Collection 15: Articles mentioning “humanities” or “literature” from ProQuest’s Ethnic NewsWatch and GenderWatch

A collection of word-frequency and other data representing 835 unique articles mentioning “humanities” or “literature” (no duplicate or close-variant documents) published mostly during 2016, 2018, and 2019 in 109 U.S. news sources gathered in ProQuest’s Ethnic NewsWatch (“ethnic and minority press”) and GenderWatch (sources gathered for “gender and women's studies, and gay, lesbian, bisexual, and transgender [GLBT] research”). WE1S and other researchers use this data to look for broad patterns and to help guide closer study.


Because the number of documents in Collection 15 is relatively low, WE1S created topic models for it with smaller numbers of topics (10, 25, 50, 75, and 100).

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.

Further Information

WE1S Collection Registry ID: 20190627_1756_us-proquest-minus-wsj
Data source: ProQuest
Collection dataset*: GitHub
Mallet topic model data files: GitHub
Topic models available (# of topics): 10, 25, 50, 75, 100
Topic model visualizations: Start page

This card created 21 Feb., 2020; rev. 21 Feb. 2020 (Alan Liu)