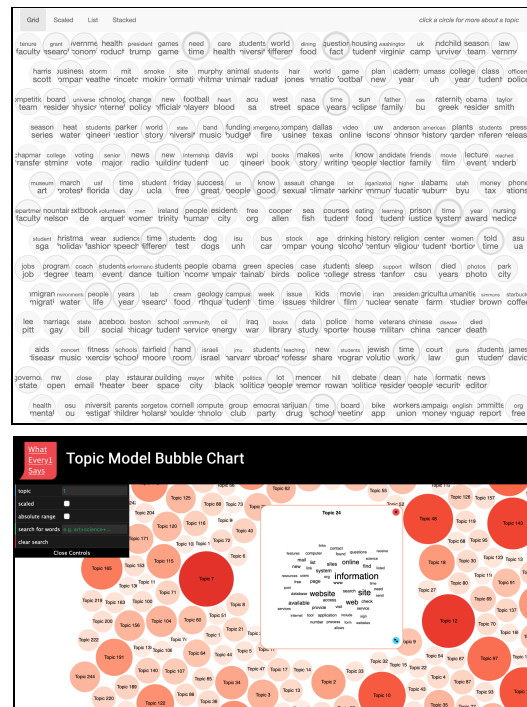# Collection 37: Articles containing the words "science" or "sciences" but that have been classified as not being about science, 1998-2018

A collection of word-frequency and other data representing 87,278 unique articles (no duplicate or close-variant documents) that contain the words "science" or "sciences" but that have not been classified as being about science published from 1998-2018 in 610 U.S. top-circulating and student newspapers and their associated blogs. The collection includes 13,628 articles from U.S. top-circulating newspapers and 73,650 articles from student newspapers. Supervised classification models have classified these articles as not being about science; this collection therefore helps WE1S understand what articles that contain the words "science" or "sciences" but that aren't about science per se are like.

News sources in Collection 37 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 599 U.S. campus newspapers, among which the top 10 sources in the collection are: *The Stanford Daily* (Stanford University), *The Marquette Tribune* (Marquette University), *The Daily Bruin* (UC Los Angeles), *The California Aggie* (UC Davis), *The Daily Cardinal* (UW Madison), *The Badger Herald* (UW Madison), *Cornell Daily Sun* (Cornell University), *Daily Californian* (UC Berkeley), *The Harvard Crimson* (Harvard), *The Daily Universe* (Brigham Young University).

Sources in Collection 37 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 37 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**: 20200607_1403_us-science-classification-results-top-newspapers-universitywire-not-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

---