

Ekspansi Query Berbasis Semantik Pada Online Public Access Catalog

Maskur^{*1}, Nirindra Primavera D. N², Nur Hayatin³

^{1,2,3}Teknik Informatika/Universitas Muhammadiyah Malang

maskur.informatika@gmail.com^{*1}, nirindra.dirga@gmail.com², noorhayatin@gmail.com³

Abstrak

Perpustakaan digital merupakan sistem penting yang berguna sebagai pemenuhan kebutuhan informasi koleksi perpustakaan. Online Public Access Catalog (OPAC) merupakan fitur yang tersedia dalam perpustakaan digital yang gunanya untuk menelusuri koleksi buku suatu perpustakaan. Pada penelitian ini dilakukan pengembangan sistem ekspansi query berbasis semantik. Tujuannya adalah untuk memperbaiki hasil pencarian agar menjadi lebih relevan. Domain yang digunakan yaitu teknologi informasi dimana data buku diperoleh dari situs buku online dan situs penerbit buku sebanyak 500 data. Tahap yang dilakukan adalah membangun ontologi sesuai data yang diperoleh. Pembentukan ontologi mengacu pada data kurikulum Teknik Informatika UMM yang mengacu pada Association for Computing Machinery (ACM). Ontologi diterapkan untuk mencari keterkaitan antara kata kunci yang dimasukkan pengguna. Untuk mendukung hasil pencarian digunakan metode cosine similarity. Cosine similarity merupakan metode untuk menghitung kemiripan antara hasil pencarian dengan kata kunci pengguna sehingga menghasilkan perankingan hasil pencarian. Tahapan pengujian sistem menggunakan precision diperoleh nilai sebesar 65,6. Hal ini menunjukkan bahwa sistem cukup mampu memberikan hasil pencarian yang relevan dengan kata kunci pencarian pengguna.

Kata kunci: OPAC, Ekspansi Query, Ontologi, Cosine Similarity

Abstract

Digital library is an important system useful as the fulfillment of information needs library collection. The Online Public Access Catalog (OPAC) is a feature available in digital libraries that lets you browse library collections. In this research, semantic based query expansion system is developed. The goal is to improve search results to be more relevant. Domain used is information technology where book data obtained from online book site and book publishing sites as much as 500 data. The stage is to build ontology according to the data obtained. The formation of ontology refers to the data of UMM Informatics Engineering curriculum which refers to the Association for Computing Machinery (ACM) Ontology is applied to find the connection between a user entered keyword. To support the search results used cosine similarity method. Cosine similarity is a method to calculate the similarity between search results with user keywords so as to generate search results ranking. Stages of testing the system using precision obtained a value of 65.6. This indicates that the system is capable enough to provide search results that are relevant to a user's search keyword.

Keywords: OPAC, Query Expansion, Ontology, Cosine Similarity

1. Pendahuluan

Perpustakaan digital merupakan perpustakaan yang memiliki koleksi buku yang sebagian besar dengan format digital yang bisa diakses melalui komputer. Perpustakaan ini berbeda dengan perpustakaan yang konvensional yang menyediakan kumpulan buku tercetak, film mikro ataupun kumpulan kaset audio dan video. Pada era yang serba teknologi, keberadaan perpustakaan digital sangat penting dalam pemenuhan kebutuhan informasi pengguna. Pada lingkungan pendidikan layanan ini dapat dirasakan manfaatnya oleh seluruh civitas akademika, karena dengan adanya perpustakaan digital dapat berbagi koleksi digital serta penggunaannya dapat meningkatkan akses elektronik.

Sistem informasi temu balik di perpustakaan digital menjadi unsur yang penting. Sistem temu balik digunakan untuk mempermudah mengakses sumber daya informasi yang tersedia di perpustakaan. Salah satu fitur yang memiliki peran penting dalam akses informasi adalah Online

Access Public Catalog (OPAC). OPAC adalah suatu sistem temu balik informasi berbasis komputer yang digunakan untuk menelusuri koleksi suatu perpustakaan atau unit informasi lainnya yang bisa diakses dimanapun [1].

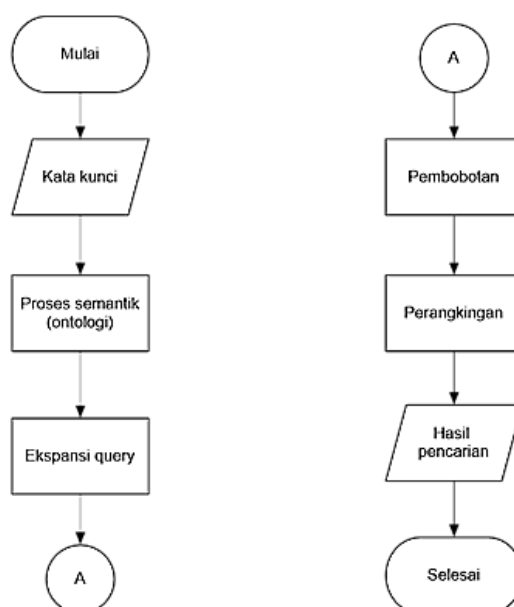
Pada penelitian Maskur dan Faiz Rizky A. (2015) tentang implementasi web semantik untuk aplikasi pencarian tugas akhir menggunakan ontologi dan cosine similarity, dikatakan bahwa pada umumnya sistem pencarian masih dengan cara mencocokkan kata kunci dengan data yang ada sehingga hasil pencarian kurang relevan. Untuk mengatasi hal tersebut, perlu ditambahkan ontologi. Ontologi berfungsi untuk mencari hubungan antara kata yang dimasukkan oleh pengguna. Selain itu ditunjang dengan cosine similarity yang digunakan untuk memberikan bobot pada setiap dokumen yang didapat, sehingga dokumen yang didapat semakin relevan dengan kata kunci [2].

Pada penelitian Adhie Tri Wahyudi dan Anita Indrasari (2012) tentang penerapan semantic web dan semantic search pada digital library online public access catalog (Digilib-OPAC) untuk meningkatkan efektivitas pencarian, dikatakan bahwa kebanyakan OPAC menggunakan teknik querying berdasarkan kata kunci sehingga pencarian kurang efektif dan tidak relevan. Untuk meningkatkan relevansi hasil pencarian maka teknologi semantic web dan semantic search diperlukan pada sistem pencarian. Hasil pengujian sistem pencarian diperoleh nilai precision dan recall yang sama besar yaitu 1:1. Hal ini memperlihatkan bahwa sistem memberikan hasil pencarian yang efektif atau lebih relevan [3]. Hasil yang di ketahui dari penelitian ini adalah relevan dan tidak relevan hasil pencarian. Penelitian ini belum dilengkapi pembobotan dan perangkaan yang mendukung proporsi jumlah dokumen yang ditemukan dan relevan dengan kebutuhan pengguna

Berdasarkan uraian diatas, maka akan dilakukan ekspansi query berbasis semantik pada *Online Public Access Catalog* (OPAC). Adanya implementasi semantik diharapkan dapat memperbaiki hasil pencarian pada OPAC menjadi lebih relevan.

2. Metode Penelitian

Pengembangan ekspansi *query* berbasis semantik pada *Online Public Access Catalog* (OPAC), menggunakan ontologi sebagai basis pengetahuan semantik. Rancangan ontologi mengacu pada data kurikulum *Association for Computing Machinery* (ACM). Langkah yang diterapkan yaitu mencari keterkaitan kata kunci yang dimasukkan pengguna pada ontologi. Setelah itu digunakan metode TF-IDF untuk melakukan *indexing* terhadap deskripsi buku. Deskripsi buku yang telah dikumpulkan terlebih dahulu dilakukan proses *pre-processing*. metode *cosine similarity* yang digunakan untuk menghitung kemiripan hasil pencarian dengan kata kunci yang dimasukkan pengguna. Sistem akan memberikan hasil pencarian buku yang relevan. Seberapa baik sistem memberikan hasil terlihat dari nilai presisi sistem.



Gambar 1. Alur Proses Ekspansi Query Berbasis Semantik Pada OPAC

Secara umum proses ekspansi query berbasis semantic pada OPAC digambarkan pada Gambar 1. Proses dimulai dengan pengguna memasukkan kata kunci pencarian buku. Selanjutnya dilakukan preprocessing, setelah didapat hasil preprocessing maka kata kunci akan dicari perluasan query pada ontologi. Setelah diperoleh hasil perluasan query, dilakukan pembobotan pada setiap dokumen menggunakan metode TF-IDF. Kemudian proses dilanjutkan dengan menghitung kemiripan tiap dokumen untuk menentukan perankingan hasil pencarian. Dimana perhitungan kemiripan menggunakan cosine similarity. Setelah diperoleh data buku maka akan ditampilkan ke pengguna.

2.1 Ekspansi Query

Menyusun ulang query (*Query Reformulation*) yang dimasukkan pengguna adalah hal yang sering dilakukan pada *information retrieval* [4]. *Query reformulation* yang sering digunakan adalah ekspansi query, yaitu dengan memanjangkan query pencarian dengan menambahkan beberapa term kedalamnya. Pada sistem ini penambahan term berdasarkan hasil pencarian pada ontologi. Query yang dimasukkan oleh pengguna pada umumnya pendek sehingga ekspansi query dapat memperkaya atau memperluas term yang nantinya akan dicocokkan dengan term pada *indexing* [5].

2.2 Semantik

Dalam kamus linguistik semantik diartikan sebagai bagian dari struktur bahasa yang berhubungan dengan makna ungkapan dan juga dengan struktur makna. Definisi kedua, semantik adalah sistem dan penyelidikan makna dan arti dalam suatu bahasa. Pengertian semantik dalam KBBI adalah ilmu tentang makna kata dan kalimat. Pengetahuan mengenai seluk-beluk dan pergeseran arti kata [6].

Secara singkat dan populer dapat dikatakan semantik adalah telaah mengenai makna. Berdasarkan definisi-definisi semantik yang telah dipaparkan, disimpulkan bahwa semantik adalah cabang linguistik yang mengkaji makna dan hubungan antar kata.

2.3 Ontologi

Istilah ontologi berasal dari kajian filsafat yang mempelajari tentang segala sesuatu yang nyata serta bagaimana mereka berhubungan. Menurut Gruber, ontologi merupakan spesifikasi baku dan jelas dari sebuah konseptualisasi [7]. Kemudian pada dunia komputer, ontologi digunakan untuk merepresentasikan istilah-istilah yang ada dalam sebuah domain pengetahuan serta hubungan antar istilah-istilah tersebut [8]. Ontologi merupakan teori hubungan makna, *property* objek, serta relasi antar objek. Ontologi banyak diterapkan pada semantik karena kemampuannya memahami makna dari suatu kata. Representasi ontologi berupa kumpulan konsep yang saling terhubung sehingga membentuk graf. Konsep disini dinyatakan dengan frasa bahasa Indonesia.

2.4 TF-IDF

Metode TF-IDF merupakan metode paling umum pada yang digunakan untuk menghitung bobot setiap kata pada dokumen. Pembobotan berpengaruh dalam menentukan kemiripan antara dokumen dengan *query*. Metode ini juga terkenal efisien, mudah dan memiliki hasil akurat. Metode ini akan menghitung frekuensi sebuah kata dalam dokumen (*TF*) dan *inverse frekuensi* dokumen yang mengandung kata tersebut (*IDF*) [9] dengan Persamaan 1.

$$W_{dt} = tf_{dt} * IDF_t \quad (1)$$

Dimana:

d : dokumen ke-d

t : kata ke-t dari kata kunci

W : bobot dokumen ke-d terhadap kata ke-t

tf : banyaknya kata yang dicari pada sebuah dokumen

IDF : Inverse frekuensi dokumen $\log_{10} \left(\frac{n}{df} \right)$

n : jumlah dokumen

df : banyak dokumen yang mengandung kata yang dicari

2.5 Cosine Similarity

Cosine Similarity adalah salah satu metode yang diterapkan untuk membandingkan dokumen. *Cosine similarity* menggunakan dua parameter vector. Vektor tersebut didasarkan pada jumlah kemiripan kata pada dokumen. Persamaan *cosine similarity* ditunjukkan pada Persamaan 2 [10].

$$Sim(D, D_i) = \cos(\theta) = \frac{\sum w_{q,j} w_{i,j}}{\sqrt{\sum w_{q,j}^2} \sqrt{\sum w_{i,j}^2}} \tag{2}$$

Dimana:

D : dokumen acuan

D_i : dokumen ke-i

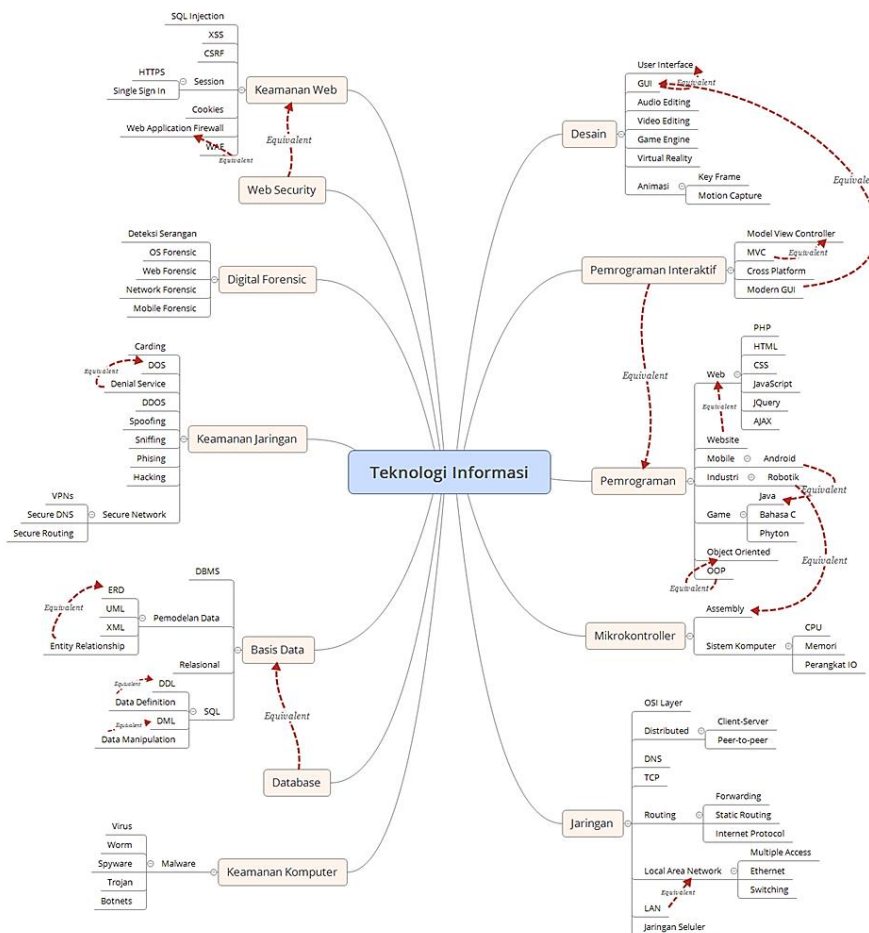
W_{q,j} : bobot *term* j pada dokumen acuan

W_{i,j} : bobot *term* j pada dokumen i

2.6 Data Penelitian

Data penelitian yang digunakan adalah koleksi buku mengenai teknologi informasi berjumlah 500 buku. Data buku yang digunakan dalam Tabel 1 adalah deskripsi dari masing-masing buku. Data tersebut diperoleh dari situs toko buku online dan situs penerbit buku.

Berdasarkan data buku yang diperoleh berkaitan dengan teknologi informasi maka selanjutnya dirancang ontologi dengan domain teknologi informasi pada Gambar 2. Perancangan ontologi terdiri dari beberapa tahap yaitu menentukan domain berupa topik data buku, mendefinisikan *class* ontologi, menyusun *subclass* dan mendefinisikan *class* yang saing *equivalent*. Beberapa *class* yang diidentifikasi yaitu basis data, desain, pemrograman dan *class* lainnya.



Gambar 2. Class Hierarchy Ontologi

Tabel 1. Data Buku

| Judul | Deskripsi |
|-----------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Konsep dan Implementasi Pemrograman Laravel 5 | Laravel adalah framework PHP yang paling populer saat ini. Laravel memiliki banyak fitur modern yang sangat membantu dalam proses pembuatan dan pengembangan web. Laravel cukup sulit untuk dikuasai, terutama bagi pemula. Karena kompleksitas yang ada didalamnya. Buku ini membantu anda mempelajari Laravel dengan cara yang mudah dan menyeluruh. |
| Administrasi SQL Server 2014 | SQL Server 2014 adalah rilis terbaru dan paling signifikan dari SQL Server sampai saat ini. Buku ini akan mengekspos beberapa fitur kunci dari SQL Server 2014 dari perspektif administrasi database. Pembahasan dalam buku mencakup- Instalasi dan Konfigurasi SQL Server 2014- Kakas dan Administrasi- Administrasi Data- Sekuriti- Fitur Baru- Pemulihan dari Musibah- Availabilitas Tinggi- Database Mirroring- Monitoring dan Optimizing. Tiap bab disertai latihan dengan petunjuk langkah demi langkah. Konsep yang sulit dijelaskan dengan kata-kata yang sederhana untuk memudahkan Anda memahaminya. |

3. Hasil Penelitian dan Pembahasan

Tabel 2. Hasil Pengujian Akurasi

| Kata kunci | Sebelum Pengembangan | | Sesudah Pengembangan | | Precision |
|------------------------------------|----------------------|------------------------|----------------------|------------------------|-----------|
| | Data ditemukan | Waktu eksekusi (detik) | Data ditemukan | Waktu eksekusi (detik) | |
| pemrograman | 61 | 0,05555 | 320 | 0,23113 | 31,3 |
| hacking | 15 | 0,01637 | 99 | 0,11936 | 63,6 |
| animasi | 4 | 0,01816 | 95 | 0,17616 | 37,9 |
| robotik | 1 | 0,01205 | 12 | 0,11248 | 75 |
| sql | 39 | 0,01261 | 127 | 0,13937 | 18,1 |
| arduino | 10 | 0,03265 | 10 | 0,20366 | 100 |
| internet | 6 | 0,01325 | 75 | 0,14038 | 74,7 |
| laravel | 3 | 0,01688 | 3 | 0,14239 | 100 |
| android | 19 | 0,0155 | 37 | 0,11545 | 100 |
| spyware | 0 | 0,0271 | 1 | 0,28493 | 100 |
| uml diagram | 0 | 0,02632 | 130 | 0,14333 | 22,3 |
| network forensic | 0 | 0,01661 | 52 | 0,14189 | 86,5 |
| malware komputer | 0 | 0,01199 | 149 | 0,14087 | 18,8 |
| serangan dos | 1 | 0,01209 | 100 | 0,17862 | 59 |
| motion capture | 0 | 0,01419 | 28 | 0,13076 | 75 |
| session https | 0 | 0,12035 | 7 | 0,13405 | 100 |
| cloud computing | 3 | 0,11162 | 8 | 0,103 | 100 |
| cms website | 1 | 0,01332 | 224 | 0,28271 | 50,9 |
| seo blog | 1 | 0,02165 | 26 | 0,13442 | 96,2 |
| video editing | 3 | 0,01552 | 105 | 0,12934 | 75,2 |
| php untuk pemula | 4 | 0,0753 | 175 | 0,15047 | 55,4 |
| css adobe | 0 | 0,01639 | 135 | 0,11768 | 57 |
| dreamweaver | | | | | |
| multithreading pada mikrokontroler | 1 | 0,01451 | 261 | 0,16071 | 32,2 |
| desain user interface | 0 | 0,01779 | 128 | 0,27008 | 64,8 |
| jaringan multiple access | 0 | 0,01391 | 101 | 0,14455 | 50,5 |
| mikrotik router os | 53 | 0,01906 | 21 | 0,32276 | 76,2 |
| basis data oracle | 3 | 0,02463 | 243 | 0,19896 | 32,5 |
| jago adobe photoshop | 0 | 0,02184 | 46 | 0,11723 | 97,8 |
| sistem operasi windows | 2 | 0,01994 | 171 | 0,14679 | 59,1 |
| menguasai microsoft access | 0 | 0,02022 | 105 | 0,1397 | 56,2 |
| | | Rata-rata | | | 65,6 |

Pengujian hasil pencarian buku dilakukan dengan menggunakan 500 data buku dimana ontologi pada penelitian digunakan sebagai acuan ekspansi query sedangkan *cosine similarity* digunakan untuk perangkian hasil pencarian buku. Selanjutnya dilakukan analisa hasil pencarian menggunakan *precision*. Pengujian presisi dilakukan untuk mengetahui tingkat ketepatan atau kecocokan sistem dalam memberikan hasil terhadap kata kunci. Skenario pengujian dilakukan dengan 3 variasi kata kunci, yaitu 1 kata kunci, 2 kata kunci dan 3 kata kunci, dimana setiap variasi menggunakan 10 kata kunci yang berbeda.

Berdasarkan hasil pengujian pada Tabel 2 dapat dilihat adanya peningkatan hasil pencarian pada sistem OPAC yang telah dikembangkan. Peningkatan hasil terjadi karena kata kunci telah mengalami ekspansi *query* dan pencarian dilakukan pada deskripsi buku. Sedangkan pencarian pada OPAC sebelumnya hanya dilakukan pada judul dan kategori buku. Selain hasil pencarian, peningkatan juga terjadi pada waktu eksekusi *query*. Waktu eksekusi meningkat karena proses pencarian melalui tahap semantik terlebih dahulu.

Menurut hasil pengujian, nilai *precision* tertinggi diperoleh dengan kata kunci laravel, android, spyware, session https, dan cloud computing dengan nilai *precision* 100 sedangkan nilai *precision* terendah diperoleh dengan kata kunci panduan SQL, yaitu 18,1. Rata-rata nilai *precision* secara keseluruhan hasil pengujian adalah 65,6. Nilai *precision* tersebut menunjukkan bahwa sistem cukup mampu memberikan hasil pencarian yang relevan dengan kata kunci yang diberikan pengguna.

4. Kesimpulan

Berdasarkan penelitian yang dilakukan mengenai pengembangan ekspansi query berbasis semantik pada Online Public Access Catalog (OPAC), dapat disimpulkan bahwa :

1. Setelah dilakukan percobaan menggunakan 30 kata kunci dengan 3 variasi pengujian yang meliputi 1 kata kunci, 2 kata kunci dan 3 kata kunci dapat disimpulkan bahwa terjadi peningkatan hasil pencarian pada sistem OPAC setelah dilakukan pengembangan ekspansi *query*.
2. Waktu eksekusi *query* pada sistem OPAC setelah pengembangan mengalami peningkatan. Hal ini terjadi karena proses pencarian terlebih dahulu melalui proses semantik.
3. Pengembangan sistem berupa ekspansi *query* berbasis semantik pada *Online Public Access Catalog* (OPAC) dengan pencarian dilakukan pada deskripsi buku mampu memberikan hasil pencarian yang lebih relevan.
4. Nilai presisi sistem dalam memberikan hasil pencarian sebesar 65,6. Hasil ini menunjukkan bahwa sistem cukup mampu memberikan hasil pencarian yang relevan sesuai informasi yang dibutuhkan pengguna.

Referensi

- [1] G. P. Putra, "Pengembangan Aplikasi Pencarian Pada OPAC (Online Public Access Catalogue) Perpustakaan Menggunakan J2ME dan Bluetooth," 2007.
- [2] Maskur and F. R. Andriansyah, "Implementasi Web Semantik Untuk Aplikasi Pencarian Tugas Akhir Menggunakan Ontologi dan Cosine Similarity," *J. Ilm. NERO*, vol. 2, no. 1, pp. 11–18, 2015.
- [3] A. T. Wahyudi and A. Indrasari, "Penerapan Semantic Web dan Semantic Search Pada Digital Library Online Public Access Catalog (DIGILIB-OPAC) Untuk Meningkatkan Efektivitas Pencarian," vol. 1, no. 1, pp. 14–22, 2012.
- [4] B. Poernomo and I. Gunawan, "Sistem Information Retrieval Pencarian Kesamaan Ayat Terjemahan Al Quran Berbahasa Indonesia Dengan Query Expansion Dari Tafsirnya," *Semin. Nas. "Inovasi dalam Desain dan Teknol."*, pp. 100–108, 2015.
- [5] B. P. Trihadnanto, "Implementasi Algoritma Naïve Bayes untuk Klasifikasi Opini pada Data Twitter dengan Ekspansi Query Menggunakan Pendekatan Sinonim," Universitas Muhammadiyah Malang, 2014.
- [6] N. Wati, "Pengukuran Kesamaan Semantik Antara Kata Berdasarkan Hasil Pencarian di Web Search Engine," Universitas Muhammadiyah Malang, 2013.
- [7] N. F. Ariyani, A. Y. Priyanto, Sarwosri, and R. Sarno, "Pemodelan Granularitas Temporal Untuk Mencari Relasi Antar Objek Warisan Budaya Indonesia Dengan Menggunakan Ontologi," *JUTI J. Ilm. Teknol. Inf.*, vol. 15, pp. 72–83, 2017.
- [8] J. Zebua and M. Mustikasari, "Aplikasi Pencarian Buku Berbasis Web Semantik Untuk Perpustakaan SMK Yadika7 Bogor," Universitas Gunadarma.

- [9] M. Nurjannah, H. Hamdani, and I. F. Astuti, "Penerapan Algoritma Term Frequency-Inverse Document Frequency (Tf-Idf) Untuk Text Mining," vol. 8, no. 3, pp. 111–114, 2013.
- [10] A. Aziz, R. Saptono, and K. P. Suryajaya, "Implementasi Vector Space Model dalam Pembangkitan Frequently Asked Questions Otomatis dan Solusi yang Relevan untuk Keluhan Pelanggan," vol. 2, no. 2, pp. 111–122, 2015.

