

Programming Applications

CSCI-GA.1123

(Note: This course will not be offered in the academic year 2017/2018.)

Course description:

This course focuses on building applications in Python using a project-based learning approach. Students will design projects using important Python packages in a variety of applied areas such as textual analysis, image processing, data analysis and data visualization, and others.

Prerequisite: *Introduction to Programming* or by passing a placement test

Overview:

This course introduces important programming concepts in Python in the context of building applications using Python. Students will design and implement four programming projects in Python in a variety of applied areas such as image processing, textual analysis, and data analysis and data visualization. Students will apply these skills in projects of their own design.

Note: As programming languages evolve, the particular language taught in this course may change.

Sample Projects

Course Unit 1: Text Mining

Sample Projects: Projects could include studies on plot, character development or manners of speech in a corpus of prose; studies on the textual/linguistic aspects of U.S. Presidents' speeches over time; examining legal documents over a specific period of time or in a specific location to check for bias or mores that reflect the times; analyzing textual data bases (Lexis/Nexus) to identify international events and relevant actors; analyzing tweets for bullying language or for the types of tweets that are likely to be retweeted.

Course Unit 2: Data Extraction and Manipulation

Sample Projects: Projects could include data extractions from governmental sites (Federal, state or city) for data analysis to study trends in education, transportation, and/or incomes; data extraction and manipulation of historical texts complemented by statistical results; retrieving historical documents and programmatically extracting data such as address or other location information in order to build a dataset for analysis as well as to cull data to be used for creating maps with GIS software.

Course Unit 3: Charts, Graphs and Data Visualization in Python

Sample Projects: Projects could include working with data about readership or literacy over time or in specific regions with appropriate charts and graphs to visualize the data; working with historical data regarding occupations and incomes with charts and graphs to visualize the trends in grouping workers by gender, race, religion; retrieving data on the growth of museum collections over time with charts and graphs to demonstrate patterns in acquisition and how they reflect curatorial decisions; visualizing

international or domestic political data via maps; visualizing relationships between campaign spending and vote outcomes by state; visualizing information relevant to current policy issues such as changes over time in climate or inequality; visualizing relationships by county in social indicators such as income, race, and outcomes such as health and education.

Course Unit 4: Image Processing in Python

Sample Projects: Projects could include a comparative study of original manuscripts by selected poets to study penmanship or handwriting; comparing illustrations for a given 19th century literary classic over time; a study of images of ancient, historical or current coins; use of satellite imaging to make inferences about agriculture or energy use in small grids or to study environmental issues such as desertification or loss of wetlands over time; analysis of facial cues in either still or moving pictures of politicians; converting handwritten numerical data in historical archives to machine readable form.

Format:

The class will meet weekly for 2.5 hours, with 1.25 hours allocated to discussion of class readings and 1.25 hours allocated to project-based work. There will be an additional weekly "virtual office hour" wherein students and the instructor can share screens as well as an in-person collaborative office hour in which students can work in small groups with the instructor or a qualified teaching assistant present for assistance.

Students will be expected to read and annotate texts before class, and to ask and answer questions of and from other students before class, using an online learning platform. Classroom interactions will be facilitated with interactive learning software.

Each project submission will require the following:

Due in one week:

1. Written proposal to contextualize the project and to identify data sources, Python packages and project goals.

Due in three weeks: (two weeks after the proposal is due)

2. Results of the project: quantitative analysis
3. Results of the project: written assessment on the interpretation of the results
4. Source data or URLs to the source data
5. Source code

Readings:

Bird, Steven and Ewan Klein, Edward Loper *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit* Published by O'Reilly, 2009.

Lutz, Mark *Learning Python: 5th Edition* Published by O'Reilly, 2013.

McKinney, Wes *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython* Published by O'Reilly, 2012.

Solem, Jan Erik *Programming Computer Vision with Python* Published by O'Reilly, 2012.

Course materials: software

1. Python <https://www.python.org>
2. Python IDE e.g. Pycharm <https://www.jetbrains.com/pycharm/>; IPython <http://ipython.org>; or other development environment of the instructor's or student's choice.
3. GitHub/git <https://github.com>
4. Python packages could include but would not be limited to: NLTK (Natural Language Toolkit: <http://www.nltk.org>), NumPy (<http://www.numpy.org>), Matplotlib (<http://matplotlib.org>), pandas (Python Data Analysis Library <http://pandas.pydata.org>) and others.