

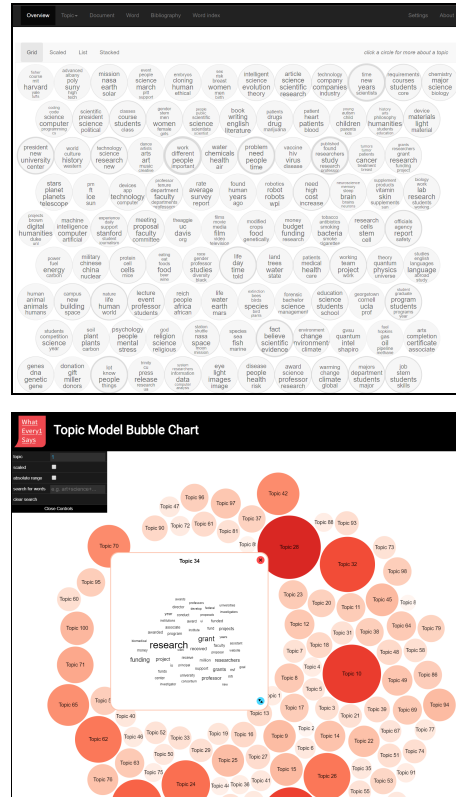
Collection 33: Articles classified as being about the humanities or science, 1998-2018

A collection of word-frequency and other data representing 13,214 unique articles (no duplicate or close-variant documents) classified as being about the humanities or science published from 1998-2018 in 507 U.S. top-circulating and student newspapers and their associated blogs. The collection includes 2,477 articles from U.S. top-circulating newspapers and 10,737 articles from student newspapers. Using supervised classification models, 2,869 articles in the collection have been classified as being about the humanities, and 10,345 articles have been classified as being about science.

News sources in Collection 33 include 15 top-circulating U.S. newspapers: *Boston Globe*, *Chicago Tribune*, *Daily News* (New York), *Dallas Morning News*, *Denver Post*, *Houston Chronicle*, *Los Angeles Times*, *New York Post*, *New York Times* (and its blogs), *Newsday* (New York), *Seattle Times*, *Star Tribune* (Minneapolis, MN), *Tampa Bay Times*, *USA Today*, *Washington Post*. Also included are documents from 491 U.S. campus newspapers, among which the top 10 sources in the collection are: *The Stanford Daily* (Stanford University), *The California Aggie* (UC Davis), *The Daily Californian* (UC Berkeley), *The Daily Bruin* (UC Los Angeles), *The Kaleidoscope* (U Alabama Birmingham), *The Tartan* (Carnegie Mellon), *Michigan Independent* (UM Ann Arbor), *The Daily Princetonian* (Princeton), *The Harvard Crimson* (Harvard), *The Dartmouth* (Dartmouth).

Sources in Collection 33 are associated with non-exclusive metadata categories that describe the nature of sources in the collection based on their explicit publication information and/or self-identification--e.g., geographical region, type of educational

institution, demographic or social group. (For the source categories of Collection 33 listed in order from those associated with the most documents to those associated with the least, see the detailed information on the collection's [start page](#).)



One of the topic models of this collection (100 topics visualized in two of the visualization interfaces in the WE1S Topic Model Observatory: [Dfr-browser](#) & [TopicBubbles](#))

WE1S Collection Registry ID:

220200515_1455_us-classification-results-top-newspapers-universitywire-hum-sci

Data source: LexisNexis (via LN Web Services Kit)

Collection deposit (term frequencies, Mallet outputs, visualization files, and other non-consumptive data):

GitHub | Zenodo (DOI: TBD)

Topic models available (# of topics): 25, 50, 100, 150, 200, 250

Topic model visualizations: [Start page](#)