# ECT 584 Web Data Mining

**Session:** Summer II

Tues / Thurs 5:45-9:00

Class: CDM 216

Office: CDM 837

**Instructor:** Jonathan F. Gemmell

Email: jonathan.gemmell@gmail.com

Office Phone: (312) 810-3167

Office Hours: Wednesday 1:30 – 3:00

**<u>***Please schedule***</u>**

**Description and Objectives:** Web mining refers to the automatic discovery of interesting and useful patterns from the data associated with the usage, content, and the linkage structure of Web resources. It has quickly become one of the most popular areas in computing and information systems because of its direct applications in e-commerce, e-CRM, Web analytics, information retrieval/filtering, Web personalization, and recommender systems. Employees knowledgeable about Web mining techniques and their applications are highly sought by major Web companies such as Google, Amazon, Yahoo, MSN and others who need to understand user behavior and utilize discovered patterns from terabytes of user profile data to design more intelligent applications. The primary focus of this course is on Web usage mining and its applications to e-commerce and business intelligence. Specifically, we will consider techniques from machine learning, data mining, text mining, and databases to extract useful knowledge from Web data which could be used for site management, automatic personalization, recommendation, and user profiling. The first half of the course will be focused on a detailed overview of the data mining process and techniques, specifically those that are most relevant to Web mining. The second half will concentrate on the applications of these techniques to Web and e-commerce data, and their use in Web analytics, user profiling and personalization. This course also counts as an advanced course for Computer Science students concentrating in AI or Data Analysis.

**Textbooks and Reading Material:** Various papers or online resources (provided in class or online).

**Recommended Books:** Data Mining Techniques for Marketing, Sales, and Customer Relationship Management, Third Edition, by Michael Berry and Gordon Linoff, John Wiley, 2011. Data Mining: Practical Machine Learning Tools and Techniques, by Ian Witten and Eibe Frank, 3rd Ed., Morgan Kaufmann, 2011. [Note: this is the WEKA book]**.** Mining the Web: Transforming Customer Data into Customer Value, by Gordon Linoff and Michael Berry, John Wiley & Sons, 2001.

**Course Management System:** DePaul University's Desire2Learn system (d2l.depaul.edu).

**News Widget:** The primary form of communication for this class will be the news widget on the D2L. Please make sure you subscribe to the widget and that DePaul has you correct email.

**Prerequisites:** PREREQUISITE(S): IT 403 AND (CSC 451 or CSC 453 or CSC 455)

**Attendance:** Attendance is not required. However, students are responsible for all material and announcements presented in class. DL students are expected to watch the lectures online.

**Grading Policy:** The final grade will be determined based on the following components: Quizzes: 10%, Readings: 10%, Project: 40%, Exam: 40%. The general grading scheme will be based on a curve. At the end of the quarter, some adjustments may be made based on overall class performance as well as signs of individual effort. Plusses and minuses will be given at the high/low ends of each grade range.

**Online Teaching Evaluation:** Evaluations are a way for students to provide valuable feedback regarding their instructor and the course. Detailed feedback will enable the instructor to continuously tailor teaching methods and course content to meet the learning goals of the course and the academic needs of the students. They are a requirement of the course and are key to continue to provide you with the highest quality of teaching. The evaluations are anonymous; the instructor and administration do not track who entered what responses. A program is used to check if the student completed the evaluations, but the evaluation is completely separate from the student's identity. Since 100% participation is our goal, students are sent periodic reminders over three weeks. Students do not receive reminders once they complete the evaluation. Students complete the evaluation online in CampusConnect.

**Academic Integrity Policy:** This course will be subject to the academic integrity policy passed by faculty. More information can be found at http://academicintegrity.depaul.edu/

**Plagiarism:** The university and school policy on plagiarism can be summarized as follows: Students in this course should be aware of the strong sanctions that can be imposed against someone guilty of plagiarism. If proven, a charge of plagiarism could result in an automatic F in the course and possible expulsion. The strongest of sanctions will be imposed on anyone who submits as his/her own work any assignment which has been prepared by someone else. If you have any questions or doubts about what plagiarism entails or how to properly acknowledge source materials be sure to consult the instructor.

**Incomplete:** An incomplete grade is given only for an exceptional reason such as a death in the family, a serious illness, etc. Any such reason must be documented. Any incomplete request must be made at least two weeks before the final, and approved by the Dean of the College of Computing and Digital Media. Any consequences resulting from a poor grade for the course will not be considered as valid reasons for such a request.

**Resources for Students with Disabilities:** Students who feel they may need an accommodation based on the impact of a disability should contact the instructor privately to discuss their specific needs. All discussions will remain confidential.

To ensure that you receive the most appropriate accommodation based on your needs, contact the instructor as early as possible in the quarter (preferably within the first week of class), and make sure that you have contacted the Center for Students with Disabilities (CSD) at:

Student Center, LPC, Suite #370
Phone number: (773)325.1677
Fax: (773)325.3720
TTY: (773)325.7296