

## Sparse Text Indexing in Small Space - DTU Orbit (09/11/2017)

### Sparse Text Indexing in Small Space

In this work, we present efficient algorithms for constructing sparse suffix trees, sparse suffix arrays, and sparse position heaps for  $b$  arbitrary positions of a text  $T$  of length  $n$  while using only  $O(b)$  words of space during the construction. Attempts at breaking the naïve bound of  $\Omega(nb)$  time for constructing sparse suffix trees in  $O(b)$  space can be traced back to the origins of string indexing in 1968. First results were not obtained until 1996, but only for the case in which the  $b$  suffixes were evenly spaced in  $T$ . In this article, there is no constraint on the locations of the suffixes. Our main contribution is to show that the sparse suffix tree (and array) can be constructed in  $O(n \log 2b)$  time. To achieve this, we develop a technique that allows one to efficiently answer  $b$  longest common prefix queries on suffixes of  $T$ , using only  $O(b)$  space. We expect that this technique will prove useful in many other applications in which space usage is a concern. Our first solution is Monte Carlo, and outputs the correct tree with high probability. We then give a Las Vegas algorithm, which also uses  $O(b)$  space and runs in the same time bounds with high probability when  $b = O(\sqrt{n})$ . Additional trade-offs between space usage and construction time for the Monte Carlo algorithm are given. Finally, we show that, at the expense of slower pattern queries, it is possible to construct sparse position heaps in  $O(n + \log b)$  time and  $O(b)$  space.

### General information

State: Published

Organisations: Department of Applied Mathematics and Computer Science , Algorithms and Logic , Technische Universität Dortmund, Weizmann Institute of Science, University of Bristol

Authors: Bille, P. (Intern), Fischer, J. (Ekstern), Gørtz, I. L. (Intern), Kopelowitz, T. (Ekstern), Sach, B. (Ekstern), Vildhøj, H. W. (Intern)

Pages: 1-19

Publication date: 2016

Main Research Area: Technical/natural sciences

### Publication information

Journal: A C M Transactions on Algorithms

Volume: 12

Issue number: 3

Article number: 39

ISSN (Print): 1549-6325

Ratings:

BFI (2017): BFI-level 2

Web of Science (2017): Indexed Yes

BFI (2016): BFI-level 2

Scopus rating (2016): SJR 1.16 SNIP 2.158 CiteScore 2.1

Web of Science (2016): Indexed yes

BFI (2015): BFI-level 2

Scopus rating (2015): SJR 1.551 SNIP 1.897 CiteScore 1.94

Web of Science (2015): Indexed yes

BFI (2014): BFI-level 2

Scopus rating (2014): SJR 1.279 SNIP 1.614 CiteScore 1.61

Web of Science (2014): Indexed yes

BFI (2013): BFI-level 2

Scopus rating (2013): SJR 1.178 SNIP 1.505 CiteScore 1.4

ISI indexed (2013): ISI indexed yes

BFI (2012): BFI-level 2

Scopus rating (2012): SJR 1.137 SNIP 1.33 CiteScore 1.21

ISI indexed (2012): ISI indexed yes

BFI (2011): BFI-level 2

Scopus rating (2011): SJR 1.063 SNIP 1.259 CiteScore 1.11

ISI indexed (2011): ISI indexed no

BFI (2010): BFI-level 2

Scopus rating (2010): SJR 1.317 SNIP 1.946

BFI (2009): BFI-level 2

Scopus rating (2009): SJR 1.505 SNIP 1.914

Web of Science (2009): Indexed yes

BFI (2008): BFI-level 1

Scopus rating (2008): SJR 1.458 SNIP 1.923

Scopus rating (2007): SJR 0.395 SNIP 0.899

Original language: English

Electronic versions:

paper\_5.pdf

DOIs:

10.1145/2836166

**Bibliographical note**

Statement for Accepted Authors manuscript: © ACM, 2016. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in ACM Transactions on Algorithms, Vol. 12, No. 3 <http://doi.acm.org/10.1145/2836166>

Source: FindIt

Source-ID: 2303831818

Publication: Research - peer-review › Journal article – Annual report year: 2016