

Munro, J. Ian; Navarro, Gonzalo; Shah, Rahul; Thankachan, Sharma V.

Ranked document selection. (English) [Zbl 1416.68064](#)

Ravi, R. (ed.) et al., Algorithm theory – SWAT 2014. 14th Scandinavian symposium and workshops, Copenhagen, Denmark, July 2–4, 2014. Proceedings. Berlin: Springer. Lect. Notes Comput. Sci. 8503, 344–356 (2014).

Summary: Let \mathcal{D} be a collection of string documents of n characters in total. The top- k document retrieval problem is to preprocess \mathcal{D} into a data structure that, given a query (P, k) , can return the k documents of \mathcal{D} most relevant to pattern P . The relevance of a document d for a pattern P is given by a predefined ranking function $w(P, d)$. Linear space and optimal query time solutions already exist for this problem.

In this paper we consider a novel problem, document selection queries, which aim to report the k th document most relevant to P (instead of reporting all top- k documents). We present a data structure using $O(n \log^\varepsilon n)$ space, for any constant $\varepsilon > 0$, answering selection queries in time $O(\log k / \log \log n)$, and a linear-space data structure answering queries in time $O(\log k)$, given the locus node of P in a (generalized) suffix tree of \mathcal{D} . We also prove that it is unlikely that a succinct-space solution for this problem exists with polylogarithmic query time.

For the entire collection see [\[Zbl 1293.68031\]](#).

MSC:

[68P20](#) Information storage and retrieval of data
[68P05](#) Data structures

Cited in **2** Documents

Full Text: [DOI](#)