Report on the TrebleCLEF Query Log Analysis Workshop 2009

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

Bettina Berendt
Katholieke Universiteit Leuven, Belguim
bettina.berendt@cs.kuleuven.be

Abstract

This report summarises a workshop organised as a part of the EU-funded TrebleCLEF project entitled "Query Log Analysis: From Research to Best Practice" held on 27-28th May 2009 at the British Computer Science 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. 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 and the presentations given by the invited speakers, and it summarises discussions held.

1 Introduction

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 [1, 2]. Logs files are being studied in several domains, both academic and commercial including: digital libraries [3], Web data mining [4, 5], information seeking and search behaviour [6, 7, 8], usability assessment [9], website design and evaluation [10], Web search evaluation [11, 12], Web search optimization [13, 14], Web Analytics [15], information visualization [16], adaptive systems and personalisation [17, 18], e-commerce [19, 20], learning to rank from implicit feedback [21] and business intelligence [22]. However, despite the obvious commercial (and research) benefits of utilising such data, many organisations collect but do not use their log file data effectively [23, 24, 25, 26]. 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 [27] Query Log Analysis Workshop, entitled "Query Log Analysis: From Research to Best Practice" was held May 27-28th 2009 at the British Computer Science Offices in London. The goal of the workshop was to provide a forum in which invited speakers from multiple disciplines could share and discuss their experiences with query (and server) log analysis. The workshop was seen as a starting point in addressing the wider goals of clarifying current research (e.g. the terminology and approaches used), collating standardised procedures and resources commonly used,

identifying common challenges, and stimulating thoughts on future directions of the field. A more detailed summary of the workshop can be found in [28, 29].

2 Workshop

2.1 Participants

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 (e.g. social, library and computer science) and business sectors. The remaining participants consisted of academics and representatives from local businesses. All attendees of the event had some level of interest and experience in conducting log analysis.

2.2 Discussion questions

Before the workshop, a number of discussion questions were circulated to the invited speakers. These questions raise several important issues facing the field of query log analysis today. 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?

2.3 Presentations

Presentations 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). Jim Jansen also provided a tutorial on Web log analysis. Slides and audio recordings of the talks and tutorial are available via the workshop website [29].

2.3.1 From server logs to query logs

Mark Levene (Birkbeck, University of London, UK) discussed three areas of his research: server log data analysis, discovery of context-topic association rules from search engine logs and the topical analysis of search engine query logs. Mark raised a number of issues on availability of log data: the lack of recent and publicly available log data; the need for additional contextual data other than logs (e.g. demographic information) and current problems with the verifiability and repeatability of experiments involving log data.

2.3.2 Query log analysis and individual differences

Nigel Ford (University of Sheffield, UK) discussed his research in investigating strategic differences in the way different types of individuals query information access systems. Nigel showed how the combination of query log analysis and responses from questionnaires (e.g. to capture user's cognitive styles) could be used to establish various patterns of user interaction and behaviour. Nigel highlighted a number of challenges for large-scale analysis and research, including: complexities in the experimental setup, under-developed measures, and coordinating and gathering personal data,.

2.3.3 Moving from description to prediction for information searching

Jim Jansen (Pennsylvania State University, USA) discussed the current state of information searching research, which he sees as mainly descriptive (e.g. reporting search trends) and lacking more predictive approaches and models. Jim discussed his own model based on treating search engine logs as information streams and considering information searching as a temporal stream (i.e. stateless). Jim pointed out that current methods (e.g. n-grams, decision trees and time series analysis) are not effective for modeling temporal streams and more research is required in this area.

2.3.4 What do click logs tell us about user's search satisfaction?

Filip Radlinski (Microsoft Research, Cambridge, UK) presented a study of search evaluation using click data, absolute metrics (to indicate how good results are) and paired comparisons (to say which results are better). His work shows that users' behaviour changes in response to results presented in different ways and log data can capture such changes. Issues raised in this talk included: the unreliability of click metrics, whether or not log data is reusable (i.e. the use of existing logs in evaluating new search engines), and that all users' interactions and not just clicks should be logged.

2.3.5 Online learning from click data

Vanessa Murdock (Yahoo! Research, Barcelona, Spain) 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. A similar approach is being used to order images returned from a Web image search engine. Similar issues arose with using click data as for textual results. Vanessa also explained some of the biases found in log files that may affect learning-to-rank from clicks, such as the user interface biasing user's behaviour – for instance, past research has shown that users prefer top-ranked results.

2.3.6 Following the trail of WorldCat users

Lynn Silipigni Connaway (Online Computer Library Center Inc., USA) presented a log analysis of WorldCat.org (free on the Web) and FirstSearch (subscription-based). Problems with analysing these logs included: volume of data; limited control over data collection; different data being captured because of modifications to the system; multiple codes/formats for search fields and incomplete information about users and access to resources (e.g. held in different transaction logs). Directions for future work included automating analysis and linking search behaviours to demographics**.**

2.3.7 Using query logs at PA Images

Dhavval Thakker (Press Association Images, UK) discussed the use of query logs to improve the performance of a commercial image search engine. An interesting point coming out of this talk is that the kinds of logs the company generate are bespoke and limited (e.g. contain no click-through data). This therefore limits the kind of analysis that can be carried out on the logs. Dhavval presented the results of an initial descriptive analysis of the query logs and indicated that the company was currently experimenting with existing Web Analytics tools (e.g. Google Analytics).

2.3.8 Query log mining @ HPC-Lab

Fabrizio Silvestri (ISTI-CNR HPC, Italy) discussed the use of query logs to produce high-performance IR systems. Fabrizio described current work including similarity-based caching, using click-through data to generate search shortcuts and the organization of indexes driven by query logs. Fabrizio discussed future areas of his research including: identifying human activities from logs (i.e. beyond identifying sessions), analysis of large (and multilingual) log files, analysis of long-term logs (spanning multiple months) and making available publicly-accessible logs to enable reproducible results.

2.3.9 Exploratory analysis needs theor[y|ies] – OR: Some answers to 14 questions

Bettina Berendt (Katholieke Universiteit Leuven, Belgium) summarized her work on analysing logs for learning about user behaviours, leveraging and/or formulating domain ontologies, and improving information design. She emphasised the importance and challenges of combining data sources, measurement and analysis methods, and background theories. Bettina suggested a number of ideas on the future challenges/directions for the field including: combining methods/ theories; analysis of interaction beyond navigation and querying; protecting privacy; time-series analysis; integration of multiple transaction logs and correlating transaction logs with user behaviour.

2.3.10 Log analysis at Essex

Udo Kruschwitz (University of Essex, UK) presented query log analysis for adaptive intranet search (academic application) and 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. Udo highlighted that the queries in logs generated by intranet search differ from general Web search (due to the focus on a particular domain).

2.3.11 Logging digital libraries

Giorgio Di Nunzio (University of Padoa, Italy) discussed the role of log analysis in digital library evaluation. Giorgio discussed work undertaken in The European Library (TEL) project in which logfile analysis was combined with HTTP server logs and questionnaires to gather information on user preferences and satisfaction. 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 (Log Analysis for Digital Societies - LADS).

2.3.12 Query classification in log file analysis: evaluation issues and user satisfaction

Thomas Mandl (University of Hildesheim, Germany) discussed work on query classification, in particular with respect to location where he is running a large-scale CLEF 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.

3 Discussions

Participants were divided into two groups and asked to discuss the questions in Section 2.2, selecting what they felt to be the most important questions and requiring immediate attention. These were:

- Generating logs for analysis and providing researchers access to logs
- Sharing solutions and approaches across different fields/disciplines
- Contributions that academics can make to log analysis
- Encouraging academic-industry collaboration

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: consideration of how log files should be made publicly available to researchers, whether log data should be gathered for specific tasks, whether there is value in general log data, how additional information can be gathered and correlated with query log data, how to address the current lack of recent and long-term data, verifiability and repeatability of experiments is important (especially if access to logs is limited), the lack of available standards for conducting experiments in log analysis which are needed to enable repeatability of experiments, click data is often used as an indicator of relevance but clicks are noisy and unreliable, the biases which inherently exist in log data need to be understood before and during analysis, methods to ensure privacy must be investigated and agreed upon and further research should be conducted to ascertain whether results from analyzing specific logs are generalisable to a wider context.

Correlating queries and clicks with user behaviour: further understanding of human behaviour is necessary to develop suitable cognitive models, human behaviour is unpredictable and difficult (if at all possible) to model, it is unclear how to map between low-level representations of user activities (queries and clicks) to the higher-level cognitive models (i.e. “meta models”), gathering personal data in a large-scale way is difficult and creates ethical issues which must be addressed by researchers, many of the behavioural measures used are not well developed, and success in predicting user behaviour varies depending on the level of analysis (e.g. at the level of individual, group or population).

Integrating query logs with other sources of user activity: multiple streams of data need to be combined to develop a richer picture of activity, the use of multiple methodologies (e.g. combining log data with questionnaire responses from individual users) is necessary to compensate for the weaknesses of log data alone, utilizing additional data sources such as social networking sites (e.g. YouTube, Flickr, Facebook) and data streams (e.g. Twitter) can be used to complement log data, how logging tools (e.g. the Lemur Toolbar) should be used to generate data and capture a richer picture of user activity, whether query log files alone are enough to represent search context (e.g. may need to capture content of Web pages, related pages, sponsored links etc. to rebuild the user’s search) and where possible the processes used to create the log data (e.g. parameter settings/biases of search engines) should be captured.

Participants concluded that an essential next step is to gather, integrate and structure the multitude of – overlapping, complementary, and sometimes even contradictory – findings from the different fields that have, often independently of one another, investigated query and other web-activity logs.

4 Acknowledgements

The workshop was supported by the TrebleCLEF project [3]: a coordinated action funded under ICT-1-4-1 Digital libraries and technology-enhanced learning; grant agreement 215231.

5 References

[1] Jansen, B.J. (2008). "Research and Methodological Foundations of Transaction Log Analysis", In: Jansen, B., Spink, A. and Taksa, I. (eds), Handbook of Research on Web Log Analysis, pp. 100-123, Hershey, PA: IGI Global.

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

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

[5] Kolari, P., and Joshi, A. (2004) Web Mining: Research and Practice, Computing in Science and Engineering, 6(4), 49-53.

[6] Mat-Hassan, M., and Levene, M. (2005) Associating search and navigation behavior through log analysis, Journal of the American Society for Information Science and Technology, 56(9): 913-934.

[7] 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.

[8] 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.

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

[10] Spiliopoulou, M. (2000), Web usage mining for Web site evaluation, Communications of the Association for Computing Machinery, 43(8), August 2000, 127-34.

[11] 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.

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

[13] 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.

[14] Robertson, S., and Zaragoza, H. (2007) On rank-based effectiveness measures and optimization, Information Retrieval, 10(3), 321-339.

[15] 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.

[16] Stephen G. Eick (2001) Visualizing online activity. Communications of the ACM 44(8), 45-50.

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

[18] Eirinaki, M., and Vazirgiannis, M. (2003) Web mining for web personalization, ACM Transactions on Internet Technology, 3(1), Feb. 2003, 1-27.

[19] I-Hsien Ting (2008) Web Mining Applications in E-commerce and E-services, Online Information Review, 32(2), 129-132.

[20] Kohavi, R., Mason, L., Parekh, R., and Zheng, Z. (2004) Lessons and Challenges from Mining Retail E-Commerce Data, Machine Learning, 57(1-2), 83-113.

[21] 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.

[22] Srivastava, J., and Cooley, R. (2003) Web Business Intelligence: Mining the Web for Actionable Knowledge, INFORMS Journal on Computing, 15(2), 191-207.

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

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

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

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

[27] EU-funded TrebleCLEF project: http://www.trebleclef.eu/

[28] Clough, P. (2009) TrebleCLEF Query Log Analysis Workshop Report, TrebleCLEF Deliverable 4.1, available online: http://ir.shef.ac.uk/cloughie/publications/qlaw2009.pdf

[29] Query Log Analysis: http://ir.shef.ac.uk/cloughie/qlaw2009/