CS 860: Patterns in Strings
Fall 2008

Instructor:	Jeffrey Shallit
Office:	DC 3134
Phone:	(519) 888-4567, x 34804
E-mail:	`shallit@cs.uwaterloo.ca`
Course website:	`http://www.cs.uwaterloo.ca/~shallit/Courses/860`
Office hours:	10:00 AM - 11:00 AM Tu Th, DC 3134
Class Meets:	11:30 AM - 12:20 PM MWF, MC 2036A

This is a course about the combinatorics and algorithmics of patterns in strings. The simplest example of a pattern is xx, which represents the squares. For example, a square in English is the word `hotshots`, which can be written as $(\texttt{hots})^2$.

We are interested in words that avoid patterns; avoiding squares is one of the simplest examples. Over a binary alphabet it is easy to see that every word of length > 3 contains a square. However, over an alphabet of three or more symbols it is possible to find arbitrarily large words with no squares. Equivalently there are infinite words over such an alphabet that avoid squares.

Given a set S of patterns, the kinds of questions we are interested in include:

- How long is the shortest string that contains all strings of length n matching a pattern of S? How can we efficiently generate such a string?

- How can we efficiently determine if a string of length n avoids S?

- How can we efficiently generate a string of length n that avoids S?

- What is the smallest alphabet size k that allows us to construct an infinite word avoiding S?

- What are good upper and lower bounds on the number of words of length n avoiding S?

- Is the number of infinite words avoiding S finite, infinite, or uncountable?

- For alphabets of cardinality smaller than k, what is the length of the longest word avoiding S?

- What is the minimum frequency of a letter occurring in an infinite word avoiding S over an alphabet of size k?

- What happens if we only want to avoid all sufficiently large members of S?

- Is there a simple description for all words avoiding S?

The kinds of patterns we will study include powers, fractional powers, abelian powers (that is, consecutive blocks that are the same up to permutation), palindromes, and commutators. There are generalizations to bigger ordinals, to multiple dimensions, to permutations, and to graphs. We will study specific results, as well as general techniques such as the Lovász local lemma. We will see dozens of good open problems, some very accessible.

There are applications of the material to number theory, algebra, music, biology, and other areas.

The only real prerequisite for the course is a certain degree of mathematical sophistication, and comfort with reading and constructing proofs. Knowledge of finite automata may be helpful, but is not essential. Experience in constructing computer experiments, using a friendly computing environment such as Maple, MATLAB, or APL, may be helpful.

The course is scheduled to meet at 11:30 AM – 12:20 PM MWF, in MC 2036A. After some introductory lectures, we will follow the following schedule:

- I will lecture on Mondays;

- We will discuss open problems on Wednesdays;

- Students will read papers and present them on Fridays.

Here is a (very) tentative outline of what we will cover:

Week 1: Introduction, basic definitions and results

Week 2: No class, instructor at a conference

Weeks 3 & 4: Overlap-free words and Fife's characterization; student presentations begin on Friday, September 26

Week 5: Fractional powers, Dejean's conjecture

Week 6: The Lovász local lemma

Week 7: Patterns in arithmetic progressions

Week 8: Abelian powers

Week 9: Repetitions in graphs

Week 10: Patterns in permutations

Week 11: Algorithmics of patterns

Week 12: Generalization to larger ordinals, enumeration, other results, applications to biology, music, etc.

Course notes will be available on the course website

`http://www.cs.uwaterloo.ca/~shallit/Courses/860`

as I write them. I expect that eventually there will be a book on the subject.

As a term project, everybody will be expected to read a small number of papers and present them. There will also be a small number of problem sets, probably 3. There will be no midterm or final. Your course mark will depend on your in-class presentation and the problem sets. The course project ground rules and suggestions for projects are on the website.

I expect the course will be very interesting and that we will solve some open problems and write some papers.

There will be no class during the second week (September 15-19), as I am away at a conference. I am available anytime my office door is open, but if it is closed, then I am either not there or having a nap. In either case, it's a waste of time to knock.