Processing APSA eJobs with `muRL`

Ryan T. Moore[*]   Andrew Reeves[†]

July 20, 2009

   The `muRL` package can scrape job listings posted on the American Political Science Association (APSA)'s website (called eJobs) and import them directly into a spreadsheet. This allows for easy integration into form letters created by `muRL` or any other mailmerge protocol the user choses.
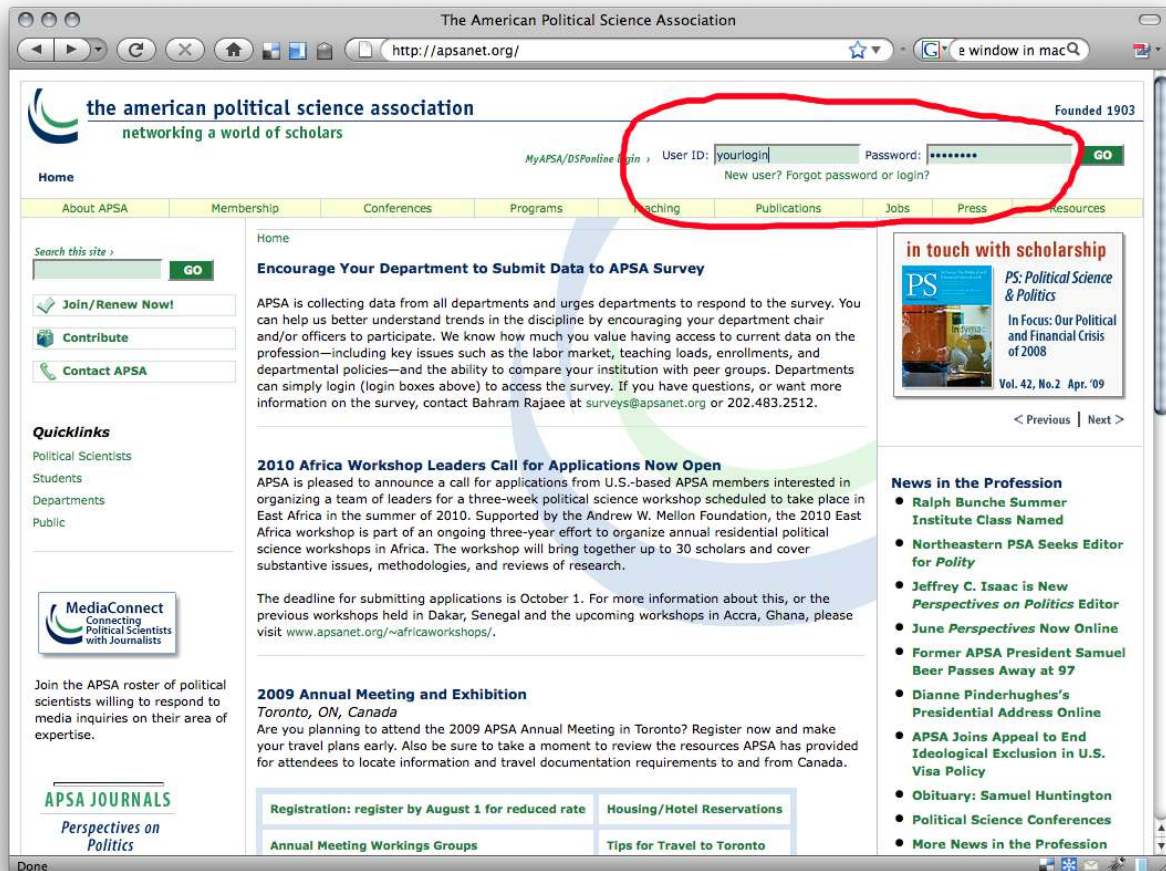
   This document describes how to import eJobs from the APSA website. The process involves downloading the html files for the jobs to which you plan to apply, opening `R` and running `apsahtml2csv`, a function in the `muRL` package.

---

   [*]Assistant Professor, Department of Political Science, Washington University in St. Lous, 241 Seigle Hall, Campus 1063, One Brookings Drive, St. Louis MO 63130; `http://rtm.wustl.edu`; `rtm@wustl.edu`; 314/935-3461.

   [†]Assistant Professor, Department of Political Science; 232 Bay State Road, Boston University, Boston MA 02215; `http://people.bu.edu/areeves`; `areeves@bu.edu`; 617/353-5284.

1 Downloading eJobs

1. Go to `http://apsanet.org` and login. You must be a member of APSA in order to access eJobs.

2. Go to the eJobs link and click on it.

3. At the eJobs interface, enter your search criteria. In the search below, we limit our search to assistant professor positions in Comparative Politics posted for all dates.

4. The search returns the following results. We click on the search results for Shepherd University.

5. We decide that we would like to apply for the job at Shepherd and so we will import the listing into the jobs spreadsheet.

6. Save the eJob as an html file. The file should be saved in a folder dedicated to all eJob listings for which the user wishes to import to the jobs spreadsheet. The user can name the file any unique name of their choosing. We recommend that the user maintain the .html file extension for each file.[1]
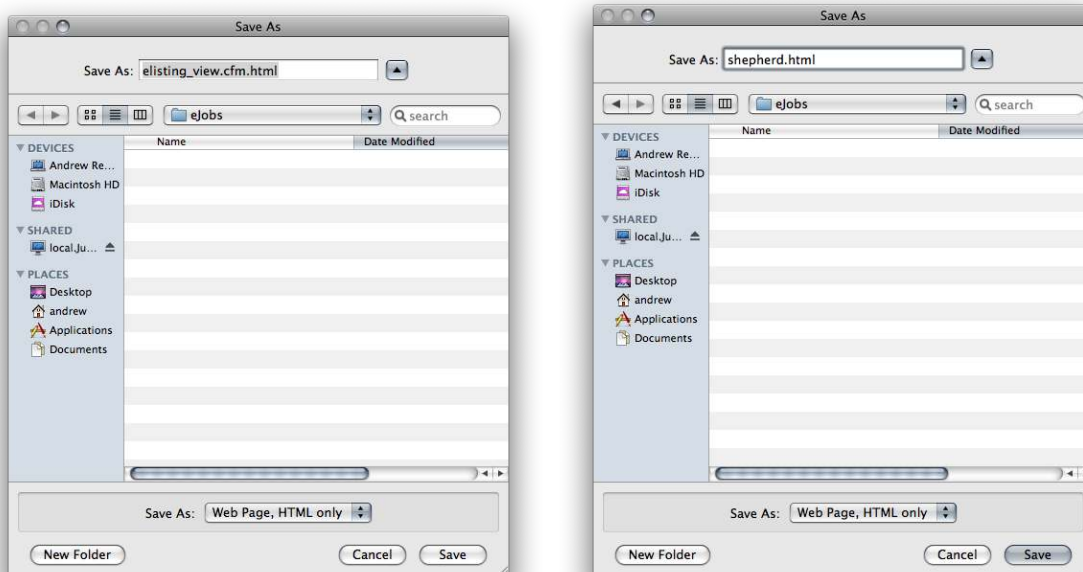
7. Repeat the previous step until you have downloaded all of the eJobs that you would like to import into the spreadsheet. Below is the folder containing all of the eJobs for which we plan to apply. Again, note that the user can specify any name although they should use a consistent file extension.

---
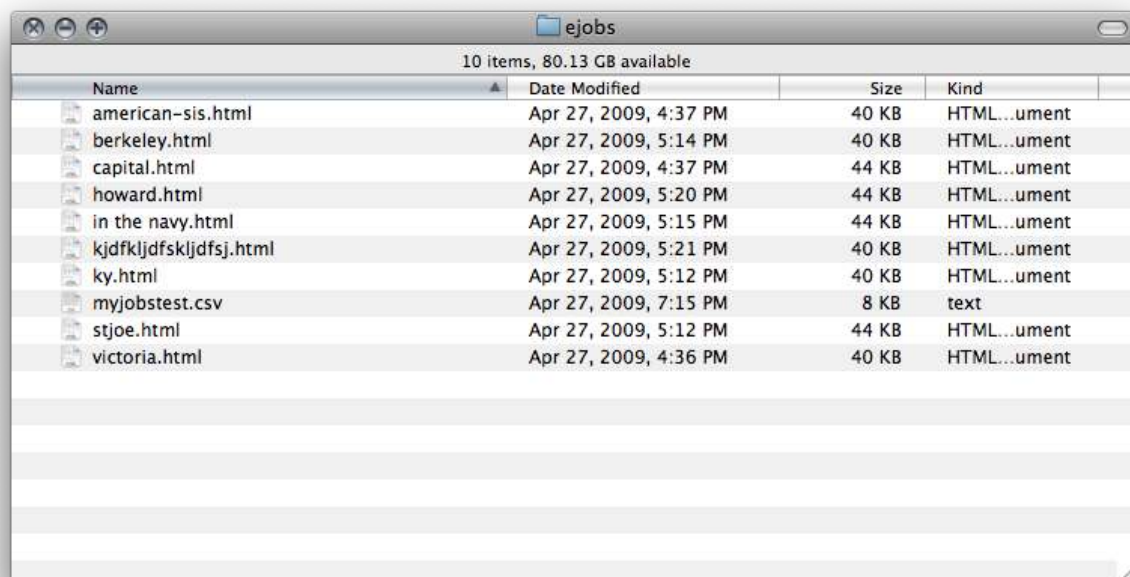
[1]In `apsahtml2csv`, the function searches for .html and .htm files by default in the (user specified) jobs folder. See the help file for `apsahtml2csv` for further information.

2 Importing eJobs into `muRL`

Once you have downloaded eJobs onto your local machine in a specified folder, you can then use `muRL` to scrap these html files and create a spreadsheet. Once you have loaded the `muRL` package in `R`, you can use the `apsahtml2csv` function to generate the spreadsheet. The function takes arguments for the location of the directory containing the saved eJob html files, the extensions of those files (*e.g.* .html or .htm) and the name and location of the spreadsheet (saved in comma seperated format) to be be generated from the files in the eJobs directory.[2]

The following code details the procedure for generating a spreadsheet called jobs.csv.

```
> library(murl)
> apsahtml2csv(directory = "eJobs", file.name = "jobs.csv", file.ext = ".html")
```

The jobs.csv file is a spreadsheet containing the details of each job position downloaded into the eJobs directory by the user. The fields include the APSA listing ID, date the job was posted, type of institution, title of the position, start date, salary, region, name of institution, contact person, address, phone, and description of the job. *Note that there is no standard method of entering the contact person, so the user must make the appropriate adjustments.* For example, the name of the contact person may be entered as *Dr. Mary Smith*, *Mary Smith*, *Mary J. Smith, PhD*, or *Professor. M. Smith*, . We recommend creating additional fields of *title*, *fname*, *lname* and manually filling these fields based on the information in the *contact* field.

3 Creating a mailmerge

Once the ejobs spreadsheet has been created using `apsahtml2csv`, we use `muRL` to create a TeX file. Again, we manually add the title (*title*), first name (*fname*), and last name (*lname*) fields based on the *contact* field scraped from ejobs. The following lines of code generate the TeX file which can be compiled to create the typeset letters.

```
> murl.ejobs <- read.murl("jobs.csv")
> write.murl(murl.ejobs)
Data stored as file 'mailmerge.tex'.
The current working directory is /Users/JohnnyF/Desktop
```

---

[2]See the help file for `apsahtml2csv` for more information.