LIS 487: Data Interoperability

Instructor: Amber Stubbs, PhD (stubbs@simmons.edu)

Lectures: Mondays, 4:30-7:30 PM, P213

Office hours: Mondays 3-4 or by appointment, P212C

Course Description:

Libraries and archives rely on data. While data is ubiquitous, the formats in which data is stored can vary widely, and the differences in formats can lead to difficulties in finding answers to question and interpreting results of searches. This class examines different data formats, and how the information they store can be transformed into other formats, and the inherent difficulties in some of these transformations. This class uses the Python programming language and related libraries to examine and transform data in a variety of formats, including .txt, CSV, XML, and JSON. By the end of the course, students will be able to write programs to perform these transformations accurately, and with awareness of potential ways that data can be lost or mistranslated.

Prerequisite: Introduction to Programming

Grading:

Weekly labs	30%
Project 1: XML	20%
Project 2: Regular expressions	20%
Project 3: Interoperability use case	20%
Participation	10%

Final grades will be assigned according to the following ranges:

Points (%) range	**Grade**
94 - 100	A
90 - 93	A-
87 - 89	B+
84 - 86	B
80 - 83	B-
77 - 79	C+
74 - 76	C
70 - 73	C-
67 - 69	D+
64 - 66	D
Below 64	F

Course learning outcomes:

- Enhance programming skills by transforming data into different, professionally applicable formats
- Evaluate the limitations and benefits of different data formats and their uses
- Create programs and procedures for ensuring accurate interoperability between data resources

Program learning outcomes: (I = introduced, E = emphasized, R = reinforced)

- Apply professional standards, tools, and best practices in the information field and across specialized areas. – R
- Communicate effectively to different audiences through use of oral, written, and visual formats across multiple media. – E
- Develop appropriate technology strategies across a range of information settings. – R
- Evaluate and create information services and/or systems to reflect and respond to the needs of diverse communities and stakeholders. – I
- Demonstrate individual and collaborative leadership ability. – I
- Be guided by professional ethics and values. – I

Technology:

This class will use the Windows Virtual Machines provided by Simmons College for work in lab. Students may use their own laptops, but they will be responsible for ensuring that their projects work on the Simmons computers.

Project 1: XML generation and validation

Students will create their own XML, schema, and XSLT files on their own data. They will write Python code to ensure that their files are valid, and to output the data in CSV format

Project 2: Regular expressions

Given multiple files as input, the students will use regular expressions and Python to extract information from the files and present it in a coherent, unified format

Project 3: Interoperability use case

Using their own data or data provided by the instructor, the student will create a full suite of programs to transform the data between at least 3 different formats of data or metadata standards.

Contact with the instructor:

Email is always the best way to get in touch with me. I will make every effort to respond to your emails within 24 hours (48 on the weekends). Please put "LIS487" in the subject of any course-related email you send me. My office hours are listed on Moodle; if those hours do not work for you then please get in touch and we can schedule an appointment.

Accommodations:

If you have a documented disability and anticipate needing accommodations in this course, it is your responsibility to register with the Disability Services office as soon as possible to ensure that requested accommodations may be implemented in a timely fashion. For more information or to request academic accommodations, contact the Disability Services Office located in Room E-108 of the Main College Building. They are available by phone at 617-521-2474 or you may email Tim Rogers at timothy.rogers@simmons.edu.

Title IX and the Simmons College Gender-Based Misconduct Policy

Title IX Federal law states that all students have the right to gain an education free of gender-based discrimination. Some examples of gender-based discrimination, as defined by this law include sexual harassment or exploitation, sexual assault, domestic/dating violence, and stalking. In compliance with Title IX, Simmons College has a 'Gender- Based Misconduct Policy' which defines these forms of misconduct, outlines College protocol and procedures for investigating and addressing incidences of gender-based discrimination, highlights interim safety measures, and identifies both on and off-campus resources. The policy and a list of resources is located here: http://internal.simmons.edu/students/general-information/title-ix/gender-based-misconduct-policy-for-students-faculty-staff-and-visitors. Additionally, the Gender-Based Misconduct Policy has a Consensual Relationships clause that prohibits intimate, romantic or sexual relationships between students, faculty, staff, contract employees of the College, teacher's assistants, and supervisors at internship/field placement sites.

Policies:

- Attendance at all lectures is mandatory.
  - Coming to class but spending the whole session on the Internet doesn't count
  - Missing classes will, therefore, affect your attendance and participation grade
- Homework will be assigned each week and due before class the following week.
  - Collaboration is encouraged but copying is cheating.
    - More specifically, you can discuss concepts and general approaches, but you shouldn't share actual code.
  - Assignments passed in after class has started will be recorded as "late" and will receive a point deduction. Assignments passed in after class has ended will not be accepted without prior permission
- All students must respect and follow the Simmons Honor Code: http://www2.simmons.edu/handbook/conduct/honor-system.php
-

Other expectations:

Lecture material should be reviewed before the next class since any questions on old material will be addressed only at the beginning of class. The reading should be done (at least skimmed over) before it's covered in lecture, and reviewed afterward. All assignments must be legible, well-formatted, on-time and relatively complete. Please, please, please make sure you observe the plagiarism rules for the take-home projects. Any students who pass in similar projects will be referred to the Honor Board.

LIS 487 – Data Interoperability
Course Calendar*

*This is a tentative schedule and may be changed during the semester. An up-to-date version will be available at the course page on Moodle.

Homeworks will usually be assigned each class and due before class the following week.

Week of:	**Topics/Events**:
September 11	Python review, file input/output, text file standards
September 18	Python libraries, CVS files
September 25	XML week 1
October 2	XML week 2 **Project 1 assigned**
October 9	**No class – Indigenous People's Day**
October 16	XML week 3
October 23	MARC and Dublin Core
October 30	Regular Expressions week 1 – professor at ASIS&T **Project 1 due**
November 6	Regular Expressions week 2 **Project 2 assigned**
November 13	JSON 1
November 20	JSON 2
November 27	Databases
December 4	Databases
December 11	Final presentations, course wrap-up