Shraddha Kulkarni (6117955457)
Bharath Chandra Thota (7232807134)

MIDTERM PROJECT REPORT

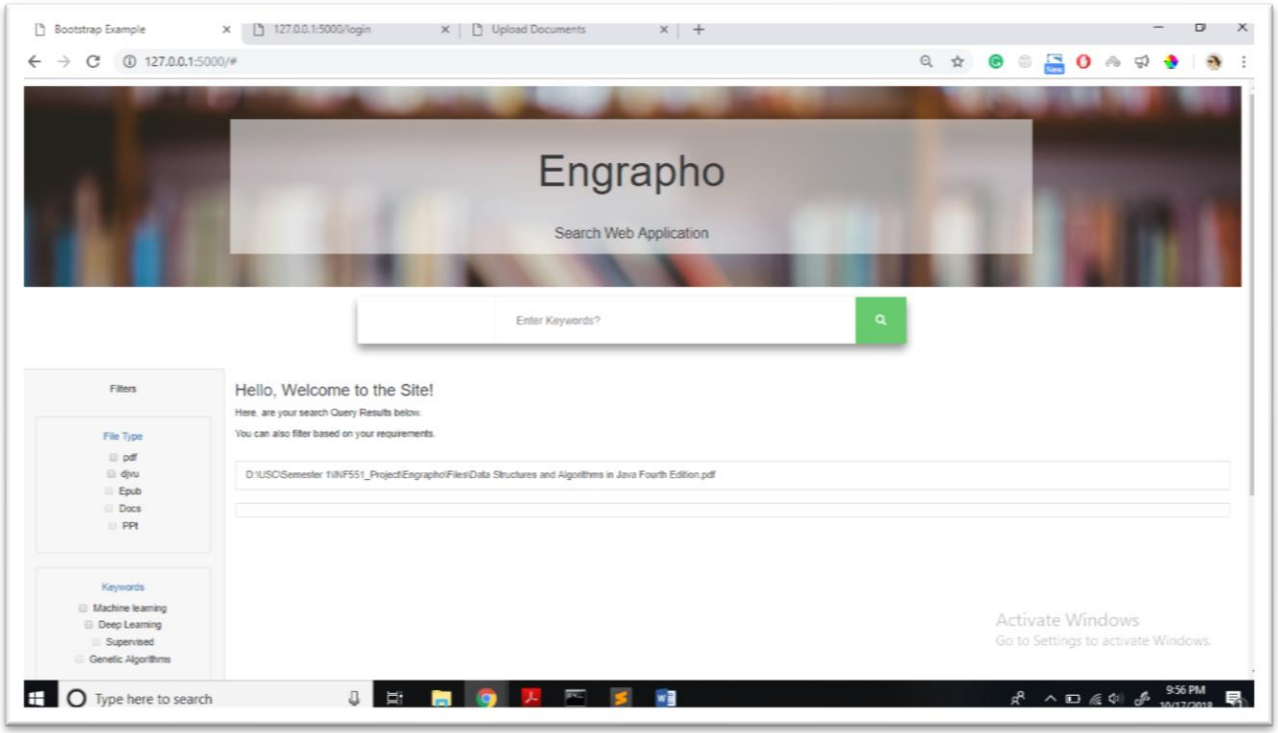
Basic Idea: Engrapho is basically a Web Application, that lets the user search on any topic, and as the search Results all kinds of documents (word, ppt, excel, pdf) which are related to that search term will be presented. Additionally, the user interface will consist of a Filter section (Faceted Search) for the most appropriate results.

Completed Task:

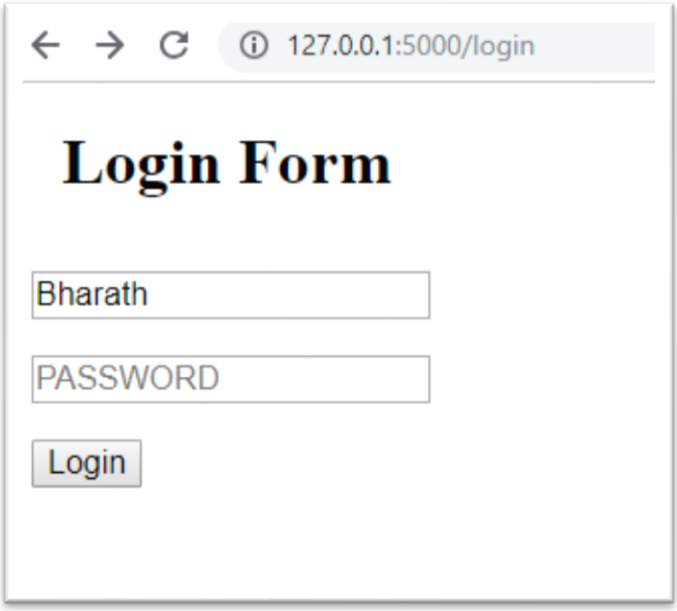
1. Documents collected and stored. (Sources: Google Scholar, SlideShare, Springer, LinkedIn)
2. Created the Flask Pipeline for document collection and Meta-Data Extraction.
3. Created the inverted Index for Keywords creation.
4. Created an UI where all the filters and results can be plugged in.

Current Progress Screenshots:

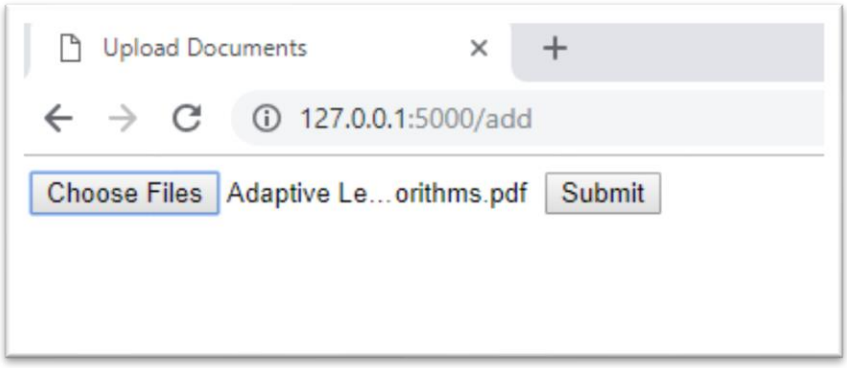
- User Interface of the Web Application



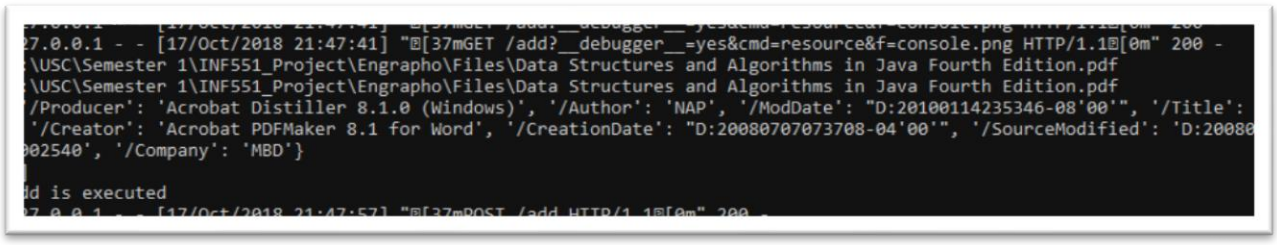
- Login to the Portal to upload Documents manually. (This UI is specifically created, so that its' easier to add new entries in the future)



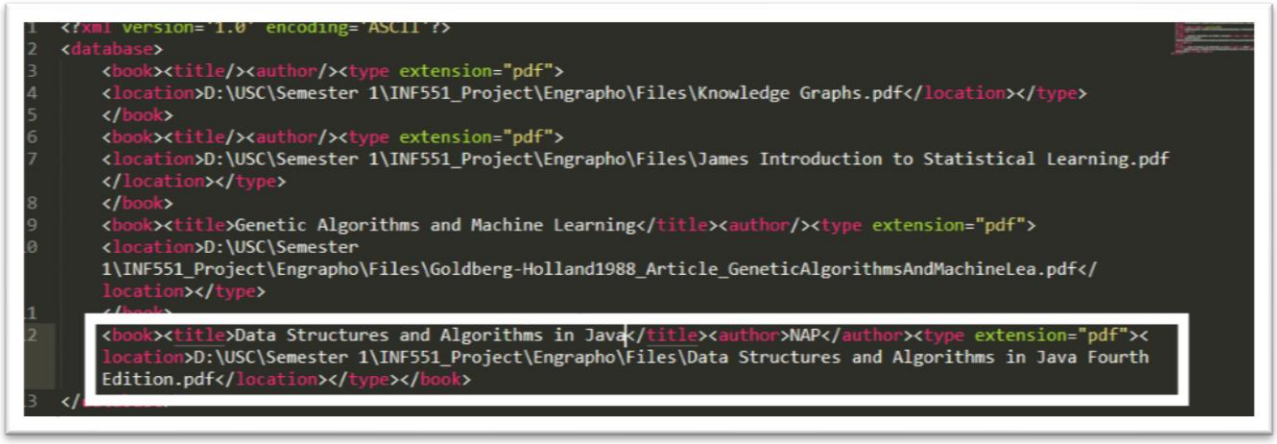
- Just Click and upload your Documents. Meta-data is automatically extracted and saved in the database.



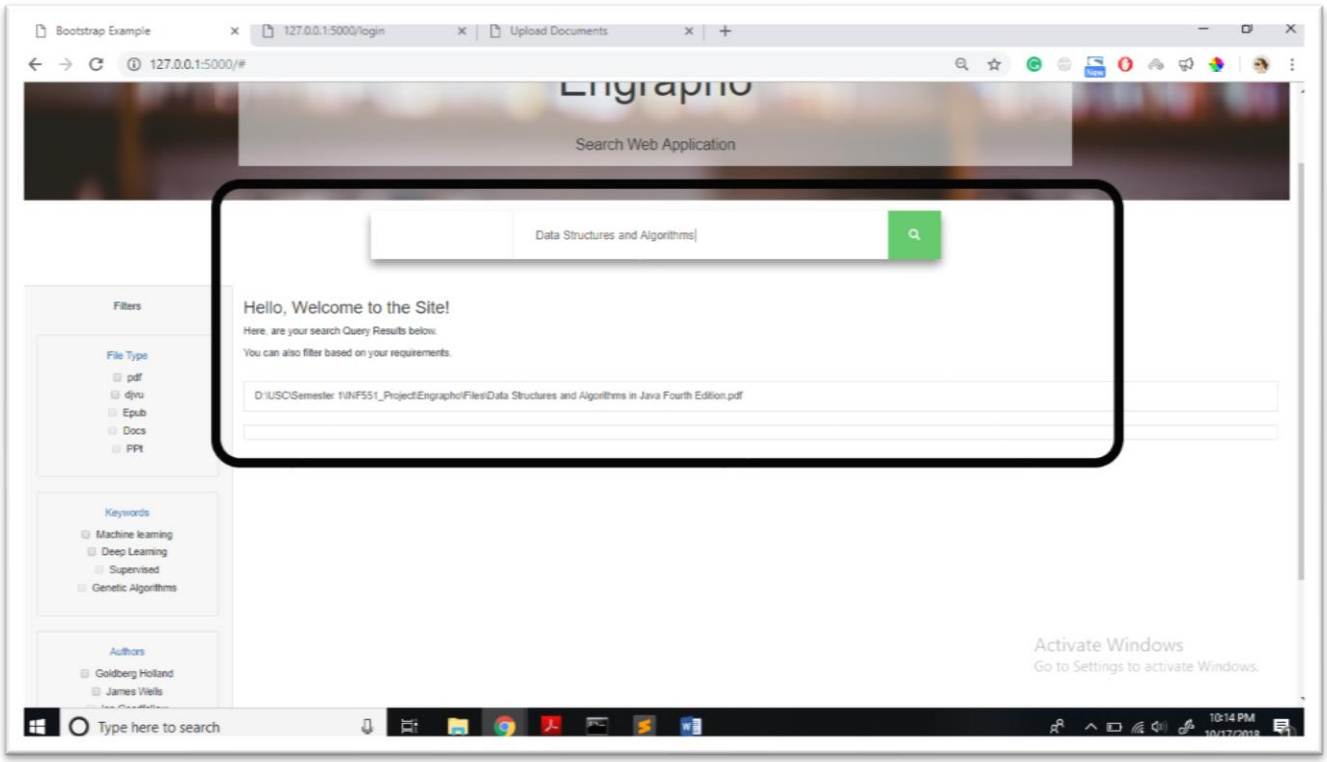
- Background processes running. As you can see the metadata is passed once the document is uploaded.



- Information of the Documents such as Title, Author, Type of File, Location saved in the XML format.



- Search the Keyword and Results will be shown in the bottom.



Shraddha Kulkarni (6117955457)
Bharath Chandra Thota (7232807134)

To be Completed after Mid-Term:

- Plug the Keywords and meta-data tags as Filter Options in the UI.
- Change the Upload pattern, instead of Locally we need to set up the app to upload the Documents to Cloud.
- Connect MongoDB with the Web App. Basic DB Schema is already created, only need to bridge the gap.
- Upload all the code to EC2 Server, so it can run on Cloud.