi basis data indeks tidak ada ketergantungan dengan proses kedua. Sedangkan proses kedua jika tidak ketemu keberadaan basis data indeks yang dihasilkan akan kembali pada proses pertama.

3. Hasil dan Pembahasan

Adapun gambaran diagram alir sistem search engine menggunakan algoritma Rocchio relevance feedback ini dapat dilihat sebagai berikut:

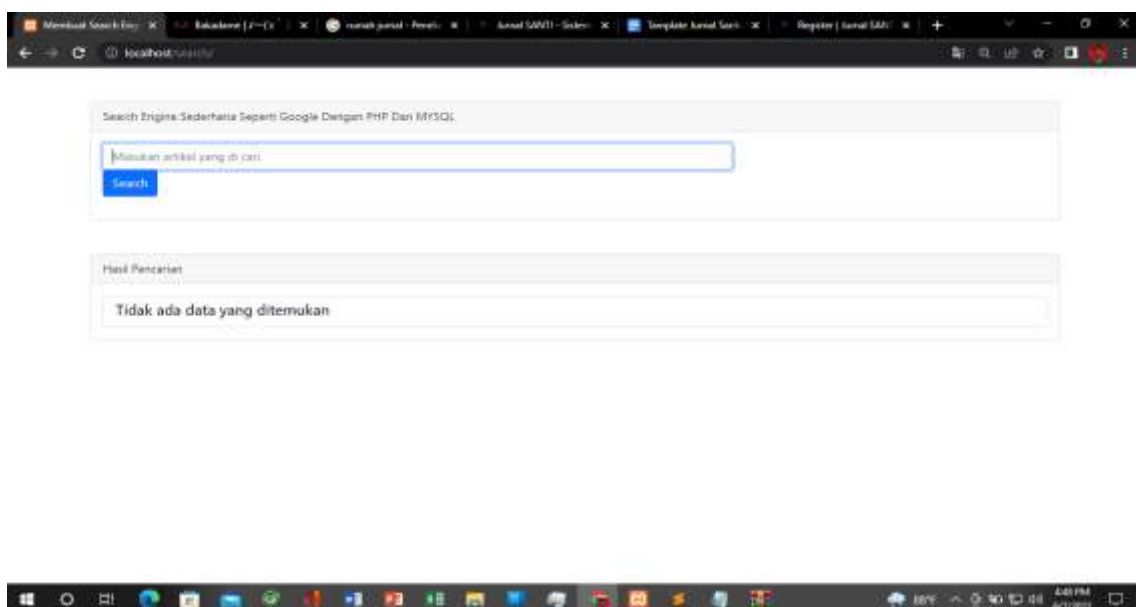


Gambar 3. Diagram Alir Algoritma Rocchio relevance feedback

1. Mencari artikel ini di dalam kolom input pencarian
2. Pada proses ini sistem melakukan pengecekan apakah hasil yang di cari ada atau tidak
3. Dan di hasilkn pencarian terbaik yang telah di inputkan

3.1 Halaman Pencarian Artikel

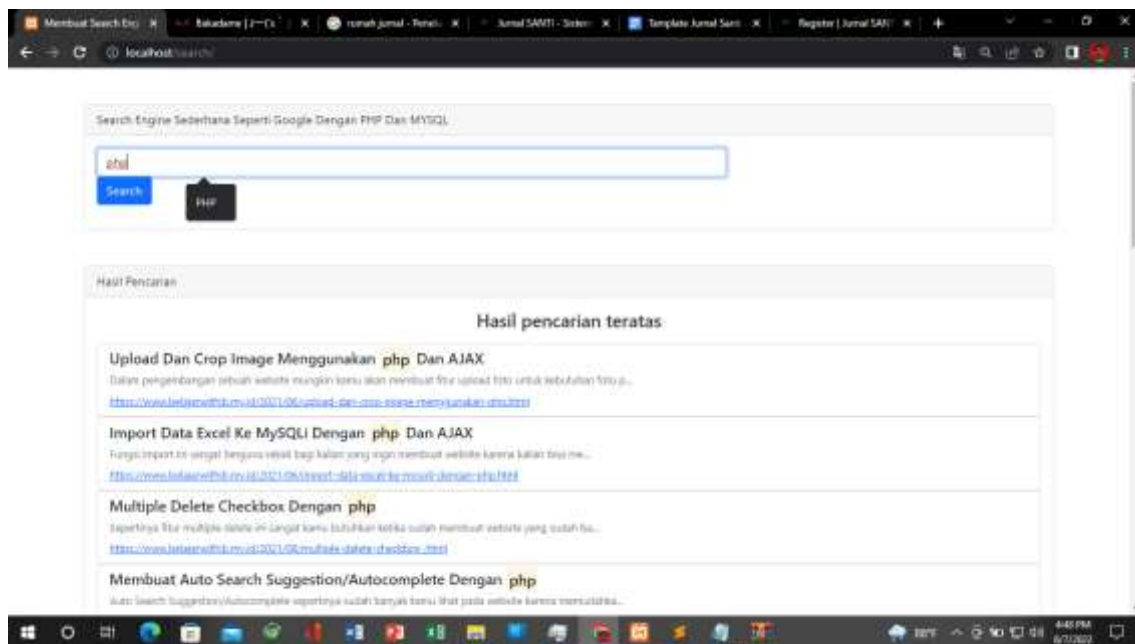
Hasil dari tampilan utama untuk pengunjung untuk melakukan pencarian. Dalam form ini pada gambar 4, user dapat mencari kata yang ada pada koleksi dokumen. Jika menekan tombol cari, maka sistem akan melakukan proses pencarian. Hasilnya akan ditampilkan dalam tabel dan diurutkan berdasarkan nilai kemiripan yang paling tinggi.



Gambar 4. Halaman Utama

3.2 Tampilan output hasil dari pencarian

Pada tampilan ini menghasilkan output yang di cari oleh pengunjung. Dalam form pada gambar 5, user dapat membaca isi dokumen dari hasil pencarian kata. Isi dokumen yang ditampilkan, tidak dapat diubah atau ditambah. Sebelum menggunakan semua fitur yang ada, user diwajibkan untuk memilih sebuah file koleksi dokumen yang mempunyai extension file (.all). File ini berisi kumpulan artikel, yang mempunyai format sama dengan format HTML.



Gambar 5. hasil output

4. Kesimpulan

Dari perancangan dan implementasi yang telah dilakukan, maka dapat dibuat kesimpulan bahwa Penerapan IR pada search engine (mesin pencari) adalah contoh sempurna bagaimana manusia sangat terbantu untuk memperoleh informasi yang dikehendaknya dengan memberi permintaan kata kunci atau bahkan sekumpulan kalimat tertentu. Search engine, berbeda dengan model sistem IR lain sebelumnya, yang cenderung kaku, tidak berkembang, dan kuno, memperlihatkan keunggulannya dalam memberi yang terbaik untuk pengguna. Proses perhitungan kemiripan dokumen dilakukan dengan mengetikkan keyword, keyword bisa terdiri dari satu kata atau lebih. Keyword yang dimasukkan juga akan diproses sama seperti dokumen.

Daftar Pustaka

- [1] I. O. Suzanti and F. A. Mufarroha, "Implementasi Relevant Feedback Menggunakan Algoritma Genetika pada Dokumen Bahasa Indonesia Implementation of Relevant Feedback Using Genetic Algorithm in Indonesian Documents," *J. Ilmu Pengetah. dan Teknol. Komun.*, vol. 23, no. 2, pp. 125–139, 2021.
- [2] Wiyanto, W. Priatna, and J. S. Hidayat, "Implementasi Term Frequency – Inverse Document Frequency (TF-IDF) Dan Vector Space Model (VSM) Untuk Pencarian Berita Bahasa Indonesia," *Pelita Teknol. J. Ilm. Inform. Arsit. dan Lingkung.*, vol. 14, no. November, pp. 119–133, 2019.

- [3] O. A. M. A. H Kara, “濟無No Title No Title No Title,” *Pap. Knowl. . Towar. a Media Hist. Doc.*, vol. 7, no. 2, pp. 107–15, 2014.
- [4] R. Mandala, “Evaluasi Efektifitas Metode Machine Learning Pada Search Engine,” *Keahlian Inform. Sekol. Tek. Elektro dan Inform.*, vol. 2006, no. Snati, pp. 11–15, 2006.
- [5] I. T. Hapsari, B. S. Andoko, and C. Rahmad, “Aplikasi Information Retrieval Untuk Pencarian Dokumen Laporan Penelitian,” *J. Inform. Polinema*, vol. 1, no. 3, p. 23, 2017, doi: 10.33795/jip.v1i3.109.
- [6] N. M. A. Lestari and M. Sudarma, “Perencanaan Search Engine E-commerce dengan Metode Latent Semantic Indexing Berbasis Multiplatform,” *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 8, no. 1, p. 31, 2017, doi: 10.24843/lkjiti.2017.v08.i01.p04.
- [7] A. R. Mulyawan and S. Alfarizi, “Perancangan Model Information Retrieval dengan Metode Cosine Similarity untuk Pencarian dan Persortiran Buku Teks Pelajaran Online,” vol. 7, no. 3, pp. 1–6, 2018.
- [8] T. G. Laksana and E. Syahara, “Analisis Perbandingan Performance Search Engine Berdasarkan 1 - 5 Suku Kata Kunci Menggunakan Strategi Phrase, Multiple Search, Pencarian Field & Penggunaan Symbol,” *Semin. Nas. Sist. Inf. Indones.*, no. November, pp. 191–199, 2018.
- [9] A. Angdresey, M. A. Lamongi, and R. Munir, “Information Retrieval System in the Bible,” *CogITo Smart J.*, vol. 7, no. 1, p. 111, 2021, doi: 10.31154/cogito.v7i1.300.111-120.
- [10] M. B. Kristanda, S. Hansun, and A. Albert, “Rancang Bangun Aplikasi UMN Library Catalog Menggunakan Metode Rocchio Relevance Feedback,” *J. Ultim. InfoSys*, vol. 9, no. 1, pp. 9–17, 2018, doi: 10.31937/si.v9i1.684.
- [11] J. Arguello, “INLS 509 : Introduction to Information Retrieval,” 2013.
- [12] M. Fitri, “Perancangan Sistem Temu Balik Informasi Dengan Metode Pembobotan Kombinasi Tf-Idf Untuk Pencarian Dokumen Berbahasa Indonesia,” *J. Sist. dan Teknol. Inf.*, vol. Vol. 1, no. 1, pp. 1–6, 2013.
- [13] F. Amin, “Implementasi Search Engine (Mesin Pencari) Menggunakan Metode Vector Space Model,” *J. Ilm. Din. Tek.*, vol. 5, no. 1, pp. 45–58, 2011.
- [14] H. Bunyamin, C. P. Negara, F. T. Informasi, and U. K. Maranatha, “Aplikasi Information Retrieval (IR) CATA Dengan Metode Generalized Vector Space Model,” *J. Inform.*, vol. 4, no. 1, pp. 29–38, 2008.
- [15] R. R. Baihaqi, “Temu Kembali Informasi pada Berita Olahraga Berbahasa Indonesia dengan Metode BM25 dan Seleksi Fitur Term Frequency (TF),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 11, pp. 4200–4206, 2020.