Temporal dynamics in information retrieval


Full text available as:
PDF Download (2MB) | Preview

Printed Thesis Information: https://eleanor.lib.gla.ac.uk/record=b3136287

Abstract

The passage of time is unrelenting. Time is an omnipresent feature of our existence, serving as a context to frame change driven by events and phenomena in our personal lives and social constructs. Accordingly, various elements of time are woven throughout information itself, and information behaviours such as creation, seeking and utilisation. Time plays a central role in many aspects of information retrieval (IR). It can not only distinguish the interpretation of information, but also profoundly influence the intentions and expectations of users’ information seeking activity. Many time-based patterns and trends - namely temporal dynamics - are evident in streams of information behaviour by individuals and crowds. A temporal dynamic refers to a periodic regularity, or, a one-off or irregular past, present or future of a particular element (e.g., word, topic or query popularity) - driven by predictable and unpredictable time-based events and phenomena.

Several challenges and opportunities related to temporal dynamics are apparent throughout IR. This thesis explores temporal dynamics from the perspective of query popularity and meaning, and word use and relationships over time. More specifically, the thesis posits that temporal dynamics provide tacit meaning and structure of information and information seeking. As such, temporal dynamics are a ‘two-way street’ since they must be supported, but also conversely, can be exploited to improve time-aware IR effectiveness.

Real-time temporal dynamics in information seeking must be supported for consistent user satisfaction over time. Uncertainty about what the user expects is a perennial problem for IR systems, further confounded by changes over time. To alleviate this issue, IR systems can: (i) assist the user to submit an effective query (e.g., error-free and descriptive), and (ii) better anticipate what the user is most likely to want in relevance ranking. I first explore methods to help users formulate queries through time-aware query auto-completion, which can suggest both recent and always popular queries. I propose and evaluate novel approaches for time-sensitive query auto-completion, and demonstrate state-of-the-art performance of up to 9.2% improvement above the hard baseline. Notably, I find results are reflected across diverse search scenarios in different languages, confirming the pervasive and language agnostic nature of temporal dynamics. Furthermore, I explore the impact of temporal dynamics on the motives behind users’ information seeking, and thus how relevance itself is subject to temporal dynamics. I find that temporal dynamics have a dramatic impact on what users expect over time for a considerable proportion of queries. In particular, I find the most likely meaning of ambiguous queries is affected over short and long-term periods (e.g., hours to months) by several periodic and one-off event temporal dynamics. Additionally, I find that for event-driven multi-faceted queries, relevance can often be inferred by modelling the temporal dynamics of changes in related information.

In addition to real-time temporal dynamics, previously observed temporal dynamics offer a complementary opportunity as a tacit dimension which can be exploited to improve IR systems. IR approaches are typically based on methods which characterise the nature of information through the statistical distributions of words or phrases. In this thesis I look to model and exploit the temporal dimension of the collection, characterised by temporal dynamics, in these established IR approaches. I explore how the temporal dynamic similarity of word and phrase use in a collection can be exploited to infer temporal semantic relationships between the terms. I propose an approach to uncover a query topic’s “chronotype” terms -- that is, its most distinctive and temporally interdependent terms, based on a mix of temporal and non-temporal evidence. I find exploiting chronotype terms in temporal query expansion leads to significantly improved retrieval performance in several time-based collections.

Temporal dynamics provide both a challenge and an opportunity for IR systems. Overall, the findings presented in this thesis demonstrate that temporal dynamics can be used to derive tacit structure and meaning of information and information behaviour, which is then valuable for improving IR. Hence, time-aware IR systems which take
temporal dynamics into account can better satisfy users consistently by anticipating changing user expectations, and maximising retrieval effectiveness over time.

**Item Type:** Thesis (PhD)  
**Qualification Level:** Doctoral  
**Additional Information:** Supported by funding from the EPSRC DTA scheme.  
**Keywords:** ir, information retrieval, search, time-aware, time, temporal dynamics  
**Subjects:**  
- QA Science > QA Mathematics > QA75 Electronic computers. Computer science  
- QA Science > QA Mathematics > QA76 Computer software  
**Colleges/Schools:**  
- College of Science and Engineering > School of Computing Science  
**Supervisor's Name:** Jose, Professor Joemon M.  
**Date of Award:** 2015  
**Depositing User:** Mr Stewart W. Whiting  
**Unique ID:** glathesis:2015-6850  
**Copyright:** Copyright of this thesis is held by the author.  
**Date Deposited:** 09 Nov 2015 10:57  
**Last Modified:** 26 Nov 2015 16:04  
**URI:** http://theses.gla.ac.uk/id/eprint/6850