Project no. 215231

TrebleCLEF

Treble-CLEF: Evaluation, Best Practices and Collaboration for Multilingual Information Access
IST: ICT-1-4-1, Digital libraries and technology-enhanced learning

Deliverable 4.1
TrebleCLEF Query Log Analysis Workshop Report

Start Date of Project: 01 January 2008
Duration: 24 Months

Organisation Name of Lead Contractor for this Deliverable: USFD

Version 1.00, July 2009

Project co-funded by the European Commission within the Seventh Framework programme

Document Information

Deliverable number:	3.1
Deliverable title:	TrebleCLEF Query Log Analysis Workshop Report
Due date of deliverable:	31/03/2009
Actual date of deliverable:	10/07/2009
Author(s):	Paul Clough
Participant(s):	USFD
Workpackage:	4
Workpackage title:	Evaluation Methodologies
Workpackage leader:	USFD
Dissemination Level:	Public
Version:	1.00 final
Keywords:	Query Log Analysis Workshop, Best Practices, Workshop, Log Analysis

History of Versions

Version	Date	Status	Author (Partner)	Description/Approval Level
1.00	10/07/2009	Final	USFD	

Abstract

This document summarises the Query Log Analysis Workshop entitled “Query Log Analysis: From Research to Best Practice” held May 27-28 at the British Computer Science (BCS) Offices in London, UK. The event involved 12 invited speakers from various academic and commercial institutions from around the world who are all involved, in some way, with query log analysis. Participants included Jim Jansen from Penn State University (USA), Bettina Berendt from Katholieke Universiteit Leuven, Belgium, Lynn Silipigni Connaway from the Online Computer Library Center, Inc. (USA), Filip Radlinski from Microsoft Research (UK) and Vanessa Murdock from Yahoo! Research (Spain). A number of other people attended the event including local businesses and academic institutions. The workshop provided a forum in which to discuss and share experiences and best practices regarding query log analysis. This report describes the event, the presentations given by the invited speakers and summarises discussions held

Table of Contents

Document Information 1
Abstract 1
Executive Summary 3
1. Introduction 4
2. Workshop details 5
4.1 Audience 5
4.2 Venue 5
4.3 Goals 5
4.4 Schedule 6
3. Presentations 7
4. Discussions 10
4.1 Questions 10
4.2 Discussion groups 11
4.3 Dissemination 12
5 Reflections 12
6 Summary 14
7 Acknowledgements 14
8 References 14
Appendix A – Attendees (full details) 17
Appendix B – Advertising flyer 20
Appendix C – Programme 21

Executive Summary

This document summarises the Query Log Analysis Workshop entitled “Query Log Analysis: From Research to Best Practice” held May 27-28 2009 at the British Computer Science (BCS) Offices in London, UK [1]. The aim of this event was to establish a forum in which invited speakers from multiple disciplines could share and discuss their experiences of analysing query logs. By involving representatives from different disciplines and selected business communities, we hoped to stimulate the cross-fertilisation of knowledge and ideas to establish best practices for conducting and utilising query logs in practical commercial settings. By inviting well-known academics we aimed to clarify current research (e.g. the terminology and approaches used), collate standardised procedures and resources commonly used, identify common challenges, stimulate thoughts on future directions of the field and create an initial community of people interested in query log analysis.

This one-and-a-half-day workshop included 12 main invited speakers who gave presentations on a variety of topics to summarise their work on log analysis and some ideas about future challenges. Speakers included Jim Jansen from Pennsylvania State University (USA), Bettina Berendt from Katholieke Universiteit Leuven, Belgium, Lynn Silipigni Connaway from the Online Computer Library Center Inc. (USA), Filip Radlinski from Microsoft Research (UK) and Vanessa Murdock from Yahoo! Research (Spain). The event was also publicised more widely (through various mailing lists and discussion boards) and attracted 9 other attendees from academia and business. In addition to the presentations, the event included a tutorial on log analysis from Dr. Jim Jansen from Pennsylvania State University (USA) at the start of the second day. Presentations at the workshop provided a useful stimulus for further discussions on current challenges and problems facing researchers conducting log analysis, and as a starting point to help transfer knowledge from academia into a commercial setting.

This report describes the organisation of the event which ran in London on May 27-28, the invited speakers and their presentations, a summary of discussions and reflections on the success of the event. The discussions generated in this workshop will help contribute to deliverable 4.2: “Best practices for test collection creation, evaluation methodologies and language processing technologies.”

1. Introduction

It is common for many online systems to record the interactions between people using the system and responses from the system itself[1] (see, e.g. [2]). This user-server transactional data, typically consisting of visitor's queries and clicks (e.g. what they looked at and selected) and system responses is known as click-stream data. This transactional data is often combined with customer-based data (e.g. purchases) and site-based data (the content and link structure of the website being navigated) to identify and analyse user activity and behaviour. These logs offer potentially valuable information for a wide range of applications (e.g. the design, personalisation and evaluation of systems), and in business can serve as a guide for decision-making processes (see, e.g. [3, 4]). Mat-Hassan & Levene [5] state the objectives of a log analysis may include: (1) to investigate a searcher's performance; (2) to establish the profile of an effective searcher; (3) to establish a user's searching characteristics; (4) to understand a user's navigational behaviour (including the number of query terms entered and the number of click-throughs viewed). Log analysis provides a new paradigm for evaluating search, and all major search engine companies exploit logs in some way to develop more effective search.

As more online services exist and more people interact with them, the analysis of log files is an important research field in its own right. Logs files are being studied in several domains, both academic and commercial including: digital libraries [6], Web data mining [7, 8, 9, 10, 11, 12], information seeking and search behaviour [5, 13, 14, 15, 16, 17]), usability assessment [18], website design and evaluation [19, 20], Web search evaluation [21, 22, 23:9-12], Web search optimisation [24, 25, 26], Web Analytics [27], information visualisation [28, 29, 30], adaptive systems and personalisation [31, 32, 33], e-commerce [34, 35], learning to rank from implicit feedback [36] and business intelligence [37]. However, despite the obvious commercial (and research) benefits of utilising such data, many organisations collect but do not use their log file data effectively. In fact, previous studies suggest that click-stream data is only being used within organisations for basic Website management activities [38, 39], and is under-used by both practitioners and academics alike [40, 41, 42]. Research in log analysis has the potential of helping organisations better understand how online services they provide are being used, but only if the research is made accessible to them.

The TrebleCLEF Query Log Analysis Workshop, entitled "Query Log Analysis: From Research to Best Practice" was held May 27-28 2009 at the British Computer Science (BCS) Offices in London, UK. This workshop constitutes deliverable 4.1 of the TrebleCLEF coordination action [43]. The goal of the workshop was to provide a forum in which invited speakers from multiple disciplines could share and discuss their experiences from analysing query logs. By involving representatives from different disciplines and selected business communities, we hoped to stimulate the cross-fertilisation of knowledge and ideas to establish best practices for conducting and utilising query logs in practical commercial settings. By inviting well-known academics we aimed to clarify current research (e.g. the terminology and approaches used), collate standardised procedures and resources commonly used, identify common challenges, and stimulate thoughts on future directions of the field.

The remainder of this deliverable is structured as follows: the setup of the workshop (audience, venue and goals; Section 2) is presented first, followed by a summary of presentations given by attendees at the workshop (Section 3), a summary of the discussions held (Section 4), reflections on the workshop (Section 5) and finally a short conclusion (Section 6).

[1]Jansen, B., Taksa, I. and Spink, A. (2009) Research and Methodological Foundations of Transaction Log Analysis http://www.igi-global.com/downloads/excerpts/8282.pdf

2. Workshop details

4.1 Audience

The workshop was restricted to a maximum of 25 people, in order to get a lively and focused discussion. Participants consisted of 12 invited speakers, selected to represent a variety of academic disciplines and business sectors. The remaining participants (9 in total) consisted of academics and representatives from local businesses (i.e. based in London). All attendees of the event had some level of interest and experience in log analysis. Table 1 summarises the invited speakers and Table 2 the additional participants (more information is provided in Appendix A).

Invited speaker	Institution	URL
Dr. Jim Jansen	Pennsylvania State University (USA)	ist.psu.edu/faculty_pages/jjansen/
Dr. Fabrizio Silvestri	ISTI-CNR (Pisa, Italy)	pomino.isti.cnr.it/~silvestr/
Dr. Lynn Silipigni Connaway	Online Computer Library Center Inc. (Ohio, USA)	www.oclc.org/research/staff/connaway.htm
Prof. Bettina Berendt	Katholieke Universiteit Leuven (Belgium)	www.cs.kuleuven.be/~berendt
Prof. Nigel Ford	University of Sheffield (UK)	www.shef.ac.uk/is/staff/ford.html
Dr. Thomas Mandl	University of Hildesheim (Germany)	www.uni-hildesheim.de/~mandl/
Dr. Filip Radlinski	Microsoft Research (Cambridge, UK)	research.microsoft.com/en-us/people/filiprad/
Prof. Mark Levene	Birkbeck, University of London (UK)	www.dcs.bbk.ac.uk/~mark
Dr. Udo Kruschwitz	University of Essex (UK)	cswww.essex.ac.uk/staff/udo/
Dr. Vanessa Murdock	Yahoo! (Barcelona, Spain)	research.yahoo.com/Vanessa_Murdock
Dr. Giorgio Di Nunzio	University of Padoa (Italy)	ims.dei.unipd.it/websites/archive/ims2009/members/dinunzio.html
Dr. Dhavval Thakker	Press Association Images (Nottingham, UK)	jaala.co.uk/

Table 1. Invited speakers at the TrebleCLEF query log analysis workshop.

To attract participants, various email distribution lists were used including SIG-IRList, the BCS mailing list and the UKeiG (the UK eInformation Group) mailing list. Appendix B shows the initial flyer created to advertise the event.

4.2 Venue

The workshop was held at the BCS (British Computer Society) London Office[2] (London Office, BCS, First Floor, The Davidson Building, 5 Southampton Street, London, WC2E 7HA) in the Wilkes Room 2. This venue was selected because it was based in Central London and was hoped to increase the likelihood of invited speakers accepting an invitation to come. Most participants stayed in the Strand Palace Hotel located next to the BCS offices.

4.3 Goals

The aim of this event was to establish a forum in which invited speakers from multiple disciplines could share and discuss their experiences from analysing query logs. By involving representatives from different disciplines and selected business communities was expected to stimulate the cross-fertilisation of knowledge and ideas to help establish best practices for conducting and utilising query

[2] http://www.bcs.org/upload/pdf/london-office-guide.pdf

logs in practical commercial settings. By inviting well-known academics we aimed to clarify current research (e.g. the terminology and approaches used), collate standardised procedures and resources commonly used, identify common challenges, and stimulate thoughts on future directions of the field. In summary, the event aimed to investigate the following: what research has been done, where is it now, where is it going?

Attendee	Institution	URL
Dr. Paul Clough	University of Sheffield (UK)	ir.shef.ac.uk/cloughie/
Dr. Mark Sanderson	University of Sheffield (UK)	dis.shef.ac.uk/mark/
Dr. Martin Braschler	Zurich University of Applied Sciences (Switzerland)	www.martin-braschler.com/
Dr. Stephen Dignum	University of Essex (UK)	privatewww.essex.ac.uk/~sandig/
Mr. Nick Luft	Royal Institution of Chartered Surveyors, London (UK)	www.rics.org
Ms. Rita Wan-Chik	University of Sheffield (UK)	www.shef.ac.uk/is/
Dr. Yan Xu	University of Sunderland (UK)	www.shef.ac.uk/is/staff/ford.html
Ms. Amy Warner	UK National Archives, London (UK)	www.nationalarchives.gov.uk/
Mrs. Stella Dextre Clarke	Luke house, London (UK)	

Table 2. Additional participants at the TrebleCLEF query log analysis workshop.

4.4 Schedule

The workshop programme (shown in Appendix C) was planned to include a series of short (30 mins) presentations by the invited speakers (see Section 3). These were aimed at stimulating discussion and speakers were asked to include details on experiences with using log files (including types of projects carried out, techniques used, problems encountered, main limitations and likely future directions of the field). The speakers were roughly grouped in the programme to reflect their various disciplines and backgrounds as follows:

- The **first group** of presentations (Mark Levene, Nigel Ford and Jim Jansen) involved people commonly associated with query log analysis;
- The **second group** (Filip Radlinski, Vanessa Murdock, Lynn Silipigni Connaway and Dhavval Thakker) represented commercial organisations;
- The **third group** (Fabrizio Silvestri, Bettina Berendt and Udo Kruschwitz) were loosely grouped to represent the web mining community;
- The **final group** (Giorgio Di Nunzio and Thomas Mandl) representing digital libraries and evaluation.

On the second day, Jim Jansen agreed to provide a 1-hour tutorial session entitled “What is Web log analysis”[3]. This was aimed to provide participants with an introduction to log analysis and encourage knowledge transfer between the academic and commercial communities. The discussion session was held in the afternoon of the 28th May. This session was organised around a series of questions which were circulated to participants beforehand in order to obtain their feedback and comments. This list of

[3] Also available from SlideShare: http://www.slideshare.net/dj9395/what-is-log-analyis

questions is discussed in Section 4. Presentations from the workshop and **audio recordings** (in MP3 format) of all talks are available from the website created for the event [1]. This website aims at providing a permanent record (along with this report) of the event.

3. Presentations

The following presentations were given during the workshop:

- **"Welcome and introduction" by Paul Clough, University of Sheffield (UK)**
  - The presentation gave a summary of the TrebleCLEF project, together with the aims and logistics of the workshop.
- **"From server logs to query logs" by Mark Levene, Birkbeck, University of London (UK)**
  - Mark Levene started out his talk by discussing four observations about the Web: the web is a complex network (but has structure); there is too much data to analyse (data to knowledge?); with mobile technologies then wherever you are your activities are logged (can be combined with query logs); and whenever you use a search engine your activities are logged. Some applications of log analysis include: recommender systems, personalisation, social search, topical search, search engine advertising, click-stream analysis and pre-fetching of web pages by Web servers. Mark also discussed three areas of his research: analysing server logs (based on variable length Markov Chain Models), mining context-topic association rules from search engine logs and query classification. A number of issues regarding log analysis were made by Mark: the lack of recent log data for the wider research; issues with verifiability and repeatability of experiments (we have the AOL and recent Microsoft logs); the availability of mobile search logs and other web log data can help understand how the Web is used; but without log data no mining can be done.
- **"Query log analysis and individual differences" by Nigel Ford, University of Sheffield (UK)**
  - Nigel Ford discussed some of his research aimed at investigating strategic differences in the way difference types of individual user query information access systems (a social science perspective). Log file analysis was combined with responses from questionnaires to establish various patterns of user interaction to build models for adaptive IR. In addition, Nigel described work on cognitive styles and possible effects on web searching, based on categorising and analysing query transformations. The aim of studying query logs (and using additional input from questionnaires) is to guide the design of interfaces for adaptive information access systems that can provide suitable help or prompts based on user's pattern of activities as exhibited in query logs.
- **"Moving from Description to Prediction for Information Searching" by Jim Jansen, Pennsylvania State University (USA)**
  - Jim Jansen began his presentation by discussing the information explosion and then carried on with the state of Web search. The main focus on the presentation was to discuss the state of information searching (actions – behavioural, affective and cognitive - employed by people when interacting with an information system), research which Jim sees as mainly descriptive and lacking more predictive approaches and models (i.e. not just describing what people do but predicting their actions). This is typically achieved with state models (e.g. Markov Models), but these will break down after a couple of transitions. Jim proposed a stateless model based on treating search engine logs as an information stream and information searching as a temporal stream. Further research work discussed included developing a time series analysis approach to model individual user's search behaviours, investigating the effect of

system branding on user perception on system performance, user modelling, modelling information searching, search engine marketing and analysis of Twitter posts using sentiment analysis. Jim pointed out that Twitter logs are freely-available to gather and analyse.

- **"What do click logs tell us about user's search satisfaction?" by Filip Radlinski, Microsoft Research Cambridge (UK)**
  - Filip Radlinski from Microsoft Research (Cambridge, UK) gave a presentation on some of his previous work involving log files conducted prior to him joining Microsoft. The talk discussed evaluation of search systems, and in particular, what click logs say about result quality and user satisfaction. Filip described a study of evaluating search with click logs and absolute metrics (to indicate how good results are) and paired comparisons (to say which results are better). This work shows how users' behaviour changes in response to results presented in different ways, and how clickstream logs can capture these changes. Metrics such as clicks per query and queries per session are absolute indicators of quality but these changes were monotonic and did not change as expected. The pairwise tests, however, were much more reliable and sensitive to differences in the results. Filip concluded his talk by considering three areas to discuss: logging interactions (all click log evaluations rely on clicks being useful but are queries and clicks giving absolute judgements?); click metrics are often noisy and unreliable (what if user's click by mistake, or click and then press the back button a short time later?); and reusability of logs (if someone else comes up with a new search engine are logs useful to them?)
- **"Online learning from click data" by Vanessa Murdock, Yahoo! Research (Barcelona, Spain)**
  - Vanessa Murdock from Yahoo! Research discussed research in which click-through data is being used to produce a ranking of adverts on a Web search engine given a specific query (i.e. predicting which adverts a user will click on). A similar approach was being used to order images brought back from a Web image search engine. She began by discussing the results of eye-tracking studies, which indicate the "golden triangle" – this explains some of the biases found in log files and will affect learning-from-click studies (i.e. learning-to-rank from clicks).
- **"Following the Trail of WorldCat Users" by Lynn Silipigni Connaway, Online Computer Library Center Inc. (USA)**
  - Lynn Silipigni Connaway from the Online Computer Library Center Inc. discussed her work on transaction log analysis to identify user behaviour of online retrieval systems (e.g. what is being accessed, the number of accesses and patterns of access). This provided an information science view on query log analysis. Lynn described her analysis of logs taken from the WorldCat.org (free on the Web) and FirstSearch (subscription-based) search engines which provide access to the WorldCat database. Lynn highlighted some of the limitations of transaction log analysis, including: massive amounts of data to analyse; analysis depends on what data is collected; capture of logs can be terminated at any time (internal organisational issues); if systems change the type of data captured may change; logs may have multiple codes/formats for search fields; log data provides incomplete information about the users and their access to resources (i.e. held in different transaction logs). Directions for future work were seen as automating the analysis (e.g. query classification and analysis of zero-hits), linking search behaviours to demographics and exploiting IP addresses.
- **"Using Query Logs at PA Images" by Dhavval Thakker, Press Association Images (Nottingham, UK)**

- Dhavval Thakker, a KTP Associate with Press Association (PA) Images (UK), discussed the use of query logs to improve the performance of their image search engine. An interesting point coming out of this talk is that the kinds of logs the PA generate are bespoke and not the usual search logs academics experiment with (i.e. they only contain query information and no click-through data). This therefore limits the kind of analysis that can be carried out on the logs. This coincided with a point made by Lynn Connaway from OCLC where technical staff had turned off log capture on one of the systems because they could not see the immediate benefit of gathering this data. Dhavval presented the results of an initial descriptive analysis of the query logs and indicated that PA was currently experimenting with Google Analytics.

– **"Query log mining @ HPC-Lab" by Fabrizio Silvestri, ISTI-CNR (Pisa, Italy)**
  - Fabrizio Silvestri provided a summary of his research in which query logs have been used to produce high performance IR systems, e.g. through informing caching, data partitioning and query routing. Fabrizio described his current work in high performance IR which includes similarity-based caching (for large-scale content-based image retrieval), using click-through data to generate search shortcuts (i.e. query recommendations) and query log driven index organisation. Future areas of research were discussed including analysing log files for identifying human activities (i.e. rather than detecting search sessions). For example, identifying from log files the activities involved in planning a trip to London. Fabrizio concluded his presentation by discussing attributes of log files he would like to be analysing: large log files, click-through data (not just queries), multilingual logs, long-term (spanning multiple months) and publicly available to enable reproducible results. Fabrizio is the author of "Mining query logs: turning search usage into knowledge" published in Foundations and Trends in Information Retrieval.

– **"Exploratory analysis needs theor[y|ies] – OR: Some answers to 14 questions" by Bettina Berendt, Katholieke Universiteit, Leuven (Belgium)**
  - Bettina Berendt structured her presentation in four main sections based on the questions provided to speakers before the event: a brief overview of her work in query log analysis; some general questions; the feasibility of query log analysis within academia and the transfer/collaboration between academia and industry. Bettina showed examples of work in which log analysis is used to help formulate domain content ontologies; used to inform demographics and user behaviour; and logs themselves are analysed descriptively, data mined and visualised. Aspects such as statistics, data-mining patterns and visualisations are techniques that can be applied across domains and for various applications. Bettina showed a number of examples in which constructs (theories) are operationalised as concrete measures (e.g. for user satisfaction). Bettina suggested a number of ideas on the future challenges/directions for the field including: combinations of methods/theories; interaction beyond navigation and querying; preservation of privacy; time-series analysis; web search advertising (as new application area); the impact of micro-blogging streams (e.g. Twitter) on analysis of Web search logs; the role of eye-tracking; integration of multiple transaction logs and correlating transaction logs with user behaviour.

– **"Log Analysis at Essex" by Udo Kruschwitz, University of Essex (UK)**
  - Udo Kruschwitz discussed research at the University of Essex on query log analysis for adaptive intranet search (academic application) and query log analysis for learning to match job seekers against best-matching jobs (commercial application). For the former application, Udo described an approach to develop a search system that makes suggestions using automatically extracted domain knowledge based on learning from users' interactions. This creates evolving domain knowledge that adjusts to users' search behaviour. Udo highlighted that the queries in the log files generated on

intranet searches differ from general Web search due to the specific domain and ended his talk by discussing a project called AutoAdapt that seeks to automatically adapt a (knowledge) domain model. Further details can be found in [44, 45].

- **"What is Web log analysis?" by Jim Jansen, Pennsylvania State University (USA)**
  - Jim Jansen gave an additional talk as a tutorial on log analysis, which he views as a part of Web Analytics [46] that covers analysis of logs from a variety of sources, e.g. Intranet, system, OPAC and search logs. Jim covered some of the theoretical grounding of query logs related to theories of behaviourism – inductive and data driven, characterized by observation of measurable behaviour. Jim gave a social science perspective to query log analysis in which understanding user behaviour is a key part analysing logs. Logs then form "trace data" for the researcher to help explain underlying theories (using inductive or deductive approaches). Jim highlighted a book he has co-edited on Web log analysis of which the first chapter is available free for download [47].
- **"Logging Digital Libraries" by Giorgio Di Nunzio, University of Padoa (Italy)**
  - Giorgio Di Nunzio gave a presentation on evaluating digital libraries and the role of log files for this purpose. Giorgio discussed work undertaken in The European Library (TEL) project in which logs file analysis was combined with HTTP server logs and questionnaires to gather user preferences and satisfaction (i.e. log files show what users do, but not why). Giorgio also discussed large-scale evaluation activities carried out in the context of the Cross Language Evaluation Forum (or CLEF), in particular the work carried out as a part of the iCLEF and LogCLEF tasks. In the latter task, one of the activities is LADS (Log Analysis for Digital Societies) in which participants to the evaluation campaign are provided with log data and user profiles to conduct their own experiments [48].
- **"Query Classification in Logfile-Analysis: Evaluation Issues and User Satisfaction" Thomas Mandl, University of Hildesheim (Germany)**
  - The final talk of the workshop by Thomas Mandl discussed work on query classification, in particular with respect to location where he is running a large-scale evaluation exercise on geographic query classification and parsing (linked to Giorgio's talk). Thomas also discussed some work carried out on user behaviours and characteristics on the Web, together with establishing user satisfaction with search based on log activity. Part of the work presented by Thomas seeks to first understand what it is that makes people satisfied with search results.

4. Discussions

4.1 Questions

Before the event, a number of initial questions to discuss at the workshop were circulated to the invited speakers for their feedback and comment on. These questions were refined to form a list used by invited speakers to structure their presentations (see, e.g. the talk by Bettina Berendt) and simulate discussions on the second day of the workshop. The resulting list of questions was as follows:

- What approaches to log analysis are used in different fields?
- What are the problems with carrying out log analysis in different fields?
- Which techniques are similar between fields/applications? (Which techniques are specific to particular applications?)
- How can we effectively transfer research into industry?

- How can researchers get access to logs? (What will stop industry from sharing logs?)
- What approaches could be used to generate logs to share within the research community?
- How generalisable are the techniques/findings of log analysis on specific logs?
- How can we evaluate approaches to log analysis? (What kind of benchmarks do we need, how do we generate them and what kind of evaluation campaign should be run?)
- What are the future challenges/directions for the field of query log analysis/mining? (e.g. eye tracking, web search advertising, time-series analysis of queries, integration of multiple transaction logs, correlating transaction logs with user behaviour)
- How can we bring researchers from different disciplines closer together?
- What are the niches and contributions that academia can make to log analysis?
- Where are areas for academic - industry collaboration?
- How can we generate funding opportunities from grant agencies in log analysis?
- Can we develop a meta-methodology that combines log analysis with other methods to provide a "truer" picture of the user - system - information interaction process?

4.2 Discussion groups

On the afternoon of day two, participants were divided into two groups and asked to discuss the questions in Section 4.1, selecting what they felt to be the most important questions. The results were then reported back to the whole group for discussion between all participants.

Group one (reported back by Mark Sanderson) highlighted two areas discussed:

- **Generating logs for analysis and providing researchers access to logs.**
  - This discussion considered how logs could be generated for research – either industry provides academics with anonymised data, or academics generate their own logs. An example of the former is the Microsoft Live Search log; an example of the latter being the Lemur Query Log Project [49] which aims to “create a database of Web search activity that will be provided to the information retrieval research community to use on current and future information retrieval research projects.” Comments were raised on whether people were actually taking part in this study and also whether the resulting logs would be biased (i.e. the value of generating logs in this way). An issue raised was privacy and whether individuals could be identified from log activities (particularly in known circles like the IR community).
- **Sharing solutions and approaches across different fields/disciplines.**
  - This discussion considered the importance of sharing knowledge across different fields or disciplines so that the same things are not re-discovered. There was a feeling that knowledge was not being shared effectively and suggestions included a new journal on log analysis, a book or publication summarising the field (published in multiple venues) and more workshops (like this one). Categorising what can be done with log analysis and generating tables of solutions was seen as more suitable than lists of past work.

Group two (reported back by Jim Jansen) highlighted the following areas as being important issues to consider:

- **Contributions that academics can make to log analysis.**

- The group discussed how academia could contribute to the field of query log analysis and suggestions included coming up with new and novel uses for log files; generating trusted sources of information and evidence to support log analysis; producing standards (e.g. for sessions and markup languages), providing education and training materials; providing basic/fundamental research.

- **Encouraging academic-industry collaboration.**
  - This discussion focused on how academics and industry might collaborate to disseminate knowledge and best practice. Suggestions included: companies (and funding bodies) supporting visiting scholars, practitioners providing guest lectures on academic courses (vice-versa); conferences focused on query log analysis and web mining; industrial research grants; internships; informal collaborations[4], NSF and EU projects with industry and networks (e.g. the EU Network of Excellence programme).

4.3 Dissemination

Details of the query log workshop can be accessed from the accompanying website [1], which contains information on participants, their talks (copies of the presentation slides and MP3 audio files of the talks can be downloaded), and links to related resources on log analysis. Following the event, publications are planned in SIGIR Forum and the BCS magazine for members: ITNOW.

5 Reflections

The event was aimed at gathering together academics (and members from industry) from different disciplines to share and discuss topics related to the area of query log analysis. The event was successful in attracting several academics, well-known for their work in log analysis and providing an environment in which to focus on and discuss the topic. The combination of presentations, social activities and discussion groups seemed to work well in stimulating interaction between attendees. The tutorial by Jim Jansen gave a useful summary of the area and all presentations are publicly-accessible from the workshop website. (The tutorial by Jim Jansen is also available from SlideShare[3].)

Various issues related to query log analysis were raised throughout the event and the presentation by Bettina Berendt[5] is a useful summary that addresses all of the questions circulated for discussion. There were many interesting areas of discussion at the event, but some of the recurring issues that were raised include the following:

- **Availability and use of log data**
  - Making logs publicly available for general use
    - The AOL logs public; should they be used?
  - Should log data be gathered for specific tasks (the value of general log data?)
  - Query logs alone are not always enough; also need demographic information
    - Can mobile logs be made available?
  - There is a lack of recent and **long-term** data (e.g. over several months)
    - AOL logs are 3 months (but should they be used?)
    - The Microsoft Live Search log is anonymised (cannot track users across days).

[4] An example of this as a result of the query log analysis workshop are the discussions underway between Bettina Berendt/ Nick Luft and Paul Clough/ Stella Clarke.
[5] http://ir.shef.ac.uk/cloughie/qlaw2009/ presentations/berendt.pdf

- Verifiability and repeatability of experiments (especially if access to logs is limited) and the availability of standards (needed to repeat experiments).
- Click data is often used as indicator of relevance, but clicks are noisy and unreliable
  - Users can click on incorrect results or by mistake
  - Users may find answers in the snippets of search results resulting in no clicks
  - Need to identify and capture alternatives too (not just the clicks)
  - The quality of data is more important than techniques
- There are biases in log data which need to be understood
  - e.g. users prefer top ranked results
  - e.g. information provided by search engines may affect users behaviours
- Need to consider methods for ensuring privacy
  - e.g. NDAs and privacy-preserving methods

- **Correlating queries and clicks with user behaviour**
  - Understanding human behaviour and developing suitable cognitive models
    - Human behaviour is unpredictable so should we try and model it, how?
  - Mapping between low-level representations of user activities (queries and clicks) and high-level cognitive models ("meta models")
  - Gathering personal data in a large-scale way is difficult (ethical issues).
  - Many of the measures used in behaviourism are not well developed (e.g. measures for users' cognitive styles).
  - Predicting user behaviour – what level of prediction?
    - Individual, group or population (similar to other fields, e.g. in authorship attribution it is easier to classify a style of writing for a group than identify an individual author).
- **Integrating query logs with other sources of user activity**
  - How can multiple streams of data be combined to build up a richer picture of activity?
  - Several talks discussed meta-methodologies
    - e.g. combining log data with questionnaire responses from individual users
  - Can sources of data, such as social networking sites (e.g. YouTube, Flickr, Facebook), and data feeds (e.g. Twitter) be used?
  - Gathering data from logging devices such as the Lemur Toolbar - useful data but may be harder to log? (need users' agreement)
  - Are the log files enough to represent search context?
    - Web pages come and go; need to also capture the pages relating to clicks to rebuild the user's search
    - Search engines will change their results and therefore need also to capture this data too before things change (relates to repeatability of experiments)
    - Need to capture the processes used to create the logs
- **Applicability of techniques and results across domains**
  - There are specific problems with log data in domains such as intranet search
  - Certain techniques will be generic; can results be generalised?

Was the workshop a success? Jim Jansen wrote this in his blog following the event:

"… Overall, the workshop was really good. One of the workshop aims was to bring together researchers and practitioners. In this regard, the workshop was heavy on the 'research' and light on the 'practice'. So, this would be an area to concentrate on in future workshops."(http://jimjansen.blogspot.com/2009/06/query-log-analysis-from-research-to.html)

6 Summary

This report has summarised the TrebleCLEF Log Analysis Workshop entitled "Query Log Analysis: From Research to Best Practice" held May 27-28 2009. This event was attended by 12 invited speakers and 9 further attendees. The workshop included presentations from well-known academics and members from industry and provided a forum in which to discuss topics related to query log analysis. The event successfully brought together people from different disciplines to stimulate discussion and cross-fertilisation of knowledge. Further details about the workshop and downloads (of slides and audio files) can be obtained from the query log analysis workshop website [1]. Following on from Jim Jansen's comment there is definitely a need for future events like the query log analysis workshop in which academics and members from industry can come together and interact with each other: to learn, share experiences and promote knowledge transfer.

7 Acknowledgements

We thank the attendees of the TrebleCLEF Query Log Analysis workshop for their presentations and participation in the discussions.

8 References

[1] TrebleCLEF Query Log Analysis Workshop: http://ir.shef.ac.uk/cloughie/qlaw 2009/

[2] Fabrizio Silvestri (2009) Mining Query Logs: Turning Search Usage Data into Knowledge. To appear in Foundations and Trends in Information Retrieval. Now Publisher.

[3] Jaideep Srivastava, Robert Cooley (2003) Web Business Intelligence: Mining the Web for Actionable Knowledge. INFORMS Journal on Computing 15(2): 191-207.

[4] Weischedel, Birgit and Huizingh, Eelko K. R. E. (2006) Website optimization with web metrics: a case study. In: Fox, Mark S. and Spencer, Bruce (eds.) Proceedings of the 8th International Conference on Electronic Commerce - ICEC 2006, Fredericton, New Brunswick, Canada. 463-470.

[5] Mazlita Mat-Hassan, Mark Levene (2005) Associating search and navigation behavior through log analysis. JASIST 56(9): 913-934.

[6] Michael D. Cooper (2001) Usage patterns of a Web-based library catalog, Journal of the American Society for Information Science and Technology, v.52 n.2, 137-148.

[7] Cooley, R., Mobasher, B., and Srivastava, J., (1999) Data Preparation for Mining World Wide Web Browsing Patterns, Journal of Knowledge and Information Systems, v1 i1, 5-32.

[8] Kohavi R.,Masand B., Spiliopoulou M., and Srivastava J. (2002) Web Mining, Data Mining and Knowledge Discovery, Volume 6, Number 1, January 2002 , 5-8(4).

[9] Pranam Kolari, Anupam Joshi, (2004) Web Mining: Research and Practice, Computing in Science and Engineering, vol. 6, no. 4, pp. 49-53.

[10] Raymond Kosala, Hendrik Blockeel, (2000) Web Mining Research: A Survey, ACM SIGKDD Explorations Newsletter, Volume 2, Issue 1.

[11] Wang Jicheng, Huang Yuan, Wu Gangshan, Zhang Fuyan. (1999) Web mining: knowledge discovery on the Web, Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999, Volume 2, 137 - 141 vol.2, 12-15.

[12] Baeza-Yates, R. and Tiberi, A. (2007) Extracting semantic relations from query logs. In Proceedings of the 13th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining (San Jose, California, USA, August 12 - 15, 2007). KDD '07. ACM, New York, NY, 76-85.

[13] Jansen, B. J. and Spink, A. 2005. How are We Searching the World Wide Web?: An Analysis of Nine Search Engine Transaction Logs. Information Processing & Management. 42(1), 248-263.

[14] Jansen, B. J. 2006. Search log analysis: What is it; what's been done; how to do it. Library and Information Science Research, 28(3), 407-432.

[15] Ford, N., Miller, D. and Moss, N. (2005). Web search strategies and human individual differences: a combined analysis. Journal of the American Society for Information Science and Technology, 56(7), 757-764.

[16] Whittle, M., Eaglestone, B., Ford, N., Gillet, V. and Madden, A. (2007) Data mining of search engine logs. Journal of the American Society for Information Science and Technology, 58(14), 2382-2400.

[17] Ricardo A. Baeza-Yates, Carlos A. Hurtado, Marcelo Mendoza, Georges Dupret: Modeling User Search Behavior. LA-WEB 2005: 242-251.

[18] Ivory, M. Y. and Hearst, M. A. (2001) The state of the art in automating usability evaluation of user interfaces. ACM Comput. Surv. 33, 4 (Dec. 2001), 470-516.

[19] Drott, M. C. (1998) Using Web server logs to improve site design. In Proceedings of the 16th Annual international Conference on Computer Documentation (Quebec, Quebec, Canada, September 24 - 26, 1998). SIGDOC '98. ACM, New York, NY, 43-50.

[20] Spiliopoulou, M. (2000), "Web usage mining for Web site evaluation", *Communications of the Association for Computing Machinery*, Vol. 43 pp.127-34.

[21] Liu, Y., Fu, Y., Zhang, M., Ma, S., and Ru, L. 2007. Automatic search engine performance evaluation with click-through data analysis. In Proceedings of the 16th international Conference on World Wide Web (Banff, Alberta, Canada, May 08 - 12, 2007). WWW '07. ACM, New York, NY, 1133-1134.

[22] T. Joachims (2002) Evaluating retrieval performance using click-through data. In Proceedings of the SIGIR 2002 Workshop on Mathematical/Formal Methods in Information Retrieval. ACM Press.

[23] Croft, B., Metzler, D., and Strohman, T. (2009) Search Engines: Information Retrieval in Practice, Publisher: Addison-Wesley.

[24] T. Joachims, Optimizing Search Engines Using Clickthrough Data, Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD), ACM, 2002.

[25] Poblete, B., and Baeza-Yates, R. (2008) Query-Sets: Using Implicit Feedback and Query Patterns to Organize Web Documents, In Proceedings of World Wide Web 2008 (WWW 2008) Conference, April 21-25, Bejing, China, 41-50.

[26] Robertson, S., and Zaragoza, H. (2007) On rank-based effectiveness measures and optimization. Inf. Retr. 10(3): 321-339.

[27] Weischedel, B. and Huizingh, E. K. (2006) Website optimization with web metrics: a case study. In Proceedings of the 8th international Conference on Electronic Commerce: the New E-Commerce: innovations For Conquering Current Barriers, Obstacles and Limitations To Conducting Successful Business on the internet (Fredericton, New Brunswick, Canada, August 13 - 16, 2006). ICEC '06, vol. 156. ACM, New York, NY, 463-470.

[28] Stephen G. Eick (2001) Visualizing online activity. Commun. ACM 44(8), 45-50.

[29] Stephen G. Eick (2002) Visual Analysis of Website Browsing Patterns. Visual Interfaces to Digital Libraries 2002, 65-80.

[30] Hong et al. (2001) WebQuilt: A proxy-based approach to remote web usability testing. ACM Trans. Inf. Syst. 19, 3 (Jul. 2001), 263-285.

[31] Brusilovsky, P., Kobsa, A., Nejdl, W. (2007) Data Mining for Personalization. In The Adaptive Web: Methods and Strategies of Web Personalization, Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.). Lecture Notes in Computer Science, Vol. 4321, PP. 90-135, Springer, Berlin-Heidelberg, 2007.

[32] Berendt, B. & Spiliopoulou, M. (2000) Analysis of navigation behaviour in web sites integrating multiple information systems. The VLDB Journal, 9, 56-75.

[33] Eirinaki, M. and Vazirgiannis, M. (2003) Web mining for web personalization. ACM Trans. Internet Technol. 3, 1 (Feb. 2003), 1-27.

[34] I-Hsien Ting (2008) "Web Mining Applications in E-commerce and E-services" Online Information Review, Vol. 32, No.2, 129-132.

[35] Ron Kohavi, Llew Mason, Rajesh Parekh, Zijian Zheng (2004) Lessons and Challenges from Mining Retail E-Commerce Data. Machine Learning 57(1-2), 83-113.

[36] Radlinski, F. and Joachims, T. (2005) Query chains: learning to rank from implicit feedback. In Proceedings of the Eleventh ACM SIGKDD international Conference on Knowledge Discovery in Data Mining (Chicago, Illinois, USA, August 21 - 24, 2005). KDD '05. ACM, New York, NY, 239-248.

[37] Jaideep Srivastava, Robert Cooley (2003) Web Business Intelligence: Mining the Web for Actionable Knowledge. INFORMS Journal on Computing 15(2), 191-207.

[38] Creese, G. (2000) Web analytics: Translating clicks into business. White Paper, Aberdeen Group, 2000; http://www.aberdeen.com.

[39] Pfenning, A. (2001) Businesses Must Pay More Attention To Website Metrics. InternetWeek, December 10.

[40] Hofacker, C. F. and J. Murphy (2005) Using Server Log Files and Online Experiments to Enhance Internet Marketing. In Contemporary Research in E-Marketing (S. Krishnamurthy), Idea Group, Hershey, PA, 226-249.

[41] Sterne, J. (2002) Web Metrics. John Wiley & Sons, Inc., New York City.

[42] Sen, A., Dacin, P. A., and Pattichis, C. (2006) Current trends in web data analysis. Commun. ACM 49, 11 (Nov. 2006), 85-91.

[43] TrebleCLEF Coordination Action: http://www.trebleclef.eu/

[44] Kruschwitz, U. and H. Al-Bakour (2005) Users Want More Sophisticated Search Assistants - Results of a Task-Based Evaluation. Journal of the American Society for Information Science and Technology (JASIST), 56(13): 1377-1393.

[45] Kruschwitz, U., N. Webb and R. Sutcliffe (2009) Query Log Analysis for Adaptive Dialogue-Driven Search. In B.J. Jansen, A. Spink and I. Taksa (eds.): Handbook of Research on Web Log Analysis, pages 389-416. Hershey, PA: IGI.

[46] Web Analytics Association: http://www.webanalyticsassociation.org/

[47] Jansen, B., Taksa, I. and Spink, A. (2009) Research and Methodological Foundations of Transaction Log Analysis, Available for download: http://www.igi-global.com/downloads/excerpts/8282.pdf

[48] LogCLEF website: http://www.uni-hildesheim.de/logclef

[49] Lemur Query Log Project: http://lemurstudy.cs.umass.edu/

Appendix A – Attendees (full details)

Invited speaker	Institution	Information about the speaker
Dr. Jim Jansen	Pennsylvania State University (USA)	Jim Jansen (ist.psu.edu/faculty_pages/jjansen/) is working on developing robust methods of data analysis that support the paradigm of log content as temporal information streams and provide predictive behaviors with reduced computational costs given the volume of data, acceptance of richer range of characteristics, and enhanced predictability modelling.
Dr. Fabrizio Silvestri	ISTI-CNR (Pisa, Italy)	Fabrizio Silvestri (pomino.isti.cnr.it/~silvestr/) has been using query logs, in the past, mainly to improve efficiency of Web IR systems. He has been involved in designing a new caching policy (SDC) tailored on Web Search Engines workload. He has also designed a data partitioning and query routing technique for better exploiting the computing capacity of Web IR systems running large clusters of PCs. Recently he has started to work on using query logs to produce query shortcuts, i.e. suggesting queries that will reduce the length of user query sessions.
Dr. Lynn Silipigni Connaway	Online Computer Library Center, Inc. (Ohio, USA)	Lynn Silipigni Connaway (www.oclc.org/research/staff/connaway.htm) has analysed transaction logs to identify discovery and access patterns demonstrated by users accessing the same database (WorldCat) through two different interfaces - First Search and WorldCat.org. WorldCat is the world's largest and most comprehensive bibliographic database, with 125 million records that represent more than 1.3 billion individual items in libraries worldwide. WorldCat is the underlying database for both FirstSearch and WorldCat.org. FirstSearch provides electronic access to dozens of databases and more than 10 million full-text and full-image articles while WorldCat.org is the web-based interface for WorldCat. The discussion will include an analysis of behaviors and patterns revealed in FirstSearch and WorldCat.org search logs as well as a comparison of the different search behaviors demonstrated in the two different interfaces of the data base. The log analyses suggest that users exhibit different behaviors and achieve different levels of success when using the different interfaces. The findings provide evidence-based information that can be used for the design and management of discovery and access tools.
Prof. Bettina Berendt	Katholieke Universiteit Leuven (Belgium)	Bettina Berendt (www.cs.kuleuven.be/~berendt) is a professor in the department of computer science at Katholieke Universiteit Leuven, Belgium. She obtained her habilitation in information systems from Humboldt University Berlin, Germany, and her Ph.D. in computer science/cognitive science from the University of Hamburg. Her research interests include Web and Social-Web mining, digital libraries, personalization and privacy, and information visualization. Bettina Berendt's work on Web log analysis has focussed on determining

		user interests by service-based log analysis, query mining, mining with background knowledge, and on investigating the impact of linguistic and cultural diversity on Web usage behaviour.
Prof. Nigel Ford	University of Sheffield (UK)	Nigel Ford's (www.shef.ac.uk/is/staff/ford.html) main research focus is on the psychology of the web searcher, and he has conducted a number of studies of web searching to support learning. This work has entailed the analysis of query logs from searchers for whom personal data are known, such as gender, age, cognitive style, and study approach. He is particularly interested in using log analysis to identify associations between human individual differences, search patterns and effectiveness, the ultimate goal of which is to build user models capable of driving adaptive behaviour in search tools.
Dr. Thomas Mandl	University of Hildesheim (Germany)	Thomas Mandl (www.uni-hildesheim.de/~mandl/) is organiser of a large-scale evaluation campaign called LogCLEF which is being run as part of the Cross Language Evaluation Forum (CLEF) 2009. He is interested in user behavior in information retrieval systems and social networks. The focus lies on user behavior in multilingual contexts. For future research, the relation between user studies and log analysis might open new perspectives. Thomas Mandl worked as a research assistant at the Social Science Information Centre in Bonn, Germany and as assistant professor at the University of Hildesheim in Germany where he is teaching in the programme International Information Management. He received a doctorate degree and a post doctoral degree (Habilitation) from the University of Hildesheim. His research interests include information retrieval, human-computer interaction and internationalization of information technology.
Dr. Filip Radlinski	Microsoft Research (Cambridge, UK)	Filip Radlinski (research.microsoft.com/en-us/people/filiprad/) is an applied researcher at Microsoft Research in Cambridge, UK where he is in the Information Retrieval group and also works for Live Search. He completed his PhD in Computer Science on learning to rank from implicit feedback at Cornell University in 2008, advised by Thorsten Joachims. His research interests focus on the practical and theoretical challenges of learning to rank, and evaluating online ranked retrieval systems, using behavioural data recorded as people use such systems. He has used search log files to learn improved ranking functions, as well as to perform interactive evaluation of search systems. In particular, he has recently studied the relative power of various metrics that can be derived from log files for evaluating the differences between ranking functions.
Prof. Mark Levene	Birkbeck, University of London (UK)	The main focus of recent work in this area by Mark Levene (www.dcs.bbk.ac.uk/~mark) has been on predicting users' navigation behaviour from server logs. Regarding query log analysis Mark has been working on query classification and applying this to topic specific analysis of query logs and for improving users' search experience.

Dr. Udo Kruschwitz	University of Essex (UK)	The log analysis research by Udo Kruschwitz (cswww.essex.ac.uk/staff/udo/) and colleagues focuses primarily on intranet collections. They have collected a query corpus of more than half a million queries submitted to a local search engine which is now being used as part of the AutoAdapt project to build adaptive domain models. They have also recently obtained funding for another KTP project with a major internet recruitment service where we will be analyzing substantial log files related to job search.
Dr. Vanessa Murdock	Yahoo! (Barcelona, Spain)	Vanessa Murdock (research.yahoo.com/Vanessa_Murdock) is a research scientist at Yahoo! Research in Barcelona. Her work focuses on leveraging click data to improve the placement of sponsored search results and contextual advertising, and images. She received her PhD in 2006 from the University of Massachusetts where she worked with Bruce Croft at the Center for Intelligent Information Retrieval.
Dr. Giorgio Di Nunzio	University of Padoa (Italy)	Giorgio Maria Di Nunzio is co-organiser of the LogCLEF evaluation task which is being run as part of the Cross Language Evaluation Forum (CLEF) 2009. His main interest in the field of log analysis concerns knowledge extraction from different sources of data: Web logs, search logs and user studies, He has been involved in the analysis of the logs of The European Library (http://www.theeuropeanlibrary.org/) since 2006. (ims.dei.unipd.it/websites/archive/ims2009/members/dinunzio.html)
Dr. Dhavval Thakker	Press Association Images (Nottingham, UK)	Dhavval Thakker (http://jaala.co.uk/) is a KTP Research Associate working on “Intelligent Image Search Engine” project in partnership with Press Association Images and Nottingham Trent University. He received a doctorate degree from the Nottingham Trent University and an MSc degree from the Brunel University, UK. His present research interests are in the areas of Natural Language Processing & Semantic Web and their application in the image annotation and retrieval processes. In the query log analysis domain, he is mainly interested in search patterns and user behaviours in image search engine systems.

Appendix B – Advertising flyer

Query Log Analysis: From Research to Best Practice

Type of Event

A one-and-a-half day brain-storming workshop to bring together academics in the field of query log analysis to share their research experiences and contribute their thoughts on emerging trends and potential best practices.

Event Details

The aim of this event is to establish a forum in which invited speakers from multiple disciplines can share and discuss their experiences from analysing query logs. By involving representatives from different disciplines and selected business communities, we hope to stimulate the cross-fertilisation of knowledge and ideas to establish best practices for conducting and utilising query logs in practical commercial settings. By inviting well-known academics we aim to clarify current research (e.g. the terminology and approaches used), collate standardised procedures and resources commonly used, identify common challenges, and stimulate thoughts on future directions of the field.

The event will be organised as one-and-a-half day event to attract researchers and practitioners and provide an environment to brainstorm and discuss ideas. The event aims to start at 2pm on the first day (lasting until approximately 4pm the following day) with the first session providing a forum for academics and invited business representatives to present their present their experiences with analysing query logs. This may take the form of a series of short talks/position papers to facilitate later discussions. The next day will pick up from the previous session and try to summarise the main points raised during the presentations. The day will then comprise a series of discussion groups between academics (and other invited guests) to address some of the themes of the workshop. The aim will be to address questions such as what research has been done and where the field is moving, problems and challenges of query log analysis, potential solutions to raised issues, and how to move from research to practice. This will provide input to form a preliminary best practice for query log analysis. This will end with a summary session and a series of key questions for panel members (selected academics) to respond to.

Dates	Venue	Number of Attendees
27-28 May 2009	BCS London Office	Approx. 30 guests

Contact

Dr. Paul Clough, University of Sheffield (UK) p.d.clough@sheffield.ac.uk

Sponsor

This event is funded by the EU-funded TrebleCLEF project: IST-215231

TrebleCLEF

Evaluation, Best Practices and Collaboration for Multilingual Information Access

Appendix C – Programme

Wednesday 27th May 2009 (Wilks room 2, BCS London Office)		
9.30-10.00	Arrival (tea and coffee)	
10.00-10.30	Welcome and introduction	Paul Clough (University of Sheffield)
10.30-11.00	Presentations	Mark Levene (Birkbeck, University of London)
11.00-11.30		Nigel Ford (University of Sheffield)
11.30-12.00		Jim Jansen Penn State University)
12.00-13.00	Lunch	
13.00-13.30	Presentations	Filip Radlinski (Microsoft Research)
13.30-14.00		Vanessa Murdock (Yahoo! Research)
14.00-14.30		Lynn Silipigni Connaway (OCLC)
14.30-15.00		Dhavval Thakker (Press Association)
15.00-15.30	Break	
15.30-16.00	Presentations	Fabrizio Silvestri (ISTI-CNR)
16.00-16.30		Bettina Berendt (Katholieke Universiteit, Leuven)
16.30-17.00		Udo Kruschwitz (University of Essex)
20.00	Workshop Dinner	Zizzi Ristorante (73-75 The Strand, WC2R 0DE) http://www.zizzi.co.uk/restaurants/93
Thursday 28th May 2009 (Wilks room 2, BCS London Office)		
08.30-9:00	Arrival (tea and coffee)	
09.00-10.00	Tutorial session	Jim Jansen (Penn State University)
10.00-10.30	Presentations	Giorgio Di Nunzio (University of Padoa)
10.30-11.00		Thomas Mandl (University of Hildesheim)
11.00-12.00	Discussion	(see discussion questions)
12.00-13.00	Lunch	
13.00-13.30	Discussion	(see discussion questions)
13.30-14.00		
14.00-14.30		
14.30-15.00		
15.00-15.30	Summary and close	Paul Clough (University of Sheffield)
15.30	Depart	