

Same Query – Different Results? A Study of Repeat Queries in Search Sessions

Johannes Leveling and Gareth J. F. Jones

Centre for Next Generation Localisation (CNGL), School of Computing
Dublin City University, Dublin 9, Ireland
{jleveling, gjones}@computing.dcu.ie

Abstract. Typically, three main query reformulation types in sessions are considered: generalization, specification, and drift. We show that given the full context of user interactions, repeat queries represent an important reformulation type which should also be addressed in session retrieval evaluation. We investigate different query reformulation patterns in logs from The European Library. Using an automatic classification for query reformulations, we found that the most frequent (and presumably the most important) reformulation pattern corresponds to repeat queries. We aim to find possible explanations for repeat queries in sessions and try to uncover implications for session retrieval evaluation.

1 Introduction

There are two different approaches to information retrieval (IR) evaluation. One is the traditional Cranfield paradigm, where IR systems are evaluated independent of any search context, i.e. the same query will yield the same, reproducible results, regardless of individual users or search context. The other approach originates from studies in human computer interaction, where evaluation is based on user interactions with a system, which poses the challenge to design these studies so that reproducible results can be obtained. There is growing interest in bridging the gap between these approaches by investigating query reformulations and evaluating IR in context (see, for example, the TREC session track [1]).

In this paper we show that repeat queries should be considered as an important query reformulation pattern in session retrieval evaluation. Typically, the main categories of query reformulations include generalization, specification, and drift [2, 1, 3]. To investigate the importance of different reformulation patterns, we perform an analysis of reformulation types in sessions. We base our analysis on user interaction logs from The European Library (TEL). He et al. [3] and Jansen et al. [4] both propose similar algorithms for automatic classification of query reformulations. We classify query reformulations following the latter algorithm, which involves examining common, added, deleted, and modified terms for two successive query formulations. To our surprise, the automatic classification of reformulations revealed that repeat queries are the most frequent reformulation pattern in sessions. Similar to the Excite logs [2], navigating multi-page result

lists appears to generate repeat queries in the TEL log. However, even if these are discarded – as in our additional analysis – repeat queries still make up the most frequent reformulation pattern. This poses a major problem for session retrieval evaluation because, to the best of our knowledge, evaluation experiments and metrics do not explicitly address repeat queries. For example, the issue of repeat queries raises such questions as: should the same result list be returned to the user for two identical successive queries in the same session? We aim to find possible explanations for repeat queries in sessions and implications for session retrieval evaluation.

2 Related Work

Most publications on repeat queries are concerned with reformulations in web search and do not concern repeat queries within sessions. For example, Teevan et al. [5] found in an analysis of web log data that 40% of all queries are re-find queries. He et al. [3] observe that the second most frequent type of reformulation in sessions is repeat queries, but they do not provide an explanation. (Note that we do not consider the most frequent type in their study, browsing, as query reformulation). Sanderson and Dumais [6] report that 80% of repeat queries in web search are navigational queries. They find that repetition of queries often occurs on the same day. The smallest time interval examined in their investigation (except for periodicity studies) is a single day.

Repeat queries may have different causes: repeat queries in different sessions typically correspond to a known-item search and occur often in web search [5]. Repeat queries in sessions are unlikely to represent known-item search, due to the different time frame (i.e. less than 30 minutes between actions), and are issued by the same user. Some log entries with repeated queries are generated automatically by a system when the user navigates the multi-page search results (e.g. Excite, see [2]) and does not actually enter a query, but clicks on a link.

Huang and Efthimiadis give an excellent overview of different query reformulation patterns in the literature [7]. In the session track at TREC 2010, the focus was on three major patterns of reformulation [1]: 1. generalization (where the original results were too narrow and the reformulation into a more general query can be achieved by deleting a term), 2. specification (where the original results are too broad and the reformulation into a more specific query can be achieved by adding a term), and 3. drifting (where the reformulation is another query aiming at a different aspect of the information need). So far, repeat queries have not been considered in the session track.

However, most of the related work does not focus on repeat queries in the same session, but on repeated queries in different sessions (e.g. in web search) and by different users [5, 6].¹ None provides possible explanations as to why these may occur within a session.

¹ Note that Spink et al. defined repeat queries as all multiple occurrences of the same query that represent requests for multi-page viewing [2].

Table 1. TEL 2009 action log statistics.

# actions (total):	1866330
# queries (total):	86981
# sessions (total):	20325
avg. # actions per session:	15.68
avg. # queries per session:	4.28

3 Session Analysis

TEL Data. The analysis described in this paper is based on the TEL 2009 action logs (queries and user interactions), which were employed at LogCLEF 2009², a task for analyzing query logs. Table 1 shows statistics on the TEL log data.

Session reconstruction. The TEL logs were preprocessed to reconstruct user sessions by grouping all actions with the same (unique) session ID together and sorting them by timestamp. We define a session as a consecutive sequence of all user interactions, presuming the start of a new session if the time between consecutive actions exceeds 30 minutes. The 30 minute time interval originates from the working definition of session reconstruction in previous editions of LogCLEF.

Thus, sessions comprise actions such as changing the interface languages, selecting the target collection, submitting a query, and viewing results. Sessions using the TEL advanced search³, sessions containing non-English queries, and sessions with less than two queries were filtered out because we did not want to over-emphasize the importance of context changes (since switches between the advanced and simple search are frequent) or language-specific aspects.

Analysis of reformulations In the context of session retrieval, we only consider directly consecutive queries as potential repeat queries and disregard all other similar queries within a session as possible repeat queries. We applied the automatic classification approach described in Jansen et al. [4] to identify the most important query reformulation patterns in sessions. We made only minor modifications to the algorithm (see [4] for details), i.e. using the edit distance for single word queries to determine if they are related and applying the classification on normalized queries, after case folding, stopword removal, and stemming. These changes were made to include spelling corrections, case folding, and morphological normalizations as query reformulations. Synonyms and abbreviations have not been explicitly considered as reformulations due to the open domain of the TEL target collections. Table 2 shows results for the reformulation classification. The most surprising result is that repeat queries correspond to the most frequent query reformulation pattern in sessions (i.e. not changing the query). So far, repeat queries in session evaluations have generally been ignored, but they raise the question as to why users should repeat queries within sessions.

² <http://www.uni-hildesheim.de/logclef/>

³ TEL supports simple (keyword) search and advanced (structured) search, which aims at different metadata fields (e.g. author, title) and employs Boolean operators.

Table 2. Frequency of query reformulation patterns in sessions.

Reformulation pattern	Frequency
# specifications:	1410
# specifications+reformulation:	584
# generalizations:	1196
# generalizations+reformulation:	590
# reformulations:	2072
# content changes:	4608
# repeat queries (with diff. context):	8942
# empty queries:	386
Σ	19788
# repeat queries (including navigation):	50963

Explaining repeat queries. One reason why repeat queries have been ignored in evaluation is the assumption that the same query should return the same results. However, if the full search context is considered, there are actually repeat queries for which different results are expected by the user, for example when a different target collection is selected or language-specific settings are changed.

There are several explanations why users repeat queries. Often, a part of the search context, which includes the viewed documents and the selected target language or collection, is changed. In web search, users typically do not change search options or choose different target collections. In general, repeat queries in web search mostly occur in different sessions, where submitting identical queries represents a known-item search [5], i.e. re-finding a result with a previous query.

One explanation for repeat queries in the *same* session is that the user’s state of mind changes after viewing result documents and he assumes that the system keeps track of his search findings and will return different results for the same query. At least some users might expect that results change for the same query, because they are unfamiliar with a search system and their mental model of the information need changes after viewing some results. For example, log analysis on the TEL logs has shown that users often modify queries regardless of query processing steps, for example by adding or deleting stopwords, which are removed by the TEL system [8]. If this mismatch between actual and expected system behavior exists – even for a small group of users – session retrieval evaluation metrics should take the search context into account (e.g. by employing user models). For example, users would expect to find the same (or very similar) results when they submit the same query on the next day to re-find previous results. When they submit the same query in the same session, they may be more interested in result diversity and expect different results. Current session evaluation does not reflect this and ignores search context.

One possible explanation for repeat queries is that users submit a query again simply by mistake. However, the number of repeat queries is much too high to account for all of them by mistakes. An explanation found in web search research is that users use previous queries as an “anchor point” to jump back to the first result page, before exploring other aspects of the topic. This explanation implies

a subsequent reformulation of the query so that each repeat should be followed by at least one reformulation on another aspect of the topic. This reformulation “meta-pattern” was not observed in the TEL logs. However, many users might navigate results and browse the same page or intermediate results several times in a session.

4 Implications for Session Retrieval Evaluation

We think that to build reusable test collections for session IR, more context of queries in sessions has to be captured. The search context includes language settings, the selected target collection, viewed documents, and user interactions other than queries (e.g. changing search options). The TEL log data already captures this information and provides additional metadata for evaluation (e.g. timestamps).

We also think that realistic data is a necessary prerequisite for research on session retrieval. So far, most (to our best knowledge: all) session retrieval models ignore repeat queries in sessions, although they are the most frequent reformulation pattern in sessions.

As part of the future work, we intend to carry out user studies to create real logs and experiments on different log data, to see if the observed results are specific to the log data used in this paper or can be observed in web search or in other domains such as enterprise search.

Acknowledgements

This research is supported by the Science Foundation Ireland (grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (<http://www.cngl.ie>).

References

1. Kanoulas, E., Clough, P., Carterette, B., Sanderson, M.: Session track at TREC 2010. In: SIGIR 2010 Workshop on Simulation of Interaction. (2010)
2. Spink, A., Wolfram, D., Jansen, B.J., Saracevic, T.: Searching the web: The public and their queries. *JASIST* **52**(3) (2001) 226–234
3. He, D., Göker, A., Harper, D.J.: Combining evidence for automatic web session identification. *Information Processing & Management* **38**(5) (2002) 727–742
4. Jansen, B.J., Booth, D., Spink, A.: Patterns of query reformulation during web searching. *ASIS&T* **60**(7) (2009) 1358–1371
5. Teevan, J., Adar, E., Jones, R., Potts, M.A.S.: Information re-retrieval: Repeat queries in Yahoo’s logs. In: SIGIR 2007. (2007) 151–158
6. Sanderson, M., Dumais, S.T.: Examining repetition in user search behavior. In: ECIR 2007. (2007) 597–604
7. Huang, J., Efthimiadis, E.N.: Analyzing and evaluating query reformulation strategies in web search logs. In: CIKM 2009. (2009) 77–86
8. Ghorab, M.R., Leveling, J., Zhou, D., Jones, G.J.F., Wade, V.: Identifying common user behaviour in multilingual search logs. In: CLEF 2009. (2010) 518–525