

Time to Ship: Some Examples from the Real-World

Omar Alonso
Microsoft Corporation
omalonso@microsoft.com

ABSTRACT

There is new and exciting research work in analyzing and exploiting information from corpora with temporal information such as news, email, or social media. We review some of the current activities around extracting temporal information from social media and present a number of examples from real-world systems. We also outline a number of open problems and potential new research areas.

Keywords

Temporal information, social media

1. INTRODUCTION

Traditionally, most of the adoption of temporal information in information retrieval has been around the identification and extraction of time data from web pages or news articles for organizing content. With the current popularity of social media, there are more opportunities to investigate how time can be used to design new search experiences.

If we look at Twitter and Facebook as the predominant examples of social networks, timelines constitute the underlying structure of how social activity is presented. By looking at user behavior in general, it is possible to detect trending topics and stories at a local and global scale. Trends are now a common feature that is well-understood by users.

Search engines like Bing also use temporal information in different ways. The Bing home page, for example, shows a relevant image that sometimes has a calendar reference and a number of news articles that are popular by some a signal. At the same time, by using social data to construct diffusion trees, it is possible to discover links that are very active in social networks. Furthermore, we can characterize the size of the trees as a measure of popularity and the structure of the trees as a measure of virality. Both can be used as signals for better link recommendation.

Topics can also be very temporal. Besides, unexpected events like an earthquake or a terrorist attack, topics described by hashtags are useful for observing topical evolution over a period of time and for capturing seasonality or recurring events. Sports (e.g., games, playoff season, etc.) and entertainment (e.g., awards, movie previews, etc.) are the most common example of recurring events that are planned in advance.

Moving away from analyzing social data at the hour level, archiving can be useful when a user examines his/her social activity and looks at a past event or topic. We present a prototype that produces social digests, an archived document from social media, along with facets for building a new search and browsing experience. We describe at a high level, the back-end pipeline that generates social digests – a condensed version of the original dataset that removes a large percentage of the noise. We show that a combination of efficient filters, duplicate removal and clustering can provide substantial savings in storage and subsequent computational costs by drastically reducing the size of the data while preserving majority of the interesting content.

Finally, we explore how temporal information from location-based social networks (LBSNs) can be used as signals. LBSNs have become widely used for sharing and consuming location information. By performing location check-ins and writing tips, users annotate a point of interest with extra information that, in aggregate, is very valuable for understanding user mobility and preference. Still, a large number of users turn to general web search engines for queries that can be considered recreational. In these cases, users typically express a query combining a specific activity type around an explicit location and time – for example, “parks for kids in NYC in winter”, or “things to do in san francisco”. We review a relevance framework for ranking points of interest in web search and show concrete examples using Bing.