PROMISE

Participative Research labOratory for Multimedia and Multilingual Information Systems Evaluation

FP7 ICT 2009.4.3, Intelligent Information Management

Inferring relevance assessments from logs

UNIPD – UvA

May 11-19, 2011

Network of Excellence co-funded by the 7th Framework Program of the European Commission, grant agreement no. 258191

Document Information

Report title:	Researcher exchange: Inferring relevance assessments from logs
Researcher Exchange date:	May 11-19, 2011
Visitor(s):	Giorgio Maria Di Nunzio, UNIPD
Host(s):	Maarten de Rijke, UvA
Preparation date:	11/07/2011
Author(s):	Giorgio Maria Di Nunzio

Table of Contents

1 Introduction

Log recording and analysis allows evaluation assessment and opens opportunities to improvements and enhanced new services. Indeed, the benefits of logging are numerous, including improving performance by recording effective evaluation data, helping in designing and testing of user interfaces, and better allocation of resources. Log analysis provides a method of collecting data from a great number of users. Logs are a reasonable and non-intrusive means of collecting user–system interaction data about the information-searching process from a large number of searchers [1]. Machine learning

Network of Excellence co-funded by the 7th Framework Program of the European Commission, grant agreement no. 258191

algorithms can successfully exploit information encoded in the actions the user perform to improve tasks such precision prediction and query triage.

2 Planned Work

The aim of this exchange is to focus on the problem of inferring relevance assessments from transaction log files. The first step will be to study similar approaches in literature. For example, [2] which study the correlation of the quality of retrieval results with the distribution in time of the documents retrieved. In [3], a relevance modeling solution to this estimation problem was adopted in order to look at the temporal information each of the top retrieved documents provide and weight this information according to the documents probability of relevance. The evaluations of the studied methods will be carried out by means of standard log datasets.

Some of the datasets that will be used are made available by the LogCLEF1 Lab at CLEF 2011. This will guarantee standard datasets which have been used in the past years and that will make the results of this study comparable to others; this study will also take adgvantage of the experience of UvA and UBER, both partners of PROMISE, as participants at LogCLEF.

3 Conducted Work

The work has been mainly focused on the definition of alternative evaluation tasks of search engines by using log data having in mind the possibility to integrate these activities in the DIRECT infrastructure. Four different alternative evaluation where identified and analyzed. The additional information required by the proposed alternative evaluation may involve several tasks, such as:

- Query classification, i.e. language, category, etc.
- Relevance assessments from clicks

[1] http://www.promise-noe.eu/mining-user-preference

- Test session, i.e. how the query evolve during a session
- Diversity of results

These proposals will be reported in the deliverables D3.2 (Specification of the evaluation infrastructure based on user requirements) of WP3 and D4.1 (First report on alternative evaluation methodology) of WP4.

An interface for the annotation of query logs has been designed and developed following, in part, of the indications discussed during this exchange.

The interface can be reached from the following URL

http://www.promise-noe.eu/mining-user-preference/logclef-2011/home

This interface is currently used by the organizers and the participants of the LogCLEF lab to create a rich dataset that would support some of the tasks previously mentioned. The interface asks each participant to annotate a query with the following information:

- annotate the language of the query (undecided or unknown are tags that can be used for a query whose language is ambiguous or unknown)
- annotate the language of the query knowing the language of the interface; in most cases the default language is English (does the information of the language of the interface of the user change your mind or help to understand the language of the query?)
- annotate the change of the query/topic within a session; a subsequent query can be the "same query" if the text of the query didn't change at all, a "generalization" if the user changed the initial query to a broader query, a "specification" if the user changed the query from a wider one to a more narrow query, a "drift" if the user changed the topic of the initial query,
- annotate the query with one or more pre-defined categories (for example, geographical location, title of a work, person, etc.).

Figure 1 and Figure 2 show an example of the current the query annotation used in LogCLEF 2011.

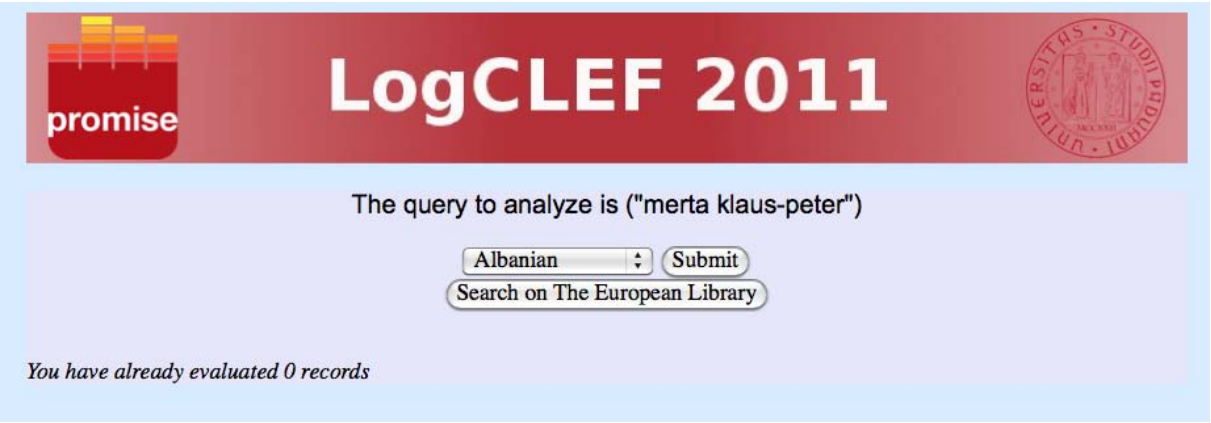

Figure 1. Query annotation interface: language of a query

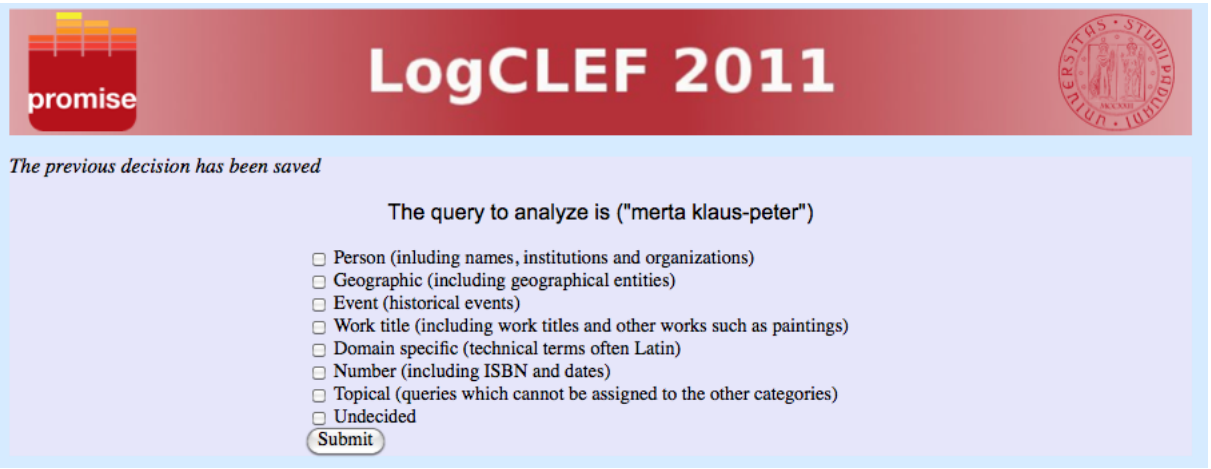

Figure 2. Query annotation interface: query categorization

4 References

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

[2] Jones, R., Diaz, F.: Temporal profiles of queries. ACM Transactions on Information Systems 25 (3) (2007)

[3] Lavrenko, V., Croft, W.B.: Relevance-based language models. In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, September 9-13, 2001, New Orleans, Louisiana, USA. ACM (2001), pp. 120–127