

Collection 21: U.S. Top Newspapers, 2000-2018 (articles mentioning humanities or science)

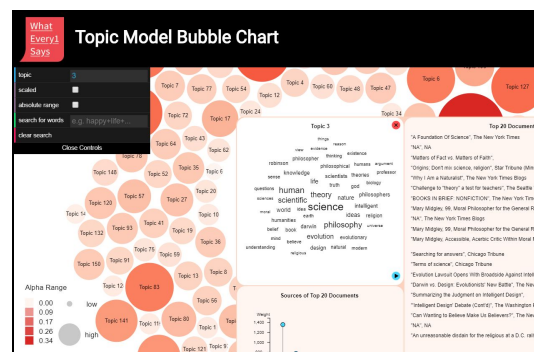
A collection of word-frequency and other data representing 28,957 unique articles mentioning “humanities” or “science(s)” (no duplicates or close variants) published during 2000-2018 in 15 top U.S. newspapers and their associated online blogs.* WE1S and other researchers use this data to look for broad patterns and help guide closer study.

Collection 21 contains data representing all 15,692 articles from its set of sources in these years mentioning “humanities” but only a sampling of the 388,691 articles mentioning “science” or “sciences” from those same sources and years. It downsamples “science(s)” articles (while maintaining the proportions of articles from particular sources and years) to achieve a 50/50 balance of articles related to the humanities and sciences. The purpose is to allow media discourse on the humanities to be studied alongside that on the sciences and not be buried so far down in the statistical pile that it cannot easily be seen in detail. Collection 21 is thus not a representation of the relative weight of discussion of the humanities and sciences but instead an aid to studying the fine features and structures of each.

News sources in Collection 21 include: *Boston Globe*, *Chicago Tribune*, *Daily News* (New York), *Dallas Morning News*, *Denver Post*, *Houston Chronicle*, *Los Angeles Times*, *New York Post*, *New York Times*

* WE1S makes available only “non-consumptive use” word frequency, topic model, and other datasets along with their visualizations. Datasets cannot be used to access, read, or reconstruct the original texts.

(and its blogs), *Newsday* (New York), *Seattle Times*, *Star Tribune* (Minneapolis, MN), *Tampa Bay Times*, *USA Today*, *Washington Post*.



Topic model of this collection (150 topics visualized in two of the visualization interfaces in the WE1S [Topic Model Observatory](#): [Dfr-browser](#) & [TopicBubbles](#))

Further Information

WE1S Collection Registry ID:

20190705_1645_us-humanities-sciences-top-newspapers

Data source: LexisNexis (via LN Web Services Kit)

Collection dataset*: GitHub

Mallet topic model data files: GitHub

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

Topic model visualizations: [Start page](#)