Boston University Metropolitan College
COMPUTER SCIENCE DEPARTMENT

Web Analytics and Mining
MET CS 688 A2

Instructor:
Tom Goulding MS, PhD
email: tlg@bu.edu
Office hours: by appointment

Class Time: Tues from 6:00pm to 9:00pm,
Location: SAR 104

Course Description

The Web Analytics and Mining course covers the areas of web analytics, text mining, web mining, and practical application domains. The web analytics part of the course studies the metrics of web sites, their content, user behavior, and reporting. Google analytics tool is used for collection of web site data and doing the analysis. The text mining module covers the analysis of text including content extraction, string matching, clustering, classification, and recommendation systems. The web mining module studies how web crawlers process and index the content of web sites, how search works, and how results are ranked.

The purpose of this course will be to explore these topics as our primary mission. Our secondary mission is to accomplish our goals by working collaboratively. Teams will be created the first day of class. Teams will work collaboratively throughout the semester in class and out of class. Homework problems, projects and classroom discussions will require extensive collaboration among team mates. The mid-term examination is a solo endeavors.

In each topic area, the methodology, including underlying assumptions and the mechanics of how it all works along with appropriate interpretation of the results, are discussed by teams in class. Teams will be heavily engaged in developing and presenting analytic concepts and solving problems in the context of real world examples.

A major topic discussion is led by the faculty member every two weeks. During the first week the class engages in an extensive discussion resulting in an understanding and application of a various web analytic concepts and tools. Problems are then assigned and the teams present the following week their solutions using the analytic tools of their choosing.

The Pedagogy:
The professor is a disciple and mathematics descendant of R.L. Moore the famous University of Texas Mathematician who was a practitioner of the Socratic method. The professor thus guides student research and engineering development projects through questions which activate inquiry, relentless experimentation and ultimately success.

For more on R. L. Moore see: . http://legacyrlmoore.org/reference/burton_jones.html

Course Prerequisites

MET CS 544 - Foundations of Analytics

Course Grading Policy

Presentations and demonstrations.

There will be approximately 5 presentations and demonstrations which are focused on applying theory and material learned in the week's modules. The 5 bi weekly team submissions should be via a Microsoft Word or Powerpoint document. The code, if any, used to generate your results should be appended to the end of your bi-weekly team presentations.

A semester long term project will be required in lieu of a final examination. Three medical research projects will be presented the first day of class. Each team will select one of the three.

The final grade for this course will be based on the following:

Deliverable	Weight
Bi Weekly presentation	50.00%
Mid Term Exam	25.00%
Term Project	25.00%

Course Topics

Week 1 - 2 / Module 1 - Web Analytics

Web Page Design Principles

Metrics

Key performance indicators

Referrers and visitors

Identifying important pages

Web site visibility

Week 3 – 4 / Module 2 - Web Analytics Tools
Using Google Analytics
Collecting data with Google Analytics
Dimensions and Segmentation
Flow visualization, navigating reports

Week 5 – 6 Module 3 - Text Mining
Preprocessing and content extraction
Searching and fuzzy string matching
Clustering text
Classification, categorization, and tagging
Question answering systems

Week 7 - 8 Module 4 - Web Mining
Web Crawlers, Indexing
Searching, precision and recall
Ranking
IoT (Internet of Things)

Week 9 – 10 Module 5 - Applications - Mining the Social Web
Twitter - trending topics, Facebook - Social Graph API
LinkedIn - Clustering colleagues, Google - Document Similarity

Week 11 – 12 Module 6 - Applications -
Metrics, telemetry, and analytics
Telemetry collection and tools
Game data analysis and visualization
Case studies

Reference Textbooks

Michael Beasley, Practical Web Analytics for User Experience: How Analytics can help you Understand your Users, Morgan Kaufmann, 2013.

Grant S. Ingersoll, Thomas S. Morton, and Andrew L. Farris, Taming Text: How to find, organize, and manipulate It, Manning Publications, 2013.

Matthew A. Russell, Mining the Social Web, 2nd Edition, O'Reilly, 2013

Magy Seif El-Nasr, Anders Drachen, Alessandro Canossa, eds., Game Analytics: Maximizing the Value of Player Data, Springer, 2013.

Course Web Site

http://learn.bu.edu

Student Conduct Code

http://www.bu.edu/met/for-students/met-policies-procedures-resources/academic-conduct-code