Introduction to DM823

Rolf Fagerberg

Fall 2010

DM823 - Motivation

String data are both common and fundamental:

> *Database, word processing, bioinformatics, and web retrieval. It is also a fundamental data model in computer science, containing e.g. integers and multi-dimensional data as special cases.*

We would like to:

- Compare strings to each other (string distances).
- Find substrings inside other strings (pattern matching).
- Find a string inside a set of strings (string dictionaries).
- Sort a set of strings (string sorting).
- Compress strings (file compression).

Subjects will be chosen based on

- Fundamentality of problem.
- Usefulness of algorithm.
- Algorithmic beauty.

DM823 - Formalities

Prerequisites: The contents of DM507 Algorithms and Data Structures and DM508 Algorithms and Complexity should be known.

Aims:
String data are very common: Database, word processing, bioinformatics, and web retrieval. It is also a fundamental data model in computer science, containing e.g. integers and multi-dimensional data as special cases.

Synopsis:
Pattern matching, exact and approximate. Suffix trees, suffix arrays, and other string data structures. String sorting. Compression algorithms. String distance algorithms.

Evaluation: Oral exam (7-point scale). Mandatory project (must be passed in order to attend the oral exam). Can be practical (implementation) or theoretical (exposition of research paper).