

Hooshmand, Sahar; Abedin, Paniz; Külekci, M. Oğuzhan; Thankachan, Sharma V.

Non-overlapping indexing – cache obliviously. (English) [Zbl 07286734](#)

Navarro, Gonzalo (ed.) et al., 29th annual symposium on combinatorial pattern matching, CPM 2018, July 2–4, 2018, Qingdao, China. Wadern: Schloss Dagstuhl – Leibniz Zentrum für Informatik (ISBN 978-3-95977-074-3). LIPIcs – Leibniz International Proceedings in Informatics 105, Article 8, 9 p. (2018).

Summary: The non-overlapping indexing problem is defined as follows: pre-process a given text $T[1, n]$ of length n into a data structure such that whenever a pattern $P[1, p]$ comes as an input, we can efficiently report the largest set of non-overlapping occurrences of P in T . The best known solution is by *H. Cohen* and *E. Porat* [Lect. Notes Comput. Sci. 5878, 1044–1053 (2009; [Zbl 1273.68097](#))]. Their index size is $O(n)$ words and query time is optimal $O(p + \text{nocc})$, where nocc is the output size. We study this problem in the cache-oblivious model and present a new data structure of size $O(n \log n)$ words. It can answer queries in optimal $O(\frac{p}{B} + \log_B n + \frac{\text{nocc}}{B})$ I/Os, where B is the block size.

For the entire collection see [[Zbl 1390.68025](#)].

MSC:

[68W32](#) Algorithms on strings

Keywords:

[suffix trees](#); [cache oblivious](#); [data structure](#); [string algorithms](#)

Full Text: [DOI](#)

References:

- [1] Alok Aggarwal and Jeffrey Scott Vitter. The input/output complexity of sorting and related problems. *{\it Commun. ACM}*, 31(9):1116-1127, 1988. doi:10.1145/48529.48535.
- [2] Alberto Apostolico and Franco P Preparata. Data structures and algorithms for the string statistics problem. *{\it Algorithmica}*, 15(5):481-494, 1996.
- [3] Michael A. Bender, Erik D. Demaine, and Martin Farach-Colton. Cache-oblivious b-trees. *{\it SIAM J. Comput.}*, 35(2):341-358, 2005. doi:10.1137/S0097539701389956.
- [4] Gerth Stølting Brodal and Rolf Fagerberg. Cache-oblivious string dictionaries. In *{\it Pro-} {\it ceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA} {\it 2006, Miami, Florida, USA, January 22-26, 2006}*, pages 581-590. ACM Press, 2006. URL: <http://dl.acm.org/citation.cfm?id=1109557.1109621>.
- [5] Hagai Cohen and Ely Porat. Range non-overlapping indexing. In Yingfei Dong, DingZhu Du, and Oscar H. Ibarra, editors, *{\it Algorithms and Computation, 20th International} {\it Symposium, ISAAC 2009, Honolulu, Hawaii, USA, December 16-18, 2009. Proceedings}*, volume 5878 of *{\it Lecture Notes in Computer Science}*, pages 1044-1053. Springer, 2009. doi:10.1007/978-3-642-10631-6_105.
- [6] Maxime Crochemore. String-matching on ordered alphabets. *{\it Theoretical Computer Science}*, 92(1):33-47, 1992.
- [7] Matteo Frigo, Charles E. Leiserson, Harald Prokop, and Sridhar Ramachandran. Cacheoblivious algorithms. In *{\it 40th Annual Symposium on Foundations of Computer Science,} {\it FOCS '99, 17-18 October, 1999, New York, NY, USA}*, pages 285-298. IEEE Computer Society, 1999. doi:10.1109/SFFCS.1999.814600.
- [8] Matteo Frigo, Charles E. Leiserson, Harald Prokop, and Sridhar Ramachandran. Cacheoblivious algorithms. *{\it ACM Trans. Algorithms}*, 8(1):4:1-4:22, 2012. doi:10.1145/2071379.2071383.
- [9] :9
- [10] Arnab Ganguly, Rahul Shah, and Sharma V Thankachan. Succinct non-overlapping indexing. In *{\it Annual Symposium on Combinatorial Pattern Matching}*, pages 185-195. Springer, 2015.
- [11] Dov Harel and Robert Endre Tarjan. Fast algorithms for finding nearest common ancestors. *{\it SIAM J. Comput.}*, 13(2):338-355, 1984. doi:10.1137/0213024.
- [12] Orgad Keller, Tsvi Kopelowitz, and Moshe Lewenstein. Range non-overlapping indexing and successive list indexing. In Frank K. H. A. Dehne, Jörg-Rüdiger Sack, and Norbert Zeh, editors, *{\it Algorithms and Data Structures, 10th International Workshop, WADS 2007, Hali-} {\it fax, Canada, August 15-17, 2007, Proceedings}*, volume 4619 of *{\it Lecture Notes in Computer} {\it Science}*, pages 625-636. Springer, 2007. doi:10.1007/978-3-540-73951-7_54.
- [13] Udi Manber and Eugene W. Myers. Suffix arrays: A new method for on-line string searches. *{\it SIAM J. Comput.}*, 22(5):935-

948, 1993. doi:10.1137/0222058.

- [14] Yakov Nekrich and Gonzalo Navarro. Sorted range reporting. In Fedor V. Fomin and Petteri Kaski, editors, *Algorithm Theory - SWAT 2012 - 13th Scandinavian Symposium and Workshops*, Helsinki, Finland, July 4-6, 2012. Proceedings}, volume 7357 of *Lecture Notes in Computer Science*, pages 271-282. Springer, 2012. doi:10.1007/978-3-642-31155-0_24.
- [15] Esko Ukkonen. On-line construction of suffix trees. *Algorithmica*, 14(3):249-260, 1995. doi:10.1007/BF01206331.
- [16] Peter Weiner. Linear pattern matching algorithms. In *14th Annual Symposium on Switching and Automata Theory*, Iowa City, Iowa, USA, October 15-17, 1973}, pages 1-11. IEEE Computer Society, 1973. doi:10.1109/SWAT.1973.13.

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.