58093 String Processing Algorithms (Autumn 2014)
Study Groups 2 (November 18)

Advance reading material. The following article describes a large number of recent algorithms for exact string matching and summarizes the results of a large experimental comparison.

> S. Faro and T. Lecroq:
> **The exact online string matching problem: A review of the most recent results**.
> ACM Computing Surveys 45, 2, Article 13 (2013), 42 pages.
> `http://doi.acm.org/10.1145/2431211.2431212`

The full results of the comparison are in the following article.

> S. Faro and T. Lecroq:
> **The Exact String Matching Problem: a Comprehensive Experimental Evaluation**.
> CoRR abs/1012.2547 (2010).
> `http://arxiv.org/abs/1012.2547`

The source code for the implementations can be found here:

> `http://www.dmi.unict.it/~faro/smart/algorithms.php`

A web site describing many older algorithms may be useful too:

> `http://www-igm.univ-mlv.fr/~lecroq/string/`

Topics for discussion. Fig. 2 on page 13:39 of the main article shows the best algorithm for each combination of the parameters pattern length and alphabet size. Each of the algorithms in the figure is assigned to one of the study groups.

Read about the algorithms assigned to you group before the study group meeting. In the meeting discuss at least the following topics and prepare to summarize the discussion to the members of the other groups.

- Describe the algorithms. How are they related to each other and to the algorithms on the lectures?
- Why are they fast in the parametric regions where they are the best?

You may also discuss additional topics, for example:

- How fast the algorithms are outside their parametric regions? How far outside they are still close to the best algorithms?
- How fast the algorithms are for natural language or for DNA sequences?

Group A: Variants of Horspool

Members

- Chinnasamy, Mohan
- Greer, Maximillian
- Hopp, Joshua
- Kruglaia, Anna
- Kukkola, Ville
- Leppänen, Jarno
- Siciliano, Gianvito

Algorithms

- TVSBS (Thathoo-Virmani-Sai-Balakrishnan-Sekar, Sect. 4.1.1)
- FJS (Franek-Jennings-Smyth, Sect. 4.1.1)
- Hash$q$ (Wu-Manber for Single Pattern, Sect. 4.1.3)

Group B: Variants of BNDM

Members

- Escoter I Torres, Llorenç
- Faghihi Berenjegani, Farbod
- Hurme, Teemu
- Malmivirta, Titti
- Obscura Acosta, Nidia
- Paasiniemi, Markus
- Viding, Jasu

Algorithms

- SBNDM-BMH/BMH-SBNDM (Sect. 4.3.1)
- FSBNDM (Sect. 4.3.1)
- SBNDMq (Sect. 4.3.6)

Group C: Sparse Text Access Algorithms

Members

- Goryachev, Vladimir
- Holmes, Nicola
- Hoya Quecedo, Jose
- Hulkko, Heidi
- Radev, Martin
- Virolainen, Herkko

Algorithms

- AOSO (Average Optimal Shift-Or, Sect. 4.3.5)
- LBNDM (Long Pattern BNDM, Sect. 4.3.5)

Group D: Other Algorithm

Members

- Heino, Lauri
- Lagus, Jarkko
- Longi, Krista
- Magnússon, Joonas
- Pitkänen, Teemu
- Puuska, Samir
- Wang, Ping

Algorithms

- SA (Shift-And, Sect. 3.3)
- EBOM (Extended Bacward Oracle Matching, Sect. 4.2.1)
- SSEF (Sect. 4.4.1)