INF 385T - Metadata Generation and Interfaces for Massive Datasets (28240)

Instructor: Unmil Karadkar (email: unmil@ischool...., Twitter: @unmil, Skype: unmil.karadkar, Google hangouts: unmilk@gmail.com)

Teaching Assistant: Henna Kim

Class meetings: UTA 1.204 Tuesdays, 3 to 6 pm

Canvas page: https://utexas.instructure.com/courses/1176193

Prerequisites

- Graduate standing
- Basic knowledge of a programming language (for example: variables, constants, statements, conditionals, loops, function calling, and return values in a modern programming language such as PHP, Python, Java, or C++)
- Some understanding of data modeling (examples: Entity-Relationship diagrams, data models, Dublin Core, METS, or data structures)
- A motivation to learn-by-doing more as necessary for your project

Introduction

Large-scale digitization projects as well as increasing quantities of born-digital materials have put enormous collections of documents and data within our reach. This is a studio-style course designed to explore techniques that will make these massive quantities of data (although, not necessarily not BIGDATA) useful to targeted demographics, for specific goals via the use of programmatic techniques. This course both draws upon and will enable you to contribute to the areas of digital libraries and archives (collections, digitization), computational techniques (database management, data mining), and user experience (interaction design, HCI). Participants in this course will work in small teams, crafting small projects that demonstrate the viability of the proposed solutions. Typical projects will involve the development and/or evaluation of parts of data pipelines--ingestion, transformation, storage, manipulation, and presentation.

Objectives

Given a data source and a demographic interested in using this data, this course will enable you to:

- identify use cases for data sets and assess the necessary metadata for supporting a use case
- use third-party libraries/programs, such as ImageMagick or RESTful APIs to generate or retrieve metadata
- design disk storage structures (such as database schemas, MongoDB documents, XML/JSON/RDF structures) for extracted metadata
- craft appropriate user interface features or widgets for supporting users in accessing the data to address the target problem
- construct workflows and data "pipes", for converting between data formats and, optionally, to user interfaces
- evaluate the developed programs, scripts, techniques, and algorithms
- work in small teams to communicate and share technical as well as functional software specifications
- conduct web searches to solve specific problems and refine your searches in response to the found information

Pedagogy

Project-based learning: We will work on real problems and projects rather than something contrived. This course will enable (rather, require) you to synthesize competencies from several other courses taught in the iSchool, for example some combination of: Database Management, Understanding Research, Interaction Design, Usability, Digital Libraries, Visualization, Information Organization, Digital Archives, Digitization or others.

The key to success in this course is the ability to augment your existing skills on the fly. For example, if you have used simple database queries before, this course will demand more complex queries. Similarly, you will need to learn new features of a programming language that you are familiar with.

Peer learning: Much of the work in this class will be conducted in groups, which will give you an opportunity to learn from and teach your team mates simultaneously. You will certainly learn from me and I expect to learn from you.

Instructor as facilitator: Often, we will encounter situations without one correct answer or where I may not know the answer. I will explain the processes that I use for learning new techniques and provide a structure for such learning-on-the-fly. You will learn how to learn.

Grading

10% - Project proposal (due Wed., Sept. 28th)
10% - User interface prototype and/or system architecture (due Wed., Oct. 12th)
20% - weekly project update reports (Individual)
30% - Project implementation
10% - Adherence to programming style guide
20% - Final poster presentation

Textbooks and resources

There is no textbook for this course. We will use a variety of online resources. Class-specific protected files will be made available via Canvas.

Academic integrity

All students are expected to abide by the University of Texas Honor code, reproduced below for your convenience.

The core values of The University of Texas at Austin are learning, discovery, freedom, leadership, individual opportunity, and responsibility. Each member of the university is expected to these values through integrity, honesty, trust, fairness, and respect toward peers and community.

Violation of academic integrity, especially plagiarism, will not be tolerated. The first infraction will result in a grade of zero for that component of the course as well as a formal reprimand in your student file for future reference. Penalty for a second violation will include failure of the course and University-level disciplinary action.

Disability accommodations

The University of Texas at Austin provides upon request appropriate academic adjustments for qualified students with disabilities. For more information, contact the Services for Students with Disabilities (SSD) at (512) 471-6259 (voice) or (512)-410-6644 (video phone). An official letter from SSD is required in order to avail academic accommodations.

Please notify me as quickly as possible if the material being presented in class is not accessible (for example, instructional videos need captioning, course packets are not readable for proper alternative text conversion, etc.).

Emergency Preparedness

Please see details in the files section.