

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

Succinct indexes for reporting discriminating and generic words. (English) [Zbl 1330.68054](#)

Moura, Edleno (ed.) et al., String processing and information retrieval. 21st international symposium, SPIRE 2014, Ouro Preto, Brazil, October 20–22, 2014. Proceedings. Berlin: Springer (ISBN 978-3-319-11917-5/pbk). Lecture Notes in Computer Science 8799, 89-100 (2014).

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 $\tau \in [1, D]$ comes 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](#))], and they proposed linear space 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. 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.

For the entire collection see [[Zbl 1298.68032](#)].

MSC:

- [68P15](#) Database theory
- [68P05](#) Data structures
- [68P20](#) Information storage and retrieval of data
- [68W32](#) Algorithms on strings

Cited in **2** Documents

Full Text: [DOI](#)