

Project Year

2007

Project Team

Beverly Silver, Professor, Sociology Department, Krieger School of Arts & Sciences; Kevan Harris, Graduate Student, Sociology Department, Krieger School & Sciences

Project Title

Automating Data Collection for Research on Global Social Change

Audience

Students taking “research practicum” courses in Sociology, which are required of all majors and open to any undergraduate with the appropriate prerequisites.

Pedagogical Issue

The Sociology Department offers a series of research practicum courses designed to give undergraduates hands-on experience with all phases of a research project including the creation of new datasets. New dataset creation is important for many projects concerned with long-term, global social change. Historically, much data has been collected on a national basis, while global data on social indicators has been lacking. Given its labor-intensive nature, it has not been feasible to give students relevant direct experience with new dataset creation in a semester-long course. The recent digitization of the full text of several major newspapers creates new possibilities for the use of these sources in the study of long-term social change. However, the search engines currently available for use with the digitized newspaper sources are too inefficient to allow for the effective completion of new dataset creation in a single semester.

Solution

Develop a web-based application to allow students to search the online historic archive of the *Times* (London) and *New York Times*. The tool will support reiterative searches of the newspaper texts for a user-specified range of years and export results analysis into an Excel spreadsheet file. Students will use search results to analyze social processes over extended time periods (e.g., labor unrests across several decades and multiple countries).

Technologies Used

Graphic Design, HTML, Web Design, Excel

Project Abstract

The Sociology Department offers a series of research practicum courses designed to give undergraduates hands-on experience with all phases of a research project—including the creation of new datasets. New dataset creation is important for many projects, especially so for projects concerned with long-term, global social change. Historically, much data has been collected on a national basis, while global data on social indicators has been lacking. Given its labor-intensive nature, it has not been

feasible to give students relevant “hands-on” experience with new dataset creation in a semester-long course. The recent digitization of the full text of several major newspapers (dating back over a century) creates new possibilities for the use of these sources in the study of long-term social change. (Paper newspaper archives had already become a widely used source for the creation of new datasets by social scientists.) However, the search engines currently available for use with the digitized newspaper sources are too inefficient to allow for the effective completion of new dataset creation (much less preliminary data analysis) in a single semester. Preliminary work by the team has resulted in the development of working prototypes of programming scripts that automate the search process for the *Times* (London) and *New York Times*, allowing for reiterative searches of the newspaper texts for a user-specified range of years, and the dumping of the resulting count totals for each year into Excel spreadsheet files. With CER support, the Technology Fellow will perfect these scripts and add a user-friendly web interface, assuring their reliability for use in a classroom context. He will also write and test a script (and web interface) that allows students to automate the classification of a subset of selected articles (saved as PDF files) according to relevant sociological variables, as a prelude to data analysis.