(articles mentioning "humanities" or "the arts")

A collection of word-frequency and other data representing 108,207 unique articles mentioning "humanities" or "the arts" (no duplicate or close-variant documents) published mostly during 1989-2019 in 1,170 U.S. news sources and their associated blogs. (5,308 articles originate from earlier years going back to 1977.)¹ The United Kingdom and Commonwealth nations often use the phrase “the arts” to refer to what the U.S. more commonly calls “the humanities and arts.”²

The top 20 news sources (ordered by number of articles in descending order) are: New York Times, Washington Post, Deseret Morning News (Salt Lake City), Houston Chronicle, Arkansas Democrat-Gazette (Little Rock), Los Angeles Times, Chronicle of Higher Education, Chicago Tribune, Newsday (New York), Houston Chronicle, Dallas Morning News, St. Louis Post-Dispatch (Missouri), Federal News Service, Arkansas Democrat-Gazette, Advocate (Baton Rouge, Louisiana), Star-Ledger (Newark, New Jersey), Boston Globe, Stanford Daily (Stanford University), Oregonian (Portland, Oregon), Hartford Courant (Connecticut). (Full sources and counts are available as a csv file in the Collection Dataset.)

In this collection, the word “humanities” occurs 136,909 times in 82,899 documents; and the phrase “the arts” 82,209 times in 42,678 documents.

¹ WE1S and other researchers use this data to look for broad patterns and to help guide closer study. 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.

² See the WE1S report by Alan Liu on “How Public Media in the US and UK Compare in Their Terminology For the Humanities.”

Further Information

WE1S Collection Registry ID: 20190621_1925_us-humanities-arts-all-no-reddit/
Data source: LexisNexis (via LN Web Services Kit)
Collection dataset: GitHub TBD
Mallet topic model data files: GitHub TBD
Topic models available (# of topics): 25, 50, 100, 150, 200, 250
Topic model visualizations: Start page