NOC:Algorithms for Big Data - Video course

COURSE OUTLINE

In this course, you will learn how to design and analyse algorithms in the streaming and property testing models of computation. The algorithms will be analysed mathematically, so it is intended for a mathematically mature audience with prior knowledge of algorithm design and basic probability theory.

Traditional algorithms work well when the input data fits entirely within memory. In many modern application contexts, however, the size of the input data is too large to fit within memory. In some cases, data is stored in large data centres or clouds and specific parts of it can be accessed via queries. In some other application contexts, very large volume of data may stream through a computer one item at a time. So the algorithm will get to see the data typically as a single pass, but will not be able to store the data for future reference. In this course, we will introduce computational models, algorithms and analysis techniques aimed at addressing such big data contexts.

COURSE DETAIL

S.No.	Lessons/Topics
1	Intro to Probability Theory
2	Tail bounds with Applications
3	Markov Chains and Random Walks
4	Randomized Algorithms against an Oblivious Adversary
5	Pairwise Independence and Universal Hashing
6	The Streaming Model
7	Approximate Counting
8	Approximate Median
9	Flajolet Martin -- Distinct Sampling
10	Alon-Mattias-Szegedy Sketch
11	Bloom Filters
12	Count-min Sketch
13	Property Testing Model
14	Local search and testing connectivity
15	Enforce and Test Techique: Biclique and Bipartiteness Testing
16	Random Walks and Testing Bipartiteness & Expansion
17	Regularity Lemma and Testing Triangle Freeness
18	Boolean Functions, BLR test for Linearity.

References:

NP-TEL

NPTEL

http://nptel.ac.in
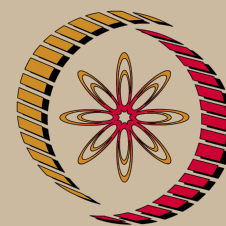
Computer Science and Engineering

Pre-requisites:

Algorithms, probability theory. Note: This course deals with analysis of algorithms in a mathematically rigorous manner.

Coordinators:

Prof. John Augustine
Department of Computer Science and EngineeringIIT Madras

1. [MU] Probability and Computing: Randomized Algorithms and Probabilistic Analysis, by Mitzenmacher and Upfal.

2. [Ron] Algorithmic and Analysis Techniques in Property Testing, by Dana Ron.

3. [CGHJ] Synopses for Massive Data: Samples, Histograms, Wavelets, Sketches, by Graham Cormode, Minos Garofalakis, Peter J. Haas and Chris Jermaine.