

Python Webcrawling and Natural Language Processing ACG 6935 - Spring 2018

Fisher School of Accounting

Syllabus version: December 26, 2017

1 General

1.1 Course objective

The course objective is to introduce first and second year PhD students in Accounting and Finance to Python. The primary focus will be on the following topics: Selenium to automate web crawling, Regular Expressions to parse HTML and SEC filings and the Natural Language Toolkit (NLTK) and SciPy libraries for natural language processing (using algorithms like LDA, SVM, etc.). The goal is to be able to automatically retrieve documents from the web and construct new variables based on the contents.

1.2 Instructor details

Instructor: Joost Impink, joost@ufl.edu

Office: 336 Gerson Hall, (352) 273-1974

Office hours: during lab sessions, or email for an appointment

1.3 Class schedule

The meeting dates are Monday (instruction) and Wednesday (lab session) in module 3. See table 1 on page 5 for a detailed (tentative) class schedule.

Monday/Wednesday, Periods 5-6 (11:45 - 1:40pm), GER 327

1.4 Textbook

We will be using online materials and no textbook. See Canvas for details.

1.5 Class website

This course has a website on Canvas (for announcements, grades, reading materials). We will use a github repository for Python code (details in class).

2 Course Requirements

2.1 Performance assessment

The grade is determined as follows:

Assignments	50%
Class participation	50%
Total	<u>100%</u>

The assignments and class participation are graded between 1 (lowest) and 10 (highest). The final grade is letter grade based on the following scale:

Grade	Score
A	9.00 - 10.00
A-	8.00 - 8.99
B+	7.00 - 7.99
B	6.00 - 6.99
B-	5.00 - 5.99
C+	4.00 - 4.99
C	3.00 - 3.99
C-	2.00 - 2.99
D	1.00 - 1.99

For example, if the grades are: 8.0 for the assignments (50%) and 9.0 for participation (50%), then the score would be: $8.0 \times 50\% + 9.0 \times 50\% = 8.5$, which corresponds to a A- (it is in the 8.00 - 8.99 score range).

2.2 UF policies for grade points

For current UF policies for assigning grade points, see <https://catalog.ufl.edu/ugrad/current/regulations/info/grades.aspx>.

2.3 Assignments

A typical class each Tuesday will consist of a discussion of an assignment, an introduction to new materials, a small in-class exercise and a discussion of a new assignment. Thursday classes consist of a lab-meeting, where students can work on the assignment and ask questions. Students need to submit their work before the following Tuesday

where the assignment will be discussed briefly. No case handed in, or attempted hand-ins after the deadline can result in failing the class.

2.4 Class sessions

Students are advised to treat class attendance as an academic appointment that must be met, much as one must meet a business appointment. Students are encouraged to attend all Thursday class sessions. Attending the Tuesday meetings (lab sessions) is optional.

Students are expected to arrive for classes prepared to meet class room obligations and to devote full attention and commitment to the work of that class. Each instruction session will consist of lecture, discussion, exercises related to the assignments.

Students are required to bring a laptop to class, but only to be used for tasks related to the class session. Emailing, texting, and working on matters unrelated to the work at hand are inappropriate behaviors because they are disrespectful and distracting to the class and to the instructor.

2.5 Policy issues

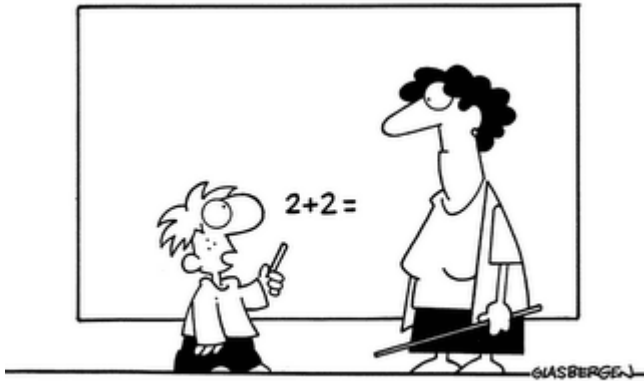
Assignments are individual efforts. Cheating is not tolerated. The University's Honor Code applies in all matters and will be enforced without limit or exception.

2.6 Students with disabilities

Students with disabilities requesting accommodations should first register with the Disability Resource Center (352-392-8565, www.dso.ufl.edu/drc/) by providing appropriate documentation. Once registered, students will receive an accommodation letter which must be presented to the instructor when requesting accommodation. Students with disabilities should follow this procedure as early as possible in the semester.

2.7 Course evaluation

Students are expected to provide feedback on the quality of instruction in this course based on 10 criteria. These evaluations are conducted online at <https://evaluations.ufl.edu>. Evaluations are typically open during the last two or three weeks of the semester, but students will be given specific times when they are open. Summary results of these assessments are available to students at <https://evaluations.ufl.edu/results/>.



"Do you want my answer in school math or corporate math?"

Table 1: Class Schedule (tentative)

Class	Date	Topic
1	Jan 8	Introduction to Python
2	Jan 10	Lab session
3	Jan 15	No class (MLK Day)
4	Jan 17	Lab session
5	Jan 22	Selenium
6	Jan 24	Lab session
7	Jan 29	Regular expressions
8	Jan 31	Lab session
9	Feb 5	NLTK
10	Feb 7	Lab session
11	Feb 12	LDA
12	Feb 14	Lab session
13	Feb 19	SVM
14	Feb 21	Lab session