

## **58093 String Processing Algorithms (Autumn 2016)**

Study Groups 4 (Thursday, December 15)

### **Group A: Applications of Suffix Trees**

#### **Members**

- Aalto, Iiro
- Björklund, Otso
- Duda, Tomás
- Mesimäki, Jerry
- Mäki, Timo

**Advance reading.** Read at least Section 8.4 (pp. 145–151) of the following book before the study group session.

Veli Mäkinen, Djamel Belazzougui, Fabio Cunial and Alexandru I. Tomescu:

**Genome-Scale Algorithm Design: Biological Sequence Analysis in the Era of High-Throughput Sequencing.**

Cambridge University Press, 2015.

<http://dx.doi.org/10.1017/CBO9781139940023>

Contact the lecturer if you have trouble accessing the material.

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

- Describe the four problems studied in the reading material (Problems 8.2–8.5).
- Describe the main ideas of the algorithms for solving the four problems.

## Group B: Applications of Enhanced Suffix Arrays 1

### Members

- Leivo, Marcus
- He, Yan
- Jääsaari, Elias
- Salonen, Sami
- Tuominen, Markus

**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. Some material from Chapter 4 will be useful too such as Definitions 4.1.3 (p. 60) and 4.3.1 (p. 86) and Algorithm 4.6. (p. 94).

Enno Ohlebusch:

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

Oldenbusch Verlag, 2013.

Chapters 4 and 5 are available at <https://www.cs.helsinki.fi/i/tpkarkka/teach/15-16/SPA/Ohlebusch/>. Contact the lecturer if you have trouble accessing the material.

**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.
- Describe the main ideas of the algorithm for the longest non-overlapping repeats problem (Sect. 5.3.5).

## Group C: Applications of Enhanced Suffix Arrays 2

### Members

- Equi, Massimo
- Hantula, Otto
- Kääriä, Erkki
- Määttä, Mikko
- Zhukova, Bella

**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. Some material from Section 4 will be useful too such as Definitions 4.1.3 (p. 60) and 4.3.1 (p. 86).

Enno Ohlebusch:

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

Oldenbusch Verlag, 2013.

Chapters 4 and 5 are available at <https://www.cs.helsinki.fi/i/tpkarkka/teach/15-16/SPA/Ohlebusch/>. Contact the lecturer if you have trouble accessing the material.

**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.
- Describe the main ideas of the algorithm for the document listing problem (Sect. 5.6.4).

## Group D: Succinct Suffix Array

### Members

- De Leydet, Rémy
- Lang, Sean
- Norri, Tuukka
- Sinkkonen, Ilari

**Advance reading.** Read at least Sections 3.2 and 3.3 (pp. 22–25) and 9.2.1, 9.2.2 and 9.2.3 (pp. 160–164) of the following book before the study group session.

Veli Mäkinen, Djamel Belazzougui, Fabio Cunial and Alexandru I. Tomescu:

**Genome-Scale Algorithm Design: Biological Sequence Analysis in the Era of High-Throughput Sequencing.**

Cambridge University Press, 2015.

<http://dx.doi.org/10.1017/CBO9781139940023>

Contact the lecturer if you have trouble accessing the material.

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

- Describe the properties of the succinct suffix array. What queries does it support? What are the time and space complexities?
- Describe the components of the succinct suffix array and the main ideas of the query algorithms.