D5.3 Development of query log analysis tools

File name	PuppyIR_PuppyIR_D5.3_v0.8.doc
Author(s)	Monica Landoni (UoS) Andreas Lingnau (UoS) Ian Ruthven (UoS) Sergio Raul Duarte (UT)
Work package/task	WP5/T5.3
Document status	Final
Version	1.0
Contractual delivery date	31/10/2011
Confidentiality	Public
Keywords	Query log analysis, interaction analysis, evaluation support
Abstract	This report presents a framework for query log analyses of children's querying behaviour and a log analysis tool for analysing interaction on novel interfaces. These approaches can be used to learn more about children's search behaviour, to learn what resources are of interest to children and help evaluate novel search systems for children.

Table of Contents

Executive Summary

PuppyIR systems are intended to help children search more effectively. In PuppyIR we are developing several, complementary approaches to understand children's search behaviour and help evaluate novel systems intended for children. D5.2 (*Development of new user centered evaluation measures*) proposed interaction-based measures for understanding performance over time, D2.4 (*Evaluation of novel interfaces in demonstrator systems*) will present a novel qualitative approach for use by young children. In this report we present a third approach based on analyzing logs, both query logs and interaction logs.

We present an investigative framework for tracking and analyzing important aspects of child querying behavior, the Behavior Characterization Framework, based on a study of childrens' online searching and then present an enhanced logging engine based on children's interaction with novel interfaces, in particular touch tables.

1 Introduction

Log analysis has a long history in Information Science and Retrieval, starting with analyses of interactions with online book catalogs (e.g. Sandore, 1993) to query log analyses on Web search engines (e.g. Baeza-Yates, 2004; Baeza-Yates, 2005; Silvestri, 2010).

There are many attractions to query log analyses: the non-intrusive nature of the data collection means that searchers are not interrupted from their normal search behaviour, large-scale data gathering is possible, which facilitates the detection of statistical patterns within the data, and it is an inexpensive method of data gathering compared to traditional qualitative techniques such as interviewing (Jansen, Taksa, & Spink, 2009).

Query log mining, the act of extracting useful information from query logs, is "*concerned with all those techniques aimed at discovering interesting patterns from query logs of web search engine with the purpose of enhancing either effectiveness or efficiency of an online service provided through the web*." (Silvestri, 2010) and, as noted there, although there has been many strides forward in developing techniques for analyzing query logs there are still many challenges. A particular challenge for PuppyIR is how to characterize children's querying behavior (Duarte Torres, Hiemstra, & Serdyukov, 2010a).

As reported in D3.3 (*Report on information extraction and mining techniques*) and in (Duarte Torres, Hiemstra, & Serdyukov, 2010b; Duarte Torres & Weber, 2011)**,** children's querying behavior displays expected and unexpected differences to adult querying behavior. For example, children do have shorter search sessions than adults which may be expected as children's needs may be less complex and their attention span shorter (Duarte & Weber, 2011). However, they also use fewer sentiment words than adults which may be unexpected as children are often more emotionally driven (Smith, Cowie, & Blades, 2003). A particular problem in investigating children's search behavior using query log mining is the lack of data on children's queries. It is particularly difficult to separate children's own queries from those by adults looking for information for use by children and queries where an adult has played a mediating role in the query creation.

Tools that can help analyze child querying and searcher behavior in real-life settings would help the development of appropriate search systems and better understand what information children would like to be able to search.

Based on studies of children's querying behavior using the Yahoo! Search engine, we propose in the next session a Behavior Characterization Framework which is a framework for collecting data on children's querying behavior for use by researchers who are evaluating search systems designed for use by children. Then in session 3 we describe a logging engine to be used to enhance the query log data by capturing more details, like interactions happening in an interface, search results picked by children, changes in query terms etc.

2 Behaviour Characterisation Framework

The Behavior Characterization Framework (BCF) has been developed (Duarte, Hiemstra, & Serdyukov, 2010b; Duarte & Weber, 2011) in order to provide information on the queries submitted to search services by children. As noted in D1.4 (*Specification Report*) the format and management of logging activities will depend upon the particular system implementation as different services using the PuppyIR framework may have different evaluation questions. The BCF provides a set of metrics from which end-users may select those appropriate.

The analysis of query logs can be performed at different levels of detail: single action level and session level. In the former, query and web clicks are analyzed in isolation. In the latter, sequences of actions are grouped using a time window. Tables 1 and 2 show the different

characteristics of the BCF which are being developed and which will be tested within the PuppyIR evaluations.

Query related metrics	**Length of queries** Younger children often submit shorter queries consisting of fewer tokens and shorter words (Duarte & Weber, 2011). Analyzing the length of queries submitted to a search service can provide information on the demographics of children using a search service. Analyzing global statistics, such as the number of unique tokens used in the queries submitted to the service, can give indications of the vocabulary size of the user group. **Metrics:** • Average number of characters used in queries • Average number of tokens used in queries • Query length distribution across all queries • Number of unique tokens used in queries **Natural language in queries** Analysing queries to understand what kind of information need is trying to be resolved can help information service providers understand what children want from their service and provide new information or new services. The use of natural language questions in particular is a useful source of information on the needs of children in the age range 10-18 (Duarte & Weber, 2011). **Metrics:** • Proportion of queries posed as questions in natural language (containing what, why, who etc) • Proportion of specific language constructs. Constructs expressing superlatives (for example, the fastest car, the tallest man) can be seen as implicit questions (what is the fastest car, who is the tallest man). • Proportion of queries that contain words that indicate the user wants to learn about a topic that is unfamiliar, e.g. the use of words such as describe, explain, or define. • Proportion of queries containing positive or negative sentiments which indicate the child wants information will a certain affective nature. This metric can be measured using available online resources (e.g., SentiStrenght)
Topic distribution of queries	The ability to know which topics are being searched can help information providers learn which topics are of interest to users. **Metrics:** • Distribution of topics searched by mapping against a list of predefined topics. In the case of specialist information providers, e.g. museums, this list will be on the main topics of the museum; in the case of general information providers the topic list can be based on generic topic lists such as Yahoo! Directory. • Distribution of entities searched. Again this is either a domain specific list for specialist providers or a generic list such as Wikipedia categories.
Click related metrics	For PuppyIR services which provide a ranked list output then standard click distribution metrics can be used to understand where on the search results page users have clicked.

	Metrics: • Rank position distribution • Number of distinct pages clicked • Number of distinct URLs and domains clicked (for web search)
Query – Click pairs	**Click entropy** can be used to understand the variation in result pages clicked from a query (Weber & Jaimes, 2011). This metric can help understand if queries are ambiguous, and perhaps indicating if users need more support through query suggestions, or the type of query, e.g. (Duarte & Weber, 2011) show how the metric can detect navigational queries.

Table 1: Non-session metrics

Sessions	Session based metrics, calculated by grouping consecutive user actions within a time window, can be useful to understand what users are trying to achieve during a search session. **Metrics:** • Session activity: Number of entries in the session (queries, clicks submitted) can provide useful information on the level of user activity within a search session. • Session duration: the total amount of time spent on searching. • Query/Click re finding (Tyler & Teevan, 2010) is the amount of repeat behavior, i.e. issuing the same query within a session or revisiting the same page within a session. Such looping behavior has been seen as a common orienteering device by children (Bilal & Kirby, 2002). • Click duration distribution which can provide indications on how long children spend on reading information as opposed to query creation and modification (Hassan & Jones & Klinkner, 2010) • Query reformulations patterns. Investigating reformulation patterns (adding terms to queries, removing terms, changing terms, etc) can provide information on where children need support in modifying queries or refining information needs.
Additional log metric analyses	Additional log based metrics can be used depending on the services provided by the search systems, including. • Query suggestions success: Number of instances in which a query suggestion was elected and led to a click (or long click) • Reading level proportion: Proportion of urls clicked classified as advanced, intermediate and basic; e.g, using Google reading level classification as is suggested in (Duartes & Weber, 2011)

Table 2: Session metrics

For each query metric mentioned in Table 1 and 2 a script will be available as part of the PuppyIR framework.

3 Enhanced data collection and analysis

The BCF provides a core set of query and click analysis tools. To provide a more elaborated data analysis we have implemented a logging engine that can be used to enhance the query log data by more details, like interactions happening in an interface, search results picked by children, changes in query terms etc. In particular, as mentioned in D4.3, multitouch logging, a largely unexplored research area, has become the focus of our efforts. The resulting enriched logging data will allow researchers and developers to generate detailed analysis not only of children's search queries but also how they interact with the results afterwards, how they interact with the interfaces, e.g. by moving fiducial objects around on a tabletop device, clicking on objects using a standard input device or even identifying objects that are never touched or moved on the table, and finally which information is re-used to refine or amend queries.

While analysing search behaviour and interaction of adult users can in certain cases be actively supported by the users themselves, children's search behaviour and interaction with information must not rely on any active support by the user.

To provide an extended support of query log and interaction analysis for interfaces in PuppyIR, we make use of a software tools developed by researchers in the Kaleidoscope NoE[1]. For this query and interaction logging we are using the Common Logfile Format (CoLoForm), a XML-based representation of user actions originally developed for interactive and collaborative applications. CoLoForm has been conjointly designed by 9 European research groups in the context of the Kaleidoscope network of excellence. A prototypical use case has been implemented for cross-system analysis of research data with two argumentation systems supported by one joint moderation framework (Harrer, Martinez-Mones, & Dimitracopoulou, 2009).

For our interface prototypes in PuppyIR we have implemented a logging engine as part of the PuppyIR open source libraries which enables developers to generate detailed logging events for every interaction within our interfaces, e.g. sending a query, search results that are used or dismissed, adding or removing tangible items to a table, moving, resizing or deleting virtual objects, complying the CoLoForm XML format (see package eu.puppyir.strath.logger and sub-packages). Events are mainly related to Java objects. This ensures that developers can very easily generate new information and include it in the log data. A detailed manual how to use the logging engine and a guideline how to generate new entries when developing interfaces with the PuppyIR framework will be part of the software documentation.

Table 3 shows an XML log entry for two different interactions with the same virtual objects, an image that has been tagged and is now used as a search term. In the first entry, the SearchImage object has been created. The log entry contains information about the user ("id", "role") and the object itself ("owner", attached EXIF data like the tag used as "query term", coordinates and size on the screen, name of loaded source file). The second entry has been created when the image was moved to another position. In (Lingnau, & Harrer, 2011) we describe an approach how additional information might be amended to the log data, e.g. to improve the information about children interacting with a tabletop device. Video recording can be used to amend the log data and add information about roles and ownership changes. This is information that can usually not be sampled automatically but might be useful for evaluation of children's search behavior and interaction. The XML data in the logfile could be easily annotated before the analysis will be accomplished.

[1] http://www.noe-kaleidoscope.org

```
<action time="1292434063521">
  <actiontype classification="create" logged="true" type="vertex" />
  <user id="Child1" role="originator" />
  <object id="114" type="SearchImage">
   <properties>
    <property name="ownerID" value="10" />
    <property name="query term" value="dodo" />
    <property name="x" value="577" />
    <property name="y" value="431" />
    <property name="width" value="169" />
    <property name="height" value="214" />
    <property name="sourcename" value="Dig52.JPG" />
   </properties>
  </object>
</action>

<action time="1292434065240">
  <actiontype classification="modifiy" logged="true" type="move/resize" />
  <user id="unknown" role="originator" />
  <object id="114" type=" SearchImage">
   <properties>
    <property name="ownerID" value="10" />
    <property name="query term" value="dodo" />
    <property name="x" value="350" />
    <property name="y" value="532" />
    <property name="width" value="184" />
    <property name="height" value="184" />
    <property name="sourcename" value="Dig52.JPG" />
   </properties>
  </object>
</action>
```

Table 3: Extract from a log file

Information captured in the CoLoForm format is well-defined and structured. Thus, a wide range of various kinds of analysis can be performed on that data. Besides simple counting and relating of children's activities in participation ratios, more sophisticated analysis tools are available that have been developed mainly in the context of the Kaleidoscope NoE but can also be used to analyse our log files.

PAnDit (Harrer, Lingnau, & Bientzle, 2009) is a tool for log file analysis that has originally been implemented to detect interaction patterns in collaboration log files using Prolog rules. But since it provides a mechanism that allows the specification of sequences of actions, including constraints and parameters, such as "the same user as in action1 performs action 3 after user2 performed action2". These patterns can be stored and re-used across different studies, populations, and scenarios. E.g., queries from different users can be analysed and compared, or behavioural patterns can be defined to analyse differences how children interacted with results from similar queries. Figure 1 shows a screenshot of the PAnDit tool. PAnDit reads the logging data generated by the PuppyIR logging framework and internally converts it into a set of Prolog facts. It contains a graphical Prolog rule generator as well as a manual Prolog rule editor to support the analysis. It is possible to combine two or more sets of Prolog rules, e.g. to first filter out log entries of special interest before running another set of Prolog rules. Example sets of Prolog rules to be used in connection with our interface prototypes will be implemented and made available as part of the PuppyIR open source.

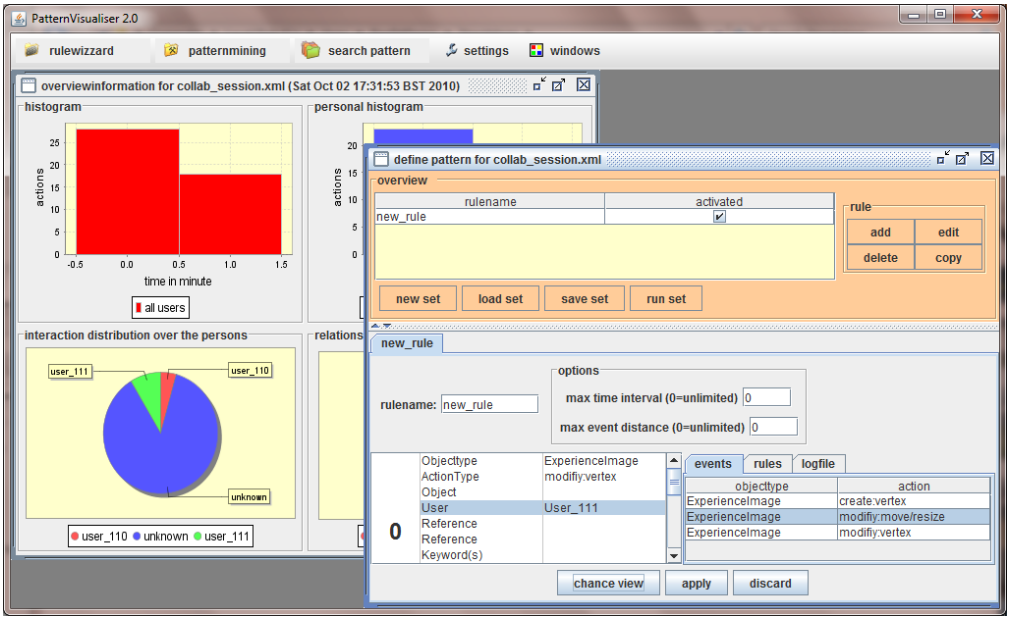

Figure 1: Screenshot of the PAnDit analysis tool

4 Conclusion

In this report, we have presented two sets of tools for analysing children's interaction with PuppyIR services. The Behavior Characterization Framework provides a core set of metrics for analysing querying and results interaction behaviour in order to learn more about how services are used. This framework can also be used to learn more about children's behaviour in general. Previous studies of children's interactive behaviour are based on small scale qualitative studies, or on systems that have not been designed for children, whereas log studies can provide richer quantitative understandings. The development of systems for use by children, and the logging tools described here, can allow for more accurate descriptions of children's behaviour when searching that can be used to understand how children at different ages conduct searches and inform the design of future tools.

The enhanced logging engine for use by novel interfaces, in particular the tangible interaction devices, provides an extensible logging approach to understand how novel interface paradigms are used by children within interactive searches. This will allow information providers to understand situations where children are collaboratively searching for information and it will allow a more detailed analysis of these activities. In combination with a synchronised timeline of video recording, interactions that normally would not be allocated to a special user could be tracked down to gain a better understanding of group behaviour. Furthermore, in situations where multiple users are working parallel with the same interface, the framework helps information providers to understand sequences and interactions on the search interfaces and gain a better understanding of interaction and behaviour patterns

References

Baeza-Yates, R. A. (2005). Applications of web query mining. *Advances in Information Retrieval, 27th European Conference on IR Research, ECIR 2005*, pp. 7–22.

Baeza-Yates, R. A. (2004). Web Mining: Applications and Techniques. ch. *Query Usage Mining in Search Engines*, pp. 307–321, Idea Group.

Bilal, D. & Kirby, J. (2002). Differences and similarities in information seeking: children and adults as web users. *Information Processing & Management*. 38.5. pp 649–670.

Harrer, A., Martinez-Mones, A., & Dimitracopoulou, A. (2009). Users' data: Collaborative and social analysis. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (Eds.), *Technology-Enhanced Learning -- Principles and Products* (pp. 175-193). Berlin: Springer.

Harrer, A., Lingnau, A., & Bientzle, M. (2009). Interaction Analysis with dedicated logfile analysis tools: a comparative case using the PAnDit tool versus manual inspection. *The Ninth IEEE International Conference on Advanced Learning Technologies* (pp. 405-407). Los Alamitos: IEEE Computer Society.

Hassan, A., Jones, R., & Klinkner, K. L. (2010). Beyond DCG: user behavior as a predictor of a successful search. In Proceedings of the third ACM international conference on Web search and data mining, WSDM pages 221–230.

Jansen, B. J., Taksa, I. and Spink, A. (2009). Research and Methodological Foundations of Transaction Log Analysis. In *Handbook of Research on Web Log Analysis*, ed. Bernard J. Jansen, Amanda Spink and Isak Taksa, pp1-16. doi:10.4018/978-1-59904-974-8.ch001

Lingnau, A., & Harrer, A. (2011). An analytic process schema for collaborative multi touch applications. *Proceedings of the 11th IEEE International Conference on Advanced Learning Technologies* (pp. 115-117). Los Alamitos, California: IEEE Computer Society. doi:10.1109/ICALT.2011.39

Duarte Torres, S., Hiemstra, D., & Serdyukov, P. (2010a). Query log analysis in the context of information retrieval for children. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (SIGIR '10). ACM, New York, NY, USA, 847-848. DOI=10.1145/1835449.1835646 http://doi.acm.org/10.1145/1835449.1835646

Duarte Torres, S., Hiemstra, D., & Serdyukov, P. (2010b). An analysis of queries intended to search information for children. In *Proceeding of the third symposium on Information interaction in context* (IIiX '10). ACM, New York, NY, USA, 235-244. DOI=10.1145/1840784.1840819 http://doi.acm.org/10.1145/1840784.1840819

Duarte Torres, S., Weber, I. (2011). What and How Children Search on the Web. *Proceedings of CIKM 2011*.

Tyler, S. K., & Teevan, J. (2010). Large scale query log analysis of re-finding. *Proceedings of the third ACM international conference on Web search and data mining*, WSDM '10, pp. 191–200.

Sandore, B. (1993). Applying the results of transaction log analysis. *Library Hi Tech*. 11. 2. Pp. 87-97.

Silvestri, F. (2010). Mining Query Logs: Turning Search Usage Data into Knowledge. *Foundations and Trends in Information Retrieval* 4(1-2): 1-174

Smith, P., Cowie, H., and Blades, M. (2003). Understanding children's developments. BlackWell.

Weber, I. & Jaimes, A. (2011). Who uses web search for what: and how. *Proceedings of the 4th ACM international conference on Web search and data mining*, WSDM '11, pp 15–24.