

**Continuous Top-k Monitoring on Document Streams**

**Abstract:**

The efficient processing of document streams plays an important role in many information filtering systems. Emerging applications, such as news update filtering and social network notifications, demand presenting end-users with the most relevant content to their preferences. In this work, user preferences are indicated by a set of keywords. A central server monitors the document stream and continuously reports to each user the top-k documents that are most relevant to her keywords. Our objective is to support large numbers of users and high stream rates, while refreshing the top-k results almost instantaneously. Our solution abandons the traditional frequency-ordered indexing approach. Instead, it follows an identifier-ordering paradigm that suits better the nature of the problem. When complemented with a novel, locally adaptive technique, our method offers (i) proven optimality w.r.t. the number of considered queries per stream event, and (ii) an order of magnitude shorter response time (i.e., time to refresh the query results) than the current state-of-the-art.