

# **An Analysis of Seasonal and Topical Queries in a Large Public Library to Support User Engagement**

Tami Oliphant and Ali Shiri (University of Alberta)

## **ABSTRACT:**

This paper reports on an investigation of user queries submitted to the Edmonton Public Library discovery system to identify seasonal and temporal queries. Using Google Analytics data and text analysis tools, this study examines how user queries change over time, particularly during summer months, weekdays and weekends.

## **1. Introduction**

The rise of increasingly sophisticated discovery systems in libraries that integrate the library catalogue with digital repositories and collections and social media has created vast volumes of interaction and transaction data that can assist researchers and practitioners in understanding user search behavior and search queries. Working in collaboration with the Edmonton Public Library (EPL), we examine seasonal and temporal queries in public libraries using Google Analytics and other open source data analysis software. We recognize the dynamics of time in this research by examining how search queries change over time (weekly, monthly, and seasonally) and within the context of broader events. While seasonal queries have been studied in the context of web search engines and information retrieval test collections, to our knowledge this is the first study investigating seasonal queries in public libraries. Specifically, we report on the characteristics of seasonal/temporal queries submitted to the discovery system of EPL within the months of June, July, and August 2014 and provide recommendations that can potentially enhance user engagement and the user experience with public libraries.

## **2. Background Context**

Studying information search behaviour of users interacting with digital information systems has a long history, dating back to the development of the first commercial bibliographic databases and online public access catalogues (Millsap & Ferl, 1993; Peters, 1993; Mat-Hassan & Levene, 2005). A recent meta-analysis of users' search behaviours studies sponsored by JISC and OCLC concluded that "a large, random sample of specific demographic groups of information seekers should be identified in order to identify how individuals engage in both the virtual and physical worlds to get information for different situations and contexts" (Connaway & Dickey, 2010). The proliferation of, and the heightened attention to, data (e.g., research data, digital data, linked data, open data, web of data, and data repositories), offers a renewed perspective of the ways in which we can understand, make sense of, and use data to improve and enhance library services and the user experience (Teets & Goldner, 2013). At the same time, the rise of more sophisticated discovery systems in libraries that integrate catalogues with other digital repositories and collections and social media applications has created vast volumes of interaction and transaction data, creating new opportunities for research and

development and new ways to understand how users search for information in this complex and multifaceted digital landscape.

Temporally dependent queries are defined by Metzler et al. (2009) as queries that are event-specific and that change over time such as “new years” and “presidential elections.” Metzler et al. go on to introduce two types of temporal queries: the ‘explicitly year qualified’ query where a year is specified in the query; and the ‘implicitly year qualified’ queries. An explicitly year qualified query is a query that contains a year such as “Scarface 1932” or “Scarface 1983” depending on which version of Scarface the user wanted. The ‘implicitly year qualified’ query is a query that does not actually contain a year, but yet the user may have implicitly formulated the query with a specific year in mind. Examples of implicitly year qualified queries include ‘olympics’ and ‘Easter’. Based on their experiment with 670 web queries, Metzler et al. concluded that more than 7% of queries belong to the implicitly year qualified query category.

In a discussion of social search and how users’ interest can support personalization of a social search system, Khodaei and Alonso (2012) provide a temporally-aware categorization of users’ social interests as follows: (a) recent interests, (b) ongoing interests, (c) seasonal interests, (d) past interests and (e) random interests. These dimensions of interests and needs can provide a useful basis for a temporal analysis of queries. Shokouhi (2011) notes that it is important for search engines to correctly identify seasonal queries and ensure that user search results are temporally reordered if necessary. He uses “Halloween” and “Christmas” as examples of seasonal queries that are derived from temporal information needs. He proposes an automatic classification of seasonal queries based on the historical frequency of queries and their conversion into time series. The distribution of queries across various periods and their seasonal peaks provide a good measure for detecting seasonal queries. Similarly, Joho et al. (2013) found that nearly 45% of users’ queries were related to either long-term interests or seasonal interests.

While numerous studies have been conducted on seasonal and temporal queries in the context of web search engines and test collections, to our knowledge this study is the first to explore seasonal queries in a public library search system. Consequently, we answer the following research questions:

1. What are the characteristics of seasonal/temporal queries in a large public library system?
  - a) Are there specific patterns or variations of users’ search terms during the months of June, July and August?
  - b) What were the popular queries at different time intervals such as weekdays and weekends?
  - c) What were the explicitly year qualified queries (e.g. new releases 2014)?
  - d) What were the implicitly year qualified queries (e.g. summer reading clubs)?
  - e) What were the popular queries in and across the three months of June, July and August 2014? (e.g. popular books or movies different months or at different times)

- f) What were the popular event-related topics across the three months? (e.g. tennis, Ukraine, ISIS, Lou Gehrig's disease (the ice bucket challenge), etc.)

## 2. How can libraries use this data to enhance the user experience?

### 3. Methodology

EPL makes use of Bibliocommons as its frontend interactive catalogue that integrates various digital resources such as e-books, catalogues, accounts, and recommendations. The EPL website and discovery system provides a single search bar with three options provided as "catalogue (default)", "website" and "articles." In order to gather user interaction and search-related data, EPL uses Google Analytics. Google Analytics gathers various statistics about EPL users, their geographic location, their search terms, and what devices and browsers were used. It also provides search term data on a daily, weekly and monthly basis. The three months of June, July, and August 2014 provide roughly 90,000 data points as follows: June (29,500 queries); July (30,000 queries); August (30,000 queries). While summer does not technically begin until June 21st, we are including June as part of the summer "season." These three months were purposively sampled for data collection for a variety of reasons: they constitute a temporal "season;" the K-12 school year ends in June and begins again after Labour Day in the beginning of September, and we wanted to explore how this transition from, and back, to school translates to search queries posed to EPL's website and catalogue; and during the summer months (including June when school is let out), public libraries offer many programs and events for children, teens, and families, and the general public partakes of festivals, movies, cultural events, travel, gardening, and other interests which would be reflected in search queries.

We have used Google Analytics to collect our data and to carry out preliminary analysis (e.g., number of searches per month, limiting to the City of Edmonton, new users vs. returning users, etc.) but further analysis requires us to use a variety of other analytical tools given the nature of public library discovery search systems that allow searching within the catalogue, website, and articles fields. Furthermore, in order to develop an analytical framework for the exploration and categorization of seasonal queries multiple approaches to data analysis are required. Thus, to address our research questions, we will make use of a combination of qualitative, quantitative and textual analysis methods. Quantitative analysis methods will focus on search term frequency and popularity as well as time and date related analyses. The topical seasonality analysis and classification of queries will be carried out using qualitative analysis methods. Due to the complex nature of this study and the different variables involved, we make use of a range of open source data transformation, text analysis and visualization applications such as OpenRefine, TAPOR, Voyant, and Many Eyes.

### 4. Significance and outcomes

The results of this study are potentially significant for the library and information science community for three primary reasons. Because this is a research project that we have undertaken in collaboration with the Edmonton Public Library, we aim to engage in

research that bridges the research/practice gap often found in LIS. Second, we are using a novel methodology in order to analyze a fairly large dataset. This initial project will provide a proof-of-concept that we can potentially use for modelling, analyzing, and visualizing EPL's "big data." Finally, and most importantly, we seek to understand seasonal information seeking behaviors to help formulate recommendations to improve search and discovery systems, collections, and services used in public libraries. This paper relates to the conference theme in two ways: time is the central dimension in our longitudinal and seasonal approach in studying user queries and the queries themselves are time bound and related to broader events happening in real-time.

## References

- Connaway, L. S., & Dickey, T. J. (2010). The digital information seeker: Findings from selected OCLC, RIN and JISC user behaviour projects. Available at: <http://www.jisc.ac.uk/media/documents/publications/reports/2010/digitalinformationseekerreport.pdf> (Accessed Sep 9, 2014).
- Khodaei, A., & Alonso, O. (2012). Temporally-aware signals for social search. In *TAIA 2012 Workshop*.
- Kulkarni, A., Teevan, J., Svore, K. M., & Dumais, S. T. (2011, February). Understanding temporal query dynamics. In *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 167-176). ACM.
- Mat-Hassan, Mazlita, & Levene, Mark. 2005. Associating search and navigation behavior through log analysis. *Journal of the American Society for Information Science and Technology*, 56(9), 913-934.
- Metzler, D., Jones, R., Peng, F., & Zhang, R. (2009, July). Improving search relevance for implicitly temporal queries. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval* (pp. 700-701). ACM.
- Millsap, Larry, & Ferl, Terry. 1993. Search patterns of remote users: An analysis of OPAC transaction logs. *Information Technology and Libraries*, 11, 321-343.
- Peters, Thomas. 1993. The history and development of transaction log analysis. *Library Hi Tech*, 42(11), 41-66.
- Shokouhi, M., & Radinsky, K. (2012, August). Time-sensitive query auto-completion. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval* (pp. 601-610). ACM.
- Teets, M., & Goldner, M. (2013). Libraries' role in curating and exposing big data. *Future Internet*, 5(3), 429-438.