CENTRAL UNIVERSITY OF KERALA
DEPARTMENT OF COMPUTER SCIENCE
M.Sc. COMPUTER SCIENCE – PROGRAMME STRUCTURE

ELECTIVES					
COURSE CODE	COURSE TITLE	CONTACT HRS/WEEK			CREDITS
		LEC	LAB	TUT	
CSC5014	Algorithms for Big Data	2	2	1	4

Lec = Lecture, Tut = Tutorial, Lab = Practical

This is a participatory, experimental, problem solving and **employability based skill development course**.

Course Objective:
The objective of the course is to provide theoretical and practical aspects of big data algorithms.

By completing this course, students will obtain the following course/learning outcomes:
1. Knowledge gained:
   (i) Theoretical concepts for developing methods and algorithms for big data
2. Skill gained:
   (ii) Critical analyzing and logic **skills in developing methods and algorithms for big data**
3. Competency gained:
   (iii) Modelling and development of big data based applications.

Prerequisites: Basic knowledge of algorithms.

Grading:

Lab implementation	– 15%
Participatory based group Project	– 10%
Assignment/Quiz/presentation	– 5%
Class Test	– 10%
Final Exam	– 60%

CSC5014 – Algorithms for Big Data

Module 1
Intro to Probability Theory: Basic definitions, conditional probability, karger's min cut algorithm, random variables, Bernoulli, Binomial, and Geometric distributions, Tail bounds with Applications: application of chernoff bound, application of chebyshev's inequality.

Module 2
Introduction to Big Data Algorithms, SAT problem, classification of States, Stationary distribution of Markov Chain, random walks on undirected graphs, introduction to streaming, Morris algorithm, reservoir sampling, approximate median. Overview of data storage, balls and bins, hashing, chain hashing, bloom filter, pair wise independence, universal hashing functions, perfect hashing.

Module 3
Heavy hitters in data stream, Random walks on linear structures, lollipop graph, cats and mouse. Estimating frequency moments, property testing frame work, testing connectivity, enforce and test introduction, testing bicyclic graph, testing bipartiteness.

Module 4
Property testing and random walk algorithms, testing if graph is bipartite using random walks, graph streaming algorithms: introduction, matching, graph sparsification. Map reduce, K-machine (aka pregel model) model.

References:

1. Michael Mitzenmacher, Eli Upfal, Probability and Computing: Randomized Algorithms and Probabilistic Analysis, Cambridge University Press, Second edition, ISBN-13: 978-1107154889, ISBN-10: 9781107154889, 2017.
2. Dana Ron, Algorithmic and Analysis Techniques in Property Testing, now publishers Inc., 2010, ISBN: 978-1-60198-318-3
3. Graham Cormode, Minos Garofalakis, Peter J. Haas and Chris Jermaine. Synopses for Massive Data: Samples, Histograms, Wavelets, Sketches. now publishers Inc., 2011, ISBN: 978-1-60198-516-3