



Research Exploration Group

Preamble

If you're a student or researcher in the humanities, this group is likely the one for you! We need you to try out our search engine on your own project. The goal here is twofold: 1) to help you identify books that are relevant to your own research; and 2) to develop compelling use cases, so that, when we take our work to the press, we can demonstrate what this search engine is capable of!

Main Goals

Unlike the other groups, your goal is to focus on what interests you most, your own research. Because our search engine is so slow it currently takes about an hour to perform a search. For that reason, it's important to e-mail Stephen Osadetz (osadetz@fas.harvard.edu) with search terms as far in advance as possible of the event. Of course, if you would rather not have your own data, you're welcome to work with other people's, or to use the sample search results that are available in our Github repository.

Other goals / instructions (in order of priority)

1. Consult with other groups about user experience, website design, useful search filters, branding, and promotional materials.
2. Develop a standard set of evaluations for passages, so you can keep track as you work through the data
3. Develop methods for determining the quality of a particular search
4. Develop new ways to use the data (compare two concepts at once; sort and filter results to judge what is most useful; combine search terms into smaller "concepts"; introduce new terms to arrive at even more precise formulations of your concept, etc.)

Performing your own search in advance

If you want to try out visualizing data relative to a project that you are working on, you may want to have us perform a search on your behalf before the hackathon. Stephen Osadetz will complete these in advance and bring your output files on a thumb drive. To have Stephen perform a search on your behalf, send him an e-mail (osadetz@fas.harvard.edu) with a few sentences describing your concept and a list of 10-20 words that you believe would cluster around that concept. For instance, a search for the concept *dog* might include such words as *dog*, *canine*, *kibble*, *tail*, etc. A few tips: we can't search for phrases, so restrict your request to single words; the search algorithm involves stemming words, so there's no need to include variants of a single words like *dog*, *dogs*, etc. The search engine treats these as the same word. Try to choose words that won't result in the same stem. Please keep in mind that at this point we are only able to search 18th-century English-language texts. At this point, search is slow – it takes one hour to do a single search – so it is unlikely we will be able to do a search for you on the day of the event.

Accessing Passages on Amazon Web Services (AWS)

We will send out instructions for how to do this in a separate e-mail, when we give you access to AWS. Each passage is given a standard numerical designator, which you can use to access its full text.

GitHub Contents (may be added to before and during the event):

- Sample search results (to be used if you choose not to do your own search in advance)
- Sample passages
- Smaller snippet of results set
- Contents of standard search filters for development
- List of potential search filters that you could develop with the Corpus_

Development

- Sample search filter for "literary and canonical authors" (we'll apply this to your results, if you submit terms for a search)
- Group Notebook
- Code Uploads

Other groups to consult

- **Website Development** – you can advise them on what you're like the website to do
- **Corpus Enrichment** (this group will be enriching the ECCO metadata, which may have the ability to improve your visualizations)
- **Design and Branding**