

Biswas, Sudip; Patil, Manish; Shah, Rahul; Thankachan, Sharma V.

Succinct indexes for reporting discriminating and generic words. (English) Zbl 1330.68055
Theor. Comput. Sci. 593, 165-173 (2015).

Summary: We consider the problem of indexing a collection \mathcal{D} of D strings (documents) of total n characters from an alphabet set of size σ , such that whenever a pattern P (of p characters) and an integer $\tau \in [1, D]$ come as a query, we can efficiently report all (i) *maximal generic words* and (ii) *minimal discriminating words* as defined below:

- maximal generic word is a maximal extension of P occurring in at least τ documents.
- minimal discriminating word is a minimal extension of P occurring in at most τ documents.

These problems were introduced by *G. Kucherov* et al. [Lect. Notes Comput. Sci. 7608, 307–317 (2012; Zbl 1330.68059)], they proposed indexes occupying $O(n \log n)$ bits with query times $O(p + \text{output})$ and $O(p + \log \log n + \text{output})$ for Problem (i) and Problem (ii) respectively. The query time for Problem (ii) is later improved to optimal $O(p + \text{output})$ by *P. Gawrychowski* et al. [Lect. Notes Comput. Sci. 8214, 129–140 (2013; Zbl 1330.68058)]. In this paper, we describe succinct indexes of $n \log \sigma + o(n \log \sigma) + O(n)$ bits space with near-optimal query times i.e., $O(p + \log \log n + \text{output})$ for both these problems.

MSC:

- 68P15 Database theory
- 68P05 Data structures
- 68P20 Information storage and retrieval of data
- 68W32 Algorithms on strings

Keywords:

succinct indexes; string searching; range queries

Full Text: [DOI](#)

References:

- [1] Belazzougui, Djamel; Navarro, Gonzalo, Alphabet-independent compressed text indexing, ACM Trans. Algorithms, 10, 4, 23, (2014) · Zbl 1325.68307
- [2] Biswas, Sudip; Patil, Manish; Shah, Rahul; Thankachan, Sharma V., Succinct indexes for reporting discriminating and generic words, (String Processing and Information Retrieval, (2014), Springer), 89-100 · Zbl 1330.68054
- [3] Chan, Timothy M., Persistent predecessor search and orthogonal point location on the word ram, ACM Trans. Algorithms, 9, 3, 22, (2013) · Zbl 1301.68236
- [4] Fadiel, Ahmed; Lithwick, Stuart; Ganji, Gopi; Scherer, Stephen W., Remarkable sequence signatures in archaeal genomes, Archaea, 1, 3, 185-190, (2003)
- [5] Fischer, Johannes; Heun, Volker, Space-efficient preprocessing schemes for range minimum queries on static arrays, SIAM J. Comput., 40, 2, 465-492, (2011) · Zbl 1222.05024
- [6] Gawrychowski, Paweł; Kucherov, Gregory; Nekrich, Yakov; Starikovskaya, Tatiana, Minimal discriminating words problem revisited, (String Processing and Information Retrieval, (2013), Springer), 129-140 · Zbl 1330.68058
- [7] Hon, Wing-Kai; Shah, Rahul; Thankachan, Sharma V.; Vitter, Jeffrey Scott, Space-efficient frameworks for top-k string retrieval, J. ACM, 61, 2, 9, (2014) · Zbl 1295.68230
- [8] Kucherov, Gregory; Nekrich, Yakov; Starikovskaya, Tatiana, Computing discriminating and generic words, (String Processing and Information Retrieval, (2012), Springer), 307-317 · Zbl 1330.68059
- [9] McCreight, Edward M., A space-economical suffix tree construction algorithm, J. ACM, 23, 2, 262-272, (1976) · Zbl 0329.68042
- [10] Raman, Rajeev; Raman, Venkatesh; Srinivasa Rao, S., Succinct indexable dictionaries with applications to encoding k-ary trees and multisets, (Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms, (2002), Society for Industrial and Applied Mathematics), 233-242 · Zbl 1093.68582
- [11] Sadakane, Kunihiro, Succinct data structures for flexible text retrieval systems, J. Discrete Algorithms, 5, 1, 12-22, (2007) · Zbl 1137.68360
- [12] Sadakane, Kunihiro; Navarro, Gonzalo, Fully-functional succinct trees, (Proceedings of the Twenty-First Annual ACM-SIAM

Symposium on Discrete Algorithms, (2010), Society for Industrial and Applied Mathematics), 134-149 · [Zbl 1288.05046](#)

- [13] Weiner, Peter, Linear pattern matching algorithms, (IEEE Conference Record of 14th Annual Symposium on Switching and Automata Theory, 1973, SWAT'08, (1973), IEEE), 1-11

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.