

ISSN 2303 - 1425

J-INTTECH

Journal of Information and Technology

Volume 05 Nomor 01, Juni Tahun 2017

J-INTTECH

Volume 05 Nomor 01, Juni Tahun 2017



STIKI

SEKOLAH TINGGI INFORMATIKA & KOMPUTER INDONESIA

Jl. Raya Tidar 100 Malang, 65146

Telp. (0341)560823, Fax (0341)562525

ISSN 2303 - 1425

J-INTTECH

Journal of Information and Technology
Volume 05 Nomor 01, Juni Tahun 2017



LEMBAGA PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT

STIKI

SEKOLAH TINGGI INFORMATIKA & KOMPUTER INDONESIA
Jl. Raya Tidar 100, Malang; Phone: 0341-560823; Fax: 0341-562525; <http://www.stiki.ac.id>; mail@stiki.ac.id

PENGANTAR REDAKSI

J-INTECH merupakan jurnal yang diterbitkan oleh Sekolah Tinggi Informatika dan Komputer Indonesia Malang guna mengakomodasi kebutuhan akan perkembangan Teknologi Informasi serta guna mensukseskan salah satu program DIKTI yang mewajibkan seluruh Perguruan Tinggi untuk menerbitkan dan mengunggah karya ilmiah mahasiswanya dalam bentuk terbitan maupun jurnal online.

Pada edisi ini, redaksi menampilkan beberapa karya ilmiah mahasiswa yang mewakili beberapa mahasiswa yang lain, yang dianggap cukup baik sebagai media pembelajaran bagi para lulusan selanjutnya.

Tentu saja diharapkan pada setiap penerbitan memiliki nilai lebih dari karya ilmiah yang dihasilkan sebelumnya sehingga merupakan nilai tambah bagi para adik kelas maupun pihak-pihak yang ingin studi atau memanfaatkan karya tersebut selanjutnya.

Pada kesempatan ini kami juga mengundang pihak-pihak dari PTN/PTS lain sebagai kontributor karya ilmiah terhadap jurnal J-INTECH, sehingga Perkembangan IPTEK dapat dikuasai secara bersama-sama dan membawa manfaat bagi institusi masing-masing.

Akhir redaksi berharap semoga dengan terbitnya jurnal ini membawa manfaat bagi para mahasiswa, dosen pembimbing, pihak yang bekerja pada bidang Teknologi Informasi serta untuk perkembangan IPTEK di masa depan.

REDAKSI

DAFTAR ISI

Implementasi Algoritma Kriptografi Elgamal pada <i>Data Text</i> <i>Binantara Parmadi</i>	01-05
<i>Game</i> Pengenalan Konsep Pemrograman Dasar Menggunakan <i>Blockly</i> Berbasis <i>Website</i> <i>Vincent Putra Gunawan</i>	06-12
Sistem Informasi Kenaikan Pangkat Guru pada UPTD Dinas Pendidikan Kecamatan Singosari..... <i>Raditias Wahana Putra</i>	13-17
<i>Game</i> Edukasi Pengenalan Lagu-Lagu Nasional Berbasis <i>Mobile</i> <i>Farul Sukrin Kanday</i>	18-23
Pengembangan Sistem Informasi Pengelolaan Aset Teknologi Informasi (Studi Kasus: STIKI Malang) <i>Francino Gigih Adi Saputro</i>	24-28
Pemanfaatan <i>Web Service</i> pada Aplikasi <i>Notifikasi</i> Pengumuman Mahasiswa (Studi Kasus: STIKI Malang) <i>I Putu Sudarma Adi Septyanto</i>	29-35
Sistem Pakar Identifikasi Hama dan Penyakit Tanaman Apel dengan Menggunakan Metode <i>Forward Chaining</i> Berbasis Android <i>Tri Mahardi Kurniawan</i>	36-42
Integrasi Sistem Informasi Pengelolaan Seminar dan <i>Workshop</i> Mahasiswa (Studi Kasus: STIKI Malang) <i>Benny Eka Atmojo</i>	43-52
Sistem Informasi Administrasi Keuangan Sekolah Berbasis Web di SMK YP 17 Selorejo - Blitar..... <i>Eka Dewi Susanti</i>	53-58
Sistem Informasi Manajemen Data Barang guna Mempercepat Proses Perhitungan dalam Proses Produksi (Studi Kasus DefraOi - Clothing)..... <i>Trenda Defra Frandisman</i>	59-63

Perancangan Tutorial Bahasa Isyarat Berbasis Android bagi Anak Tuna Rungu	64-70
<i>Ita Kumala Wardani</i>	
Sistem Informasi Administrasi Lembaga Sertifikasi Profesi STIKI Malang untuk Pengelolaan Sertifikasi TIK.....	71-77
<i>Fuad Hasan Perdana Putra</i>	
<i>Virtual Tour</i> Berbasis 3D untuk Pengenalan Kampus STIKI Malang.....	78-82
<i>Ajib Trimannula</i>	
Tutorial Pengenalan Warna Berbasis Android dengan Menggunakan Macromedia Flash CS6	83-88
<i>Penta Galih Registrara</i>	
Sistem Informasi Perencanaan Jadwal di Asia Hardware Berdasarkan <i>Material Requirement Planning</i>	89-92
<i>Astutik Puji Afianti</i>	
Sistem Pakar Penentuan Jenis Penyakit Ayam dengan Metode <i>Forward Chaining</i> Berbasis Android.....	93-103
<i>Fida Wiji Lestari</i>	
Aplikasi <i>Game</i> Sejarah Maang dengan Memanfaatkan <i>Corona Game Engine</i> Berbasis Android.....	104-113
<i>Julio Menahemi Psalmoi</i>	
Penerapan Teknik <i>Webscraping</i> dan <i>Vector Space Model</i> pada Mesin Pencari Lowongan Kerja.....	114-118
<i>Andriansyah Dwi Wardana</i>	
Sistem Pendukung Keputusan Seleksi Siswa Berprestasi di SMK PGRI 3 Malang Menggunakan Metode <i>Weighted Product</i> (WP).....	119-124
<i>Muhammad Faisal</i>	
Game 3D Punakawan Guna Mengenalkan Tokoh Punakawan dan Cerita Bagong Labuh Berbasis Android	125-131
<i>Bijahika Maulana Kohri Rijal</i>	

ISSN 2303 - 1425

J-INTECH

Journal of Information and Technology

Volume 05 Nomor 01, Juni Tahun 2017

- Pelindung** : Ketua STIKI
- Penasehat** : Puket I, II, III
- Pembina** : Ka. LPPM
-
- Editor** : Subari, S.Kom, M.Kom
- Section Editor** : Daniel Rudiaman S.,ST, M.Kom
-
- Reviewer** : Dr. Eva Handriyantini, S.Kom, M.MT.
Evi Poerbaningtyas, S.Si, M.T.
Laila Isyriyah, S.Kom, M.Kom
Anita, S.Kom, M.T.
-
- Layout Editor** : Nira Radita, S.Pd., M.Pd
Muh. Bima Indra Kusuma

Penerapan Teknik Webscraping dan Vector Space Model pada Mesin Pencari Lowongan Kerja

Andriansyah Dwi Wardana

Program Studi Teknik Informatika, Sekolah Tinggi Informatika & Komputer Indonesia (STIKI)
Malang
Email: andredwi93@gmail.com

ABSTRAK

Hal yang melatar belakangi Tugas Akhir ini adalah untuk mendapatkan informasi lowongan kerja terkini seringkali pengguna harus membuka beberapa website sekaligus untuk menemukan informasi tentang lowongan kerja yang sesuai. Sehingga, ini memakan waktu untuk membuka website-website tersebut satu persatu. Maka dengan pembuatan website ini pencari lowongan kerja tidak perlu membukanya satu persatu. Karena website ini mengambil semua informasi lowongan kerja terbaru dari website resmi penyedia lowongan kerja dan meletakkannya dalam satu halaman website, sehingga tidak memakan banyak waktu.

Kata Kunci: *website, lowongan kerja, web scraping, vector space model*

1. PENDAHULUAN

Meningkatnya kebutuhan akan informasi mendorong manusia untuk mengembangkan teknologi-teknologi baru agar pengolahan data dan informasi dapat dilakukan dengan mudah dan cepat. Salah satu teknologi yang sedang berkembang dengan pesat saat ini adalah teknologi informasi atau komputer, teknologi internet. Dengan adanya internet akan mempermudah dan mempercepat proses pengolahan data, mencari informasi dan lain-lain.

Dengan menggunakan internet pengguna juga dapat mencari lowongan kerja sesuai dengan kriteria. Untuk mendapatkan informasi tentang lowongan kerja tersebut seringkali pengguna harus membuka beberapa website sekaligus untuk menemukan informasi tentang lowongan kerja yang sesuai. Sehingga, ini memakan waktu untuk membuka website-website tersebut satu persatu.

Web Scraping adalah proses pengambilan sebuah dokumen semi terstruktur dari internet, umumnya berupa halaman-halaman web dalam bahasa markup seperti HTML atau XHTML, dan menganalisa dokumen tersebut untuk diambil data tertentu dari halaman tersebut untuk digunakan bagi kepentingan lain.

Pada *Information Retrieval System* terdapat beberapa metode yang digunakan dalam pencarian salah satunya adalah dengan merepresentasikan proses pencarian menggunakan Model Ruang Vektor (*Vector Space Model*). Model ruang vektor dibuat berdasarkan pemikiran bahwa isi dari

dokumen ditentukan oleh kata-kata yang digunakan dalam dokumen tersebut. Model ini menentukan kemiripan (*similarity*) antara dokumen dengan *query* dengan cara merepresentasikan dokumen dan *query* masing-masing ke dalam bentuk vektor.

Dengan adanya kemajuan teknologi yang semakin pesat, maka dari itu alasan pembuatan aplikasi berbasis website ini adalah untuk mempermudah pencari kerja dalam mencari lowongan kerja terbaru yang sebelumnya pencari kerja harus membuka website lowongan kerja satu persatu untuk menemukan lowongan pekerjaan yang sesuai.

Maka dengan website ini pencari lowongan kerja tidak perlu membukanya satu persatu. Karena website ini mengambil semua informasi lowongan kerja terbaru dari website resmi penyedia lowongan kerja dan meletakkannya dalam satu halaman website, sehingga tidak memakan banyak waktu.

2. METODOLOGI PENELITIAN

a. Analisa Masalah

Masalah sebelum menggunakan teknik *web scraping* dan *vector space model* dalam melakukan pencarian lowongan kerja adalah sebagai berikut:

1. Pada sebuah website lowongan kerja terdapat dokumen html yang berisi banyak konten untuk diambil informasinya.
2. Di beberapa website lowongan kerja tidak dapat mencari dengan memasukkan lebih dari satu kata kunci. Contoh kata kunci lebih dari satu kata:

“supervisor programmer teknik mesin Malang Surabaya”.

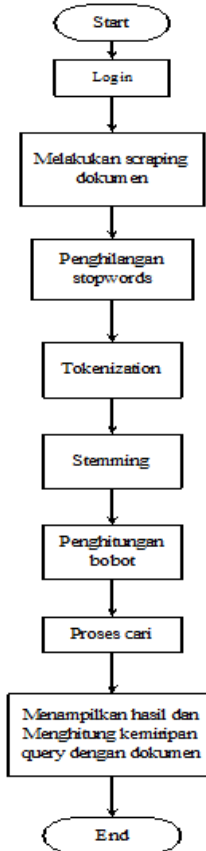
Dari masalah diatas maka solusi untuk pencari kerja dalam mencari kerja lewat website adalah sebagai berikut:

1. Menggunakan teknik *web scraping* untuk mengambil beberapa konten yang dibutuhkan misalnya judul lowongan kerja, deskripsi lowongan kerja, url lowongan kerja dll.
2. Menggunakan *vector space model* dalam pencarian lowongan kerja lewat website sehingga dapat mencari lowongan kerja meskipun kata kunci yang dimasukkan lebih dari satu kata kunci.

b. Perancangan

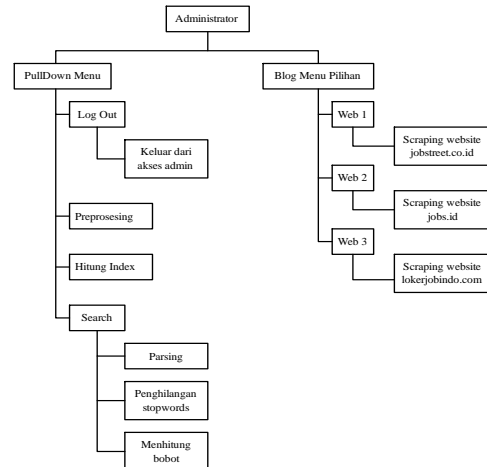
Alur Proses

Sistem ini nantinya akan dibagi menjadi tiga bagian besar, yaitu proses *web scraping* yang berhubungan dengan pengambilan informasi-informasi lowongan kerja terbaru, *indexing* yang berhubungan dengan informasi-informasi dan *query* yang berhubungan dengan pengguna. Pengaturan *web scraping* dan *indexing* berbasis teks menjadi kumpulan indeks istilah akan dilakukan oleh administrator. Sedangkan proses *query* pengguna akan direpresentasikan melalui pengiriman kata kunci berupa teks untuk diproses menjadi *query* yang dapat digunakan oleh sistem ini untuk mencari informasi yang tersimpan di *database*.



Gambar 1. Alur Proses

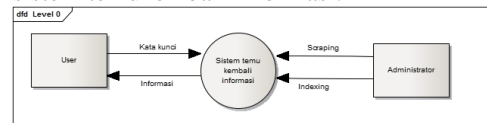
Desain Menu



Gambar 2. Struktur Menu

Data Flow Diagram

Gambar 3 adalah *Context Diagram* dari sistem temukembali informasi.

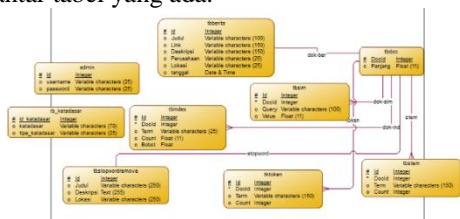


Gambar 3. Context Diagram Sistem Temu Kembali Informasi

Dari bentuk diagram konteks sistem temu kembali informasi, maka bentuk DFD level 1 dari sistem temu kembali informasi dapat dilihat pada gambar 3. Pada gambar 3. DFD level 1 dapat dilihat 3 proses utama dalam sistem temu kembali informasi, yaitu proses *Scraping* yang diberi nama *Web Scraping*, proses *Indexing* yang diberi nama mesin indeks dan proses *Query* yang diberi nama proses cari.

Perancangan Database
Entity Relationship Diagram

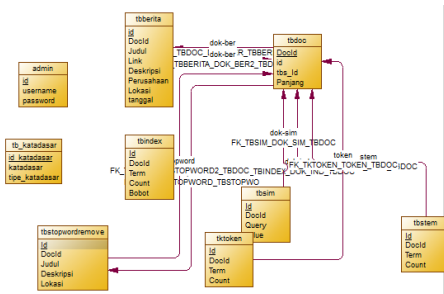
Gambar 4. merupakan ERD dari sistem temu kembali informasi. Gambar tersebut menampilkan seluruh hubungan antar tabel yang ada.



Gambar 4. ERD Sistem Temu Kembali Informasi

Physical Model

Bentuk *physical model* dari ERD sistem temu kembali informasi dapat dilihat pada Gambar 5.

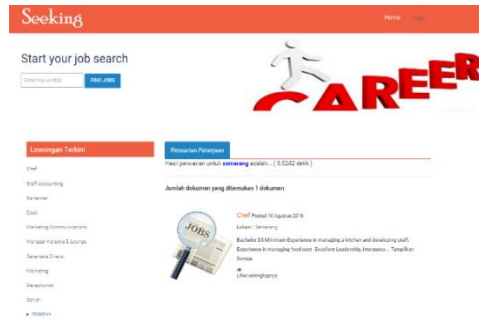


Gambar 5. Physical Model dari ERD Sistem Temu Kembali Informasi

3. HASIL DAN PEMBAHASAN

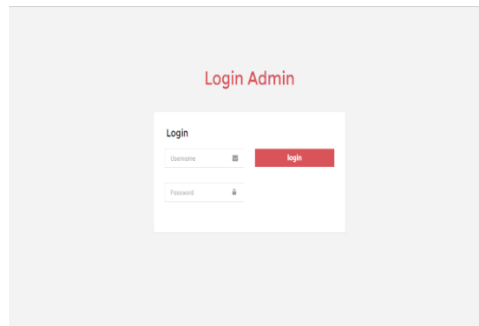
a. Implementasi

Pada Segemen program disusun sebagai berikut: Halaman Utama User, Login Admin, Halaman Utama Admin, Halaman Hasil Scraping Admin, Halaman Hasil Penghilangan Stopword Admin, Halaman Hasil Tokenisasi Admin, Halaman Hasil Stemming Admin, Halaman Penghitungan Bobot, Halaman Pencarian Admin, Halaman Hasil Kemiripan Kata Kunci Dengan Hasil Dokumen.



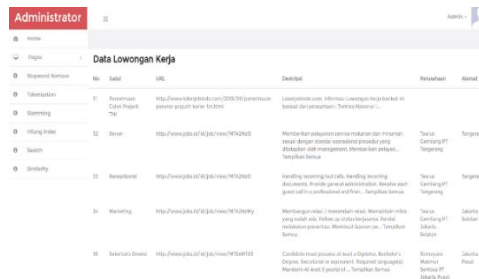
Gambar 6. Halaman Utama User

Sebelum menambahkan dokumen, administrator harus melakukan login ke dalam sistem dengan memasukkan *username* dan *password*. Sistem selanjutnya akan melakukan verifikasi *username* dan *password* yang dimasukkan oleh administrator tadi dengan cara mencari nama *user* dan *password* tersebut pada tabel admin di dalam *database*.



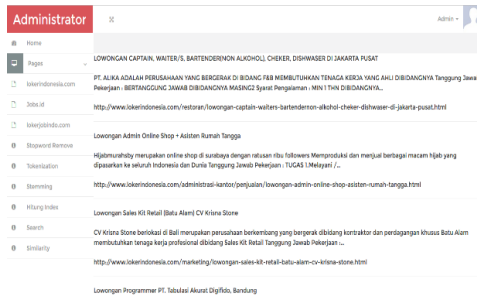
Gambar 7. Login Admin

Halaman utama admin merupakan fasilitas admin untuk menginputkan, mengupdate, maupun menghapus data.

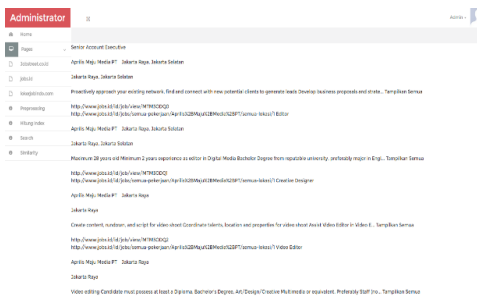


Gambar 8. Halaman Utama Admin

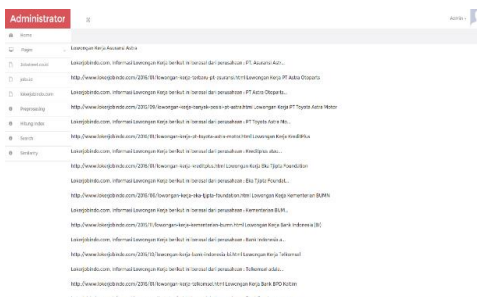
Merupakan halaman hasil *scraping* data atau lowongan pekerjaan terkini yang berasal dari halaman website lowongan kerja.



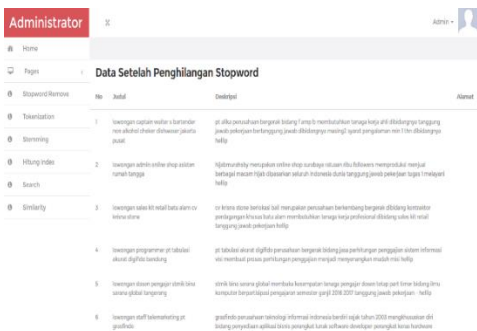
Gambar 9. Halaman Hasil Scraping LokerIndonesia.com



Gambar 10. Halaman Hasil Scraping Jobs.id

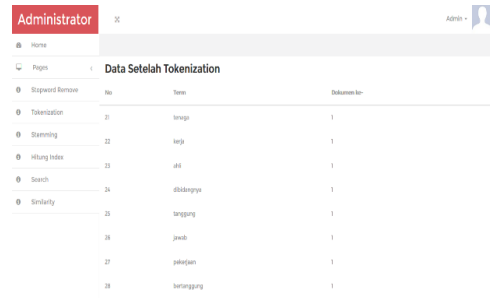


Gambar 11. Halaman Hasil Scraping Lokerjobindo.com



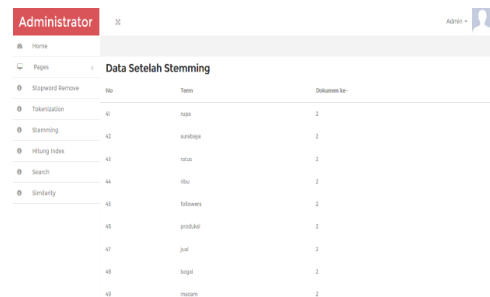
Gambar 12. Halaman Hasil Penghilangan Stopword

Hasil dari penghilangan stopword nantinya akan dipecah menjadi per-kata seperti di bawah ini.



Gambar 13. Halaman Hasil Tokenisasi

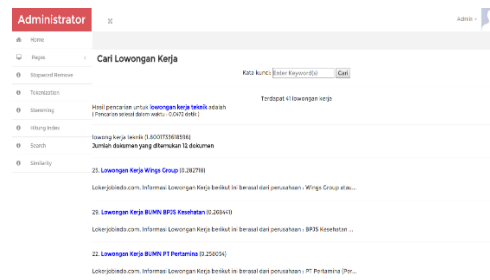
Setelah di pecah menjadi per-kata maka setiap kata berbahasa Indonesia yang memiliki imbuhan maka akan dijadikan menjadi kata dasar seperti ini.



Gambar 14. Halaman Hasil Stemming

b. Pembahasan Hasil Percobaan

Pada sistem temu kembali informasi ini maka pada saat dilakukan percobaan pencarian oleh admin maka akan ditemukan dokumen hasil dari pencarian dengan kata kunci yang sesuai.



Gambar 15. Halaman Administrator

Kemudian setelah dilakukan percobaan pencarian oleh admin maka dari setiap kata kunci akan memiliki nilai kemiripan dengan dokumen yang ditemukan untuk menentukan dokumen mana yang paling relevan dengan kata kunci yang dimasukkan. Sehingga dapat diketahui berapa nilai kemiripan antara kata kunci yang dimasukkan pengguna dengan hasil yang ditemukan.

stopword removal	No	Query	Doc-Id	Sim
Tokenization	1	keang ke-pa beak	25	0.20278
Stemming	2	keang ke-pa beak	26	0.20240
Hitung bobot	3	keang ke-pa beak	22	0.20254
Search	4	keang ke-pa beak	23	0.20229
Selanjutnya	5	keang ke-pa beak	27	0.20289
	6	keang ke-pa beak	28	0.20289
	7	keang ke-pa beak	24	0.20428
	8	keang ke-pa beak	28	0.20430
	9	keang ke-pa beak	27	0.20789
	10	keang ke-pa beak	30	0.20823
	11	keang ke-pa beak	31	0.20822

Gambar 16. Halaman Administrator

4. SIMPULAN

Berdasarkan hasil penelitian yang berupa pengembangan sistem pencarian dengan metode temu-kembali informasi dapat diambil sebuah kesimpulan yaitu:

1. Sistem temu-kembali informasi yang dibuat dapat mencari informasi dari isi file di dalam dokumen yang disimpan di dalam *database*.
2. Proses pengindeksan dokumen didalam sistem temu-kembali informasi yang dikembangkan melalui beberapa tahapan pemrosesan teks, yaitu penghilangan *stopwords*, tokenization, stemming dan penghitungan nilai bobot setiap kata yang akan dijadikan indeks. Sedangkan untuk proses pencariannya juga melalui beberapa tahapan proses yang hampir sama dengan proses pengindeksan, yaitu penghilangan *stopwords*, stemming, cek frasa dan yang terakhir adalah penghitungan fungsi kesamaan antara kata kunci yang dimasukkan oleh *user* dengan dokumen yang ada di dalam *database*.
3. Stemming yang dilakukan adalah terhadap kata yang berbahasa Indonesia saja.

5. REFERENSI

- [1] Anhar. (2010). *Panduan Menguasai PHP & MySQL Secara Otodidak*. Jakarta: Mediakita.
- [2] Bunyamin, H. (2011). "Algoritma Umum Pencarian Informasi Dalam Sistem Temu Kembali Informasi Berbasis Metode Vektorisasi Kata dan Dokumen." *Jurnal Informatika UKM*. 1(2).
- [3] Dini dan Kemal. (2014). "Aplikasi Pencarian Kata Dasar Dokumen

Berbahasa Indonesia Dengan Metode Stemming Porter Menggunakan PHP & MYSQL." *Jurnal Informatika*. 1(8).

- [4] Jasman, Mira Musrini dan Wildan Denny. (2013). "Implementasi Metode Generalized Vector Space Model Pada Aplikasi Information Retrieval." *Jurnal Informatika*. 1(4).
- [5] Manning, Christopher D., Prabhakar Raghavan dan Hinrich Schutze. (2009). *An Introduction to Information Retrieval*. England: Cambridge University Press.
- [6] Pressman, R.. (2005). *SOFTWARE ENGINEERING: A Practitioner's Approach Sixth Edition*. New York: McGraw-Hill.
- [7] Solichin, A. (2005). *Pemrograman Web dengan PHP dan MySQL*. Jakarta: Universitas Budi Luhur.
- [8] Turland, M. (2010). *Php architect's Guide to Web Scraping with PHP*. Toronto: Marco Tabini & Associates, Inc.
- [9] Ward, J. (2013). *Get Up and Running With the Basic Techniques of Web Scraping Using PHP*. Birmingham-Mumbai.

