Department of Computer Science and Information Systems

MSc Information & Web Technologies
Proposal Form (2011/2012)

1. Proposal

The student should complete parts 1(a) and 1(b). They should put their supervisor's name and date the proposal was agreed with their supervisor in the last box below. They should **send the completed electronic copy of this form to the programme administrator (thomas@dcs.bbk.ac.uk) and upload it on Blackboard** no later than 16 January 2012 (deadline for Year-2 Part-time students) and 12 March 2012 (deadline for Full-time students). Expand the boxes where necessary.

(a) Student details

Name: Artur Jach	**Address:**
Tel (home): 07704983570	124B Portnall Rd
E-mail: *arthurjach@yahoo.co.uk*	London W9 3BQ

(b) Project details

Title:
The analysis of organic click-through-rates (CTR) from different Google search result positions

Objectives:

Delivering a report presenting interesting and useful findings about differences in average click through rates from different organic (non-paid) Google search result positions.

This will be based on the analysis of previously gathered, stored and mined data. Data will be anonymous and taken from web analytics suites of real business.

What makes this project unique is the fact that the data will also be aggregated and/or sliced using different metrics and dimensions (e.g. branded vs generic keyword).

Description:

In order to calculate average click through rates for different positions in search results we need to know:

- the keyword and its current (daily) position in Google results
- traffic which the keyword managed to drive to the site each day
- estimate of maximum traffic which the keyword could drive to the site each day

Title:
The analysis of organic click-through-rates (CTR) from different Google search result positions

Having this data we can calculate the actual share of traffic (CTR) for this particular keyword from that particular position in search results. These can be then averaged and manipulated using different aggregation filters and dimensions.

Problem addressed:

Currently there are no detailed organic click-through rate reports backed up by any solid and real data out there on the web. There have been some good attempts made, e.g. http://www.optify.net/guides/organic-click-through-rate-curve but they don't split CTRs by brand vs generic search terms or by sectors, e.g. finance, travel, etc.

Required Activities:

Data retrieval - in order to calculate average CTR from different positions in search results following data will be retrieved on a daily basis over a period of minimum 2 weeks:

- Selecting a sample and then the full set of keywords
- Keyword and traffic data (retrieved from clients' accounts using web analytics APIs)
- Estimated keyword search volume in Google (retrieved manually from the Google Adwords Tool or automated using their API - paid for service)
- Daily keyword ranking in Google (through existing software/browser plugins or self-developed automated rank checking script). The amount of base data used for analysis will be mainly influenced by the number of keywords which rankings can be checked for each day.

Data storage:

- all retrieved data will be cleansed and uploaded and stored in MySQL database.
- the database will be carefully designed in order to enable efficient mining
- each record in the database will have additional and optional attributes which may be needed during the mining stage (to find interesting CTR facts): industry sector, generic or brand name keyword, result with organic and sponsored link vs result with organic listing only.
- database will be periodically checked for data consistency

Data mining:

- In order to mine and analyse the data an open source data mining tool will be used (e.g. RapidMiner)
- Getting familiar with and testing chosen data mining tools
- In case of issues with integrating data sources stored in MySQL database with data mining tools, an alternative solution will be put in place and efficient SQL queries will parse the data and store resulting click through rates

Title:
The analysis of organic click-through-rates (CTR) from different Google search result positions

- Average click through rates for every position on the first page in Google will be calculated
- Resulting data will be aggregated or split using different interesting dimensions: industry sector, generic or brand name keyword, result with organic and sponsored link vs result with organic listing only.

Desirable Activities:

- In order to make the findings even more interesting other data representation dimensions could be considered to show differences in click through rates: localisation, device usage, demographics (though difficult to obtain and associate).
- One activity which would be very useful would be analysing how Google personalised search results affect click through rates. This activity is only desirable as currently there is no obvious algorithmic way of doing this, but it may become possible in near future.
- Data representation through a web interface (using PHP to retrieve results from MySQL database).

Method & relevance to MSc modules:

PHP (Object Oriented Programming. Search Engines & Web Navigation)

MySQL and data mining tools like RapidMiner (Advanced Database Management, Data Warehousing, Data Mining)

Various APIs (Information Retrieval)

Advanced Web Rankings (ranking checking tool, but only an alternative to self developed PHP script)

MS Excel

XHTML, JS, CSS (previously studied at Birkbeck BSc modules)

Work plan:

Proposal, Project Scope & Functional Specifications Agreement - deadline 16th Jan 2012

Testing data retrieval APIs using the sample set of keywords - deadline 29th Feb 2012

Writing PHP script for checking keyword rankings and running a “stress test” on it to find out the maximum number of keywords it can check the ranking for - deadline 31st Mar 2012

Project break (exams) – 15th Apr - 15th Jun 2012

Database design and writing and testing a script for uploading retrieved data to the database - deadline 30th Jun 2012

Data mining and storing click through rate results - deadline 22nd Jul 2012

Title: The analysis of organic click-through-rates (CTR) from different Google search result positions
Testing application using dummy data - deadline 31st Jul 2012 Running application with real data daily for minimum 2 weeks - deadline 15th August 2012 Project report - deadline 17th Sep 2012
College equipment required: None

Supervisor: Dr Dell Zhang	**Date agreed:** 15/01/2012