Current Methods for Web-Based Data Collection and Analysis using Python

University of Bristol, December 13th - 15th, 2017

Delivered by Andrew Leone

(Professor and Warren C. Johnson Chair in Accounting, University of Miami)

COURSE DESCRIPTION

This three-day course is intended for PhD students and faculty seeking to update their knowledge of current methods for data collection and analysis. As more and more source data finds its way onto the Internet, the need for skills to efficiently extract and analyse these data increases. The course is aimed at accounting and finance researchers looking to analyze data found in SEC filings at the SEC Edgar site. Additional data sources to be covered include PDF documents and websites.

PRELIMINARY COURSE OUTLINE

- Extract data from SEC filings using Python.

- Extract basic data from SEC filings using Python Regular Expressions.

- Export extracted data to a database (MySQL) for later analysis.

- Apply algorithms from computational linguistics (e.g., readability and tone measures) to text using Python modules.

- Read data directly from popular statistics packages (e.g., Stata).

- Advance understanding of SQL (Structured Query Language). SQL is a powerful query language for working with relational databases (e.g., SAS, MySQL). The great advantage of SQL is that it can be used across software platforms. For example, SQL can be applied within SAS, Stata, Python, Excel, Microsoft Access, and MySQL. In this class, you will learn the fundamentals of SQL.

VENUE AND REGISTRATION

The course will be held at the University of Bristol, UK. Places are limited in order to facilitate questions and interaction. Participants need to bring their own laptop computers.

The cost is £450 for PhD students and £650 for faculty (excluding accommodation).

To register, or for more information, please contact Mark Clatworthy Mark.Clatworthy@bristol.ac.uk, cc Sharon Bryant (Sharon.Bryant@bristol.ac.uk).

The deadline for registration is Sunday November 19th, 2017.