CS418/518 Project – Milestone 2 – Project B (ETD Digital Library)

In this milestone, you will build the search function in your search engine using Elasticsearch.

Base requirements

1. The website should index all documents in the ETD500 repository. The JSON object should look like below

   {“etdid”:1,

   “title”:”some aspects of radiation induced nucleation in water”,

   “author”:” tso, chih-ping”,

   “year”:1970,

   “university”:” massachusetts institute of technology”,

   “program”:”nuclear engineering”,

   ”degree”:”m.s.”,

   ”advisor”:”neil e. todreas”,

   ”abstract”:”the content of abstract… ”,

   ”pdf”:”1.pdf”,

   ”wikifier_terms”:[{“term”:”nucleation”,”url”:”https://en.wikipedia.org/wiki/Nucleation” }, {“term”:”radiation”,”url”:”https://en.wikipedia.org/wiki/Radiation”}]}
2. Users can query the search engine without logging in.
3. The search engine accepts a text query in the search box and return results on the search engine result page (SERP).
4. Each item in SERP should link to a page for a document.
5. The search engine should display the number of returned items on top of search results.
6. The search engine should show the actual keywords after filtering and a search box on the top.
7. The search results should be paginated.
8. The query terms should be highlighted in the search results.
9. Use Wikifier to obtain Wikipedia terms appearing in ETD abstracts and store the terms and the Wikipedia URLs in the JSON object above.
10. When a user clicks a search result, he should see a summary page showing all metadata fields (including the abstract) of an ETD.
11. A regular user should be able to insert a new document, including key metadata (title, author, year, advisor, university, degree, program) and upload the PDF.
12. Search the title of the new paper and it should show up in the search results.

Submission

All source codes and should be committed to GitHub by noon, Thursday, November 10, 2022. The report should be submitted to Canvas by November 15, 2022.