Text Queries on Document Time Series

Rosie Jones
Yahoo! Research
USA

Summary:

Many documents, such as emails and news stories, have timestamps, corresponding to the creation date, or date the information was sent. Using the timestamp, each document can be represented as a point on a timeline. We can then represent the entire document collection as a function of the number of documents at each point. A text query on a collection of documents on a timeline selects the subset of documents that are relevant to the query. The resulting set of documents can also be represented on a timeline. For example, a query for "earthquake" would lead to documents clustered in peaks on the time line, near each earthquake described in the news. We describe experiments on ways of using this timeline to identify temporal query ambiguity, as well as predict when queries are likely to lead to poor search results.