

Abedin, Paniz; Hooshmand, Sahar; Ganguly, Arnab; Thankachan, Sharma V.

The heaviest induced ancestors problem: better data structures and applications. (English)

Zbl 07549531

Algorithmica 84, No. 7, 2088-2105 (2022)

Summary: Let \mathcal{T}_1 and \mathcal{T}_2 be two rooted trees with an equal number of leaves. The leaves are labeled, and the labeling of the leaves in \mathcal{T}_2 is a permutation of those in \mathcal{T}_1 . Nodes are associated with weight, such that the weight of a node u , denoted by $W(u)$, is more than the weight of its parent. A node $x \in \mathcal{T}_1$ and a node $y \in \mathcal{T}_2$ are induced, iff their subtrees have at least one common leaf label. A *heaviest induced ancestor* query $\text{HIA}(u_1, u_2)$ with input nodes $u_1 \in \mathcal{T}_1$ and $u_2 \in \mathcal{T}_2$ asks to output the pair (u_1^*, u_2^*) of induced nodes with the highest combined weight $W(u_1^*) + W(u_2^*)$, such that u_1^* is an ancestor of u_1 and u_2^* is an ancestor of u_2 . This is a useful primitive in several text processing applications. Gagie et al. (Proceedings of the 25th Canadian Conference on Computational Geometry, CCCG 2013, Waterloo, Ontario, Canada, 2013) introduced this problem and proposed three data structures with the following space-time trade-offs: (i) $O(n \log^2 n)$ space and $O(\log n \log \log n)$ query time, (ii) $O(n \log n)$ space and $O(\log^2 n)$ query time, and (iii) $O(n)$ space and $O(\log^{3+\epsilon} n)$ query time. Here n is the number of nodes in both trees combined and $\epsilon > 0$ is an arbitrarily small constant. We present two new data structures with better space-time trade-offs: (i) $O(n \log n)$ space and $O(\log n \log \log n)$ query time, and (ii) $O(n)$ space and $O(\log^2 n / \log \log n)$ query time. Additionally, we present new applications of these results.

MSC:

68P05 Data structures

68W32 Algorithms on strings

Keywords:

data structure; string algorithms; orthogonal range queries

Full Text: DOI

References:

- [1] Abedin, P., Hooshmand, S., Ganguly, A., Thankachan, S.V.: The heaviest induced ancestors problem revisited. In: Navarro, G., Sankoff, D., Zhu, B. (eds.) Annual Symposium on Combinatorial Pattern Matching, CPM 2018, July 2-4, 2018 - Qingdao, China, LIPIcs, vol. 105, pp. 20:1-20:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2018). doi:10.4230/LIPIcs.CPM.2018.20 · Zbl 07286746
- [2] Aluru, S.: Handbook of Computational Molecular Biology. Chapman and Hall/CRC (2005) · Zbl 1115.68387
- [3] Amir, A., Boneh, I.: Locally maximal common factors as a tool for efficient dynamic string algorithms. In: Navarro, G., Sankoff, D., Zhu, B. (eds.) Annual Symposium on Combinatorial Pattern Matching, CPM 2018, July 2-4, 2018 - Qingdao, China, LIPIcs, vol. 105, pp. 11:1-11:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2018). doi:10.4230/LIPIcs.CPM.2018.11 · Zbl 07286737
- [4] Amir, A., Boneh, I., Charalampopoulos, P., Kondratovsky, E.: Repetition detection in a dynamic string. In: Bender, M.A., Svensson, O., Herman, G. (eds.) 27th Annual European Symposium on Algorithms, ESA 2019, September 9-11, 2019, Munich/Garching, Germany, LIPIcs, vol. 144, pp. 5:1-5:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2019). doi:10.4230/LIPIcs.ESA.2019.5 · Zbl 07525442
- [5] Amir, A., Charalampopoulos, P., Iliopoulos, C.S., Pissis, S.P., Radoszewski, J.: Longest common factor after one edit operation. In: Fici, G., Sciortino, M., Venturini, R. (eds.) String Processing and Information Retrieval - 24th International Symposium, SPIRE 2017, Palermo, Italy, September 26-29, 2017, Proceedings, Lecture Notes in Computer Science, vol. 10508, pp. 14-26. Springer (2017). doi:10.1007/978-3-319-67428-5_2 · Zbl 1454.68196
- [6] Amir, A., Charalampopoulos, P., Pissis, S.P., Radoszewski, J.: Longest common substring made fully dynamic. In: Bender, M.A., Svensson, O., Herman, G. (eds.) 27th Annual European Symposium on Algorithms, ESA 2019, September 9-11, 2019, Munich/Garching, Germany, LIPIcs, vol. 144, pp. 6:1-6:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2019). doi:10.4230/LIPIcs.ESA.2019.6 · Zbl 07525443
- [7] Amir, A.; Charalampopoulos, P.; Pissis, SP; Radoszewski, J., Dynamic and internal longest common substring, Algorithmica, 82, 12, 3707-3743 (2020) · Zbl 1494.68314 · doi:10.1007/s00453-020-00744-0
- [8] Amir, A., Kondratovsky, E.: Searching for a modified pattern in a changing text. In: Gagie, T., Moffat, A., Navarro, G., Cuadros-Vargas, E. (eds.) String Processing and Information Retrieval - 25th International Symposium, SPIRE 2018,

- Lima, Peru, October 9-11, 2018, Proceedings, Lecture Notes in Computer Science, vol. 11147, pp. 241-253. Springer (2018). doi:10.1007/978-3-030-00479-8_20
- [9] Bender, M.A., Farach-Colton, M.: The LCA problem revisited. In: Gonnet, G.H., Panario, D., Viola, A. (eds.) LATIN 2000: Theoretical Informatics, 4th Latin American Symposium, Punta del Este, Uruguay, April 10-14, 2000, Proceedings, Lecture Notes in Computer Science, vol. 1776, pp. 88-94. Springer (2000). doi:10.1007/10719839_9 · Zbl 0959.68133
- [10] Brodal, G.S., Jørgensen, A.G.: Data structures for range median queries. In: Dong, Y., Du, D., Ibarra, O.H. (eds.) Algorithms and Computation, 20th International Symposium, ISAAC 2009, Honolulu, Hawaii, USA, December 16-18, 2009. Proceedings, Lecture Notes in Computer Science, vol. 5878, pp. 822-831. Springer (2009). doi:10.1007/978-3-642-10631-6_83 · Zbl 1273.68096
- [11] Chan, T.M., Larsen, K.G., Patrascu, M.: Orthogonal range searching on the RAM, revisited. In: Proceedings of the 27th ACM Symposium on Computational Geometry, Paris, France, June 13-15, 2011, pp. 1-10 (2011). doi:10.1145/1998196.1998198. · Zbl 1283.68139
- [12] Charalampopoulos, P., Gawrychowski, P., Pokorski, K.: Dynamic longest common substring in polylogarithmic time. In: Czuma, A., Dawar, A., Merelli, E. (eds.) 47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, July 8-11, 2020, Saarbrücken, Germany (Virtual Conference), LIPIcs, vol. 168, pp. 27:1-27:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2020). doi:10.4230/LIPIcs.ICALP.2020.27
- [13] Chockalingam, S.P., Thankachan, S.V., Aluru, S.: A parallel algorithm for finding all pairs k-mismatch maximal common substrings. In: West, J., Pancake, C.M. (eds.) Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC 2016, Salt Lake City, UT, USA, November 13-18, 2016, pp. 784-794. IEEE Computer Society (2016). doi:10.1109/SC.2016.66
- [14] Demaine, ED; Landau, GM; Weimann, O., On cartesian trees and range minimum queries, *Algorithmica*, 68, 3, 610-625 (2014) · Zbl 1360.68378 · doi:10.1007/s00453-012-9683-x
- [15] Farach, M.: Optimal suffix tree construction with large alphabets. In: 38th Annual Symposium on Foundations of Computer Science, FOCS '97, Miami Beach, Florida, USA, October 19-22, 1997, pp. 137-143. IEEE Computer Society (1997). doi:10.1109/SFCS.1997.646102
- [16] Ferragina, P.; Manzini, G., Indexing compressed text, *J. ACM*, 52, 4, 552-581 (2005) · Zbl 1323.68261 · doi:10.1145/1082036.1082039
- [17] Fischer, J.; Heun, V., Space-efficient preprocessing schemes for range minimum queries on static arrays, *SIAM J. Comput.*, 40, 2, 465-492 (2011) · Zbl 1222.05024 · doi:10.1137/090779759
- [18] Funakoshi, M., Mieno, T.: Minimal unique palindromic substrings after single-character substitution. In: Lecroq, T., Touzet, H. (eds.) String Processing and Information Retrieval - 28th International Symposium, SPIRE 2021, Lille, France, October 4-6, 2021, Proceedings, Lecture Notes in Computer Science, vol. 12944, pp. 33-46. Springer (2021). doi:10.1007/978-3-030-86692-1_4
- [19] Funakoshi, M., Nakashima, Y., Inenaga, S., Bannai, H., Takeda, M.: Longest substring palindrome after edit. In: Navarro, G., Sankoff, D., Zhu, B. (eds.) Annual Symposium on Combinatorial Pattern Matching, CPM 2018, July 2-4, 2018 - Qingdao, China, LIPIcs, vol. 105, pp. 12:1-12:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2018). doi:10.4230/LIPIcs.CPM.2018.12 · Zbl 07286738
- [20] Funakoshi, M., Nakashima, Y., Inenaga, S., Bannai, H., Takeda, M.: Faster queries for longest substring palindrome after block edit. In: Pisanti, N., Pissis, S.P. (eds.) 30th Annual Symposium on Combinatorial Pattern Matching, CPM 2019, June 18-20, 2019, Pisa, Italy, LIPIcs, vol. 128, pp. 27:1-27:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2019). doi:10.4230/LIPIcs.CPM.2019.27 · Zbl 07286738
- [21] Funakoshi, M.; Nakashima, Y.; Inenaga, S.; Bannai, H.; Takeda, M., Computing longest palindromic substring after single-character or block-wise edits, *Theor. Comput. Sci.*, 859, 116-133 (2021) · Zbl 07310533 · doi:10.1016/j.tcs.2021.01.014
- [22] Gagie, T., Gawrychowski, P., Nekrich, Y.: Heaviest induced ancestors and longest common substrings. In: Proceedings of the 25th Canadian Conference on Computational Geometry, CCCG 2013, Waterloo, Ontario, Canada, August 8-10, 2013. Carleton University, Ottawa, Canada (2013). http://cccg.ca/proceedings/2013/papers/paper_29.pdf
- [23] Grossi, R.; Vitter, JS, Compressed suffix arrays and suffix trees with applications to text indexing and string matching, *SIAM J. Comput.*, 35, 2, 378-407 (2005) · Zbl 1092.68115 · doi:10.1137/S0097539702402354
- [24] Gusfield, D., Algorithms on Strings, Trees, and Sequences - Computer Science and Computational Biology (1997), Cambridge: Cambridge University Press, Cambridge · Zbl 0934.68103 · doi:10.1017/CBO9780511574931
- [25] Harel, D.; Tarjan, RE, Fast algorithms for finding nearest common ancestors, *SIAM J. Comput.*, 13, 2, 338-355 (1984) · Zbl 0535.68022 · doi:10.1137/0213024
- [26] JáJá, J., Mortensen, C.W., Shi, Q.: Space-efficient and fast algorithms for multidimensional dominance reporting and counting. In: Fleischer, R., Trippen, G. (eds.) Algorithms and Computation, 15th International Symposium, ISAAC 2004, Hong Kong, China, December 20-22, 2004, Proceedings, Lecture Notes in Computer Science, vol. 3341, pp. 558-568. Springer (2004). doi:10.1007/978-3-540-30551-4_49 · Zbl 1116.68629
- [27] Nekrich, Y., Navarro, G.: Sorted range reporting. In: Fomin, F.V., Kaski, P. (eds.) Algorithm Theory - SWAT 2012 - 13th Scandinavian Symposium and Workshops, Helsinki, Finland, July 4-6, 2012. Proceedings, Lecture Notes in Computer Science, vol. 7357, pp. 271-282. Springer (2012). doi:10.1007/978-3-642-31155-0_24 · Zbl 1347.68343
- [28] Sadakane, K.: Succinct representations of lcp information and improvements in the compressed suffix arrays. In: Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms, January 6-8, 2002, San Francisco, CA, USA., pp. 225-232 (2002). <http://dl.acm.org/citation.cfm?id=545381.545410> · Zbl 1093.68578
- [29] Sadakane, K., Navarro, G.: Fully-functional succinct trees. In: Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, USA, January 17-19, 2010, pp. 134-149 (2010). doi:10.1137/1.9781611973075.13 · Zbl 1288.05046

- [30] Sleator, D.D., Tarjan, R.E.: A data structure for dynamic trees. In: Proceedings of the 13th Annual ACM Symposium on Theory of Computing, May 11-13, 1981, Milwaukee, Wisconsin, USA, pp. 114-122 (1981). doi:10.1145/800076.802464
- [31] Thankachan, S.V., Chockalingam, S.P., Aluru, S.: An efficient algorithm for finding all pairs k-mismatch maximal common substrings. In: Bourgeois, A.G., Skums, P., Wan, X., Zelikovsky, A. (eds.) Bioinformatics Research and Applications - 12th International Symposium, ISBRA 2016, Minsk, Belarus, June 5-8, 2016, Proceedings, Lecture Notes in Computer Science, vol. 9683, pp. 3-14. Springer (2016). doi:10.1007/978-3-319-38782-6_1
- [32] Urabe, Y., Nakashima, Y., Inenaga, S., Bannai, H., Takeda, M.: Longest lyndon substring after edit. In: Navarro, G., Sankoff, D., Zhu, B. (eds.) Annual Symposium on Combinatorial Pattern Matching, CPM 2018, July 2-4, 2018 - Qingdao, China, LIPIcs, vol. 105, pp. 19:1-19:10. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2018). doi:10.4230/LIPIcs.CPM.2018.19 · Zbl 07286745
- [33] Weiner, P.: Linear pattern matching algorithms. In: 14th Annual Symposium on Switching and Automata Theory, Iowa City, Iowa, USA, October 15-17, 1973, pp. 1-11 (1973). doi:10.1109/SWAT.1973.13
- [34] Willard, DE, Log-logarithmic worst-case range queries are possible in space $\theta(n)$, Inf. Process. Lett., 17, 2, 81-84 (1983) · Zbl 0509.68106 · doi:10.1016/0020-0190(83)90075-3
- [35] Zhou, G., Two-dimensional range successor in optimal time and almost linear space, Inf. Process. Lett., 116, 2, 171-174 (2016) · Zbl 1347.68105 · doi:10.1016/j.ipl.2015.09.002
- [36] Ziv, J.; Lempel, A., A universal algorithm for sequential data compression, IEEE Trans. Inf. Theory, 23, 3, 337-343 (1977) · Zbl 0379.94010 · doi:10.1109/TIT.1977.1055714

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.