

Pattern Matching Algorithms

Overview

Having grown accustomed to the wealth of the entire internet at our fingertips it is hard to remember that a mere 25 years ago, an encyclopedia on a CD was the height of information technology available to the individual! It is astounding that a few decades ago billions of dollars were invested by governments and industry for mapping the first human genome whereas today one can purchase his DNA on a flash drive for a few thousand dollars.

Some of the key algorithmic tools in these endeavors were pattern matching algorithms. Pattern matching problems are among the oldest in computer science. For example, Celera Genomics used *shotgun sequencing* to complete the first sequenced genome. Every text editor, from *Office* to *Emacs* uses the Boyer-Moore algorithm. Yet, the area is still a fertile ground for very active current research. Part of its appeal is in its many application domains, such as text editing, image processing, music analysis, or molecular biology. Some concrete examples follow. Google uses pattern matching algorithms for indexing images. The app *Soundhound* for detecting music, requires heterogeneous indexing. Pattern matching algorithms are used for detecting malware, such as from the SNORT database, in data streams. Indeed, pattern matching algorithms served as building blocks for major projects in the past and will continue to be invaluable in the future.

Another important aspect of the field is that pattern matching has produced or incorporated some novel and powerful algorithmic techniques and is likely to continue to be a catalyst for new data structures and algorithmic techniques. These techniques are general, transportable, and scalable, thus can be used to solve myriad tasks in heterogeneous searching and indexing.

Course participants will learn these topics through lectures and assignments. The classes will be transcribed by the students.

Modules	A: Matching : Aug 28 – Sep 4 B: Indexing : Sep 4 – Sep 7 Number of participants for the course will be limited to fifty.
You Should Attend If...	<ul style="list-style-type: none">▪ a researcher, graduate student, or an advanced undergraduate student familiar with data structures and algorithms, and interested in pattern matching algorithms,▪ you are an engineer who is interested in learning state of the art methods in pattern-matching for that you do (could be, but not limited to: computational biology, image processing, music analysis, text editing, and so on).
Fees	The participation fees for taking the course is as follows: Industry/Research Organizations: 3000 INR Academic Institutions: 2000 INR Students: 1500 INR The above fee includes all instructional materials, access to the computer labs (during specified times), and 24-hour free internet facility. The participants will be provided with accommodation on payment basis, subject to availability. The above does not include charges for registration on the GIAN portal.

The Faculty



Professor Amihood Amir is a Professor at the Department of Computer Science at Bar-Ilan University and Research Professor at the Johns Hopkins University. He is one of the international leaders in Pattern Matching, an area in which he wrote over 150 papers and book chapters. He is one of the members of the Steering Committee of CPM – the leading international conference in Pattern Matching.

Course Co-ordinator

Prof. Neeldhara Misra

Phone: +91 79 2395 2490

E-mail: neeldhara.m@iitgn.ac.in

<http://www.gian.iitkgp.ac.in/>