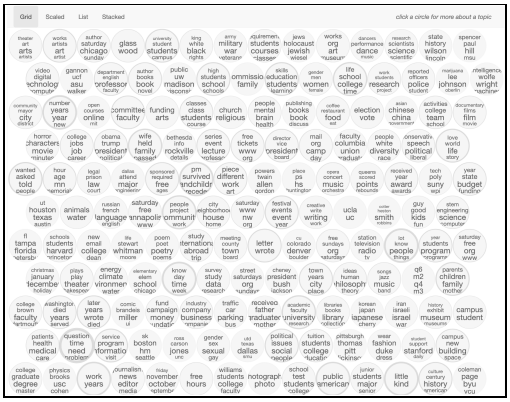
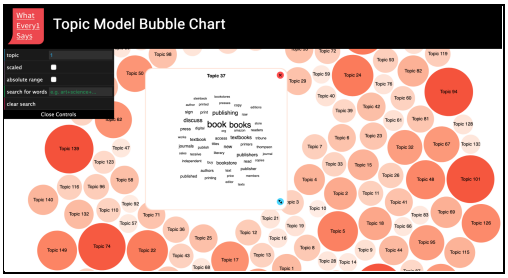
Collection 36: Articles containing the word “humanities” but that have been classified as not being about the humanities, 1998-2018

A collection of word-frequency and other data representing 27,362 unique articles (no duplicate or close-variant documents) that contain the word "humanities" but that have not been classified as being about the humanities published from 1998-2018 in 545 U.S. top-circulating and student newspapers and their associated blogs. The collection includes 13,309 articles from U.S. top-circulating newspapers and 14,053 articles from student newspapers. Supervised classification models have classified these articles as not being about the humanities; this collection therefore helps WE1S understand what articles that contain the word "humanities" but that aren't about the humanities per se are like.

News sources in Collection 36 include 15 top-circulation 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 530 U.S. campus newspapers, among which the top 10 sources in the collection are: *The Stanford Daily* (Stanford University), *The Tartan* (Carnegie Mellon), *The Dartmouth* (Dartmouth), *The Harvard Crimson* (Harvard), *The Daily Princetonian* (Princeton), *The Brown Daily Herald* (Brown), *The Daily Cardinal* (UW Madison), *The Daily Titan* (California State University, Fullerton), *The Daily Bruin* (UC Los Angeles), *The Daily Universe* (Brigham Young University).

Sources in Collection 36 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 (150 topics visualized in two of the visualization interfaces in the WE1S Topic Model Observatory: Dfr-browser & TopicBubbles)

WE1S Collection Registry ID: 20200522_1900_us-humanities-classification-results-top-newspapers-universitywire-not-hum

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

This card created 31 May 2021 (Lindsay Thomas)