

# Rancang Bangun Visualisasi Peta Kerjasama Peneliti Berdasarkan Topik Penelitian pada Sistem Repositori Peneliti menggunakan Pemodelan Graf

Agus Nugroho, Umi Laili Yuhana, dan Diana Purwitasari

Teknik Informatika, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

e-mail: yuhana@if.its.ac.id

**Abstrak**— Bila suatu negara ingin bersaing di dunia global yang persaingannya semakin ketat sekarang ini, maka penelitian dan pengembangan harus menjadi perhatian utama. Penelitian di masa depan akan menjadi peluang bagi kemajuan suatu bangsa. Informasi mengenai peta kerjasama peneliti menjadi penting bagi para peneliti untuk memprediksi peluang kerjasama dengan peneliti lain. Sistem Repositori Peneliti Institut Teknologi Sepuluh Nopember (ITS) merupakan sistem informasi yang secara khusus memberikan informasi kepada para peneliti, mahasiswa maupun masyarakat umum seputar dunia penelitian di ITS. Aplikasi ini memiliki salah satu fitur yang memberikan informasi kepada pengguna mengenai peta kerjasama peneliti di ITS. Pengguna dapat melihat kedekatan kerjasama peneliti berdasarkan topik penelitiannya. Peta kerjasama peneliti ini divisualisasikan menggunakan pemodelan graf, sehingga pengguna akan mudah melihat kedekatan-kedekatan para peneliti. Setiap peneliti direpresentasikan dengan lingkaran (*node*) pada graf, sedangkan hubungan kedekatan antar peneliti direpresentasikan dengan garis (*edge*). Semakin besar *node*, berarti semakin banyak penelitian yang telah dilakukan oleh peneliti tersebut, dan sebaliknya. Selain itu, warna *node* menunjukkan area penelitian peneliti. Semakin tebal *edge*, berarti semakin dekat hubungan kerjasama antar *node* (peneliti). Fitur visualisasi peta kerjasama peneliti berdasarkan topik penelitian ini dapat menyediakan informasi peta kerjasama peneliti di ITS yang dibutuhkan pengguna. Data penelitian yang diolah berasal dari 11.824 dokumen penelitian tugas akhir dan 746 peneliti, dengan batas *threshold* hubungan kerjasama berdasarkan kemiripan topik sebesar 0,009. Informasi peta kerjasama peneliti yang ditampilkan dapat membantu pengguna untuk mengetahui perkembangan penelitian yang ada di ITS.

**Kata Kunci**— Institut Teknologi Sepuluh Nopember, Graf, Peta Kerjasama Peneliti, Topik Penelitian, Visualisasi

## I. PENDAHULUAN

Perguruan Tinggi merupakan salah satu lembaga penelitian yang paling berpengaruh terhadap kemajuan dunia penelitian di suatu negara. Hal ini karena banyaknya penyumbang publikasi penelitian berasal dari Perguruan Tinggi. Saat ini untuk menggiatkan dunia penelitian di Indonesia, Institut Teknologi Sepuluh Nopember (ITS) Surabaya sedang mengembangkan sebuah sistem informasi yang akan menunjang dunia penelitian khususnya di lingkungan ITS. Sistem informasi tersebut yaitu Sistem

Repositori Peneliti. Sistem Repositori Peneliti atau disebut juga dengan Resources ITS (RESITS), nantinya akan menjadi versi terbaru dari Web Personal Dosen yang sebelumnya sudah ada [1]. Pada Web Personal Dosen belum tersedia fitur yang secara khusus memberikan informasi tentang kerjasama antar peneliti di ITS. Sistem Repositori Peneliti secara khusus akan memberikan informasi kepada masyarakat seputar dunia penelitian yang ada di ITS. Beberapa fitur yang terdapat dalam sistem informasi tersebut yaitu pengguna dapat melakukan pencarian peneliti dengan kriteria tertentu, melihat daftar publikasi jurnal penelitian terakhir, melihat hubungan kerjasama antar peneliti, serta fitur lainnya.

Pada saat ini, bentuk representasi visual kerjasama (jaringan) yang dinilai paling modern adalah menggunakan visualisasi berbentuk graf [2]. Dengan pemodelan graf, informasi kerjasama akan lebih mudah dipahami dibandingkan dengan bentuk visualisasi lainnya seperti grafik, tabel, dan *listview*. Dalam sebuah visualisasi berbentuk graf, *node* biasanya diwakili oleh bentuk lingkaran dengan ukuran dan warna berbeda-beda yang menunjukkan karakteristik masing-masing *node*. Dan setiap *edge* yang diwakili oleh garis biasanya memiliki panjang dan ketebalan yang berbeda-beda pula, sesuai karakteristik hubungan antar *node* yang dihubungkan oleh *edge* tersebut.

Untuk itu, muncul sebuah gagasan untuk mengembangkan sebuah perangkat lunak yang dapat digunakan para peneliti, mahasiswa, maupun masyarakat umum untuk mengetahui peta kerjasama peneliti yang ada di ITS. Agar mudah dipahami, peta kerjasama antar peneliti divisualisasikan menggunakan pemodelan graf. Perangkat lunak yang akan dibuat ini akan menjadi bagian dari fitur Sistem Repositori Peneliti. Perangkat lunak ini hanya akan berfokus pada peta kerjasama antar peneliti berdasarkan topik penelitian yang diolah dari dokumen penelitian tugas akhir. Topik penelitian setiap peneliti diekstraksi dari dokumen penelitian tugas akhir yang pernah dibuat oleh peneliti. Kemudian hubungan antar peneliti didapatkan berdasarkan kemiripan topik penelitian serta kerjasama yang pernah dilakukan bersama.

## II. TINJAUAN PUSTAKA

### A. Zend Search Lucene Framework Library

Zend Framework merupakan salah satu kakas kerja bahasa pemrograman PHP yang berbasis *OOP (Object Oriented Programming)* serta bersifat *open source*. Fokus utama dari kakas kerja ini adalah untuk membangun aplikasi web dan *web service* yang lebih aman, reliabel, dan modern. Selain itu juga telah didukung *API* untuk beberapa vendor seperti Google, Amazon, Yahoo!, Flickr. Beberapa fitur terbaru Zend Framework antara lain: dukungan *AJAX* melalui *JSON*, *Search-Lucene*, *Syndication*, *Web Services*, dan *OOP PHP5*.

Zend Framework melakukan pengindeksan terhadap dokumen-dokumen yang akan digunakan sebagai kamus data pencarian, dan melakukan pencarian dalam indeks yang telah dibentuk. Indeks merupakan semacam kamus data yang akan digunakan pada mesin pencari yang akan dibuat, yaitu tempat dimana dokumen-dokumen diekstrak informasi (misal: tautan, judul, isi dokumen) kemudian diurutkan menjadi satu kesatuan yang terindeks. Zend Framework juga memberikan fitur pencarian menggunakan *query parser*, dimana dengan *query parser* hasil pencarian dapat disesuaikan dengan keinginan, misal menambahkan operator *AND*, *OR*, dll [3].

### B. Vis.js

Vis.js merupakan kerangka kerja visualisasi berbasis web. Kerangka kerja ini dibangun sedemikian rupa sehingga mudah digunakan, dapat menangani data dinamis dalam jumlah besar, serta dapat memanipulasi data dan interaksi dengan data. Kerangka kerja ini terdiri dari komponen *DataSet*, *Timeline*, *Graph*, dan *Graph 3D*. Vis.js dikembangkan oleh Almende B.V, yang merupakan bagian dari CHAP. Vis.js dapat berjalan baik pada Chrome, Firefox, Opera, Safari, IE9+, dan kebanyakan mobile browsers [4].

Kerangka kerja ini merupakan bahasa pemrograman yang memberikan kemampuan tambahan terhadap bahasa HTML dengan mengijinkan pengekseskuan perintah perintah di sisi pengguna (*web client*). Kerangka kerja ini dapat memodelkan *dataset* dengan pemodelan graf berbasis web.

### C. CodeIgniter

CodeIgniter merupakan kerangka kerja aplikasi web yang bersifat terbuka. Kerangka kerja ini digunakan untuk membangun situs web PHP yang dinamis. Kerangka kerja ini menyediakan berbagai macam library untuk kebutuhan pada umumnya, rancangan antarmuka yang sederhana dan struktur yang logis untuk mengakses *library* yang telah disediakan. Kerangka kerja CodeIgniter menggunakan *Model-View-Controller* [5].

### D. Algoritma Dice's Coefficient

*Dice's coefficient* digunakan untuk menghitung persamaan dari dua sampel data. Persamaan *Dice's coefficient* dapat dilihat pada Persamaan 2.1.

$$s = \frac{2|X \cap Y|}{|X| + |Y|} \quad (2.1)$$

dengan,

$|X \cap Y|$  = jumlah yang dikerjakan X dan Y

$|X|$  = jumlah yang dikerjakan X

$|Y|$  = jumlah yang dikerjakan Y

Contoh

$X = \{aa, bb, cc, dd\}$

$Y = \{ee, ff, gg, dd\}$

$S = (2 \cdot 1) / (4 + 4) = 0.25$

### E. Algoritma TF-IDF (Term Frequency - Inversed Document Frequency)

Algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF) adalah suatu metode pemberian bobot hubungan suatu *term* (kata) terhadap dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu *Term Frequency* (TF) merupakan frekuensi kemunculan *term* (t) pada setiap dokumen (d), dan *Inverse Document Frequency* (IDF) merupakan nilai bobot suatu *term* dihitung dari seringnya *term* muncul dibeberapa dokumen.

Pada metode ini pembobotan kata dalam sebuah dokumen dilakukan dengan mengalikan nilai TF dan IDF. Pembobotan diperoleh berdasarkan jumlah kemunculan *term* dalam dokumen (TF) dan jumlah kemunculan *term* pada seluruh dokumen (IDF). Bobot suatu *term* semakin besar jika *term* tersebut sering muncul dalam suatu dokumen dan semakin kecil jika *term* tersebut muncul dalam banyak dokumen [6]. Bobot TF-IDF sebuah *term* pada dokumen tertentu dapat dihitung menggunakan Persamaan 2.2.

$$TFIDF(t_k, d_j) = TF(t_k, d_j) \cdot IDF(t_k) \quad (2.2)$$

dengan,

$t_k$  = *term* ke-k

$d_j$  = dokumen ke-j

$TF(t_k, d_j)$  = *term frequency term* ke-k pada dokumen ke-j

$IDF(t_k)$  = *inverse document frequency term* ke-k

$TFIDF(t_k, d_j)$  = bobot *term* ke-k pada dokumen ke-j

Untuk menghitung bobot IDF setiap *term*, digunakan Persamaan 2.3.

$$IDF(t_k) = \log \frac{N}{df(t)} \quad (2.3)$$

dengan,

$N$  = jumlah dokumen

$df(t)$  = banyaknya dokumen yang mengandung *term* (t)

$IDF(t_k)$  = *inverse document frequency term* ke-k

Untuk mendapatkan TF setiap dokumen digunakan Persamaan 2.4.

$$TF(t_k, d_j) = f(t_k, d_j) \quad (2.4)$$

dengan,

$TF(t_k, d_j)$  = *term frequency term* ke-k pada dokumen ke-j

$f(t_k, d_j)$  = frekuensi munculnya *term* ke-k pada dokumen ke-j

### F. Algoritma Vector Space Model

*Vector Space Model* adalah suatu model yang digunakan untuk mengukur kemiripan antara suatu dokumen dengan suatu *query*. Pada model ini, *query* dan dokumen dianggap sebagai vektor-vektor pada ruang  $n$ -dimensi, dimana  $n$  adalah jumlah dari seluruh *term* yang ada dalam *leksikon*. *Leksikon* adalah daftar semua *term* yang ada dalam indeks. Salah satu cara untuk mengatasi hal tersebut dalam model *vector space* adalah dengan cara melakukan perluasan vektor. Proses perluasan dapat dilakukan pada vektor *query*, vektor dokumen, atau pada kedua vektor tersebut.

$$\cos \theta \rightarrow \text{sim}(d_j, q) = \frac{d_j \cdot q}{\|d_j\| \times \|q\|} = \frac{\sum_{i=1}^N w_{ij} \times w_{iq}}{\sqrt{\sum_{i=1}^N w_{ij}^2} \times \sqrt{\sum_{i=1}^N w_{iq}^2}} \quad (2.5)$$

dengan,

$\|d_j\|$  = vektor dokumen ke- $j$

$\|q\|$  = vektor kata kunci (*query*)

$W_{ij}$  = bobot TF-IDF dokumen ke- $j$  *term* ke- $i$

$W_{iq}$  = bobot TF-IDF kata kunci (*query*) *term* ke- $i$

$\text{sim}(d_j, q)$  = nilai similaritas antara dokumen ke- $j$  dengan kata kunci (*query*)

Pada algoritma *vector space model* menggunakan rumus untuk mencari nilai *cosinus* sudut antara dua vektor dari setiap bobot TF-IDF dokumen (WD) dan bobot TF-IDF dari kata kunci (WK). Rumus yang digunakan terdapat pada Persamaan 2.5.

## III. ANALISIS DAN PERANCANGAN

### A. Analisis Permasalahan

Institut Teknologi Sepuluh Nopember (ITS) Surabaya merupakan salah satu Perguruan Tinggi yang aktif memberikan kontribusi di dunia penelitian Indonesia. Saat ini dari peneliti-peneliti di ITS, tentunya sudah banyak aktivitas penelitian yang dilakukan, salah satunya penelitian tugas akhir. Aktivitas penelitian tugas akhir tersebut setiap tahunnya kian bertambah dan menghasilkan dokumen penelitian tugas akhir yang tersimpan dalam sebuah basis data. Sebenarnya, hubungan kerjasama penelitian antar peneliti bisa didapatkan dari dokumen penelitian yang tersimpan dalam basis data tersebut. Namun hingga saat ini masih belum ada representasi visual yang merepresentasikan hubungan-hubungan kerjasama antar peneliti tersebut.

Dari dokumen penelitian tugas akhir yang tersimpan, dapat dilakukan suatu pengolahan dokumen tugas akhir untuk mendapatkan keterkaitan antar peneliti. Misalnya saja, semakin sering peneliti-peneliti melakukan bimbingan bersama terhadap penelitian tugas akhir, maka hubungan kerjasama antar peneliti tersebut semakin dekat. Selain itu, juga bisa didapatkan keterkaitan antar peneliti berdasarkan topik penelitian berdasarkan hasil pengolahan ekstraksi dokumen tugas akhir. Dari hubungan-hubungan antar peneliti tersebut dapat dilakukan pemetaan kerjasama peneliti, yaitu suatu bentuk representasi visual hubungan-hubungan kerjasama penelitian dari seluruh peneliti yang ada. Representasi visual hubungan kerjasama antar peneliti merupakan sebuah inovasi untuk memudahkan para peneliti maupun masyarakat mengetahui adanya hubungan kerjasama antar peneliti. Selain itu, peluang-

peluang kerjasama penelitian baru juga bisa didapatkan dari representasi visual hubungan kerjasama antar peneliti tersebut.

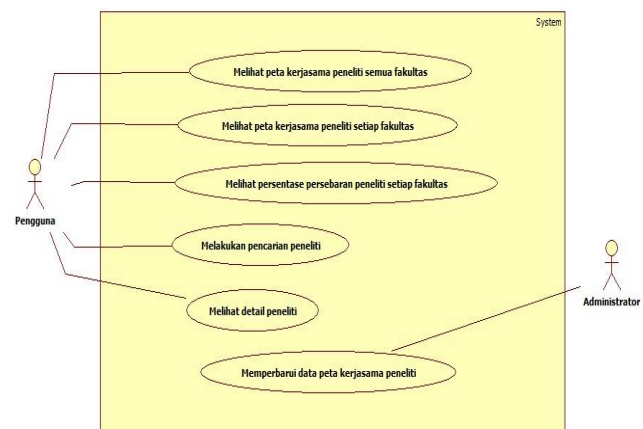
Salah satu media visualisasi modern untuk pemetaan jaringan yang paling banyak digunakan saat ini yaitu visualisasi berbentuk graf. Visualisasi graf ini juga bisa digunakan untuk memvisualisasikan pemetaan kerjasama peneliti. Dengan ini, visualisasi berbentuk graf menjadi media visualisasi yang representatif karena memudahkan pengguna untuk melihat sebuah pemetaan penelitian yang berasal dari data yang cukup kompleks dengan hanya melihat graf. Dari visualisasi graf tersebut dapat dilihat peneliti-peneliti yang sering bekerjasama dengan peneliti lain ataupun peneliti-peneliti yang memiliki kemiripan topik penelitian dengan peneliti lain. Oleh karena itu, sistem yang dibangun pada penelitian ini menggunakan representasi visual berbentuk graf.

### B. Analisis Sistem

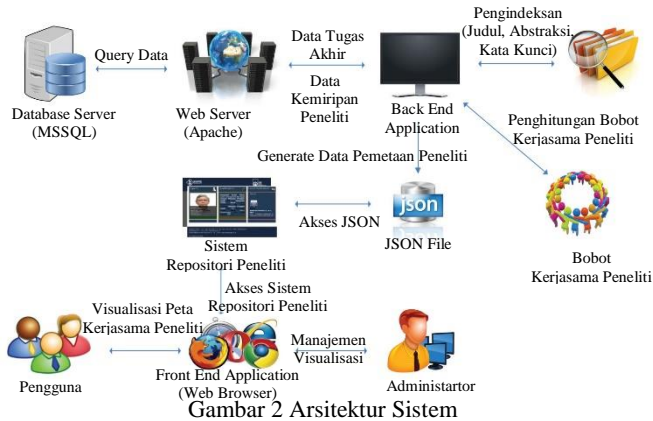
Perangkat lunak yang dibangun pada penelitian ini merupakan aplikasi berbasis web. Aplikasi ini terbagi menjadi dua modul utama, yaitu modul visualisasi peta kerjasama peneliti dan modul manajemen visualisasi peta kerjasama peneliti. Modul visualisasi peta kerjasama peneliti merupakan modul yang diakses oleh pengguna, sedangkan modul manajemen visualisasi merupakan modul yang diakses oleh administrator. Adapun kebutuhan fungsional aplikasi ini secara keseluruhan dapat dilihat pada Gambar 1.

### C. Perancangan Sistem

Adapun perancangan sistem ini tergambar pada Gambar 2 mengenai arsitektur sistem. Aplikasi ini merupakan salah satu fitur dari Sistem Repositori Peneliti, untuk keperluan visualisasi peta kerjasama peneliti, sistem ini terdapat visualisasi peta kerjasama yang diakses oleh pengguna dan manajemen visualisasi yang dikelola oleh administrator. Aplikasi pada modul visualisasi menampilkan peta kerjasama peneliti berdasarkan topik penelitian peneliti. Sedangkan aplikasi pada modul manajemen visualisasi digunakan oleh administrator untuk memperbarui data peta kerjasama peneliti jika terdapat dokumen penelitian maupun peneliti baru, ini dilakukan agar hubungan kerjasama kemiripan topik antar peneliti yang ditampilkan pada modul visualisasi sesuai dengan data penelitian yang berasal dari basis data.



Gambar 1 Diagram Kasus Penggunaan



Gambar 2 Arsitektur Sistem

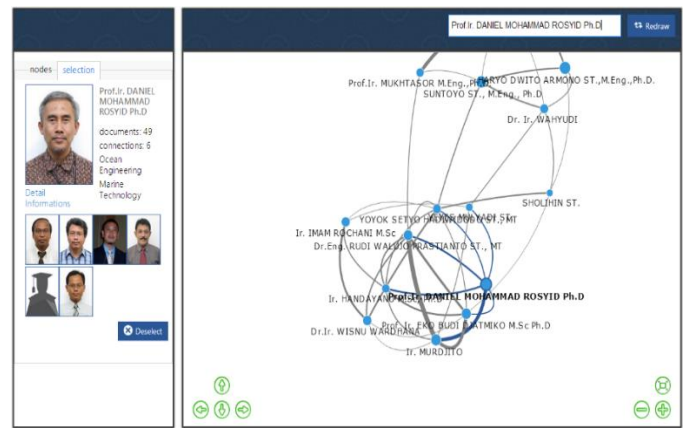
Data peta kerjasama peneliti diolah dari data penelitian tugas akhir yang diambil dari basis data SQL Server menggunakan *web server*. Kemudian pada bagian *back end* aplikasi dilakukan proses pengindeksan data penelitian tugas akhir dan penghitungan bobot kerjasama peneliti. Setelah didapatkan bobot kerjasama peneliti, bobot kerjasama peneliti tersebut di-*generate* dalam bentuk *file* notasi graf JSON yang kemudian diakses oleh modul visualisasi. Selanjutnya pada bagian *front end* aplikasi pengguna dapat mengakses modul visualisasi peta kerjasama peneliti menggunakan *web browser*. Dengan menggunakan *web browser* pula administrator mengakses modul manajemen visualisasi untuk memperbarui data peta kerjasama peneliti yang tersimpan dalam *file* JSON. Manajemen visualisasi berisi tahapan-tahapan yang dilakukan pada *back end* aplikasi untuk mengolah data penelitian hingga terbentuk *file* JSON berisi notasi graf peta kerjasama peneliti.

#### D. Perancangan Proses

Secara garis besar proses yang dibangun pada sistem terbagi menjadi tiga proses, yaitu proses pembentukan vektor peneliti, proses penghitungan bobot kerjasama berdasarkan topik, dan proses visualisasi. Namun secara mendetail, proses-proses yang dibangun pada sistem ini adalah proses pengindeksan data penelitian tugas akhir, proses menghilangkan *stopwords*, proses melakukan *stemming*, proses pembobotan kata, proses penghitungan bobot relasi topik antar peneliti, proses penghitungan bobot kerjasama penelitian antar peneliti, proses penghitungan bobot kerjasama peneliti berdasarkan topik penelitian, proses pemodelan graf dalam bentuk JSON, dan proses visualisasi graf berbasis web. Secara mendetail pada proses visualisasi graf berbasis web adalah proses inisiasi visualisasi graf, proses menampilkan detail peneliti dan mengurutkan peneliti-peneliti yang mirip topik penelitiannya, proses menghitung dan menampilkan persentase persebaran peneliti, proses pencarian peneliti, dan menampilkan peta kerjasama peneliti setiap area peneliti.

#### E. Implementasi Sistem

Pada tahap ini dilakukan pembangunan aplikasi visualisasi peta kerjasama peneliti beserta modul manajemen visualisasinya.



Gambar 3 Antarmuka visualisasi peta kerjasama peneliti menampilkan detail informasi peneliti

Aplikasi ini dibangun dengan bahasa pemrograman PHP dengan kerangka kerja Zend Search Lucene Framework untuk pengolahan kata, kerangka kerja Vis.js untuk visualisasi graf berbasis web, dan kerangka CodeIgniter pada sisi web. Antarmuka yang dibangun yaitu antarmuka visualisasi dan antarmuka manajemen visualisasi. Beberapa tampilan implementasi pada antarmuka visualisasi dapat dilihat pada Gambar 3.

## IV. KESIMPULAN

Aplikasi visualisasi peta kerjasama peneliti yang dibangun dapat menampilkan peta kerjasama peneliti berdasarkan topik penelitian menggunakan pemodelan graf berbasis web. Data penelitian yang diolah berasal dari 11.824 dokumen penelitian tugas akhir dan 746 peneliti, dengan batas *threshold* hubungan kerjasama berdasarkan kemiripan topik sebesar 0,009. Aplikasi juga dapat menampilkan data peta kerjasama peneliti yang dapat berubah-ubah sesuai dengan data penelitian terbaru. Selain itu, aplikasi juga dapat menampilkan peneliti-peneliti berdasarkan kemiripan topik penelitiannya.

## DAFTAR PUSTAKA

- [1] Institut Teknologi Sepuluh Nopember, "Personal Dosen ITS," Institut Teknologi Sepuluh Nopember, [Online]. Available: <http://personal.its.ac.id/>. [Accessed 25 June 2014].
- [2] C. Belter, "Information Today, Inc. ITI Blogs," NOAA Central Library, [Online]. Available: <http://www.infotoday.com/online/may12/Belter-Visualizing-Networks-of-Scientific-Research.shtml>. [Accessed 20 June 2014].
- [3] Zend Technologies Ltd. , "Query Language - Zend\_Search\_Lucene - Zend Framework," Zend Technologies Ltd. , 2005. [Online]. Available: <http://framework.zend.com/manual/1.10/en/zend.search.lucene.query-language.html>. [Accessed 3 March 2014].
- [4] A. B.V, "vis.js is a dynamic, browser-based visualization library," 2010. [Online]. Available: <http://visjs.org/>. [Accessed 24 June 2014].
- [5] EllisLab, Inc., "What is CodeIgniter?," EllisLab, Inc., 2012. [Online]. Available: <http://ellislab.com/codeigniter>. [Accessed 24 June 2014].
- [6] T. Rollek'e, T. Tsikrika and G. Kazai, "A General Matrix Framework for Modelling Information Retrieval," [Online]. Available: [http://cir.dcs.unipannon.hu/cikkek/rolleke\\_paper.pdf](http://cir.dcs.unipannon.hu/cikkek/rolleke_paper.pdf). [Accessed 25 June 2014].