

**Shah, Rahul; Sheng, Cheng; Thankachan, Sharma V.; Vitter, Jeffrey Scott**

**Top- $k$  document retrieval in external memory.** (English) [Zbl 1394.68129](#)

Bodlaender, Hans L. (ed.) et al., Algorithms – ESA 2013. 21st annual European symposium, Sophia Antipolis, France, September 2–4, 2013. Proceedings. Berlin: Springer (ISBN 978-3-642-40449-8/pbk). Lecture Notes in Computer Science 8125, 803-814 (2013).

Summary: Let  $\mathcal{D}$  be a given set of (string) documents of total length  $n$ . The top- $k$  document retrieval problem is to index  $\mathcal{D}$  such that when a pattern  $P$  of length  $p$ , and a parameter  $k$  come as a query, the index returns those  $k$  documents which are most relevant to  $P$ . We present the first non-trivial external memory index supporting top- $k$  document retrieval queries in optimal  $O(\frac{p}{B} + \log_B n + \frac{k}{B})$  I/Os, where  $B$  is the block size. The index space is almost linear  $O(n \log^* n)$  words.

For the entire collection see [\[Zbl 1270.68017\]](#).

**MSC:**

[68P20](#) Information storage and retrieval of data

[68P05](#) Data structures

[68W40](#) Analysis of algorithms

Cited in **10** Documents

**Full Text:** [DOI](#)