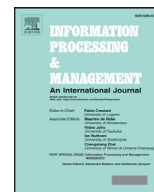




Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

Temporal information searching behaviour and strategies

Hideo Joho^{a,*}, Adam Jatowt^b, Roi Blanco^c^a Research Center for Knowledge Communities, Faculty of Library, Information and Media Science, University of Tsukuba, Japan^b Department of Social Informatics, Graduate School of Informatics, Kyoto University, Japan^c Yahoo Labs, Barcelona, Spain

ARTICLE INFO

Article history:

Received 23 September 2014

Revised 3 February 2015

Accepted 27 March 2015

Available online 22 April 2015

Keywords:

Temporal information retrieval

Information searching behaviour

Search strategies

User study

ABSTRACT

Temporal aspects have been receiving a great deal of interest in Information Retrieval and related fields. Although previous studies have proposed, designed and implemented temporal-aware systems and solutions, understanding of people's temporal information searching behaviour is still limited. This paper reports the findings of a user study that explored temporal information searching behaviour and strategies in a laboratory setting. Information needs were grouped into three temporal classes (Past, Recency, and Future) to systematically study their characteristics. The main findings of our experiment are as follows. (1) It is intuitive for people to augment topical keywords with temporal expressions such as *history*, *recent*, or *future* as a tactic of temporal search. (2) However, such queries produce mixed results and the success of query reformulations appears to depend on topics to a large extent. (3) Search engine interfaces should detect temporal information needs to trigger the display of temporal search options. (4) Finding a relevant Wikipedia page or similar summary page is a popular starting point of past information needs. (5) Current search engines do a good job for information needs related to recent events, but more work is needed for past and future tasks. (6) Participants found it most difficult to find future information. Searching for domain experts was a key tactic in Future search, and file types of relevant documents are different from other temporal classes. Overall, the comparison of search across temporal classes indicated that Future search was the most difficult and the least successful followed by the search for the Past and then for Recency information. This paper discusses the implications of these findings on the design of future temporal IR systems.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Predicting the future and remembering the past are very common cognitive processes of humans. Same as in real life, we should expect many search activities to be of a strong temporal character. Previous studies (Joho, Jatowt, & Roi, 2013; Nunes, Ribeiro, & David, 2008) have confirmed this fact and elucidated a relatively high number of search intents and queries that center on information associated to particular temporal scopes such as future or past. Regarding future-oriented search, users often need to know more about planned events, forecasted trends, possible scenarios, speculations, predictions and so on. This kind of information can effectively help them to be better prepared for events to come or tasks to be undertaken. Imagine a user who wants to purchase shares or futures of a particular company, and another user who plans a visit in Kyoto.

* Corresponding author.

E-mail addresses: hideo@slis.tsukuba.ac.jp (H. Joho), adam@dl.kuis.kyoto-u.ac.jp (A. Jatowt), roi@yahoo-inc.com (R. Blanco).

Any predictions or speculations about the company's future, its plans or forecasts and any information about forthcoming cultural events in Kyoto would be valuable in these scenarios. Similarly, the study of the past helps us to explain the present, to learn the background of the current course of actions or to form opinions on any trends and changes occurring over time. In general, we should be able to easily imagine a multitude of reasons on why someone would need to find future or past-related information.

However the core interest of society centers on the present or, at least, on the near past. The Web, for example, abounds in rather up-to-date information that is mainly about "now" (Fetterly, Manasse, Najork, & Wiener, 2003; Ntoulas, Cho, & Olston, 2004). The information about more distant past or future tends to be buried in the wealth of data on the current topics and events. Intuitively, it may be difficult for an average user to extract content that is related to the future and to the past, especially, to distant future and past, for arbitrary queries. Such information is often scattered across many documents and expressed in a large number of different ways and is thus likely difficult to be extracted, merged, processed and understood by searchers. We would then expect the state-of-the-art search engines to offer some kind of support for those users who try to seek for time-related information. The academic community has already started developing methods for enabling the temporal search (Berberich, Bedathur, Alonso, & Weikum, 2010; Campos, Jorge, Dias, & Nunes, 2012; Jones & Diaz, 2007; Jatowt, Au Yeung, & Tanaka, 2013; Kanhabua, Blanco, & Matthews, 2011; Kanhabua & Nørvåg, 2010; Li & Croft, 2003; Metzler, Jones, Peng, & Zhang, 2009). What is still missing is the knowledge of how searchers actually filter out present-related information when they have search intents related strongly to either the past of the future. Empowered with such knowledge it will be easier to reason about the required level of support searchers should receive or about any temporal mechanisms that ought to be implemented for enabling effective temporal search. In this work we aim to fill in this gap and shed light on the actual behaviour of searchers who wish to find past or future-related information. In particular, we conduct controlled settings experiments with 30 participants who are asked to perform searches on variety of topics on the Web to find information related to particular time scopes. We then analyse their behaviour as well as feedback regarding the tasks and their difficulties. We report a large number of observations that have not been known to the community and which could have considerable implications on the design of temporal search mechanisms and search interfaces for facilitating retrieval of time-related information.

The remainder of the paper is organised as follows. Section 2 provides the literature survey in temporal information retrieval. Section describes the design of user study we performed to capture information seeking behaviour of temporal search. Section presents the results of the user study. Section 5 discusses the implications of these findings on the design of future temporal search engines, followed by a conclusion and the outline of future directions.

2. Related work

Temporal Information Retrieval (e.g., Alonso, Baeza-Yates, Strötgen, & Gertz, 2011; Berberich et al., 2010; Campos, Dias, & Jorge, 2011; Campos, Dias, Jorge, & Jatowt, 2014; Campos et al., 2012; Joho, Jatowt, & Blanco, 2014a; Joho et al., 2013; Jones & Diaz, 2007; Kanhabua & Nørvåg, 2010; Mazur, 2012; Metzler et al., 2009; Nunes et al., 2008) has increasingly been gaining much interest in the IR community. This subarea of information retrieval focuses on temporal aspects of search, treating time as crucial facet for determining document relevance. Prior work mainly focused on either estimating temporal features of documents (Jatowt et al., 2013; Strötgen, Alonso, and Gertz, 2012; Strötgen, Gertz, and Popov, 2010, temporal aspects of queries (Campos et al., 2011; Jones & Diaz, 2007; Metzler et al., 2009; Nunes et al., 2008) or on matching temporalities of queries with ones of documents for realising effective time-aware retrieval (Anand, Bedathur, Berberich, & Schenkel, 2011; Kanhabua & Nørvåg, 2012; Kanhabua et al., 2011; Li & Croft, 2003).

According to a study performed on the AOL query dataset (Nunes et al., 2008) about 1.5% of queries are explicit temporal queries, that is, they contain an explicit temporal expression. Examples of such queries are: "Poland 1940s", "Olympics 2016" or "most popular songs 2000s". A subsequent study (Campos et al., 2011) revised this number to about 1.21% queries after excluding some false positive temporal expressions (e.g., "Excel 2007", "Honda civic 2004"). Searchers also issue implicit temporal queries that are related to time despite lacking apparent temporal expressions (e.g., "Einstein childhood", "WWII major battles", "USA debt size", "Rio de Janeiro Olympics") (Jones & Diaz, 2007).

Metzler et al. (2009) concluded that about 7% of queries have a certain temporal intent. Considering the popularity of Web search, this rate translates to a remarkable number of unique searches. Some methods have been already proposed to automatically classify queries into different temporal classes with the underlying aim of improving search results once query temporal intent is known (see, for example, Campos et al., 2012; Jones & Diaz, 2007; Kanhabua & Nørvåg, 2010). For a more extensive overview of existing approaches the reader may refer to Alonso et al. (2011) and Campos et al. (2014).

Given the relatively large amount of temporal queries, search engines should handle them in a way which appropriately considers particular temporal classes to prevent potential mismatch like returning past-related information for a query with obvious future-focused intent. In addition, search engines could offer support for finding information from particular temporal class (e.g., by query recommendation, by elucidating temporal aspects of snippets). For this to be effective, however, a prior comprehensive analysis of the way in which users search for time-sensitive information is needed.

Although previous works focused on ranking documents for temporal queries and on estimating temporal intents underlying user queries, little has been done to uncover the actual search patterns and behaviour of users who seek information of temporal character. The only previous study in this regard that we are aware of was conducted as an online questionnaire

Download English Version:

<https://daneshyari.com/en/article/514960>

Download Persian Version:

<https://daneshyari.com/article/514960>

[Daneshyari.com](https://daneshyari.com)