Jurnal Informatika dan Teknologi Interaktif
Volume 3, Nomor 1, April 2026

DOI : 10.63547/jiite.v3i1.124
e-ISSN : 3048-1023

Sistem Pencarian Literatur Lintas Bahasa Berbasis API Crossref

Ahmad Syahrudin[1], Devan Regiana[2], Fachri Yanuar Nuriawan[3], Firdaus Adel[4], Muhammad Zaky Ramadhani[5]

[1,2,3,4,5]Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Purwokerto, Indonesia
surel:[1]ahmad.syahrudin03@gmail.com,[2]devanregiana12345@gmail.com,[3]yanuarfachri@gmail.com,[4]firdausadel@gmail.com,[5]mzrramadhani@gmail.com,[6]yanuarfachri@gmail.com

Info Artikel

Sejarah artikel:
Diterima 26-01-2026
Revisi 25-02-2026
Diterima 10-03-2026

Kata kunci:
Cross-Language Information Retrieval
API Crossref
String Similarity
Automatic Query Translation
Manajemen Literatur.

ABSTRAK

Perbedaan bahasa antara kueri pengguna dan dokumen ilmiah sering menjadi kendala dalam pencarian literatur internasional, sehingga menurunkan efektivitas dan efisiensi penelusuran referensi. Penelitian ini bertujuan untuk mengembangkan sistem temu balik informasi lintas bahasa (*Cross-Language Information Retrieval/CLIR*) yang mengintegrasikan *Automatic Query Translation*, API Crossref sebagai sumber metadata, dan algoritma *string similarity* untuk pemeringkatan hasil pencarian. Metode penelitian yang digunakan adalah kuantitatif dengan pendekatan eksperimental. Data diperoleh melalui pengambilan metadata artikel ilmiah dari Crossref, meliputi judul, penulis, tahun publikasi, dan DOI. Kueri uji dirancang dalam Bahasa Indonesia dan Bahasa Inggris untuk mengevaluasi efektivitas penerjemahan otomatis dan relevansi hasil pencarian. Sistem melakukan praproses teks, penerjemahan kueri, perhitungan tingkat kemiripan antara kueri dan judul artikel menggunakan algoritma *Sequence Matcher*, serta pemeringkatan hasil pencarian. Hasil penelitian menunjukkan bahwa integrasi API Crossref dengan algoritma *string similarity* mampu menyajikan daftar jurnal nasional dan internasional secara terstruktur berdasarkan tingkat relevansi, serta meningkatkan efisiensi pencarian literatur dibandingkan pencarian manual. Meskipun demikian, pendekatan berbasis *string similarity* masih memiliki keterbatasan dalam menangkap makna semantik yang kompleks, sehingga diperlukan pengembangan lanjutan untuk meningkatkan akurasi relevansi hasil pencarian.

Penulis yang sesuai:
Ahmad Syahrudin
Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Purwokerto
Email: ahmad.syahrudin03@gmail.com

1. PENDAHULUAN

Aksesibilitas terhadap metadata bibliografi terbuka saat ini menjadi fondasi utama bagi terciptanya ekosistem penelitian yang transparan dan inklusif. Crossref berperan sentral dalam menyediakan metadata terbuka yang memungkinkan analisis bibliometrik dilakukan secara lebih luas, meskipun tantangan besar masih ditemukan pada variasi kelengkapan data di antara berbagai penerbit, terutama untuk jenis publikasi non-jurnal [1].

Untuk mengatasi fragmentasi data tersebut, pengembangan pangkalan data seperti OpenCitations Meta menjadi sangat krusial karena mampu menghubungkan berbagai publikasi melalui identifikasi persisten yang unik.

Sistem ini tidak hanya memfasilitasi integrasi data dari sumber yang berbeda seperti Crossref dan PubMed, tetapi juga menetapkan identitas baru bagi sumber daya yang sebelumnya tidak memiliki pengenal global, guna memperkuat jaring-jaring sitasi ilmiah [2].

Selain infrastruktur data global, peran repositori universitas di tingkat institusional tidak dapat diabaikan dalam mendukung strategi sains terbuka. Pemanfaatan repositori ini memerlukan modernisasi teknologi dan kebijakan yang selaras dengan prinsip-prinsip internasional agar hasil riset lokal dapat terarsip secara sistematis dan mudah dijangkau oleh komunitas global [3]. Dalam operasionalisasi sistem manajemen pengetahuan di institusi pendidikan, efektivitas pencarian informasi sering kali terhambat oleh kesalahan pengetikan kueri pengguna. Penggunaan algoritma Levenshtein distance dalam sistem berbasis web terbukti mampu meningkatkan akurasi hasil pencarian hingga mencapai skor kemiripan yang tinggi, sehingga meminimalkan hambatan teknis dalam akses informasi akademik [4].

Analisis lebih mendalam mengenai kemiripan dokumen juga terus dikembangkan melalui perbandingan berbagai algoritma pencocokan string dan ekstraksi kata. Studi pada dokumen hukum agama seperti Fatwa MUI menunjukkan bahwa kombinasi teknik Cosine Similarity dengan TF-IDF memberikan performa terbaik dalam mengidentifikasi keterkaitan antar dokumen, yang menjadi aspek penting dalam manajemen basis data yang kompleks [5]. Tantangan dalam temu kembali informasi menjadi semakin dinamis ketika melibatkan lingkungan lintas bahasa dan multibahasa. Inisiatif seperti NeuCLIRBench menyediakan koleksi evaluasi modern untuk mengukur kemajuan sistem dalam memproses kueri bahasa Inggris terhadap dokumen dalam bahasa-bahasa yang memiliki struktur kompleks, guna menjembatani kesenjangan linguistik dalam riset global [6]. Untuk mendukung sistem informasi lintas bahasa tersebut, pengembangan Neural Machine Translation (NMT) yang ditingkatkan melalui penggunaan bahasa perantara (pivot) menjadi solusi efektif bagi bahasa-bahasa dengan sumber daya rendah. Pendekatan ini memungkinkan kualitas terjemahan yang lebih baik sehingga informasi ilmiah tetap dapat diproduksi dan dikonsumsi meskipun terdapat batasan data pelatihan bahasa [7].

Meskipun teknologi mesin terjemahan berkembang pesat, interaksi manusia dalam proses post-editing tetap menjadi elemen penting. Riset menunjukkan bahwa bagi penerjemah pemula, bantuan mesin terjemahan sangat membantu dalam mengurangi beban kerja mental dan mempercepat durasi pengerjaan, walaupun pemahaman mendalam terhadap nuansa budaya dan semantik masih tetap memerlukan pengawasan manual yang teliti [8]. Di sisi lain, integrasi teknologi kecerdasan buatan seperti Vision Transformers (ViT) mulai diaplikasikan untuk menyelesaikan masalah penautan rekaman pada dokumen hasil pemindaian (OCR) yang berisik. Dengan mengukur kemiripan karakter secara visual, teknologi ini mampu mengatasi ambiguitas karakter yang sering terjadi pada naskah-naskah kuno atau dokumen dengan kualitas cetak rendah, yang selama ini menjadi kendala dalam digitalisasi literatur [9].

Akhirnya, efektivitas seluruh sistem informasi ini sangat bergantung pada bagaimana antarmuka bibliografi dirancang untuk menarik minat pengguna, khususnya generasi muda. Transformasi dari sistem pencarian pasif menjadi mesin keterlibatan yang aktif melalui desain partisipatif dan metadata visual yang kaya menjadi kunci agar perpustakaan digital tetap relevan bagi peneliti di masa depan [10]. Selain pengembangan koleksi evaluasi, upaya untuk membangun fondasi yang kokoh dalam sistem Cross-Language Information Retrieval (CLIR) dilakukan melalui pengembangan *baseline* saraf (*neural baselines*) yang sederhana namun efektif. Dengan memanfaatkan arsitektur multi-tahap yang melibatkan proses *ranking* awal dan *reranking*, penggunaan model bahasa multibahasa kini memungkinkan proses pencarian dokumen dalam satu bahasa menggunakan kueri dari bahasa lain dengan tingkat keberhasilan yang lebih tinggi. Implementasi kerangka kerja konseptual yang dapat direproduksi ini menjadi sangat penting sebagai standar acuan bagi kemajuan riset di masa depan, terutama dalam menangani dokumen dalam bahasa Persia, Rusia, dan Tionghoa secara lebih akurat [11].

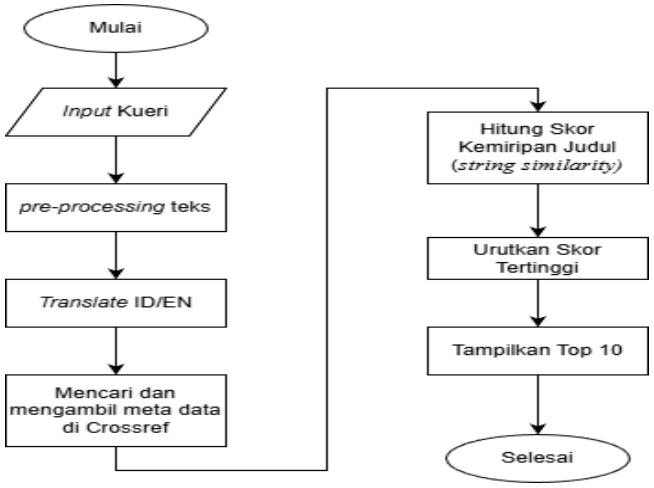
2. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimental dalam lingkup sistem temu balik informasi lintas bahasa (Cross-Language Information Retrieval/CLIR). Sistem dirancang untuk secara otomatis melakukan penerjemahan kueri dan pencocokan relevansi judul kueri menggunakan kerangka kerja evaluasi modern yang mendukung skenario monolingual maupun lintas bahasa [12]. Pendekatan ini dipilih untuk mengukur efektivitas algoritma pencocokan string (string similarity) dalam menyaring basis data jurnal internasional secara bersamaan. Mekanisme ini mengadopsi strategi cascading yang memprioritaskan presisi pada tahap awal melalui pencocokan

identitas unik sebelum beralih ke kriteria yang lebih longgar, termasuk penggunaan pipeline khusus untuk menangani rekonsiliasi referensi dengan metadata yang tidak lengkap [13].

Data penelitian diperoleh melalui metode arsip digital dengan memanfaatkan metadata artikel ilmiah yang tersedia pada layanan Crossref dan OpenCitations melalui API resmi [13]. Data yang dikumpulkan meliputi judul artikel, nama penulis, tahun publikasi, dan Digital Object Identifier (DOI). Pengambilan data dilakukan secara otomatis menggunakan skrip yang terintegrasi dengan API, yang dirancang secara kokoh (robust) dengan menyertakan mekanisme penanganan kesalahan (error handling) serta manajemen batasan permintaan (rate limiting) untuk menjaga stabilitas akses terhadap endpoint penyedia data [14]. Langkah ini memastikan bahwa proses ekstraksi data dalam jumlah besar tetap mematuhi kebijakan penggunaan platform sekaligus menjaga integritas data yang dikumpulkan.

Populasi dalam penelitian ini mencakup seluruh metadata artikel ilmiah yang terindeks dalam layanan Crossref, dengan perhatian khusus pada keragaman bahasa dan kualitas metadata multibahasa yang tersedia [15]. Sampel penelitian dipilih menggunakan teknik purposive sampling, yakni artikel yang diperoleh dari hasil pencarian menggunakan kueri uji lintas bahasa pada topik tertentu. Kueri uji dirancang dalam beberapa bahasa berbeda untuk mengevaluasi efektivitas mekanisme CLIR dan kemampuan sistem dalam menangani ambiguitas linguistik serta perbedaan kualitas metadata antar budaya, sehingga dapat membedakan dokumen yang relevan dari dokumen yang tidak relevan secara akurat [12], [15].



Gambar 1. *FlowChart* Sistem

2.1 Penjelasan Alur Sistem (Flowchart)

Alur kerja sistem pencarian literatur lintas bahasa ini dirancang secara berurutan untuk memastikan data yang diambil relevan dan terstruktur. Berikut adalah penjelasan tiap tahapannya:

1. Input Kueri: Proses dimulai ketika pengguna memasukkan kata kunci (keyword) pencarian dalam Bahasa Indonesia atau Bahasa Inggris ke dalam sistem.
2. Pre-processing Teks: Kueri yang dimasukkan melalui tahap pembersihan awal untuk menghilangkan karakter yang tidak diperlukan agar proses pencocokan lebih akurat.
3. Translate ID/EN: Sistem melakukan deteksi bahasa secara otomatis. Jika kueri dalam Bahasa Indonesia, sistem akan menerjemahkannya ke Bahasa Inggris, dan sebaliknya, menggunakan pustaka machine translation agar pencarian dapat menjangkau literatur internasional dan nasional secara bersamaan.
4. Mencari dan Mengambil Metadata di Crossref: Kueri yang telah diterjemahkan dikirim ke API Crossref untuk menarik metadata artikel ilmiah seperti judul, penulis, tahun, dan DOI.
5. Hitung Skor Kemiripan Judul (String Similarity): Metadata yang didapat kemudian diproses menggunakan algoritma Sequence Matcher. Algoritma ini membandingkan kesamaan karakter antara kueri pengguna dengan judul artikel yang ditemukan untuk menentukan tingkat relevansinya.
6. Urutkan Skor Tertinggi: Sistem melakukan pemeringkatan otomatis berdasarkan persentase kecocokan (match percentage) dari skor yang paling tinggi ke yang terendah.

7. Tampilkan Top 10: Sebagai tahap akhir, sistem menyajikan 10 artikel terbaik untuk masing-masing bahasa (Indonesia dan Inggris) sebagai referensi literatur yang paling relevan bagi pengguna.

Implementasi algoritma yang digunakan memerlukan penetapan parameter teknis yang spesifik guna menjamin stabilitas pengambilan data. Konfigurasi variabel operasional yang diterapkan dalam kode program, mulai dari batasan jumlah permintaan data hingga versi pustaka yang digunakan, dirincikan pada tabel di bawah ini.

Tabel 1. Parameter konfigurasi sistem pencarian jurnal

hal	Indikator	Nilai
1	Limit Request Data Internasional	500
2	Limit Request Data Indonesia	1000
3	Batas Waktu (Timeout) Koneksi	10 detik
4	Batas Penampilan Hasil Teratas Tiap Bahasa	10 artikel
5	Versi Pustaka Penerjemah	4.0.0-rc1

Implementasi sistem ini melibatkan penetapan parameter teknis yang spesifik guna menjamin stabilitas dan efisiensi dalam pengambilan data dari API Crossref. Berdasarkan Tabel 1, sistem dibatasi untuk mengambil maksimal 500 data jurnal internasional dan 1000 data jurnal nasional dalam satu kali proses pencarian. Perbedaan limit ini disesuaikan dengan perkiraan volume data dan kebutuhan relevansi hasil pencarian pada basis data Crossref.

Selain itu, ditetapkan batas waktu (timeout) koneksi selama 10 detik untuk mencegah terjadinya kegagalan sistem akibat respons server yang lambat, sehingga integritas pengambilan data tetap terjaga. Penggunaan pustaka penerjemah versi 4.0.0-rc1 dipilih karena kestabilannya dalam melakukan Automatic Query Translation (AQT), yang menjadi komponen krusial dalam mekanisme Cross-Language Information Retrieval (CLIR) pada penelitian ini.

3. HASIL DAN PEMBAHASAN

Pada bagian ini disajikan hasil implementasi sistem pencarian jurnal lintas bahasa yang dikembangkan, disertai pembahasan mengenai kinerja sistem berdasarkan parameter yang telah ditetapkan. Hasil penelitian dianalisis untuk mengevaluasi efektivitas integrasi Automatic Query Translation, API Crossref, dan algoritma string similarity dalam meningkatkan relevansi pencarian literatur ilmiah.

3.1. Hasil Implementasi Sistem Pencarian Jurnal

Hasil penelitian menunjukkan bahwa sistem pencarian jurnal lintas bahasa yang dikembangkan berhasil diimplementasikan sesuai dengan rancangan metode yang diusulkan. Sistem mampu menerima masukan kata kunci (keyword) dalam Bahasa Indonesia maupun Bahasa Inggris, kemudian secara otomatis mendeteksi bahasa input dan melakukan penerjemahan kueri menggunakan pustaka googletrans.

Setelah proses penerjemahan, sistem mengirimkan kueri ke layanan Crossref API untuk memperoleh metadata artikel ilmiah. Metadata yang diambil meliputi judul artikel, tahun publikasi, serta tautan Digital Object Identifier (DOI). Berdasarkan hasil pengujian, sistem mampu mengambil hingga 500 data jurnal internasional dan 1000 data jurnal nasional sesuai dengan parameter konfigurasi yang telah ditetapkan.

Hasil pencarian kemudian diproses menggunakan algoritma string similarity berbasis Sequence Matcher untuk menghitung tingkat kemiripan antara kueri dengan judul artikel. Setiap artikel diberikan skor relevansi dalam rentang 0 hingga 1, yang selanjutnya dikonversi menjadi persentase kecocokan (match percentage). Sistem menampilkan 10 artikel teratas dengan skor relevansi tertinggi untuk masing-masing bahasa.

3.2. Analisis Pemeringkatan Berdasarkan String Similarity

Pemeringkatan hasil pencarian dilakukan berdasarkan nilai kemiripan string antara kueri dan judul artikel. Algoritma Sequence Matcher digunakan karena efisien dalam membandingkan kesamaan karakter antara teks tanpa memerlukan proses pelatihan model. Pendekatan ini sesuai untuk sistem temu balik informasi berbasis metadata yang menitikberatkan pada presisi pencocokan judul.

Hasil pengujian menunjukkan bahwa artikel dengan susunan kata kunci yang lebih mendekati kueri pengguna memperoleh skor relevansi yang lebih tinggi. Hal ini membuktikan bahwa algoritma string similarity efektif digunakan sebagai mekanisme penyaringan awal (initial filtering) pada sistem pencarian literatur ilmiah lintas bahasa.

Namun demikian, pendekatan berbasis string similarity masih memiliki keterbatasan dalam menangkap makna semantik yang lebih dalam. Artikel dengan topik relevan tetapi menggunakan istilah sinonim cenderung memperoleh skor yang lebih rendah. Oleh karena itu, hasil penelitian ini dapat menjadi dasar pengembangan lanjutan dengan integrasi metode semantic similarity.

3.3. Pembahasan Integrasi CLIR dan API Crossref

Integrasi Automatic Query Translation, Crossref API, dan string similarity memungkinkan sistem untuk melakukan pencarian literatur ilmiah lintas bahasa secara otomatis dan terstruktur. Dibandingkan pencarian manual, sistem ini mampu meningkatkan efisiensi waktu pencarian serta mengurangi beban kognitif pengguna dalam memahami perbedaan bahasa kueri.

Dengan memanfaatkan metadata terbuka dari Crossref, sistem dapat diandalkan sebagai alternatif pencarian referensi akademik yang legal dan terstandarisasi. Penyajian hasil terpisah antara jurnal internasional dan jurnal berbahasa Indonesia juga memberikan fleksibilitas bagi pengguna dalam memilih sumber referensi yang sesuai dengan kebutuhan penelitian.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem pencarian literatur ilmiah lintas bahasa (*Cross-Language Information Retrieval/CLIR*) yang mengintegrasikan *Automatic Query Translation*, API Crossref, dan algoritma *string similarity*. Sistem mampu mengatasi perbedaan bahasa kueri dengan melakukan penerjemahan otomatis serta menyajikan hasil pencarian jurnal nasional dan internasional secara terstruktur berdasarkan tingkat relevansi. Hasil pengujian menunjukkan bahwa algoritma *Sequence Matcher* efektif digunakan untuk pemeringkatan awal berbasis pencocokan judul, sehingga meningkatkan efisiensi pencarian literatur dibandingkan metode manual. Meskipun demikian, pendekatan berbasis *string similarity* masih memiliki keterbatasan dalam menangkap makna semantik, sehingga pengembangan lanjutan dapat dilakukan dengan mengintegrasikan metode *semantic similarity* untuk meningkatkan akurasi hasil pencarian.

REFERENSI

[1] N. Jan van Eck and L. Waltman, “Crossref as a source of open bibliographic metadata.” [Online]. Available: https://tinyurl.com/3zk5nvvf.

[2] A. Massari, F. Mariani, I. Heibi, S. Peroni, and D. Shotton, “OpenCitations Meta.” [Online]. Available: https://opencitations.net/index/croci

[3] M. M. KULYK and A. GLADYŠEVA, “Role of University Repositories in the Formation of Open Science Infrastructure and Strategy at the Institutional Level: Domestic and Foreign Experience (Using the Example of Lithuania),” *University Library at a New Stage of Social Communications Development. Conference Proceedings*, no. 10, pp. 323–332, Dec. 2025, doi: 10.15802/unilib/2025_347706.

[4] Moch. Firmansyah and S. Deswana, “Application of the Levenshtein Algorithm for Optimizing Search Accuracy in a Web-Based Knowledge Management System,” *JUSIFO (Jurnal Sistem Informasi)*, vol. 10, no. 2, pp. 99–106, Dec. 2024, doi: 10.19109/jusifo.v10i2.21951.

[5] M. Fahmi, S. A∗, G. Adiatmaja, and B. Hidayaturrohman, “Calculation of Similarity between MUI Fatwas: A Comparison of Word Extraction and String-Matching Algorithms,” 2025.

[6] D. Lawrie *et al.*, “NeuCLIRBench: A Modern Evaluation Collection for Monolingual, Cross-Language, and Multilingual Information Retrieval,” Nov. 2025, [Online]. Available: http://arxiv.org/abs/2511.14758

[7] D. A. Sulistyo, A. P. Wibawa, D. D. Prasetya, and F. A. Ahda, “An enhanced pivot-based neural machine translation for low-resource languages,” *International Journal of Advances in Intelligent Informatics*, vol. 11, no. 2, pp. 258–274, May 2025, doi: 10.26555/ijain.v11i2.2115.

[8] Y. Yang, R. Liu, X. Qian, and J. Ni, “Performance and perception: machine translation post-editing in Chinese-English news translation by novice translators,” *Humanit. Soc. Sci. Commun.*, vol. 10, no. 1, Dec. 2023, doi: 10.1057/s41599-023-02285-7.

[9] X. Yang, A. Arora, S.-Y. Jheng, and M. Dell, “Quantifying Character Similarity with Vision Transformers,” May 2023, [Online]. Available: http://arxiv.org/abs/2305.14672

[10] G. Nuriddinova and R. Turgunbaev, “Optimizing bibliographic interfaces for youth engagement.”

[11] J. Lin *et al.*, “Simple Yet Effective Neural Ranking and Reranking Baselines for Cross-Lingual Information Retrieval,” Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.01019

[12] J. Mayfield *et al.*, “Synthetic Cross-language Information Retrieval Training Data,” Apr. 2023, [Online]. Available: http://arxiv.org/abs/2305.00331

[13] M. Guenci, “A pipeline for matching bibliographic references with incomplete metadata: experiments with Crossref and OpenCitations.” [Online]. Available: https://www.crossref.org/blog/amendments-to-membership-terms-to-open-reference-

[14] J. R. Harper, “Automated Extraction and Maturity Analysis of Open Source Clinical Informatics Repositories from Scientific Literature.”

[15] D. Donathan Ii, M. Nason, M. Tullney, J. Shi, and J. P. Alperin, “Evaluating Multilingual Metadata Quality in Crossref Funder acknowledgement.”