

PROSIDING

PENINGKATAN DAYA SAING BANGSA MELALUI
PENGEMBANGAN DAN PEMANFAATAN
TEKNOLOGI INFORMASI



2013

SEMINAR TEKNIK INFORMATIKA DAN SISTEM INFORMASI

PROSIDING

SeTISI 2013

**Seminar Teknik Informatika dan Sistem Informasi
Fakultas Teknologi Informasi
Universitas Kristen Maranatha**

Bandung, 6 April 2013

PROSIDING

SeTISI 2013 Seminar Teknik Informatika dan Sistem Informasi Fakultas Teknologi Informasi, Universitas Kristen Maranatha

Editor: **Robby Tan**

Desain Sampul: **Laurentius Risal**

Penerbit:

Maranatha University Press (MUP)

Jl. Prof. Drg. Suria Sumantri, MPH No. 65

Bandung 40164

Cetakan pertama, 2013

Hak cipta dilindungi undang-undang

Perpustakaan Nasional: Katalog Dalam Terbitan (KDT)

Prosiding SeTISI 2013: Prosiding Seminar Teknik Informatika dan Sistem Informasi 2013

Peningkatan Daya Saing Bangsa Melalui Pengembangan dan Pemanfaatan Teknologi Informasi / editor: Robby
Tan, Bandung, MUP, 2013

299 hlm, 21 × 29,7 cm

ISBN 978-602-98685-3-1

KOMITE

Pelindung

Rektor Universitas Kristen Maranatha

Penanggung Jawab

Dekan Fakultas Teknologi Informasi
Universitas Kristen Maranatha

Ketua Pelaksana

Dr. Andi Wahyu Rahardjo Emanuel, BSEE., MSSE.

Komite Program

Dr. Andi Wahyu Rahardjo Emanuel, BSEE., MSSE. (UKM)

Ir. Dana Indra Sensuse, MLIS, Ph.D. (UI)

Dr. Ir. Husni Setiawan Sastramihardja, M.T. (ITB)

Ito Wasito, Ph.D. (UI)

Ir. Kridanto Surendro, M.Sc., Ph.D. (ITB)

Dr. Ir. Mewati Ayub, M.T. (UKM)

Dr. dr. Oerip S. Santoso, M.Sc. (ITB)

Drs. Retantyo Wardoyo, M.Sc., Ph.D. (UGM)

Dr. Ir. Rila Mandala, M.Eng. (ITB)

Dra. Sri Hartati, M.Sc., Ph.D. (UGM)

Yenni M. Djajalaksana, Ph.D. (UKM)

Komite Pelaksana

Radiant Victor Imbar, S.Kom., M.T.

Doro Edi, S.T., M.Kom.

Tanti Kristanti, S.T., M.T.

Hendra Bunyamin, S.Si., M.T.

Hapnes Toba, M.Sc.

Yenni M. Djajalaksana, Ph.D.

Robby Tan, S.T., M.Kom.

Maresha Caroline Wijanto, S.Kom.

Laurentius Risal, S.T.

Meliana Christianti J., S.Kom., M.T.

Daniel Jahja Surjawan, S.Kom., M.T.

Diana Trivena Yulianti, S.Kom., M.T.

Tjatur Kandaga, S.Si., M.T.

Sendy Ferdian, S.Kom.

KATA PENGANTAR

Puji syukur kami haturkan ke hadirat Tuhan Yang Maha Kuasa karena kasih dan rahmat-Nya maka Seminar Teknik Informatika dan Sistem Informasi 2013 (SeTISI 2013) dapat dilaksanakan.

Seminar Teknik Informatika dan Sistem Informasi 2013 (SeTISI 2013) merupakan seminar nasional kedua yang dilaksanakan oleh Fakultas Teknologi Informasi Universitas Kristen Maranatha setelah pada tahun 2011 yang lalu telah terselenggara SeTISI 2011. Adapun tema yang kami usung pada seminar ini adalah “Peningkatan Daya Saing Bangsa Melalui Pengembangan dan Pemanfaatan Teknologi Informasi”. Seminar ini merupakan ajang bertukar pikiran dan pemberian sumbangsih dari para pakar dan akademisi yang pada gilirannya nanti bisa memberikan andil dalam peningkatan daya saing bangsa Indonesia di ajang regional maupun global.

Hingga batas waktu penerbitan naskah yang telah ditentukan, kami menerima 52 karya ilmiah yang dapat dipresentasikan dalam SeTISI 2013 ini. Adapun bidang keilmuan dari karya-karya ilmiah ini mencakup Rekayasa Perangkat Lunak, Multimedia, Jaringan, Keamanan Informasi, Sistem Cerdas, dan Sistem Informasi.

Panitia mengucapkan banyak terima kasih kepada Universitas Kristen Maranatha, Komite Program, Panitia Pelaksana, Keynote Speaker, sponsor dan seluruh peserta yang berpartisipasi aktif memberikan dukungan sehingga SeTISI 2013 dapat terlaksana dengan baik.

Akhir kata, Panitia mengucapkan selamat datang bagi seluruh peserta dan pemakalah SeTISI 2013 di kampus Universitas Kristen Maranatha. Semoga kita semua selalu dalam perlindungan dan bimbingan dari Tuhan Yang Maha Kuasa.

Bandung, 6 April 2013
Ketua Panitia SeTISI 2013

Dr. Andi Wahyu Rahardjo Emanuel, BSEE., MSEE.

SAMBUTAN DEKAN

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa, karena kasih dan rahmat-Nya maka Seminar Teknik Informatika dan Sistem Informasi 2013 (SeTISI 2013) yang diselenggarakan oleh Fakultas Teknologi Informasi, Universitas Kristen Maranatha dapat terlaksana pada hari ini. SeTISI 2013 merupakan seminar nasional kedua yang diselenggarakan oleh Fakultas Teknologi Informasi.

Kami mengharapkan SeTISI 2013 ini dapat dimanfaatkan sebagai salah satu sarana untuk publikasi ilmiah dari karya penelitian yang dilakukan oleh dosen/peneliti dari Universitas Kristen Maranatha dan perguruan tinggi lainnya, khususnya yang memiliki bidang penelitian teknik informatika dan sistem informasi. Melalui SeTISI 2013 ini, gagasan atau hasil penelitian yang telah diperoleh dapat disebarluaskan dan dipublikasikan, sehingga peneliti, akademisi, dan praktisi dapat saling bertukar informasi di bidang teknologi informasi, serta dapat memberi sumbangsih bagi kemajuan ilmu di bidang teknologi informasi di Indonesia.

Atas terselenggaranya SeTISI 2013 ini, kami menghaturkan banyak terima kasih kepada berbagai pihak yang telah berperan serta sehingga seminar dapat terlaksana dengan baik, khususnya kepada Komite Program, yaitu Ir. Kridanto Surendro, M.Sc., Ph.D. (ITB), Dr. dr. Oerip S. Santoso, M.Sc. (ITB), Dr. Ir. Husni Setiawan Sastramihardja, M.T. (ITB), Dr.Ir. Rila Mandala, M.Eng (ITB), Drs. Retantyo Wardoyo, M.Sc., Ph.D. (UGM), Dra. Sri Hartati, M.Sc, Ph.D (UGM), Ir. Dana Indra Sensuse, MLIS., Ph.D. (UI), dan Ito Wasito, Ph.D. (UI). Ucapan terima kasih kami sampaikan juga kepada seluruh panitia pelaksana serta pemakalah yang telah berpartisipasi dalam diseminasi karya ilmiah ini.

Selamat mengikuti SeTISI 2013, semoga kegiatan ini dapat membantu meningkatkan daya saing bangsa Indonesia, khususnya dalam pengembangan dan pemanfaatan teknologi informasi. Kiranya Tuhan memberkati dan menyertai kita semua.

Bandung, 6 April 2013

Dr. Ir. Mewati Ayub, M.T.
Dekan Fakultas Teknologi Informasi
Universitas Kristen Maranatha

DAFTAR ISI

KOMITE	i
KATA PENGANTAR	ii
SAMBUTAN DEKAN	iii
DAFTAR ISI	iv
Penggunaan Metode <i>Paper Prototype</i> untuk Melakukan Inspeksi <i>Usability</i> pada Aplikasi Berbasis <i>Web</i> (Studi kasus: Sistem Informasi Akademik Universitas)	
R. Sandhika Galih A.	1
Perancangan <i>Game “Onion Boy”</i> Berbasis Android untuk Melatih Kecepatan dan Fleksibilitas	
Irene A. Lazarusli, Aditya R. Mitra, Kenny Saputra	6
Algoritma Penggantian <i>Cache</i> sebagai Optimalisasi Kinerja pada <i>Proxy Server</i>	
Suandra Eka Saputra, Timotius Witono	12
<i>Website</i> Perhitungan Angka Kecukupan Gizi Anak	
Pratiwi Chandraningsih, Diana Trivena Yulianti	18
Pengamanan Jalur Komunikasi Internet Menggunakan PPTP (<i>Point-to-Point Tunnelling Protocol</i>)	
I Made Mustika Kerta Astawa, Claudia Dwi Amanda	24
<i>Sentiment Classification</i> Menggunakan <i>Machine Learning</i>: Metode Naïve-Bayes dan <i>Support Vector Machines</i> (Studi kasus: <i>movie reviews imdb.com</i>)	
Hendra Bunyamin, Tjatur Kandaga	29
Analisis IT <i>Governance</i> pada Layanan Teknologi Informasi Perguruan Tinggi Berbasis IT <i>Service Management</i>	
Aradea	37
<i>Monogame Framework</i> sebagai Salah Satu <i>Framework</i> Alternatif pada Mata Kuliah Pemrograman <i>Game</i>	
Erico Darmawan Handoyo, Sulaeman Santoso	43
Penerapan <i>SMS Gateway</i> untuk Pengingat dan Rekomendasi di Rental Komik Daruma	
Teddy Marcus Zakaria, Inwan Aditya Halim	47
Penerapan Algoritma Bayesian <i>Classification</i> untuk Pemberian Harokat pada Kalimat Bahasa Arab	
Maliki Ahmad Nur, Irfan Maliki	53
<i>Website</i> Penyedia Informasi Pariwisata di Kota Bandung Menggunakan <i>Ruby on Rails</i>	
Resky Bagja Sunjaya, Robby Tan	58
<i>E-Services Customer Management System</i> Unit Pelayanan PT. XYZ	
Eka Widhi Yunarso	65
Analisis Perbandingan Unjuk Kerja Protokol TCP, UDP, dan SCTP Menggunakan Simulasi Lalu Lintas Data Multimedia	
Rinda Tri Yuniar Anggraeni, Jusak, Anjik Sukmaaji	72
<i>Best Practices for Choosing Non-Intrusive but Effective CAPTCHAs</i>	
Setia Budi	78
Deteksi Otomatis Perubahan Pustaka API dengan Solusi Sistem Repositori Kode Sumber dan Revisi API Pustaka Perangkat	

Aditya Ideawan, Siti Rochimah	83
Metodologi Pengembangan Sistem Informasi Berbasis <i>Web</i> Menggunakan Pendekatan <i>Software Engineering</i>	
Shelvy Arini, Wahyudianto	89
Rancang Bangun Desain <i>Game</i> Cagar Budaya Kota Semarang bagi Anak Usia 9-10 Tahun sebagai Bagian dari Media Edukatif Nasional dan Wujud Sosialisasi Peninggalan Sejarah	
Dzuha Hening Yanuarsari.....	95
Analisis, Perancangan, dan Implementasi Aplikasi Kalender Akademik Fakultas Teknologi Informasi	
Danny Aguswahyudi, Meliana Christianti J.	101
Menuju Perencanaan Persediaan Obat Berbasis <i>Data Mining</i> pada Instalasi Farmasi Rumah Sakit	
Zainudin Zukhri, Sri Hartati	106
Sistem <i>E-Learning</i> pada Sekolah Menengah Atas Menggunakan Bahasa Pemrograman PHP	
Yustecia Andika Efdom, Doro Edi.....	112
Pengukuran Tingkat Kematangan Tatakelola TI <i>Domain Acquire and Implement (AI)</i> di Politeknik Telkom	
Heru Nugroho	118
Sistem Informasi Penjualan Pembelian Akuntansi dengan Sistem Pengambilan Keputusan <i>Trend Moment</i> untuk Menganalisa Peramalan Penjualan Barang	
Radiant Victor Imbar, Rizky Ananda	123
Pengoptimalan Penerapan Algoritma Genetik dalam Masalah Penjadwalan Sidang	
Mewati Ayub, Andi Irvan Widjaja	131
Kajian Faktor-Faktor Penunjang Peranan Strategis TIK untuk Menunjang Pembelajaran di Perguruan Tinggi	
Hilyah Magdalena.....	136
Pembangkitan Animasi Struktur Data Sederhana melalui Pemetaan Kode Program	
Aditya R. Mitra.....	142
Pembobotan Fitur Tekstual dengan Inferensi Metaheuristik untuk Pengurutan Jawaban	
Hapnes Toba, Setia Budi	147
Perencanaan Arsitektur <i>Enterprise</i> untuk Mendukung Strategi Pengembangan Sistem Informasi (Studi Kasus: PT. ABC)	
Paramita Mayadewi	153
Perumusan Strategi dan Kebijakan Teknologi Informasi untuk Usaha Kecil Menengah (UKM) di Indonesia	
Novi Sofia Fitriasari	159
Aplikasi Pengelolaan Soal Latihan Berbasis <i>Web Bimbel Link</i>	
Dodi Sulistio, Maresha Caroline Wijanto.....	166
Perbandingan Efektifitas Model Pembelajaran <i>Hybrid</i> dan Non Konvensional Mata Kuliah Kewirausahaan Berbasis Multimedia	
R. Reza El Akbar	171
Studi Kasus Evolusi Proyek Perangkat Lunak <i>Open Source Weka</i>	
Andi Wahyu Rahardjo Emanuel.....	175

Aplikasi Pemesanan Perhiasan dan Perhitungan Hasil Produksi (Studi Kasus: Toko Emas Macan)	
Andreanto Abeth Saputra, Daniel Jahja Surjawan.....	180
Sistem Pemodelan Perpindahan <i>Terminal-User</i> secara Terpola untuk Mengukur Pola Perubahan <i>Throughput</i> pada Topologi MANET	
S.N.M.P. Simamora, T. Juhana, Kuspriyanto, N. R. Bagjarasa.....	186
Tren Kebutuhan Kompetensi Kerja Teknologi Informasi di Pasar Kerja Industri Indonesia	
Yenni Merlin Djajalaksana, Tiur Gantini	192
Aplikasi Sistem Keperawatan Rumah Sakit Paru dr. H. A. Rotinsulu	
Ricardo Manarintar Simarmata, Daniel Jahja Surjawan.....	198
<i>Filter-based Feature Selection</i> pada Kategorisasi Artikel Berita Berbahasa Indonesia	
Yan Puspitarani.....	204
Implementasi Politepedia sebagai Portal <i>Knowledge Management System</i> pada Politeknik Telkom	
Suryatiningsih, Dhea Shavera.....	210
Analisis Keamanan Informasi Alat Pembayaran Transaksi <i>E-Commerce</i>	
Husni Mubarak, Aradea, Ismail Salam.....	215
Analisis dan Desain Kebutuhan Fungsionalitas Sistem Persediaan Obat di Apotek	
Inne Gartina Husein.....	222
Model Rancangan Sistem Informasi Persediaan Barang: Studi Kasus STMIK Atma Luhur	
Elly Yanuarti.....	226
Analisis dan Simulasi Pemodelan <i>Cellular Automata (CA)</i> dan Algoritma Optimasi <i>Artificial Bee Colony (ABC)</i> dalam Penjadwalan Lampu Lalu Lintas	
Zenfrison Tuah, Dede Rohidin, Z.K. Abdurahman	231
Analisa Kesenjangan Tatakelola Teknologi Informasi untuk Proses Pengelolaan TI Menggunakan COBIT (Studi Kasus: Pemerintah Daerah Kabupaten Bandung)	
Dede Rohidin.....	237
Pengembangan Perangkat Lunak Asesmen Kerja Tim	
Fariska Zakhralatifa Ruskanda.....	242
Sistem Informasi “<i>Backpack-Traveler System</i>” pada Platform Android dengan Memanfaatkan Framework kSOAP2	
Ryan Permana, Djoni Setiawan K.....	247
Implementasi Politeidroid sebagai Solusi Akses Informasi Akademik bagi Mahasiswa Politeknik Telkom	
Dedy Rahman Wijaya, Irfani Arief, Mirza Febrian Ekaputra.....	253
Pengembangan Perangkat Lunak <i>New Queuing System</i> di Bank	
Maniah.....	257
Implementasi <i>Kinect</i> untuk <i>Future Kindergarten</i>	
Yahdi Siradj.....	262
Peran Bioinformatika dalam Penelitian Kanker	
Teresa Liliana Wargasetia	266
<i>Swarm Intelligence Bee Colony</i> Menggunakan Teori <i>Chaos</i> pada Permasalahan Psikologi Emosi	
Widyastuti Andriyani, Retanty Wardoyo	270

Optimalisasi Proses Komputasi melalui Pengaturan Penyeimbangan Beban Sumber Komputasi dengan Perpaduan Algoritma <i>Genetic</i> dan <i>Tabu Search</i> di Lingkungan Komputasi <i>Grid</i> Irfan Darmawan, Kuspriyanto, Yoga Priyana, Ian Yosep M.E	274
Implementasi Algoritma Rivest-Shamir-Adleman (RSA) untuk Keamanan Data pada Sistem Informasi Berbasis <i>Web</i> (Studi Kasus: Universitas X) Tanti Kristanti, Nurul Amanda	280
Analisis Perbandingan Unjuk Kerja Algoritma <i>Congestion Control</i> pada TCP Tahoe, Reno dan SACK (<i>Selective Acknowledgment</i>) Yuliana Wahyu Putri Utami, Jusak, Anjik Sukmaaji.....	286

Pembobotan Fitur Tekstual dengan Inferensi Metaheuristik untuk Pengurutan Jawaban

Hapnes Toba^{#1}, Setia Budi^{*2}

[#]Program Studi D3 Teknik Informatika, Universitas Kristen Maranatha
Jalan Suria Sumantri 65 Bandung 40164

¹hapnes.toba@itmaranatha.org

^{*}Program Studi S1 Sistem Informasi, Universitas Kristen Maranatha
Jalan Suria Sumantri 65 Bandung 40164

²setia.budi@itmaranatha.org

Abstract — In this paper, an approach to construct a dynamic feature weighting in question answering system is proposed. In order to find a set of optimum weights for a question and its answer (candidate) pair(s), a combination of genetic algorithm and logistic regression technique are introduced in our approach. As part of our experimental study, a standard dataset of non-factoid question and answer pairs, called UIUC, is used as a testing-case. The experiment results show that our approach can produce a set of feature weights which leads to a better performance compared to the conventional static weighting.

Keywords — algoritma genetika, metaheuristik, pengurutan jawaban, regresi logistik sistem tanya jawab

I. PENDAHULUAN

Sistem tanya jawab (STJ) adalah sebuah penerapan temu balik informasi yang diharapkan dapat memberikan sebuah kandidat jawaban secara langsung. Ditinjau dari sisi arsitektur perangkat lunak, STJ memiliki empat komponen utama [1], yaitu: 1. analisis pertanyaan; 2. pembentuk kueri; 3. mesin pencarian; 4. validasi jawaban. Salah satu faktor yang dapat menentukan performa STJ adalah kemampuan sistem untuk memprediksi pada bagian teks yang manakah jawaban tersirat. Dengan kata lain, sebelum validasi jawaban dilakukan, sistem harus terlebih dahulu memastikan bahwa kandidat-kandidat jawaban baik berada pada urutan awal hasil temu balik informasi.

Penelitian-penelitian terkini perihal pengurutan kandidat jawaban banyak mengandalkan pembelajaran empiris dengan menerapkan teknik pembelajaran mesin saat menentukan bobot fitur. Beberapa teknik pembelajaran mesin yang sering diterapkan antara lain: *support vector machine* (SVM) [2], *perceptron* [1, 3], dan regresi logistik [4]. Melalui pembelajaran mesin tersebut, setiap fitur yang digunakan akan mendapatkan bobot yang tetap sesuai dengan model yang dihasilkan dari pembelajaran. Salah satu kelemahan dengan pendekatan pembelajaran mesin seperti ini adalah: dimungkinkannya ketidakcocokkan pada saat bobot dipakai dalam lingkungan data yang sifatnya berbeda.

Untuk inilah diperlukan adanya semacam pendekatan yang memungkinkan bahwa bobot dapat dihasilkan secara

dinamis mengikuti karakteristik dari obyek yang sedang dipelajari, dalam hal ini adalah pasangan pertanyaan dengan (kandidat) jawabannya. Pembobotan dinamis seperti ini diharapkan dapat meniru cara manusia dalam menentukan jawaban, yaitu dengan melakukan inferensi dari karakteristik jawaban secara independen kasus per kasus, tanpa dipengaruhi karakteristik jawaban lainnya [5, 6].

Dengan mengacu pada beberapa permasalahan dan potensi sebagaimana tertulis di atas, maka dalam makalah ini penulis mengajukan beberapa hal untuk dijadikan sebagai pertanyaan riset:

1. Bagaimana mempelajari karakteristik sebuah pasangan pertanyaan dan jawaban?
2. Bagaimana komposisi bobot untuk setiap pasangan dapat diidentifikasi secara unik?
3. Bagaimana pengaruh karakteristik yang teridentifikasi pada hasil pengurutan jawaban?

Untuk menjawab pertanyaan riset di atas, di dalam penelitian akan digunakan pendekatan metaheuristik [7]. Perhitungan secara metaheuristik menjamin bahwa untuk setiap nilai yang hendak dicapai - misalnya: maksimum skor yang menggambarkan kekuatan relasi antara sebuah pertanyaan dengan kandidat jawabannya - akan terdapat komposisi nilai fitur yang optimum sehingga nilai akhir yang hendak dicapai juga optimum [8]. Pendekatan dengan dinamisasi bobot seperti ini, diharapkan memiliki nilai lebih dibandingkan dengan konsep pembelajaran mesin yang biasa, yaitu dapat menghasilkan bobot tanpa adanya supervisi (*unsupervised learning*).

Dalam kaitannya dengan STJ, kita dapat berhipotesa bahwa jika sebuah pasangan pertanyaan-jawaban dipandang sebagai pasangan yang unik dan baik, maka akan dihasilkan sebuah komposisi bobot yang optimum, yang menggambarkan derajat "kebaikan" antara pertanyaan dengan jawaban. Dengan cara seperti ini diharapkan bahwa semua kandidat jawaban baik akan berada pada urutan yang lebih baik (di urutan atas), dibandingkan dengan kandidat yang kurang baik.

II. METODOLOGI

A. Inferensi sebagai Problem Optimasi

Dalam penelitian ini, inferensi didefinisikan sebagai sebuah problem metaheuristik untuk mencari bobot fitur yang optimum sehingga dapat mengidentifikasi karakteristik setiap pasangan pertanyaan dengan jawaban. Kerangka kerja metaheuristik sebagaimana diuraikan dalam [7], menjamin bahwa persoalan-persoalan kombinatorial, seperti halnya mencari komposisi nilai optimum dari beberapa parameter, akan dapat diselesaikan. Unsur utama yang perlu diperhatikan untuk sebuah metaheuristik adalah representasi persoalan ke dalam sebuah rangkaian formula yang dianggap cocok untuk merepresentasikan persoalan.

Kerangka kerja umum untuk persoalan metaheuristik dapat dilihat dalam Algoritma 1.

Cari sebuah himpunan solusi awal θ_0 ,
dan buat nilai $k = 0$

Repeat:

1. Identifikasi solusi bertetangga $N(\theta_k)$ dari solusi-solusi saat ini.
2. Pilih kandidat solusi $\{\theta_c\} \in N(\theta_k)$ dari solusi bertetangga.
3. Terima kandidat dan himpunan $\theta_{k-1} = \theta_c$ atau tolak, dan buat nilai $\theta_{k-1} = \theta_k$
4. Naikkan nilai $k = k + 1$

Until kondisi terminasi dipenuhi

Algoritma 1 Kerangka Kerja Metaheuristik.

Beberapa contoh metaheuristik yang sering digunakan adalah: tabu search, algoritma genetika, dan simulated annealing. Dalam penelitian ini, digunakan pendekatan algoritma genetika dengan representasi obyektif tunggal (single objective). Salah satu kelebihan algoritma genetika adalah kemampuannya untuk menelaah dengan ukuran yang relatif luas di mana pendekatan pencarian linear yang konvensional dinilai tidak lagi relevan untuk diaplikasikan. Algoritma genetika memiliki tiga operator utama dalam melakukan proses pencarian, yaitu: seleksi, persilangan dan mutasi. Operator mutasi memungkinkan sebuah proses pencarian terhindar dari keterpakuan pada nilai optimum lokal tertentu [7, 8].

Implementasi algoritma genetika yang digunakan dalam penelitian ini adalah NSGAI [8] dari pustaka Java bernama jMetal [9]. Konfigurasi operator genetika yang digunakan adalah sebagai berikut:

- seleksi = BinaryTournament2
- persilangan = SBXCrossover, dengan peluang 0.7 [17]
- mutasi = PolynomialMutation, dengan peluang 0.1 [17]

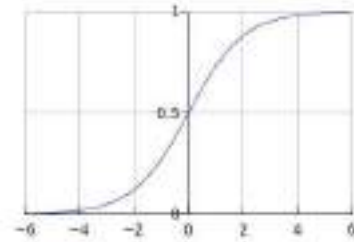
Batasan jumlah generasi dan juga ukuran maksimum dari tiap populasi yang dihasilkan di tiap generasi adalah: generasi: 1000, populasi: 50. Konfigurasi ini mengacu pada penelitian [10], yang memberikan hasil positif untuk optimasi kueri pada mesin temu balik.

A. Inferensi dengan Regresi Logistik

Dalam pembahasan di atas disebutkan bahwa dalam penelitian ini digunakan representasi obyektif tunggal. Dalam hal ini, optimasi ditujukan untuk memaksimalkan nilai fungsi logistik sebagai kombinasi linear hasil perkalian bobot dengan nilai fitur. Fungsi logistik banyak dimanfaatkan untuk mempelajari karakteristik data dengan memberikan contoh positif/ negatif, misalnya dalam [4, 6].

Dalam pembelajaran dengan fungsi logistik, sesuatu dianggap "baik" jika nilai fungsi sama dengan atau di atas 0.5, dan "buruk" jika nilai fungsi di bawah 0.5. Karakteristik inilah yang dimanfaatkan dalam penelitian, yaitu dengan mengasumsikan bahwa setiap pertanyaan akan berpasangan dengan kandidat jawaban yang "baik". Dengan bertumpu pada asumsi ini, dibangunlah sebuah mesin inferensi yang bertugas untuk menghasilkan komposisi nilai parameter yang optimal, sehingga dapat memaksimalkan nilai dari fungsi logistik.

Konfigurasi regresi logistik yang dilakukan dapat dilihat dalam Gambar 1. Regresi logistik dalam mesin inferensi berperan sebagai agregator, $h_\theta(x)$, dari beberapa fitur pengukuran pada STJ. Dengan demikian akan dihasilkan sebuah nilai tunggal yang kemudian akan dioptimasi oleh algoritma genetika sebagai sebuah representasi obyektif tunggal.



Fungsi Logistik

$$h_\theta(x) = g(\theta^T x); \quad g(z) = \frac{1}{1+e^{-z}}$$

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

Asumsi prediksi "y = 1" jika $h_\theta(x) \geq 0.5$

Asumsi prediksi "y = 0" jika $h_\theta(x) < 0.5$

Regresi Logistik

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m \text{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_\theta(x^{(i)})) \right]$$

Optimasi: $\min J(\theta)$

$$\text{score}(Q, A) = \max \sum_{i=1}^n J(\theta) * f_i(Q, A) + b$$

Gambar 1 Konfigurasi Fungsi Logistik Dan Peran Mesin Inferensi Untuk Memaksimalkan Nilai Skor

Dalam konfigurasi di Gambar 1, dapat dilihat pula bahwa dengan meminimalkan fungsi jarak $J(\theta)$ pada saat regresi dilakukan, maka pada mesin inferensi, sebagaimana dibahas pada bagian II.A, diharapkan akan diperoleh suatu komposisi bobot optimal θ .

B. Pemilihan Fitur Tekstual

Terkait dengan pengukuran kekuatan relasi antara suatu pertanyaan dengan suatu pasangan jawaban, terdapat bermacam-macam fitur yang dapat digunakan dan dikombinasikan satu dengan lainnya [4, 11, 12]. Dalam penelitian ini, fitur-fitur pengukuran yang dipilih adalah fitur-fitur tekstual yang umum digunakan dalam *Community-based Question Answering* (CQA), seperti: *Yahoo!Answer*. Dalam CQA, seorang pengguna dapat mengajukan pertanyaan ataupun menjawab pertanyaan yang diajukan. Dengan demikian CQA memiliki kemampuan inferensi yang lebih dalam dibandingkan STJ konvensional karena jawaban langsung diberikan oleh manusia, dan bukan hasil temu balik informasi [11, 12].

Alasan utama digunakannya fitur-fitur tekstual CQA adalah untuk menjamin bahwa mesin inferensi dapat "mendekati" kemampuan manusia dalam memberikan/menilai sebuah jawaban. Dalam [12] diusulkan penggunaan 31 fitur tekstual yang secara intrinsik mengidentifikasi kualitas sebuah jawaban.

Dari 31 fitur tekstual yang tersedia, kemudian akan dilakukan proses seleksi lebih lanjut sehingga akan dihasilkan sekumpulan fitur yang memiliki signifikansi dalam proses pengukuran kekuatan relasi pasangan pertanyaan dan jawaban. Dalam penelitian ini digunakan dua tahap seleksi fitur, dengan mengkombinasikan:

1. Kuadrat *Chi* (X^2) yang berbasis statistik
2. Algoritma genetika yang berbasis evaluasi korelasi sub-himpunan fitur [13].

Implementasi seleksi fitur dilakukan dengan memanfaatkan pustaka yang terdapat dalam perangkat penggalan data Weka [14]. Algoritma seleksi fitur dapat dilihat pada Algoritma 2.

Hasil akhir seleksi fitur dengan menggunakan Algoritma 2 di atas dapat dilihat pada Tabel I.

TABEL I
HASIL AKHIR SELEKSI FITUR BESERTA BOBOTNYA UNTUK DATASET UIUC
(PEMBAHASAN TENTANG DATA PENELITIAN DAPAT DILIHAT PADA BAGIAN IIIB)

Fitur Terpilih	Nilai Bobot
Jarak kosinus 1 gram	0.34
Jumlah kata unik beririsan	0.30
Jumlah kata benda pada jawaban	0.12
Jumlah kata kerja pada pertanyaan	0.08
Jumlah kata penting pada jawaban	0.07
Jarak kosinus 2 gram	0.06
Jumlah kata benda pada pertanyaan	0.03

Bagi dataset ke dalam komposisi seimbang (50%-50% baik-buruk)

Tahap 1:

For Each: dataset dengan komposisi seimbang

1. Lakukan seleksi fitur kuadrat *Chi* (Weka)
2. Hitung jumlah kemunculan sebuah fitur dengan nilai bobot di atas 0
3. Lakukan seleksi fitur algoritma genetika (Weka)
4. Hitung jumlah kemunculan sebuah fitur di dalam subset dengan % di atas 0

End For

Pilih fitur dengan nilai kemunculan di atas 15 dalam setiap algoritma seleksi fitur

Tahap 2:

1. Hitung rata-rata nilai bobot dari setiap fitur yang terpilih dalam setiap algoritma seleksi fitur
2. Kombinasikan secara linear nilai bobot dari setiap algoritma seleksi fitur
3. Pilih fitur dengan nilai bobot di atas 0.025 setelah

Algoritma 2. Seleksi Fitur Sebagai Kombinasi Antara Pendekatan Statistik Dan Evaluasi Sub Himpunan.

III. RANCANGAN EKSPERIMEN

A. Kerangka Kerja Eksperimen

Kerangka kerja yang dilakukan dalam eksperimen dapat dilihat pada Gambar 2. Dalam Gambar 2 terlihat bahwa untuk melakukan pengurutan ulang jawaban diperlukan adanya hasil temu balik informasi beserta *judgment/*penilaian relevansi hasil temu balik. Untuk saat ini relevansi yang digunakan diambil dari penelitian [15].



Gambar 2 Kerangka Kerja Eksperimen.

Ekstraksi fitur dilakukan dengan menggunakan fitur-fitur tekstual kualitas pasangan pertanyaan-jawaban pada [12]. Seleksi fitur akan dilakukan setelah ekstraksi fitur (lihat juga bagian II.C). Hasil seleksi fitur inilah yang akan dioptimasi dengan algoritma genetika untuk mencari bobot fitur optimal pada setiap pasangan pertanyaan-jawaban, dan dipakai untuk mengurutkan ulang hasil temu balik informasi.

B. Data Penelitian

Data penelitian diambil dari penelitian [15], yaitu sebuah dataset untuk STJ *open domain* dengan tipe pertanyaan kompleks dalam bahasa Inggris. Pertanyaan kompleks disini mengacu pada pertanyaan-pertanyaan non-faktoid, yaitu

pertanyaan mengenai: definisi (*What is ...?*), prosedur (*How ...?*), dan sebab-akibat (*Why ...?*). Di dalam *dataset* terdapat 136 pertanyaan dengan 9558 kandidat jawaban dalam ukuran kalimat, dan 2591 kandidat jawaban dalam ukuran paragraf. Sebagaimana dilaporkan dalam [15], penilaian relevansi untuk *dataset* ini dilakukan oleh manusia dengan mengacu pada urutan hasil temu balik STJ bernama *YourQA* pada korpus AQUAINT-2 [2]. Tipe pertanyaan dalam *dataset* didominasi oleh tipe definisi, sejumlah 123 pertanyaan.

Setiap pertanyaan cenderung memiliki kemungkinan jawaban pada lebih dari satu kalimat/ paragraf. Namun dalam *dataset* ini, proporsi relevansi lebih cenderung pada kalimat/ paragraf yang tidak relevan, dengan perbandingan antara jawaban "relevan/tidak relevan" adalah: 309/9558 ($\pm 3\%$) untuk kalimat dan 106/2591 ($\pm 4\%$) untuk paragraf. Dengan pertimbangan inilah sebenarnya dalam melakukan seleksi fitur perlu dilakukan dengan proporsi seimbang (lihat bagian II.C).

C. Cara Evaluasi dan Sistem Pemanding

Evaluasi performa dilakukan dengan menggunakan dua ukuran metrik, yaitu TRR (*total reciprocal rank*), dan T1-RR (*1st answer reciprocal rank*) [16]. *Reciprocal rank* (RR) memberikan proporsi urutan kemunculan jawaban relevan untuk suatu kueri tertentu, dan difinisikan dalam rumus (1).

$$RR = 1/Rank \dots (1)$$

Secara definisi metrik TRR memberikan nilai jumlah total RR untuk semua kueri, dan metrik T1-RR memberikan nilai jumlah total RR untuk jawaban pertama yang relevan untuk semua kueri.

Sistem pembandingan yang digunakan untuk menilai performa metode optimasi meliputi:

1. *YourQA* [2, 15];
2. Pembobotan dengan regresi logistik biasa
 - a. *Dataset* seimbang (proporsi 50%-50%);
 - b. *Dataset* natural (proporsi: lihat bag. IIIA);
3. *Bayesian Analogical Reasoning* (BAR) [4].

Terkait dengan kebutuhan adanya pelatihan pada sistem pembandingan nomor 2 dan 3, maka *dataset* dibagi ke dalam kelompok data pelatihan dan data pengujian secara acak. Data pelatihan menggunakan 75% dari keseluruhan data (132 pertanyaan; 7132 kalimat; 1.922 paragraf). Data pengujian menggunakan 25% dari data (34 pertanyaan; 2426 kalimat; 669 paragraf).

IV. HASIL UJI COBA

Dalam bagian mengenai hasil uji coba ini disampaikan hasil tercapai dalam eksperimen. Hasil TRR dapat dilihat pada Tabel II.

Beberapa hal yang dapat dilihat dalam Tabel II adalah:

- Untuk jawaban dalam bentuk paragraf: hasil TRR dengan optimasi bobot (16.17) setara dengan sistem pembandingan *YourQA* (16.25), dan jauh mengungguli performa dari sistem lainnya. Hal ini menunjukkan bahwa optimasi bobot (sesuai fitur yang digunakan)

memiliki potensi untuk dapat menemukan jawaban benar dengan meniru kecenderungan manusia dalam menilai jawaban.

TABEL II
HASIL EKSPERIMEN UNTUK TRR

TRR (<i>Total Reciprocal Rank</i>)		
Sistem	Kalimat	Paragraf
YourQA	9.12	16.25
Bal. LR	12.25	12.95
Nat. LR	14.61	15.12
BAR	4.76	11.33
Opt. Bbt.	7.63	16.17

- Untuk jawaban dalam bentuk kalimat: hasil TRR dengan optimasi bobot (7.63) masih belum dapat mengungguli *YourQA* (9.12) ataupun hasil pembelajaran dengan regresi logistik (12.25 untuk data seimbang dan 14.61 untuk data natural). Namun, metode optimasi bobot mampu melampaui salah satu metode termutakhir untuk pengurutan ulang hasil temu balik informasi, yaitu BAR [4].

Hasil eksperimen untuk T1-RR dapat dilihat pada Tabel III. Beberapa hal yang dapat dilihat pada Tabel III adalah:

- Untuk jawaban dalam bentuk paragraf: hasil T1-RR dengan optimasi bobot (9.75) melampaui performa semua sistem lainnya. Hal ini menunjukkan bahwa dengan optimasi bobot, kemungkinan bahwa jawaban pertama yang relevan berada pada urutan atas lebih besar jika dibandingkan dengan metode lainnya.
- Untuk jawaban dalam bentuk kalimat: hasil T1-RR dengan optimasi bobot (4.52), seperti halnya dengan hasil TRR, masih belum dapat mengungguli performa metode lainnya, kecuali BAR (1.26).

TABEL III
HASIL EKSPERIMEN UNTUK T1-RR

T1-RR (<i>1st ans Reciprocal Rank</i>)		
Sistem	Kalimat	Paragraf
YourQA	6.06	9.67
Bal. LR	6.31	7.45
Nat. LR	7.05	8.50
BAR	1.26	4.02
Opt. Bbt.	4.52	9.75

Dalam bagian mengenai hasil uji coba ini disampaikan pula sebuah contoh pasangan pertanyaan dengan kandidat jawaban yang berupa paragraf, sebagai berikut:

- "*What is Wimbledon?*"
- "*Wimbledon became the first \$5-million tennis tournament with the announcement of a 23% increase in prize money.*" *That boosted the purse to \$5.3 million, an increase of \$1 million, for this summer's event.* *The previous record for a tennis tournament was last summer's \$4.4 million at the US Open.* *Wimbledon officials said they expect the Americans to top their amount when the prizes for the tournament in New York in September are announced next month.*"

Dalam Tabel IV, terdapat tujuh fitur yang dioptimasi sebagaimana dijabarkan dalam bagian II.A dan II.B.

TABEL IV
CONTOH OPTIMASI BOBOT

	Cosine_1_Grams	Overlap_1_Grams	Answer_number_nouns	Query_number_verbs	Answer_number_non_stop_words	Query_number_nouns
Nilai Fitur Kalimat	0.14	1	8	1	9	1
Bobot Opt.	0.9484	0.9986	0.9998	0.9996	0.9983	0.9962
Norm. Bobot	0.1395	0.1469	0.1471	0.147	0.1469	0.1465
Nilai Fitur Paragraf	0.1601	1	61	2	34	1
Bobot Opt.	0.911	0.8888	0.9981	0.9828	0.9982	0.8509
Norm. Bobot	0.1526	0.1489	0.1672	0.1647	0.1672	0.1426

Hasil optimasi bobot untuk setiap fitur dapat dilihat pada tabel pada Tabel IV. Untuk bentuk jawaban berupa kalimat (berupa kalimat yang bergaris bawah pada butir contoh di atas), dapat dilihat pada baris kedua. Untuk jawaban berupa paragraf dapat dilihat pada baris keempat. Nilai optimasi dipilih untuk berada dalam cakupan $[0, \dots, 1]$. Normalisasi dari hasil optimasi tersebut - yaitu dengan membuat jumlah bobot keseluruhan sama dengan 1 - dapat dilihat pada baris ketiga untuk bentuk jawaban berupa kalimat, dan pada baris keenam untuk bentuk jawaban berupa paragraf.

Komposisi bobot pada baris ketiga dan keenam ini adalah hasil akhir optimasi yang kemudian dipakai untuk menghasilkan sebuah skor akhir yang menggambarkan nilai "kebaikan" sebuah pasangan pertanyaan-jawaban. Sebagai contoh untuk jawaban dalam bentuk paragraf hasil akhir dari skor adalah 16.39, yang diperoleh dengan menerapkan rumus (2).

$$\text{score}(Q, A) = \sum_{i=1}^7 w_i * \text{feat}(Q, A)_i \dots (2)$$

$$\begin{aligned} \text{score}(Q, A) &= (0.1526 * 0.1601) + (0.1489 * 1) \\ &\quad + (0.1672 * 61) + (0.1647 * 1) \\ &\quad + (0.1672 * 34) + (0.1426 * 1) \\ &\quad + (0.0567 * 0) = \mathbf{16.39} \end{aligned}$$

Contoh perubahan urutan yang terjadi untuk setiap metode pengurutan untuk contoh pada Tabel IV, dapat dilihat pada Gambar 3. Dalam Gambar 3 ini dapat dilihat pengaruh dari setiap metode pengurutan berdasarkan model / rumus dan bobot yang digunakan. Dalam contoh ini, terlihat bahwa urutan yang diperoleh dengan metode optimasi bobot (19) memiliki kesetaraan urutan (17) dengan metode *YourQA* untuk bentuk jawaban kalimat, dan jauh mengungguli (urutan 1 vs. 17) untuk bentuk jawaban berupa paragraf.

Metode	Kalimat	Paragraf
YourQA	19	17
Bal. LR	14	17
Nat. LR	6	17
BAR	5	4
Opt. Bbt.	17	1

$$\text{sim}(Q, A) = \alpha \cdot \text{bow}(Q, A) + \beta \cdot \text{ngram}(Q, A) + \gamma \cdot \text{char}(Q, A) + \delta \cdot \text{hdCM}(Q, A)$$

$$\text{sim}(Q, A) = \log P(1^{19} = 1|1^{19}, 5, 1^1 = 1) - \log P(1^{19} = 1|1^{19})$$

$$\text{sim}(Q, A) = \max \sum_{j=1}^n f_j(\theta) * f_j(Q, A) + b$$

Gambar 3 Contoh Pengaruh Model Pada Urutan Jawaban.

V. DISKUSI SINGKAT TERKAIT HASIL

Dengan melihat pada hasil eksperimen sebagaimana dituliskan pada bagian IV, beberapa hal yang menarik untuk diulas adalah:

- Optimasi bobot dengan tujuh fitur utama yang telah terseleksi lebih cocok untuk digunakan pada jawaban dalam bentuk paragraf. Salah satu faktor yang mungkin menjadi penyebab adalah fitur-fitur yang terpilih lebih banyak menekankan pada "jumlah" karakteristik tertentu, misalnya: jumlah kata penting pada jawaban. Hal ini sangat terpengaruh pada "panjang jawaban, yaitu: jumlah kata dalam jawaban", yang sangat variatif untuk jawaban dalam bentuk kalimat, namun lebih stabil untuk jawaban dalam bentuk paragraf. Salah satu hal yang mungkin untuk dilakukan guna mengatasi hal ini adalah dengan menormalisasi jumlah berdasarkan panjang jawaban atau panjang pertanyaan kebutuhan.
- Optimasi bobot tidak memerlukan data pembelajaran (*unsupervised*), karakteristik pasangan pertanyaan dengan jawaban ditelaah secara unik untuk setiap pasangan. Hal ini akan menguntungkan apabila terdapat ketidakseimbangan jumlah data, misalnya: apabila jumlah jawaban relevan jauh lebih sedikit dibanding jumlah jawaban yang tidak relevan. Salah satu hal yang menunjukkan bahwa pendekatan optimasi bobot sangat menjanjikan adalah hasil pengurutan temu balik yang lebih baik dibanding sistem *YourQA* untuk bentuk jawaban berupa paragraf. Dalam sistem *YourQA* bobot untuk fitur-fiturnya dianggap optimal untuk pasangan pertanyaan-jawaban yang bersifat faktoid dengan bentuk jawaban berupa kalimat [2], dan bersifat statik (lihat juga rumus *YourQA* pada Gambar 3), dalam hal ini nilai $\alpha = 0.6$, $\beta = 0.2$, dan $\gamma = \delta = 0.1$.
- Cakupan nilai optimasi bobot yang dihasilkan sangat rapat, dan masih sulit untuk dapat menganalisis fitur yang memiliki kontribusi dalam menentukan keunikan pasangan pertanyaan dan jawaban.

VI. SIMPULAN DAN KEBERLANJUTAN

Beberapa simpulan yang dapat dituliskan dari hasil eksperimen terkait dengan pertanyaan riset adalah:

- Karakteristik sebuah pasangan pertanyaan dan jawaban dapat dipelajari dari fitur-fitur terpilih yang menunjukkan kualitas "kebaikan" relasi antara keduanya. Dalam penelitian ini digunakan 31 fitur

instrinsik, yang kemudian diseleksi menjadi 7 fitur utama.

2. Komposisi bobot untuk setiap pasangan pertanyaan dan jawaban dapat diidentifikasi dengan menerapkan konsep metaheuristik yang bertugas untuk mencari nilai optimum dari setiap bobot pada fitur. Dalam penelitian digunakan algoritma genetika NSGAI dan fungsi logistik sebagai sebuah fungsi obyektif yang akan dimaksimalkan nilainya.
3. Dari hasil eksperimen pada *dataset* UIUC terlihat bahwa pendekatan optimasi bobot dapat mempengaruhi hasil pengurutan kandidat jawaban, dan berpotensi untuk menggantikan peran pembobotan yang bersifat statik.

Untuk keberlanjutan penelitian, beberapa hal yang masih berpotensi untuk diteliti adalah:

1. Melihat pengaruh kombinasi fitur-fitur berbasis linguistik dan statistik pada pengurutan kandidat jawaban. Pada penelitian saat ini masih terkonsentrasi pada fitur-fitur berbasis statistik.
2. Meneliti pengaruh sebaran dan normalisasi pada nilai fitur serta bobot selama proses pencarian komposisi bobot ideal. Misalnya dengan membandingkan pengaruh nilai bobot antara $[0, \dots, 1]$ dan $[0, \dots, 10]$ terhadap hasil pengurutan temu balik informasi.
3. Menghasilkan sebuah model yang mampu mengagregasi komposisi bobot optimum pada beberapa *dataset* sehingga dapat mengurangi kompleksitas algoritma genetika pada saat mencari komposisi bobot ideal.
4. Memanfaatkan optimasi bobot dalam melakukan proses inferensi analogi [5], yang mungkin dapat berperan untuk menemukan strategi menjawab pertanyaan untuk tipe-tipe pertanyaan faktoid yang memerlukan jawaban tunggal, misalnya pertanyaan-pertanyaan yang terkait dengan pencarian lokasi, nama orang ataupun waktu.

DAFTAR PUSTAKA

- [1] M.W. Bilotti, J. Elsas, J. Carbonell, & E. Nyberg, "Rank Learning for Factoid Question Answering with Linguistic and Semantic Constraints," *Prosiding CIKM'10*, 2010.
- [2] S. Quarteroni, & S. Manandhar, "Designing an Interactive Open-Domain Question Answering System," *Natural Language Engineering*, vol. 15 no. 1, pp. 73-95, 2009.
- [3] M.A. Pasca, & S.M. Harabagiu, "High Performance Question/Answering," *Prosiding SIGIR'01*, 2001.
- [4] R. Silva, K. Heller, Z. Ghahramani, & E.M. Airolidi, "Ranking Relations Using Analogies in Biological and Information Networks," *The Annals of Applied Statistics*, vol. 4, no. 2, pp. 615-644, 2010.
- [5] R.J. Sternberg, "Component Processes in Analogical Reasoning," *Psychological Review*, vol. 84, no. 4, pp. 353-378, 1977.
- [6] J.N. Marewski, & L.J. Schooler, "Cognitive Niches: An Ecological Model of Strategy Selection," *Psychological Review*, vol. 118, no. 3, pp. 393-437, 2011.
- [7] S. Ólafsson, "Metaheuristics," in *Nelson and Henderson (eds.). Handbook on Simulation, Handbooks in Operations Research and Management Science VII*, Elsevier, 633-654, 2006.
- [8] K. Deb, A. Pratap, S. Agarwal, & T. Meyarivan, "A Fast and Elitist Multiobjective Genetic Algorithm: NSGAI," *IEEE Trans. on Evolutionary Computation*, vol. 6 no. 2, 2002.
- [9] J.J. Durillo, A.J. Nebro, & E. Alba, "The jMetal Framework for Multi-Objective Optimization: Design and Architecture," *Prosiding CEC'10*, 2010.
- [10] J. Tiedemann, "Improving Passage Retrieval in Question Answering using NLP," *Progress in Artificial Intelligence LNCS* vol. 3808, pp. 634-646, 2005.
- [11] E. Agichtein, C. Castillo, D. Donato, A. Gionis, & G. Mishne, "Finding High-Quality Content in Social Media," *Prosiding WSDM'08*, 2008.
- [12] H. Toba, M. Zhao-Yan, M. Adriani, & C.T. Seng, "Discovering High-Quality Answer in CQA Archives using a Hierarchy of Classifiers," *submitted to Information Sciences*, Feb 2013.
- [13] H. Liu, & R. Setiono, "A Probabilistic Approach to Feature Selection," *Prosiding ICML'96*, 1996.
- [14] I.H. Witten, E. Frank, & M.A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed., Morgan Kaufmann Publishers, 2011.
- [15] S. Quarteroni & A. Moschitti, "A Comprehensive Resource to Evaluate Complex Open Domain Question Answering," *Prosiding LREC'10*, 2010.
- [16] D. Radev et al., "Probabilistic Question Answering on the Web," *J. American Society for Information Sci. & Tech.*, vol. 56, no.6, pp. 571-583, April 2005.
- [17] Y.F. Li, M. Xie, & T.N. Goh. "A Study of Project Selection and Feature Weighting for Analogy-based Software Cost Estimation," *The Journal of Systems and Software*, vol. 82, pp. 241-252, 2009.