CAS 722 – Computing Patterns in Strings
Course Outline

Neerja Mhaskar & Bill Smyth

January 4, 2019

Instructors: Dr. Neerja Mhaskar & Prof. Bill Smyth
Email: (pophlin, smyth)[at]mcmaster[dot]ca

Course Outline:

A string is just a sequence (linked list) of letters drawn from some alphabet. Thus it is a very simple data structure, perhaps deceptively so, since it is important in numerous areas of Computer Science and of science in general: compilers, data compression, cryptography, computer security, computational biology. A computer program (source or object code) is a string; so are

* the human genome;

* a text file;

* the contents of a computer's main memory;

* a book;

* the e-mail messages received at a web site;

* and many many more objects in the real world.

Generally speaking, one searches for **patterns** in strings:

* **specific**, such as finding selected phrases in a text file or in the day's e-mail messages;

* **generic**, in particular repeating substrings or sequences in a coded message or in a DNA sequence;

* **intrinsic**, those that are always there, such as the periods of a string and its prefixes.

To a mathematician, a string is a **combinatorial object**, and the subject matter of this course is really a subset of **combinatorial algorithms**. The emphasis is very much on algorithms — those that search for specific, generic and intrinsic patterns. The treatment however is mathematical, and students will find themselves constantly solving mathematical problems and making presentations of their results to their classmates. Graduate courses based on the book have so far been given at McMaster (every year 1997–2015) and at various other universities around the world.

Prerequisites:

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

Textbook:

- Bill Smyth. Computing Patterns in Strings

Marking Scheme:

- Problem Solving - 10%

- Final Project: 90%

Problem Solving:

Problem solving is 10% of your grade. It includes solving the assigned exercise problems listed at the back of the chapters at home and presenting your solution in class. Beware, some of the exercises are open problems! (All the more reason to solve them.)

Project:

Credit in the course is primarily based on a project due at end April; the title and abstract of the project are settled through discussion with the instructors no later than end January. Typically projects are unsolved problems of interest to the student and the lecturers.

The course is based on one of the instructor's book of the same name, published in 2003.

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.