Computing Patterns in Strings – CAS 722
Fall 2021, Course Outline

September 14, 2021

Instructor: Prof. Neerja Mhaskar
Email: pophlin[at]mcmaster[dot]ca
Office Hours: TBA. However depending on the class size, I
might just schedule one on one meetings if needed.
Course URL: `https://avenue.mcmaster.ca`

Lectures and Tutorials:

- Online Lectures: Monday 9:30 – 12:30 on MS Team

Prerequisite(s):

Good undergraduate preparation in discrete mathematics, algorithms & data
structures, and complexity theory.

Calendar Description/Course Outline:

This course deals with algorithms for finding "patterns" in strings, patterns
of three main kinds: specific, generic, and intrinsic. The importance of ap-
proximate patterns and algorithms which identify them is made clear. In
addition various important algorithms and data structures proposed to com-
pute regularities in strings are discussed. Applications to Bioinformatics,
Data compression, Information retrieval and other scientific areas are em-
phasized.

Course Objectives:

The study of string algorithms is increasingly important in computer science, primarily due to its many applications, in computational biology, data compression, information retrieval, and other areas. Over the last 40 years a variety of elegant algorithms have been discovered. The objective of this course is to provide students with some sense of this mathematical diversity while at the same time making clear the subject's relevance to applications.

See Calendar Description.

Resources:
Below are some good books on string algorithms for your reference. The lectures slides, however would be your main material to follow.

1. Computing Patterns in Strings, Bill Smyth. Pearson (2003).

2. Algorithms on Strings, Maxime Crochemore, Christophe Hancart & Thierry Lecroq, Cambridge University Press (2007)

3. Algorithms on Strings, Trees & Sequences, Dan Gusfield, Cambridge University Press (1997)

4. Introduction to Computational Molecular Biology, Joao Carlos Setubal & Joao Meidanis, PWS Publishing (1997)

5. String Searching Algorithms, Graham A. Stephen, World Scientific Publishing (1994)

All of these books are available in the Thode Science Library.

Some useful files for testing purposes can be picked up at the Prof. Smyth's website (Simon's collection of test strings).

Grading:

- Problem Solving (includes individually assigned problems, and in-class presentation of your solution): 20%

- Review and present two published papers: 15% each

- Project: 50% each

<u>Problem Solving</u>: Problem solving includes solving the assigned problems at home and presenting your solution in class.

<u>Review and present two published papers</u>: You may review and present any two papers of your choice from the list of papers given by the instructor.

<u>Project</u>: Typically projects are unsolved (research and coding) problems of interest to the student and the instructor. At the beginning of the term many options for choosing your project will be given. You may choose a project from this list to work on, or come up with your own interesting project to work on!

Please note:

The instructors and university reserve the right to modify elements of the course during the term. The university may change the dates and deadlines for any or all courses in extreme circumstances. If either type of modification becomes necessary, reasonable notice and communication with the students will be given with explanation and the opportunity to comment on changes. It is the responsibility of the student to check their McMaster email and course websites weekly during the term and to note any changes.