Aplikasi Pencarian Hadis Berbahasa Indonesia Menggunakan Metode Jaccard Similarity

Septiyawan Rosetya Wardhana[1], Ilyasa Nanda Rahmadianto[2], Gusti Eka Yuliastuti[3]
Institut Teknologi Adhi Tama Surabaya
e-mail: rossywardhana@itats.ac.id, ilyasananda16@gmail.com, gustiekay@itats.ac.id

ABSTRACT

Hadiths are all the words, actions, agreements (taqrir), characteristics, good behavior, and life journey of a person, both before being appointed as a messenger and after. Along with the rapid development of technology today, many applications have been created to support the learning of the Al-Quran and Hadith, which are often referred to as digital Al-Quran and Hadith, but there are still many digital Al-Quran and Hadith learning applications that have been created, so that existing hadith search applications often require devices with special specifications that are not owned by most people. This study aims to develop a web-based Indonesian-language hadith search application using the Jaccard Similarity method, which measures the degree of similarity between search keywords and hadith texts. The dataset used includes 658 hadith from the books of Sahih Al-Bukhari, Sahih Muslim, Sunan Ibn Majah, and Jami' At-Tirmidzi. The implementation of this method allows for a more efficient and accurate hadith search. The test results show that this application has an average precision for K=10 is 0.76 for single query and 0.57 for multi-query. Meanwhile, the average precision for K=20 is 0.52 for single query and 0.45 for multi-query. The decrease in precision from K@10 to K@20 shows more relevant search results at K@10. This application is expected to facilitate the public in finding and verifying hadith, as well as being a useful tool in learning and understanding hadith.

Kata kunci: *Hadith, Search System, Jaccard Similarity, Precision @K.*

ABSTRAK

Hadis merupakan semua kata, tindakan, persetujuan (taqrir), sifat, perilaku baik, dan perjalanan hidup seseorang, baik sebelum diangkat sebagai rasul maupun setelahnya. Seiring dengan perkembangan teknologi yang begitu cepat saat ini, banyak aplikasi yang telah diciptakan untuk mendukung pembelajaran Al-Quran dan Hadis, yang sering disebut sebagai Al-Quran dan Hadis digital, namun masih banyak aplikasi pembelajaran Al-Quran dan Hadis digital yang telah diciptakan, sehingga aplikasi pencarian hadis yang ada sering memerlukan perangkat dengan spesifikasi khusus yang tidak dimiliki oleh sebagian besar masyarakat. Penelitian ini bertujuan untuk mengembangkan aplikasi pencarian hadis berbahasa Indonesia berbasis website menggunakan metode Jaccard Similarity, yang mengukur tingkat kemiripan antara kata kunci pencarian dengan teks hadis. Dataset yang digunakan mencakup 658 hadis dari kitab Shahih Al-Bukhari, Shahih Muslim, Sunan Ibnu Majah, dan Jami' At-Tirmidzi. Implementasi metode ini memungkinkan pencarian hadis yang lebih efisien dan akurat. Hasil pengujian menunjukkan bahwa aplikasi ini memiliki rata-rata precision untuk K=10 adalah 0.76 untuk single query dan 0.57 untuk multi query. Sementara itu, rata-rata precision untuk K=20 adalah 0.52 untuk single query dan 0.45 untuk multi query. Penurunan precision dari K@10 ke K@20 menunjukkan hasil pencarian yang lebih relevan pada K@10. Aplikasi ini diharapkan dapat memudahkan masyarakat dalam menemukan dan memverifikasi hadis, serta menjadi alat bantu yang bermanfaat dalam mempelajari dan memahami hadis.

Kata kunci: Hadis, Sistem Pencarian, Jaccard Similarity, Precision @K.

PENDAHULUAN

Seiring dengan perkembangan teknologi yang cepat saat ini, banyak aplikasi yang telah diciptakan untuk mendukung pembelajaran Al-Quran dan Hadis, yang sering disebut sebagai Al-Quran dan Hadis digital, sehingga mempermudah proses pembelajarannya. Hadis, sebagai bahan

pelajaran dalam Islam, dianggap sebagai sumber kedua setelah al-Qur'an. Pengertian hadis umumnya merujuk pada segala hal yang diatribusikan kepada Nabi Muhammad SAW., termasuk kata-kata, tindakan, persetujuan, serta ciri-ciri fisik dan mentalnya, yang terjadi baik sebelum maupun sesudah masa kenabiannya[1].

Masalah pencarian hadis melalui aplikasi (hadis digital) yang ada, seringkali membutuhkan perangkat dengan spesifikasi khusus yang mungkin tidak dimiliki oleh sebagian besar masyarakat. Disisi lain menjadi perhatian penting di era digital saat ini, dimana masih ada masyarakat yang mencari hadis secara manual dan pastinya menghabiskan waktu yang lama untuk menemukan hadis yang relevan yang akan dicarinya. Oleh karena itu, membangun sistem pencarian hadis melalui website menjadi alternatif yang lebih inklusif. Sistem pecarian hadis memiliki peran penting dalam mencegah penyebaran informasi palsu dan memastikan ketersediaan hadis-hadis yang valid dan terverifikasi. Dengan kemudahan akses ke sumber-sumber terpercaya, masyarakat dapat melakukan verifikasi keaslian hadis-hadis yang ditemukan, menjaga integritas dan keakuratan pengetahuan agama di era digital yang penuh dengan informasi yang belum tentu benar.

Sistem pencarian informasi merupakan alat buatan manusia yang menyatukan komponen seperti orang, perangkat keras, perangkat lunak, jaringan komputer dan komunikasi data, serta basis data. Fungsinya adalah untuk mengumpulkan, memproses, dan menyajikan informasi yang berguna bagi pengguna dalam mendukung operasi organisasi dan pengambilan keputusan[2].

Sistem pencarian hadis yang akan dibangun nantinya menggunakan metode Jaccard Similarity. Metode Jaccard Similarity menggunakan kesamaan himpunan elemen untuk mengevaluasi tingkat kemiripan antara dua set data. Kelebihan utama metode Jaccard Similarity terletak pada kemampuannya untuk mengidentifikasi kesamaan berdasarkan elemen-elemen dalam dua set, tanpa terlalu bergantung pada kompleksitas algoritma seperti pada mesin pencari konvensional. Dalam konteks pemrosesan data besar dan pencarian informasi yang lebih spesifik, metode Jaccard Similarity dianggap lebih efektif karena fokus pada kesamaan elemen. Ini membantu menghasilkan hasil pencarian yang lebih akurat dan spesifik.

Oleh karena itu dalam penelitian ini diusulkan implementasi Metode Jaccard Similarity untuk mengembangkan aplikasi pencarian hadis Bahasa Indonesia berbasis website. Pencarian ini dilakukan dengan cara mengukur sejauh mana kemiripan query dengan dokumen yang terkait. Manfaat dari penelitian ini adalah untuk membatu memudahkan akses masyarakat menemukan hadis-hadis yang relevan dengan topik yang akan di cari, sehingga nantinya dapat dengan lebih baik memverifikasi keaslian hadis-hadis yang mereka temukan dan pengetahuan agama dapat tersebar secara lebih merata di seluruh masyarakat.

TINJAUAN PUSTAKA

Hadis

Hadis mencakup semua kata-kata, tindakan, persetujuan (taqrir), sifat, perilaku baik, dan perjalanan hidup seseorang, baik sebelum diangkat sebagai rasul maupun setelahnya[3].

Kitab hadis adalah kumpulan informasi dari beberapa hadis yang digabungkan dalam satu kitab atau buku. Sebagai karya tulis, hadis sebagai buku berfokus pada merekam berita mengenai perkataan, tindakan, dan sikap Nabi Muhammad SAW sebagai Rasul. Fungsi utama hadis adalah sebagai penguat dan penjelas terhadap teks atau hukum islam yang masih bersifat umum dan belum dibahas secara rinci dalam Al-Qur'an. Sebagai pelengkap Al-Qur'an, kitab-kitab hadis menjadi rujukan penting bagi umat islam dalam mengikuti syari'at islam. Kitab-kitab hadis yang disusun oleh ulama hadis memiliki ragam ciri, mulai dari sistematika penulisan, kualitas hadis, hingga isi pembahasan hadis itu sendiri. Contoh kitab hadis terkenal pada abad ke-3 hijriyah adalah Al-Jami' al-Shahih Bukhari dan Al-Jami' al-Shahih Muslim, sedangkan kitab

As-Sunan seperti Al-Sunan At-Tirmudzī dan Al-Sunan Ibnu Majah juga turut memberikan kontribusi pada perkembangan hadis[4].

Mesin Pencarian

Mesin Pencarian (Search Engine) bertujuan untuk memahami cara pengguna memanfaatkannya dan memberikan jawaban terbaik dari berbagai teks dalam koleksinya sesuai dengan pencarian pengguna (query). Dengan demikian, mesin pencarian dapat menyajikan jawaban yang paling sesuai[5]. Sistem pencarian menggunakan teknik information retrieval yang membandingkan dokumen dengan query pengguna dan menampilkan daftar hasil yang diurutkan berdasarkan relevansinya. Tingkat keakuratan dan kecepatan dalam pencarian sangat dipengaruhi oleh metode atau algoritma pencarian yang digunakan.

Text Mining

Text mining merupakan proses ekstraksi informasi yang relevan dari teks dengan tingkat kualitas yang tinggi. Kualitas informasi tersebut umumnya dapatdicapai dengan memerhatikan pola dan tren yang terbentuk melalui analisis statistik. Dalam proses text mining, terjadi pengembangan nilai atau bobot yang diberikan pada kata-kata yang terdapat dalam dokumen tertentu. Penentuan bobot ini bergantung pada metode yang digunakan dalam proses tersebut[6].

Text Preprocessing

Preprocessing merupakan langkah awal dalam klasifikasi teks yang bertujuan mempersiapkan data teks sebelum dimasukkan ke proses berikutnya. Proses ini memiliki dampak positif yang signifikan terhadap peningkatan kinerja sistem. Pada tahap ini, data teks diubah menjadi bentuk yang lebih optimal, menghasilkan informasi teks yang berkualitas dan siap digunakan dalam tahapan berikutnya[7]. Beberapa langkah umum yang sering dilakukan dalam proses *text preprocessing* meliputi: *case folding, tokenizing, stopwords,* dan *stemming.*

Jaccard Similarity

Jaccard Similarity merupakan suatu algoritma dengan tujuan membandingkan dokumen dan mengukur nilai kesamaan (similarity) antara dua objek atau dokumen. Kesamaan Jaccard didefinisikan sebagai ukuran irisan dibagi oleh ukuran gabungan dari himpunan sampel[8]. Nilai similaritas dokumen yang tertinggi dapat dianggap bahwa dokumen tersebut paling similar, atau memiliki banyak kesamaan[9].

Jaccard Similarity atau Koefisien Jaccard adalah suatu algoritma yang digunakan untuk membandingkan dua sampel, seperti dokumen, berdasarkan kata-kata yang terdapat di dalamnya. Fungsi utama Jaccard Similarity adalah untuk mengevaluasi sejauh mana dua dokumen mirip satu sama lain dan menghitung nilai kemiripan antara keduanya[10]. Jaccard similarity dapat dirumuskan sebagai berikut:

$$Similarity\ (X,Y) = \frac{|X \cap Y|}{|X \cup Y|} \ldots\ldots (1)$$

Dalam rumus diatas dijelaskan, X dan Y mewakili dua dokumen berbeda, di mana X adalah himpunan elemen (misalnya, kata-kata unik atau fitur) dalam Dokumen 1, dan Y adalah himpunan elemen dalam Dokumen 2. Notasi |X∩Y| menggambarkan jumlah elemen yang ada di kedua himpunan, sementara |X∪Y| menunjukkan jumlah total elemen unik dalam kedua himpunan.

Keterangan :
X = Dokumen 1
Y = Dokumen 2

Precision @K

Precision @K adalah metode untuk mengevaluasi seberapa baik sistem pencarian dalam menyediakan hasil yang relevan dengan preferensi pengguna di dalam K item teratas dari daftar hasil pencarian. Evaluasi ini membantu dalam memahami sejauh mana sistem dapat memenuhi kebutuhan pengguna dalam hal relevansi hasil pencarian[11].

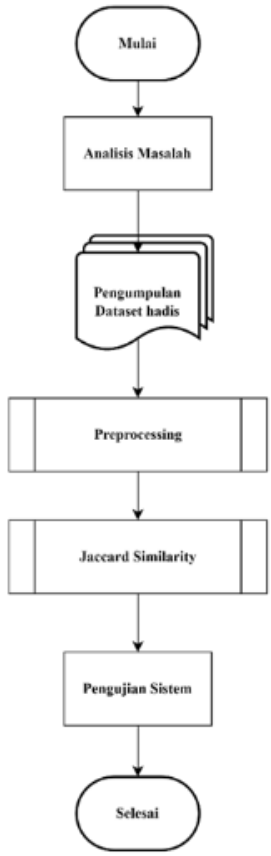
Precision @K digunakan untuk mengukur keakuratan sistem. Evaluasi dilakukan dengan mengevaluasi sejauh mana dokumen yang dikembalikan oleh sistem sesuai (relevant) dengan query pengguna, dengan mengambil sampel dari hasil teratas seperti top-k[12].

$$Precision@K = \frac{Jumlah\ item\ relevan\ pada\ @K}{@K} \text{ ..... (2)}$$

METODE

Gambaran Umum

Pada penelitian ini, dibuat sebuah aplikasi pencarian hadis Bahasa Indonesia berbasis website. yang nantinya akan digunakan oleh pengguna untuk mendapatkan hadis-hadis yang relevan dan sesuai dengan yang dicari oleh pengguna. Untuk membuat sistem tersebut terdapat tahapan yang diperlukan seperti analisis masalah, pengumpulan dataset hadis, proses preprocessing, proses jaccard similarity, dan proses pengujian sistem. Berikut merupakan gambaran umum tahapan penelitian yang digunakan, seperti pada Gambar 1, yaitu diagram tahapan penelitian.



Gambar 1. Tahapan Penelitian

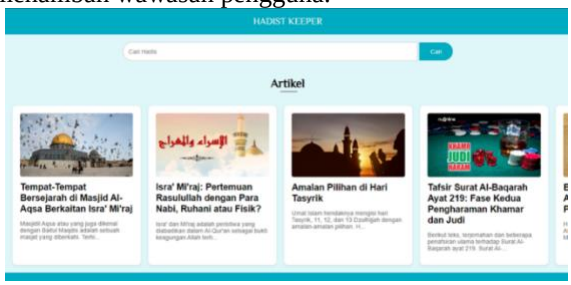
HASIL DAN PEMBAHASAN

Implementasi

Peneliti melakukan implementasi terhadap algortima yang sudah diterapkan, berdasarkan tahapan penelitian, dihasilkan implementasi sistem pencarian hadis berbahasa indonesia berbasis website sebagai berikut :

1. Tampilan Halaman Pencarian Hadis (Landing Page)

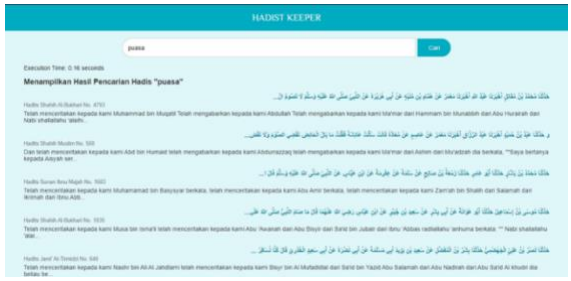
Halaman ini merupakan halaman pertama yang akan muncul saat pengguna akan melakukan pencarian atau bisa disebut dengan landing page. Dalam halaman ini terdapat bar search yang nantinya dapat digunakan pengguna untuk mencari query yang diiginkan, kemudian terdapat beberapa artikel untuk menambah wawasan pengguna.



Gambar 2. Halaman Pencarian Hadis (Landing Page)

2. Tampilan Halaman Hasil Pencarian Hadis

Halaman ini merupakan hasil pencarian hadis dari pengguna, pada halaman ini akan muncul 10 dokumen hasil pencarian berdasarkan query yang di inputkan oleh pengguna, dimana hasil dari 10 dokumen teratas yang tampil pada halaman ini dihasilkan dari implementasi metode Jaccard Similarity.



Gambar 3. Halaman Hasil Pencarian

3. Tampilan Halaman Detail Hadis

Halaman ini merupakan halaman detail hadis, dimana ketika pengguna memilih salah satu hadis yang ada dalam halaman hasil pencarian hadis akan di direct ke halaman detail hadis ini. Pada halaman detail hadis berisikan Judul atau kitab hadis kemudian terdapat Bahasa arab, cendekiawan, dan juga terjemahannya.



Gambar 4. Halaman Detail Hadis

Pengujian

Dalam penelitian ini pengujian sistem pencarian hadis menggunakan metode Jaccard Similarity dilakukan dengan pengujian precision @K. Penggunaan precision @K dalam pengujian sistem pencarian hadis berbahasa Indonesia menekankan pada relevansi dan kualitas hasil pencarian yang paling mungkin diakses oleh pengguna.

Pada proses pengujian ini dilakukan oleh 10 pengguna (users), dimana pengguna memasukkan single query yang dicarinya ke dalam sistem pencarian hadis dengan masing-masing K = @10 dan K = @20.

No	Pengguna	Query	K=10	K=20
1	Pengguna 1	Niat	0.7	0.35
2	Pengguna 2	Manusia	1.0	1.0
3	Pengguna 3	Benci	0.8	0.4
4	Pengguna 4	idul adha	0.1	0.05
5	Pengguna 5	Haji	0.9	0.45
6	Pengguna 6	Zina	0.8	0.4
7	Pengguna 7	Makan	1.0	1.0
8	Pengguna 8	Nikah	1.0	0.5
9	Pengguna 9	Sabar	0.3	0.15
10	Pengguna 10	Puasa	1.0	0.9
Rata-Rata			0.76	0.52

Untuk perhitungan nya diambil pengguna pertama :

Precision @10 untuk pengguna 1 = $\frac{7}{10} = 0.7$

Precision @20 untuk penggguna 1 = $\frac{7}{20} = 0.35$

Selanjutnya dilakukan proses pengujian kembali dimana pengguna memasukkan multi query yang dicarinya ke dalam sistem pencarian hadis dengan masing-masing K = @10 dan K = @20.

No	Pengguna	Query	K=10	K=20
1	Pengguna 1	doa yang telah dikabulkan	0.2	0.1

2	Pengguna 2	manusia wajib atasnya shadaqah	0.2	0.1
3	Pengguna 3	jual beli	0.3	0.15
4	Pengguna 4	menyembelih hewan	0.4	0.2
5	Pengguna 5	Haji dan umrah	1.0	0.6
6	Pengguna 6	dosa besar	1.0	1.0
7	Pengguna 7	makanan kurma	1.0	1.0
8	Pengguna 8	makkah dan madinah	1.0	1.0
9	Pengguna 9	bersabar terhadap ujian	0.5	0.25
10	Pengguna 10	puasa sunnah asyura	0.1	0.1
		Rata-Rata	0.57	0.45

Untuk perhitungan nya diambil pengguna pertama :

Precision @10 untuk pengguna 1 = $\frac{2}{10} = 0.2$

Precision @20 untuk penggguna 1 = $\frac{2}{20} = 0.1$

Berdasarkan pengujian sistem pencarian menggunakan Precision@K, hasilnya menunjukkan bahwa rata-rata precision untuk K=10 adalah 0.76 untuk single query dan 0.57 untuk multi query. Rata-rata precision yang tinggi menunjukkan bahwa hasil pencarian untuk single query cukup relevan. Ini berarti bahwa dokumen di posisi 1 hingga 10 sebagian besar relevan dengan query yang diinputkan pengguna. Sedangkan rata-rata precision yang lebih rendah untuk multi query menunjukkan bahwa hasil pencarian kurang relevan dibandingkan dengan single query. Hal ini mungkin disebabkan oleh terbatasnya dataset yang ada sehingga kesulitan dalam menemukan dokumen yang memenuhi semua kata kunci dalam multi query secara bersamaan, sehingga lebih banyak dokumen yang hanya relevan sebagian.

Sementara itu, rata-rata precision untuk K=20 adalah 0.52 untuk single query dan 0.45 untuk multi query. Precision menurun saat k meningkat dari 10 ke 20, menunjukkan bahwa dokumen di posisi 11 hingga 20 kurang relevan dibandingkan dengan posisi 1 hingga 10. Ini adalah indikasi bahwa relevansi hasil pencarian berkurang seiring dengan meningkatnya jumlah hasil yang dikembalikan. Precision @20 yang lebih rendah dibandingkan precision @10 menunjukkan penurunan relevansi yang lebih tajam untuk multi query. Hal ini menunjukkan bahwa dengan meningkatnya jumlah hasil, semakin banyak dokumen yang kurang relevan masuk dalam hasil pencarian. Berdasarkan hasil pengujian sistem yang telah dilakukan, maka hasil yang relevan dan ditampilkan pada sistem dengan metode jaccard similarity terdapat pada K@10.

KESIMPULAN

Dari hasil penelitian Sistem Pencarian Hadis Berbahasa Indonesia menggunakan Metode Jaccard Similarity, penulis mengambil kesimpulan bahwa sistem ini berhasil dikembangkan dengan menggunakan bahasa pemrograman Python dengan framework Flask. Metode Jaccard Similarity telah diimplementasikan pada sistem pencarian hadis berbahasa Indonesia dengan pengujian precision @k. Sistem pencarian hadis yang dikembangkan mendapatkan hasil rata-rata precision untuk K=10 adalah 0.76 untuk single query dan 0.57 untuk multi query. Sementara itu, rata-rata precision untuk K=20 adalah 0.52 untuk single query dan 0.45 untuk multi query. Penurunan precision dari K@10 ke K@20 menunjukkan hasil pencarian yang lebih relevan pada K@10.

DAFTAR PUSTAKA

[1] Andariati Leni, "HADIS DAN SEJARAH PERKEMBANGANNYA Leni Andariati."

[2] A. W. Pratama and A. Dores, "Arfianto Wahyu Pratama dan Ardiasnyah Dores SISTEM INFORMASI PENCARIAN PENGEPUL BARANG BEKAS DI KOTA TANGERANG BERBASIS WEBSITE," Teknologi Informatika dan Komputer, vol. 9, no. 1, 2018.

[3] M. Mukhtar, S. Al-Hadis, and F. Al-Hadis, "SYARH AL-HADIS DAN FIQH AL-HADIS (Upaya Memahami dan Mengamalkan Hadis Nabi)."

[4] Algifri Muqsit Jabar, "MEMBAHAS KITAB HADIS (KitabṢaḥiḥ al-Bukhari dan Sunan al-Turmudzi)."

[5] R. Ridlo Baihaqi, "Temu Kembali Informasi pada Berita Olahraga Berbahasa Indonesia dengan Seleksi Fitur Term Frequency dan Metode BM25," 2020. [Online]. Available: http://j-ptiik.ub.ac.id

[6] A. Deolika and E. Taufiq Luthfi, "ANALISIS PEMBOBOTAN KATA PADA KLASIFIKASI TEXT MINING," Jurnal Teknologi Informasi, vol. 3, no. 2, 2019.

[7] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," JURNAL MEDIA INFORMATIKA BUDIDARMA, vol. 5, no. 2, p. 406, Apr. 2021, doi: 10.30865/mib.v5i2.2835.

[8] D. Riyanto and A. Azis, "Implementation of the Jaccard Similarity Algorithm on Answer Type," International Journal of Informatics and Information Systems, vol. 5, no. 2, pp. 76–83, 2022.

[9] S. Utomo, I. Much Ibnu Subroto, and A. Riansyah, "Deteksi Plagiat Tugas Akhir dengan Metode Jaccard Similarity," Jurnal Transistor Elektro dan Informatika (TRANSISTOR EI, vol. 4, no. 2, p. 2022.

[10] A. Yudhana, dan Iif Alfiatul Mukaromah, P. Studi Teknik Elektro, U. Ahmad Dahlan, A. Dahlan, and U. Jl Soepomo, "IMPLEMENTASI DETEKSI PLAGIARISME MENGGUNAKAN METODE N-GRAM DAN JACCARD SIMILARITY TERHADAP ALGORITMA WINNOWING." [Online]. Available: https://ejournal.undip.ac.id/index.php/transmisi

[11] S. Rahmawati, D. Nurjanah, and R. Rismala, "Analisis dan Implementasi pendekatan Hybrid untuk Sistem Rekomendasi Pekerjaan dengan Metode Knowledge Based dan Collaborative Filtering," Indonesian Journal on Computing (Indo-JC), vol. 3, no. 2, p. 11, Sep. 2018, doi: 10.21108/indojc.2018.3.2.210.

[12] I. Mutia Ayudita and P. Pandu Adikara, "Sistem Pencarian Jurnal Ilmiah Cross Language dengan Metode Vector Space Model (VSM)," 2018. [Online]. Available: http://j-ptiik.ub.ac.id