Collection 38: Reddit on the Humanities
(Reddits mentioning "humanities," "liberal arts," or "the arts" from 2006-2019)

WE1S's Collection 38 (C-38) includes 124,340 Reddit comments longer than 225 words from 2006 to 2019 containing the terms *humanities*, *liberal arts*, or *the arts*.

Explanation: WE1S's rationale and methodology for collecting Reddit posts to study public discourse about the humanities (especially by students) is explained in the blog post by WE1S's lead Reddit researcher Raymond Steding: ["A Digital Humanities Study of Reddit Student Discourse about the Humanities"](#) (2019).

As Steding notes, he initially collected 3.3 terabytes of Reddit data from 2006 to 2018 (approximately five billion comments) by downloading it from [pushshift.io](#) in [JSON format](#).[1] Data for 2019 was later added.

This data was filtered to retain only comments containing at least one of the terms *humanities*, *liberal arts*, or *the arts*. Then comments under 225 words were subtracted to improve the coherence of WE1S's topic models of the collection.
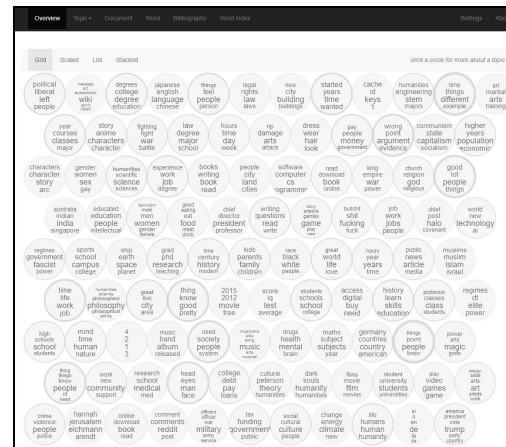
WE1S did not weed out duplicates in this collection; nor filter posts and comments based on Reddit user "karma" scores.[2]

Related collections:

- [**Corpus-A**](#): subset of C-38 in which Reddit comments are reduced to those also including one or more of the terms *student*, *major*, or *college* (designed to focus on student discussion of the humanities).

- [**C-39**](#): Version of Corpus-A with addition of Reddit comments from 2019.

Topic model of this collection ([C-38.100](#), 100 topics) visualized in Dfr-browser

"That doesn't even make sense and it's your (uneducated) opinion, not a fact. A degree in anything doesn't guarantee you a job in anything - you do realize that the majority of people don't work in careers directly related to their majors, right? But don't take my word for it: [[link](#)] Oh and in case you're wondering about the facts, According to the 2011 numbers, the median starting salary for philosophy BAs is $39,800 and the median mid-career salary is $75.600. This puts it at #2 among the humanities majors…."

Example of a Reddit from this collection ([full text](#))

Further Information

WE1S Collection Registry ID: 20200716_2221_reddit-all-3-humanities-keyword-search-terms

Data Source: [Reddit](#) (via data harvested at [pushshift.io](#))

Collection dataset: [TBD]

Mallet topic model data files: [TBD]

Topic models available (# of topics): 100, 200

Topic model visualizations: [Start page](#)

---

[1] Pushshift is a "big-data storage and analytics project" providing APIs for searching and downloading Reddit and also downloadable raw data (comments) from Reddit. (See [Pushshift FAQ](#).)

[2] "When posts or comments get upvoted, that user gains some karma" (Reddit, ["What is Karma?"](#))