Probabilistic Methods in Computer Science and Applications
http://www.cs.purdue.edu/homes/spa/cs590a.html

Professor: W. Szpankowski
E-mail: spa@cs.purdue.edu
Course Hours: M-F, 11:00–14:30 in Room 201
Consulting: M-F, 14:30–15:30

Overview. In this course we discuss probabilistic and analytic methods that are often used in designing and analyzing (randomized) algorithms and data structures. Donald E. Knuth in Randomization and Religion confessed: “... If somebody would ask me, what in the last 10 years, what was the most important change in the study of algorithms I would have to say that people getting really familiar with randomized algorithms had to be the winner”. Following the general precept: Give a man a fish and you feed him for the day; teach him to fish and you feed him for his lifetime, we will concentrate on methodology. In the first part of the course, we discuss several techniques, and every method is illustrated by a variety of specific problems that arose from algorithms and data structures, mostly on strings. Our choice of algorithms stems from the fact that there has been a resurgence of interest in algorithms on sequences and their applications in computational biology, security, multimedia compression, and information theory.

Approximate Course Outline

0. Data structures and algorithms on sequences (digital trees, Lempel-Ziv data compression algorithms, pattern matching algorithms, shortest common superstring, string editing problem).

1. Review of probability (expectation, moments, variance, Markov and Chebyshev inequalities, type of probabilistic convergence).

2. Probabilistic models (memoryless sources, Markov sources, mixing and stationary sources)

3. The first and second moment methods (Lovasz local lemma, height in tries and suffix trees).

4. Large Deviations (Chernoff bound, Azuma’s inequality)


6. Randomized Algorithms (Las vegas vs Monte Carlo algorithms, fingerprinting techniques, randomized selection).


8. Applications (data compression, graph compression, twitter classification, data mining, bioinformatics, security).

Books used in the course:


Course Notes: I plan to run this course as a seminar with lectures, problems solving, and student presentations at the end of the class.