SNESTIK

Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika

https://ejurnal.itats.ac.id/snestik dan https://snestik.itats.ac.id

Informasi Pelaksanaan :
SNESTIK III - Surabaya, 11 Maret 2023
Ruang Seminar Gedung A, Kampus Institut Teknologi Adhi Tama Surabaya

Informasi Artikel:
DOI : 10.31284/p.snestik.2023.4262

Prosiding ISSN 2775-5126

Fakultas Teknik Elektro dan Teknologi Informasi-Institut Teknologi Adhi Tama Surabaya
Gedung A-ITATS, Jl. Arief Rachman Hakim 100 Surabaya 60117 Telp. (031) 5945043
Email : snestik@itats.ac.id

Pencarian Dokumen Jurnal Skripsi dengan Menggunakan Metode Positive Impact Factor Query (PIFQ)

Tutuk Indriyani, Septiyawan Rosetya Wardhana, Muhammad Fakhrul Amin

Institut Teknologi Adhi Tama Surabaya

e-mail: tutuk@itats.ac.id

ABSTRACT

To facilitate the search for student thesis journals using an information retrieval system, users are made easy in finding information from a large amount of data. Information retrieval in the current era has been widely applied in various fields, such as information on the internet and also information stored locally. Document search is one of the implementations of using information retrieval. proposed a new weighting method on words based on Positive Impact Factor Query (PIFQ). The result is that calculations using the PIFQ method can outperform calculations using the TF-IDF method. How can the PIFQ method be applied to an information retrieval system for searching Informatics Engineering thesis documents that match the queries and topics entered by the user. The goal is to find the thesis document that matches the query entered by the user. The benefit is as a means for Informatics Engineering students at the Adhi Tama Surabaya Institute of Technology in finding thesis documents that are appropriate to their topic. a system designed to recommend ITATS student journal data using the Term Frequency - Inverse Document Frequency and Positive Impact Factor Query methods is proven to be able to produce relevant search results for users with the highest rating rating of 13.38. In the third test scenario, several journals get a value of 0 for calculations using the TF-IDF method. This is of course a deficiency in the system that has been designed because it is not in accordance with the wishes of the researcher.

Keywords: *Information Retrieval*; *Positive Impact Factor Query*; *Term Frequency - Inverse Document Frequency*

ABSTRAK

Untuk memudahkan dalam pencarian jurnal skripsi pada mahasiswa menggunakan sistem temu kembali informasi, pengguna dimudahkan dalam menemukan suatu informasi dari data yang sangat banyak. *Information retrieval* pada era sekarang telah banyak diterapkan pada berbagai bidang, seperti informasi dalam jaringan internet dan juga informasi yang tersimpan secara lokal. Pencarian dokumen merupakan salah satu dari implementasi penggunaan *information retrieval*. diusulkan sebuah metode pembobotan baru pada kata yang berbasis *Positive Impact Factor Query* (PIFQ). Hasilnya perhitungan dengan menggunakan metode PIFQ dapat mengungguli perhitungan dengan menggunakan metode TF-IDF. Bagaimana metode PIFQ dapat diterapkan pada sistem temu kembali informasi untuk pencarian dokumen skripsi Teknik Informatika yang sesuai dengan *query* dan topik yang dimasukkan oleh pengguna. Tujuannya untuk mencari dokumen skripsi yang sesuai dengan *query* yang dimasukkan oleh pengguna. Manfaatnya sebagai sarana untuk mahasiswa Teknik Informatika Institut Teknologi Adhi Tama Surabaya dalam mencari dokumen skripsi yang sesuai dengan topiknya. Sistem yang dirancang untuk merekomendasikan data jurnal mahasiswa ITATS menggunakan metode *Term Frequency - Inverse Document Frequency* dan *Positive Impact Factor Query* terbukti mampu menghasilkan hasil pencarian yang relevan bagi pengguna dengan nilai tertinggi dari perangkingannya 13.38. Pada skenario pengujian yang ketiga, beberapa jurnal mendapatkan nilai 0 untuk perhitungan dengan menggunakan metode TF-IDF. Hal ini tentu saja menjadi kekurangan pada sistem yang telah dirancang karena tidak sesuai dengan keinginan peneliti.

Kata kunci: *Information retrieval*; *Positive Impact Factor Query*; *Term Frequency - Inverse Document Frequency*;

PENDAHULUAN

Information Retrieval System atau Sistem Temu Kembali Informasi merupakan sebuah sistem terstruktur yang berguna untuk menemukan kembali (*retrieve*) sebuah informasi yang relevan dengan *keyword* atau *query* yang berasal dari masukan pengguna. Informasi yang didapatkan kembali dari sebuah sistem temu kembali informasi dapat berupa data teks, citra, dan sebagainya. Dengan sistem temu kembali informasi, pengguna dimudahkan dalam menemukan suatu informasi dari data yang sangat banyak. *Information retrieval* pada era sekarang telah banyak diterapkan pada berbagai bidang, seperti informasi dalam jaringan internet dan juga informasi yang tersimpan secara lokal. Pencarian dokumen merupakan salah satu dari implementasi penggunaan *information retrieval* [1][2].

Pada penelitian terdahulu, para peneliti telah banyak menggunakan berbagai macam metode untuk pencarian dokumen guna mendapatkan hasil yang akurat. Salah satu metode yang pernah digunakan adalah *Positive Impact Factor* (PIF). Dr. Emmanuel M. beserta koleganya (2013) melakukan penelitian dengan menggunakan metode PIF. Dalam jurnal yang berjudul "*A Novel Scheme for Term Weighting in Text Categorization : Positive Impact Factor*" metode PIF diuji keakuratannya dengan metode yang lainnya, seperti Binary, TF, TF-IDF, TF-RF, TF-CHI. Dalam penelitian ini metode PIF juga digabung dengan metode IDF. Hasilnya metode PIF, PIF-IDF mampu mengungguli rata – rata akurasi dari metode yang lain dengan nilai diatas 90%. [1][3]

Para peneliti dari Institut Teknologi Sepuluh Nopember juga pernah menggunakan metode PIF dalam melakukan penelitiannya. Dalam jurnal yang berjudul "*Term Weighting Based on Positive Impact Factor Query for Arabic Fiqh Document Ranking*", diusulkan sebuah metode pembobotan baru pada kata yang berbasis *Positive Impact Factor Query* (PIFQ)[4]. Hasilnya

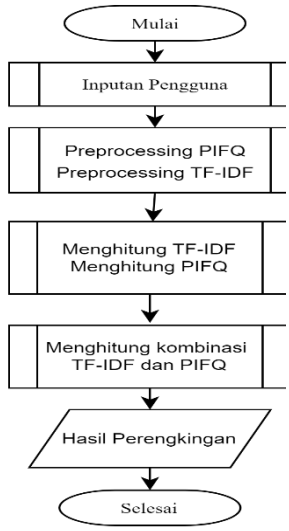
perhitungan dengan menggunakan metode PIFQ dapat mengungguli perhitungan dengan menggunakan metode TF-IDF. (Rizka W. Sholikah, Dhian S. Y. Kartika, Agus Zainal Arifin, dan Diana Purwitasari, 2017). Sedangkan Septyawan dkk pada tahun 2015, melakukan penelitian dengan menggabungkan kedua metode *Latent Semantic Indexing* dan pembobotan berbasis preferensi pengguna dikombinasikan dengan ICF untuk melakukan perangkingan dalam sebuah dokumen [1][4]. Berdasarkan latar belakang yang telah disebutkan diatas, maka diusulkan sistem temu kembali informasi untuk pencarian dokumen menggunakan pembobotan kata berdasarkan *Positive Impact Factor Query* [5]

METODE

Dalam sistem yang akan dibuat, akan terdapat beberapa macam proses didalamnya. Langkah pertama pengguna memasukan *query* yang diinginkan. Selanjutnya *Query* dan *dataset* yang berupa dokumen jurnal ilmiah yang digunakan sebagai objek penelitian akan melalui tahap *preprocessing*, yang bertujuan untuk menghindari data yang tidak sempurna, *noise* pada data, dan data-data yang tidak konsisten. Sehingga data – data tersebut menjadi bersih dari *noise*, bentuk kata yang lebih seragam, dan volume kata yang tidak dibutuhkan berkurang. Selanjutnya kata-kata yang telah melalui *preprocessing* akan dihitung dengan menggunakan metode yang telah ditentukan [6].

Metode *TF* digunakan untuk menghitung frekuensi *key term* yang berada dalam dokumen. Sedangkan *IDF* digunakan untuk memberi nilai tinggi pada kata yang jarang muncul pada setiap dokumen yang digunakan sebagai *dataset*. Berdasarkan metode *PIF*, *PIFQ* digunakan untuk menghitung keberadaan *key term* pada setiap kategori atau topik jurnal ilmiah yang bertujuan untuk meningkatkan relevansi pencarian dokumen antara *query* dengan hasil pencarian [7].

Selanjutnya hasil dari perhitungan yang menggunakan dua metode tadi dilakukan perhitungan kombinasi untuk memberi bobot baru pada setiap kata. Hal ini bertujuan untuk melakukan perangkingan pada setiap dokumen. Semakin banyak kata pada dokumen yang mengandung *key term,* maka semakin relevan dokumen tersebut dengan *query* yang dimasukan oleh pengguna. Perancangan sistem adalah proses pembuatan sebuah laporan perencanaan sistem yang mendukung tujuan dari penelitian ini. Perencanaan sistem bertujuan untuk menyusun rencana secara matang, sehingga diharapkan dapat menghasilkan sebuah sistem yang sesuai dengan perencanaan yang telah disusun. Berikut perencanaan sistem pada penelitian ini Gambar 1.



Gambar 1. Algoritma desain sistem

Pada Gambar 1 menggambarkan tentang alur dari sistem yang akan dibangun. Pada langkah pertama, pengguna memasukan *query*. Selanjutnya, dilakukan *preprocessing* untuk mengolah masukan dari pengguna dan *preprocessing* untuk data dokumen jurnal ilmiah. Kemudian dilakukan perhitungan *TF-IDF* untuk pembobotan kata pada data dokumen. Lalu dilakukan perhitungan *PIFQ* untuk mencari kata yang sesuai dengan *key term*. Langkah terakhir dilakukan perhitungan kombinasi dari hasil perhitungan *TF-IDF* dan *PIFQ* sehingga menghasilkan bobot baru pada masing - masing kata, lalu dokumen - dokumen tersebut diurutkan berdasarkan nilai yang tinggi[8].

Masukkan Pengguna

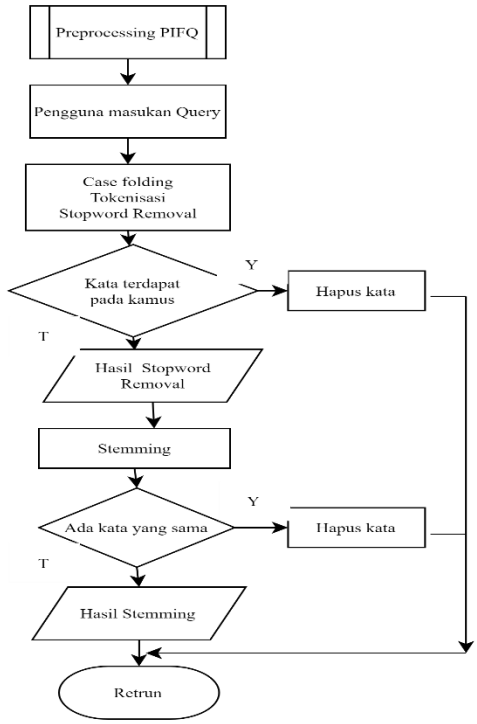
Masukan pengguna adalah data yang berupa *query* yang dimasukan oleh pengguna de dalam sistem melalui kolom pencarian untuk diproses lebih lanjut. Berikut bagan alur masukan pengguna dapat ditunjukkan pada Gambar 2.

Gambar 2. Algoritma Inputan Pengguna

Penjelasan Gambar 2. menjelaskan mengenai alur masukan pengguna yang nanti akan diolah oleh sistem agar dapat memberi hasil pencarian dokumen jurnal ilmiah. Untuk masukan pengguna berupa *query,* yaitu beberapa kata yang digunakan untuk mencari dokumen yang dinginkan. Dalam *query* terdapat kata yang menjadi *key term*, dimana *key term* adalah kata yang memiliki nilai informasi dan bobot yang lebih tinggi dibanding dengan kata lain.

Preprocessing untuk *PIFQ*

Preprocessing merupakan proses yang menghilangkan beberapa kata dalam teks data yang terindikasi menimbulkan *noise* dan *missing value* yang dapat menghambat kinerja sistem. Berikut adalah bagan alur *preprocessing* untuk *PIFQ* ditunjukkan pada Gambar 3.

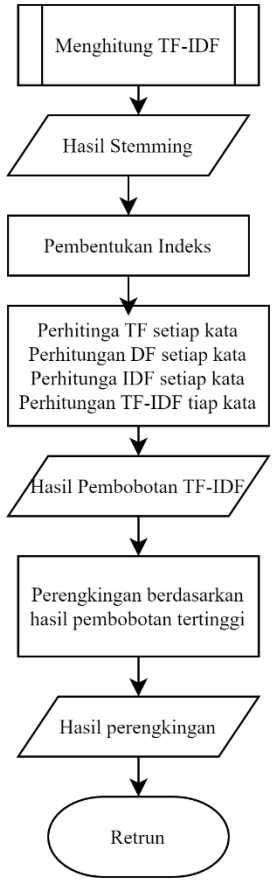


Gambar 3. Algoritma proses preprosessing PIFQ

Pada Gambar 3 menjelaskan tentang tahap *preprocessing* untuk *PIFQ*. Pada tahap awal, sistem menjalankan proses *Case Folding* yang bertujuan untuk merubah semua kata yang berhuruf kapital menjadi huruf kecil. Tahap kedua dilakukan proses tokenisasi yang bertujuan untuk memotong *string input* berdasarkan tiap kata yang menyusunnya, sehingga merubah data teks yang awalnya merupakan data yang berupa kalimat menjadi data yang berupa kata. Tahap ketiga dilakukan proses *stopword removal* yang bertujuan untuk menghilangkan kata-kata yang tidak diperlukan dalam proses, sehingga dapat mempersingkat waktu pemrosesan data teks dan dapat meningkatkan kinerja sistem. Pada tahap akhir *preprocessing*, dilakukan proses *Stemming* yang bertujuan merubah sebuah kata yang berimbuhan menjadi kata dasar saja.

Hitung *TF – IDF*

TF-IDF (Term Frequency-Inverse Document Frequency) adalah metode *statistic numeric* yang menggambarkan seberapa pentingnya sebuah kata dalam sebuah dokumen atau korpus. Berikut bagan alur hitung *TF-IDF* dapat ditunjukkan pada Gambar 4.

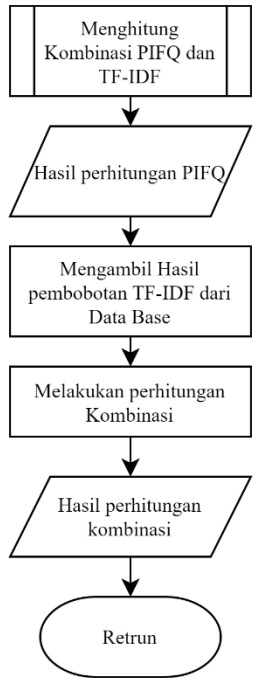


Gambar 4. Algoritma menghitung TF-IDF

Pada Gambar 4. langkah pertama adalah hasil *preprocessing* pada langkah sebelumnya dijadikan indeks setiap kata (*Term*). Lalu dilakukan perhitungan *Term Frequency* pada setiap dokumen. Selanjutnya menghitung *Document Frequency* pada setiap dokumen. Lalu dilakukan perhitungan *Inverse Document Frequency* pada setiap dokumen. Kemudian dilakukan perhitungan *TF – IDF* pada setiap dokumen yang akan menghasilkan bobot kata. Langkah selanjutnya adalah melakukan perangkingan dokumen berdasarkan hasil pembobotan tertinggi.

Kombinasi Hitung *PIFQ* dan *TF – IDF*

Pada bagian ini akan dilakukan perhitungan kombinasi hasil dari perhitungan *PIFQ* dengan hasil dari perhitungan *TF – IDF* yang bertujuan untuk memberi bobot baru pada dokumen jurnal. Sehingga pada hasil pencarian akan menampilkan data dokumen yang relevan. Berikut adalah bagan alurnya yang ditunjukkan pada Gambar 5.



Gambar 5. Algoritma menghitung kombinasi

Penjelasan Gambar 5. tentang menghitung kombinasi nilai *PIFQ* dan *TF – IDF*, pertama sistem mengambil hasil hitung *PIFQ* pada proses sebelumnya. Lalu sistem juga akan mengambil hasil dari pembobotan *TF –IDF* dan dilakukan proses hitung kombinasi. Sehingga proses hitung kombinasi menghasilkan pembobotan baru pada data dokumen.

HASIL DAN PEMBAHASAN

Proses Ekstraksi Data

Pada penelitian ini akan dilakukan studi kasus kecil untuk menguji rancangan sistem. Dimulai dari awal hingga mendapatkan hasil pencarian dokumen jurnal yang relevan dengan masukan pengguna. Langkah awal adalah mengumpulkan data jurnal yang akan diolah diambil dari jurnal Informatika yang berada di berbagai repository kampus – kampus di Indonesia. Data yang diambil sebanyak 6 dokumen jurnal bagian judul dan abstrak saja, dan satu *query* sebagai masukan pengguna. Data jurnal dibagi menjadi tiga kategori penjurusan, yaitu Artificial Intelligence (AI), Rekayasa Perangkat Lunak (RPL), dan Jaringan Komputer (Jarkom).

Tabel 1. Tabel Data Jurnal Kategori RPL, Jaringan dan AI

No	Peneliti	Judul dan Tahun	Isi Abstrak
1	Deny Adhar (Jaringan)	Implementasi Algoritma Des (Data Encryption Standard) Pada Enkripsi Dan Dekripsi Sms Berbasis Android (2019)	Perkembangan teknologi membuat lahirnya sebuah alat yang merupakan cikal bakal terbentuknya Short Messaging Service (SMS) yaitu pager. Namun, dengan kebutuhan masyarakat pesan tersebut sehingga pesan sms benar - benar aman…
2	Arafat (Jaringan)	Sistem Pengamanan Pintu Rumah Berbasis Internet Of Things (Iot) Dengan Esp8266 (2016)	Berbagai permasalahan selalu mengancam keamanan pemilik rumah seperti pencurian. Hal ini menyebabkan diperlukannya sistem keamanan rumah. Sistem keamanan berfungsi memberikan informasi yang terjadi di dalam rumah. Internet of Things (IoT) membuat
3	Muhammad Susilo, Rezki Kurniati, Kasmawi (RPL)	Rancang Bangun Website Toko Online Menggunakan Metode Waterfall (2018)	Website Toko online dapat memudahkan pelaku usaha atau toko untuk mempromosikan produknya dan mempermudah konsumen untuk mendapatkan informasi tentang produk-produk yang dimiliki penjual atau toko…
4	Jati Sasongko, Dwi Agus Diartono (RPL)	Rancang Bangun Sistem Informasi Manajemen Surat (2009)	Banyak jumlah surat yang dibuat dan diterima di bagian administrasi tata usaha, sehingga pencarian data akan menjadi tidak efisien dalam hal waktu dan tenaga dengan menggunakan sistem manual.
5	Aprilia Sulistyohati, Taufiq Hidayat (AI)	Aplikasi Sistem Pakar Diagnosa Penyakit Ginjal Dengan Metode Dempster Shafer (2008)	Sistem pakar adalah sistem berbasis komputer yang menggunakan pengetahuan, fakta, dan teknik penalaran dalam memecahkan masalah yang biasanya hanya dapat dipecahkan oleh seorang pakar dalam bidang tertentu…..
6	Hengki Tamando Sihotang (AI)	Perancangan Aplikasi Sistem Pakar Diagnosa Diabetes Dengan Metode Bayes (2017)	Deteksi secara dini dapat meminimalisir resiko kematian penderita Diabetes Mellitus serta untuk penanganan awal penyakit ini. Proses deteksi yang ada selama ini masih manual, ..

Dari Tabel 1. dilakukan proses *preprocessing* PIFQ (*Positive Impact Factor Query)* yaitu mengubah huruf menjadi huruf tak kapital (*Case Folding),* memecah kalimat menjadi kata (Tokenisasi), menghapus kata yang tidak dibutuhkan (*Stopword Removal)* dan mengubah kata menjadi kata dasar (*Stemming).* Setelah itu dilakukan proses yang merubah huruf kapital menjadi huruf kecil. Pada penelitian ini *query* masukan pengguna akan diubah menjadi huruf kecil (*case folding*) kemudian dilakukan Tokenisasi yang bertujuan memecah kalimat menjadi token atau

kata. Pada penelitian ini *query* masukan pengguna akan dipecah menjadi kata. Selanjutnya dilakukan perhitungan PIFQ hasil ditunjukan pada Tabel 2.

Tabel 2. Hasil proses perhitungan PIFQ

Kategori	Frekuensi kata kunci	Nilai Rata-Rata *PIFQ*
AI	5 kata	1,2108
Jarkom	3 kata	1,1139
RPL	5 kata	1,2108

Setelah dilakukan tahapan tahapan dari proses untuk mencari dokumen jurnal skripsi, kemudian dilakukan perangkingan dokumen.

Perangkingan Dokumen

Pada tahap ini akan dilakukan proses akhir yaitu perangkingan dokumen jurnal bagi pengguna. Perangkingan dokumen dilakukan dengan cara mengurutkan dokumen dengan nilai tertinggi dari hasil perhitungan kombinasi yang telah didapat dari proses sebelumnya dapat ditunjukkan pada Tabel 3.

Tabel 3. Tabel Perangkingan Dokumen

No	Judul Jurnal	Kategori	Nilai Hitung Kombinasi
1	Aplikasi Sistem Pakar Diagnosa Penyakit Ginjal Dengan Metode Dempster-Shafer	AI	13,3783
2	Perancangan Aplikasi Sistem Pakar Diagnosa Diabetes Dengan Metode Bayes	AI	12,532
3	Rancang Bangun Website Toko Online Menggunakan Metode Waterfall	RPL	4,590284
4	Sistem Pengamanan Pintu Rumah Berbasis Internet Of Things (Iot) Dengan Esp8266	Jarkom	3,846171
5	Rancang Bangun Sistem Manajemen Surat Jurnal	RPL	3,060189
6	Implementasi Algoritma Des (Data Encryption Standard) Pada Enkripsi Dan Dekripsi Sms Berbasis Android	Jarkom	0

Pada Tabel 3 nilai yang memiliki kombinasi yang paling tinggi yang disarankan dari proses pencarian data

KESIMPULAN

Dari tiga skenario hasil uji coba yang dilakukan, sistem yang dirancang untuk merekomendasikan data jurnal mahasiswa ITATS menggunakan metode *Term Frequency - Inverse Document Frequency* dan *Positive Impact Factor Query* terbukti mampu menghasilkan hasil pencarian yang relevan bagi pengguna. Pada skenario pengujian yang ketiga, beberapa jurnal mendapatkan nilai 0 untuk perhitungan dengan menggunakan metode TF-IDF. Hal ini tentu saja menjadi kekurangan pada sistem yang telah dirancang karena tidak sesuai dengan keinginan peneliti.

DAFTAR PUSTAKA

[1] Wardhana, S. R., Yunianto, D. R., Arifin, A. Z., & Purwitasari, D. Pembobotan Kata Berbasis Preferensi Dan Hubungan Semantik Pada Dokumen Fiqih Berbahasa Arab. Jurnal Teknologi Informasi Dan Ilmu Komputer, 2(2), 132–137. 2015

[2] Yuhefizar, HA Mooduto, Rahmat Hidayat. *Cara Mudah Membangun Website Interaktif Menggunakan Content Management System Joomla Edisi Revisi*. Jakarta: Elex Media Komputindo 2009.

[3] Emmanuel. M, S. M. Khatri dan Ramesh. B. D.R., *"A Novel Scheme for Term Weighting in Text Categorization : Positive Impact Factor"* dalam IEEE International Conference on Systems, Man, and Cybernetics (hlm. 2292-2297). Pune: IEEE Computer Society. 2013.

[4] Hemalatha, I., & Govardhan,, A. "*Preprocessing the Informal Text for Efficient Sentiment Analysis*" dalam International Journal of Emerging Trends & Technology in Computer Science (hlm. 58-61). Bhimavaram: IJETTCS 2012.

[5] Python Software Foundation, *"What is Python? Executive Summary"*, https://www.python.org/doc/essays/blurb/, diakses pada 11 Maret 2021 pukul 21.35 WIB.

[6] Dwi Putranti, Noviah, Edi Winarko. *Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropydan Support Vector Machine.* dalam Indonesia Journal of Computing and Cybernatics Sytems (hlm. 91-100). Yogyakarta: Universitas Gadjah Mada. 2014

[7] Yuhefizar, HA Mooduto, Rahmat Hidayat. *Cara Mudah Membangun Website Interaktif Menggunakan Content Management System Joomla Edisi Revisi*. Jakarta: Elex Media Komputindo. 2009.

[8] Rizka W. Sholikah, Dhian S. Y. Kartika, Agus Zainal Arifin, dan Diana Purwitasari. 2017. "*Term Weighting Based on Positive Impact Factor Query for Arabic Fiqh Document Ranking"* dalam Journal of Computer Science and Information (hlm. 29-35). Surabaya: Institut Teknologi Sepuluh Nopember.