

58093 String Processing Algorithms (Autumn 2014)

Study Groups 4 (December 11)

Group A: (P)LCP Array Construction

Members

- Chinnasamy, Mohan
- Escoter I Torres, Llorenç
- Faghihi Berenjegan, Farbod
- Goryachev, Vladimir
- Longi, Krista
- Siciliano, Gianvito

Advance reading. Read at least Sections 1,2 and 4 (and Lemma 1 in Section 3) of the following paper before the study group session.

Juha Kärkkäinen, Giovanni Manzini, Simon J. Puglisi:

Permuted Longest-Common-Prefix Array.

Proceedings of the 20th Symposium on Combinatorial Pattern Matching (CPM), pp. 181-192, 2009.

http://dx.doi.org/10.1007/978-3-642-02441-2_17

Topics for Discussion. Discuss at least the following topics in your group. Prepare to summarize the discussion for members of the other groups.

- The PLCP array. How is it related to the LCP array?
- Lemma 1. Compare it to Lemma 4.9 on the lectures.
- The Φ algorithm. Compare it to Algorithm 4.10 on the lectures.
- The algorithm based on irreducible LCPs. Describe the main ideas.

You may also discuss other topics, for example:

- The compact representations of the PLCP array.
- Experimental performance of the algorithms.

Group B: Suffix Array Construction

Members

- Greer, Maximillian
- Hurme, Teemu
- Kruglaia, Anna
- Kukkola, Ville
- Lagus, Jarkko
- Viding, Jasu

Advance reading. Read at least Sections 1-3 of the following paper before the study group session.

Juha Kärkkäinen, Dominik Kempa:

Engineering a Lightweight External Memory Suffix Array Construction Algorithm.

Proceedings of the 2nd International Conference on Algorithms for Big Data (ICABD), pp. 53-60, 2014.

<http://ceur-ws.org/Vol-1146/paper9.pdf>

Topics for Discussion. Discuss at least the following topics in your group. Prepare to summarize the discussion for members of the other groups.

- The SAscan algorithm. Describe the main ideas.
- The *partial* suffix array. How does it differ from a normal suffix array?
- The gap array. What is its role in the algorithm?
- The gt-bitvector. Why is it needed?

You may also discuss other topics, for example:

- Further details of the algorithm.
- Experimental performance of the algorithm.

Group C: Finding Repeats

Members

- Heino, Lauri
- Hopp, Joshua
- Hoya Quecedo, Jose
- Radev, Martin
- Virolainen, Herkko
- Wang, Ping

Advance reading. Read at least Section 5.3 excluding Subsection 5.3.6 (pp. 138-156) of the following book before the study group session. You may skip the parts that are not relevant to the discussion topics.

Enno Ohlebusch:

Bioinformatics Algorithms: Sequence Analysis, Genome Rearrangements, and Phylogenetic Reconstruction.

Oldenbusch Verlag, 2013.

[http://www.uni-ulm.de/in/theo/m/ohlebusch/
book-bioinformatics-algorithms.html](http://www.uni-ulm.de/in/theo/m/ohlebusch/book-bioinformatics-algorithms.html)

Some basic definitions are given in Section 4 (for example Definitions 4.1.3 and 4.3.1).

Topics for Discussion. Discuss at least the following topics in your group. Prepare to summarize the discussion for members of the other groups.

- Describe all the repeat finding problems for which an algorithm is given in the reading material.
- Maximal repeats (Sect. 5.3.3). Describe the main ideas of the algorithm for maximal repeats (Alg. 5.15).

You may also discuss other topics, for example algorithms for the other problems.

Group D: Comparing Multiple Strings

Members

- Hulkko, Heidi
- Leppänen, Jarno
- Nidia, Obscura
- Paasiniemi, Markus
- Pitkänen, Teemu
- Puuska, Samir

Advance reading. Read at least Section 5.6 excluding Subsections 5.6.6 and 5.6.7 (pp. 206–227) of the following book before the study group session. You may skip the parts that are not relevant to the discussion topics.

Enno Ohlebusch:

Bioinformatics Algorithms: Sequence Analysis, Genome Rearrangements, and Phylogenetic Reconstruction.

Oldenbusch Verlag, 2013.

[http://www.uni-ulm.de/in/theo/m/ohlebusch/
book-bioinformatics-algorithms.html](http://www.uni-ulm.de/in/theo/m/ohlebusch/book-bioinformatics-algorithms.html)

Some basic definitions are given in Section 4 (for example Definitions 4.1.3 and 4.3.1).

Topics for Discussion. Discuss at least the following topics in your group. Prepare to summarize the discussion for members of the other groups.

- Describe all the problems for which an algorithm is given in the reading material.
- Document frequency (Sect. 5.6.3). Describe the main ideas of the algorithm for the document frequency problem (Alg. 5.33).

You may also discuss other topics, for example algorithms for the other problems.