Research on American News

Working hours: 1 or 2 hours per day

Current Goal: acquire at least one US news source dated back to a relatively long time ago(2000s)

Research Journal

01/22/2019:

Talked to professor Molly, started the research, set the goal, set up time track

Write email to Annelise Sklar to ask about database

Possible alternative, find a jpg archive, and use text recognizer to convert the newspaper image into plain text – Don’t know if that would take too long…

Event Registry – dating only back to 2014

01/24/2019

* Receive Email from Annelise Sklar, investigate 3 databases that she mentioned:
  + (Note: consistent article number is an evidence to show that the collection of certain newspaper source is comprehensive)
  + Source1: NewsBank:
    - Major Newspaper search result:
      * New York Daily News
        + 1995-2017
        + consistent number
      * USA Today
        + 1987-2019, peaked around 1988-1990 with 100,000~130,000 articles, then all other years are around 30,000~80,000, might suggest incomprehensive collection
      * Dallas Morning News
        + 1984-2016
        + Consistent journal number
      * The Seattle Times
        + 1985-2019
        + Consistent number
      * The Denver Post
        + 1989-2019
        + Consistent number
      * NewsDay
        + 2007-2019
        + Consistent number
    - Data Download Process:
      * No convenient batch download method
      * Manual download:
        + Can only select 20 article at a time
        + Download format is pdf
      * Possible automated process
        + Easy to automate(auto selecting “select all article” box and then download)
  + Source 2: Nexis Uni
    - * Only selecting 20 article at a time
  + Source 3: Factiva
    - * Selecting 100 articles at a time

01/28/2019

* Reply email and acquire further information
* Read other papers to find out what are their way of acquiring the data
  + Not necessarily need the full body of text? Maybe it depends on how we want to look at the text?
* New resources: Newslibrary, Alliance of Audited Media

02/05/2019

Send email to Prof.Roberts

Reply Annelise

2/11/2019

Resources:

We do indeed have databases with full text of news articles in plain text. The three big ones are:

[Nexis Uni](http://uclibs.org/PID/20081)

Thousands of full-text sources including newspapers, wires, transcripts, newsletters, magazines, trade journals; also industry news, company information, law reviews, court decisions. Coverage dates vary. Former name: Academic Universe.

[Access World News](http://uclibs.org/PID/98452)

Dates vary. More than 1500 sources worldwide, including 650 in the U.S. Largely newspapers,local news video for the U.S.; newspapers for Canada, UK and other English-speaking countries. For the rest of the world mostly wire services, BBC news digests.

[Factiva](http://uclibs.org/PID/32497)

Factiva.com provides access to over 8,000 sources including: \* 1,000+ Global and local newspapers \* 3,000+ Journals and magazines \* 30,000+ Company Reports Factiva.com delivers content in 22 languages.

Write email to Professor Molly,

Factiva – Batch download of 100 articles per time,

|  |  |  |
| --- | --- | --- |
| Newspaper Title | # articles related to China | Accessible time |
| USA Today | 20974 | 1987-2019 |
| The New York Daily News | 8161 | 1993-2019 |
| The New York Times | 115188 | 1980-2019 |
| The Wall Street Journal | 90738 | 1979-2019 |
| The Washington Post | 71071 | 1984-2019 |
| New York Post | 8530 | 1997-2019 |
| New York Daily News | 8161 | 1993-2019 |
| The Boston Globe | 30156 | 1987-2019 |
| Tampa Bay Times | 22960 | 1986-2019 |