Temporal Dynamics in Information Retrieval

Stewart Whiting
School of Computing Science,
University of Glasgow,
Scotland, UK.
stewh@dcs.gla.ac.uk

Abstract

Since time is an omnipresent feature of our existence, many elements of time are embedded in information itself, and related behaviours such as creation, seeking and utilisation. In IR, time can distinguish the interpretation of information, and 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 emerge in IR. This thesis explores temporal dynamics from the perspective of (i) query popularity and meaning, and (ii) word use and relationships over time. In particular, I consider how real-time temporal dynamics in information seeking should be supported for consistent user satisfaction over time, and moreover, how previously observed temporal dynamics offer a complementary dimension which can be exploited to inform more effective IR systems.

Uncertainty about user expectations is a perennial problem for IR systems, further confounded by changes over time. Addressing this, IR systems can either assist the user to submit an effective query (e.g., error-free and descriptive), or better anticipate what the user is most likely to want in relevance ranking. I first explore methods to always help users formulate queries with time-aware query auto-completion capable of suggesting both recent and always popular queries. I propose and evaluate several novel approaches, and demonstrate state-of-the-art performance of up to +9.2% improvement above existing baselines for diverse search scenarios in different languages. 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 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-driven temporal dynamics. Finally, I find that for many event-driven multi-faceted queries, relevance can often be inferred by modelling the temporal dynamics of changes in related information.

IR approaches are typically based on methods which characterize the nature of information through the statistical distributions of words and phrases. I model and exploit the temporal dimension of the collection, captured 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. Experiments demonstrate that 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, this thesis demonstrates that temporal dynamics can be used to derive tacit structure and meaning of information and information behaviour, which is valuable for improving time-aware IR system effectiveness.

Available from: http://theses.gla.ac.uk/6850/