

BibWrangleR

Brian E. Perron, Ph.D.

December 19, 2014

Overview

This document provides an overview of a suite of scripts that is designed to facilitate bibliometric analyses of a variety of databases that are not readily amenable to analyses due to the shape of the output. These scripts represent a major advance in bibliometric analyses, because it allows a wide range of journal article meta-data to be analyzed. The scripts *wrangle* the data into proper shape using the R programming language – hence, the name, **BibWrangleR**. Users who are not familiar with the R programming language can easily process the files, which can then be output into a standard *.csv format for analysis in statistical packages like Stata, SAS, and SPSS.

The scripts support the wrangling of three different databases: PubMed, PsychInfo (via EbscoHost), and Citation Reports from Web of Science. In addition to wrangling the output of search results into a shape for easy analysis, functionality is also provided to link the results across multiple databases. Thus, a highly refined search can be performed in PubMed or PsychInfo, and those search results can then be linked to the Citation Reports from Web of Science.

To functionality of the scripts will be demonstrated with a simple research question: To what extent are new or translated measures published in social work journals subsequently used in research? Because this question is an example to show the functionality, this example will focus on all the measures from a single journal, *Research on Social Work Practice*. However, keep in mind that the user can easily scale the search to include as many journals that are indexed by the database being analyzed.

Step 1. Data Collection

The actual data collection involves the use of databases, so the BibWrangleR functions are used to process the search research, not collect any data. It is assumed that the reader is skilled at performing database queries, preferably using advanced search functions.

Obtain Data from PsychInfo (EbscoHost)

The first step of the process is to perform a search on the database of interest. For this example, PsychoInfo is used as the target database. This journal has been indexed in PsychInfo since the journal was initially launched in 1990, so the search within this journal will be comprehensive. A simple EbscoHost query can be performed by specifying the journal using the source (SO) operator, followed by a series of other search terms with Boolean operators. It is assumed the reader knows how to perform advanced search queries to narrow results and, therefore, will not be covered here.

After the search is performed in EbscoHost, the search results need to be exported. This is done by selecting the *Share* button, and then *Export results*. The *Export results* gives the user a number of different formatting options for the search results. The required for PsychInfo results (with the EbscoHost system) to be processed by BibWrangleR must be the *Generic bibliographic management format*. The search results for this example produced 176 unique articles, and the format of the output are a standard text (*.txt) file:

```

ebsco.txt
TI- Psychometric properties of a Korean measure of person-directed care in nursing homes.
PS- Publication Status: Fully Published
AU- Choi, Jae-Sung
AU- Lee, Minhong
AF- Choi, Jae-Sung, Department of Social Welfare, Yonsei University, Seoul, South Korea
AF- Lee, Minhong, minhongsw@gmail.com, Department of Social Welfare, Dong-Eui University,
    Busan, South Korea
AD- Lee, Minhong, Department of Social Welfare, Dong-Eui University, 176 Eomgwang Rd.,
    Busan Jin-Gu, Busan, South Korea, 614-714, minhongsw@gmail.com
S0- Research on Social Work Practice
VI- 24
IP- 6
SD- Nov, 2014
YR- 2014
PM- Nov, 2014
PG- 676-684
PC- 9
SP- 676
PU- US : Sage Publications
PN- 9
SN- 1049-7315, Print
SN- 1552-7581, Electronic
D0- 10.1177/1049731513509897
LA- English
KP- person-directed care
KP- psychometric properties
KP- nursing home
KP- assessment tool
KP- South Korea
SU- Measurement; Nursing Homes; Psychometrics; Test Reliability; Test Validity

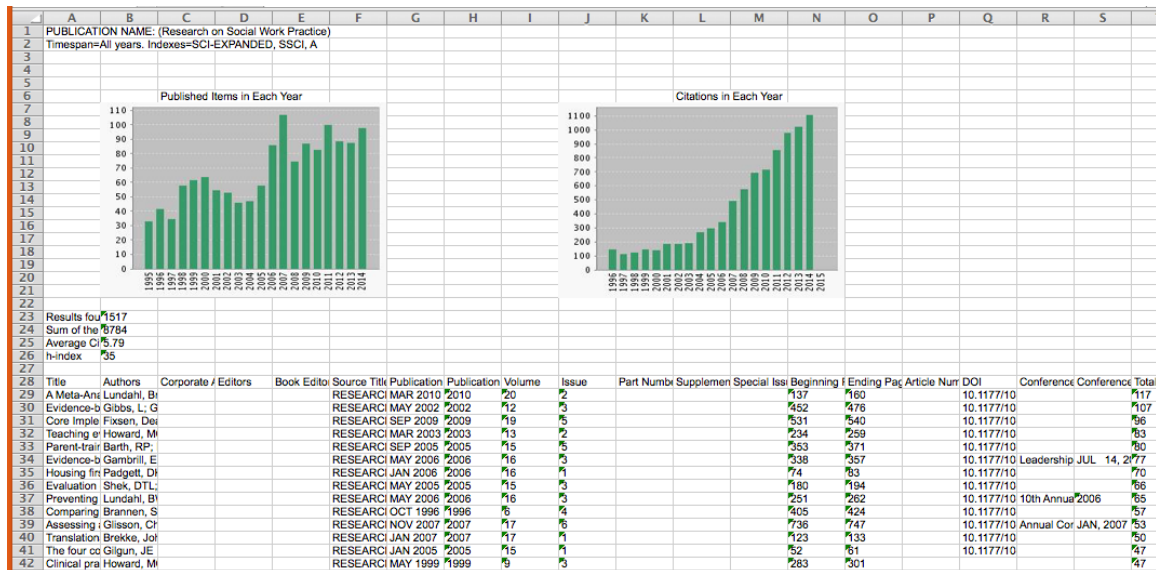
```

The reader will need to save this *txt* file in a folder that does not contain any other *.txt* files.

Obtain Citation Reports (Web of Science)

Citation reports can easily be produced from the Web of Science. For this example, the entire history of *Research on Social Work Practice* is obtained and citation reports are processed. The citations reports should be exported as **.xls* files. This particular query resulted in 1,573 articles, but the web interface limits the user to processing only 500 reports at a time. Thus, four separate reports are exported, and all four files are saved in the same folder as the text file produced from EbscoHost.

Provided below is a snapshot of the file to be processed. It should be apparent that this file is not amenable for analysis for common statistical software packages.



Step 2: Wrangle the Data

The process of getting the data into shape for analysis – i.e., *wrangling* – is straightforward with the the `BibWrangleR`. It is assumed that the reader has installed the latest version of the R. The first step involves loading `BibWrangler`. To do, simply specify the path for the file that contains the `BibWrangleR` function. The `source` function reads the file called `BibWrangleR.R`, and the remainder of the path specifies the exact location of the file.

```
source("/Users/beperron/Git/BibWrangleR/Functions/BibWrangleR.R")
```

Then, the user needs to specify the path of the folder that contains the data files. This can be done using the dropdown menu in R or programmatically with `setwd` function.

```
setwd("/Users/beperron/Git/BibWrangleR/Files2Process")
```

Now, the user processes the data files by calling the `BibWrangleR` function. The user has the option of producing CSV files by changing the `csv` argument to `TRUE`.

```
#RSWP.measures <- BibWrangleR.f(csv = FALSE)
```