

INFO-533 Project 3 (100 points)

Deadline: May 01, 2024 11:59 pm in Brightspace

Please remember to include the following statement at the beginning of your submitted assignment and SIGN it. Your assignment won't be graded without the signed statement.

"We have done this project/assignment completely on our own. We have not copied it, nor have we given our solution to anyone else. We understand that if we are involved in plagiarism or cheating, we will have to sign an official form that we have cheated and that this form will be stored in our official university record. We also understand that we will receive a grade of 0 for the involved assignment and our grade will be reduced by one level (e.g., from A to A- or from B+ to B) for our first offense, and that we will receive a grade of "F" for the course for any additional offense of any kind."

For the project of this course, you are going to develop a basic **Information Retrieval** system or Search Engine. This system involves storing the documents by performing indexing and retrieving those documents based on a given query. You will be developing this whole system in 3 parts.

Project 3 is a team project with up to 4 members.

For project 3, you will be using your project 1 and project 2 to create a full-fledged web application which will work almost similar to a search engine. Your project 3 should support the following requirements:

1. An Input functionality which should take multiple document files:
 - a. Similar to project 1, you will take documents, perform preprocessing and store the terms with positional indexing. You may use the same storing format as in the project 1
2. A Search functionality that will take free search query (such as phrase query) as input:
 - a. Similar to project 2, you will take string as input and retrieve documents that satisfy the search query. (Complete match, ignore casing)
3. A document viewer:
 - a. Users should be able to view the retrieved documents with the searched query highlighted in the document. You may show any additional information which you deem necessary or will be helpful for the user.

You may refer [An Example UI](#) for the project.

Note: You can use Django (or java) for this project as many of you are already familiar with it. Your code should be properly commented. Your code will also be tested against other documents and queries.

Bonus: Bonus will be given related to the following functionalities (or features) only after completion of above functionalities/features (up to 50%).

1. PDF and TEXT will be considered as basic functionality and in case the project supports uploading & searching for Various Image types, URLs, HTML, etc. then it will be considered for bonus.
2. Storage of positional indexes and documents (for example: types of databases like elastic search, Mongo, etc.).
3. Additional information displayed in the document viewer (presentation of your retrieved documents as a result of search) to the user.

All your code will be subject to check for similarity with other submissions. So you are advised not to look at the other team's code.

Submission:

1. Your Django project (zipped) with a readme/txt file with instructions about how to run and operate your project.
 2. A PDF with the Honesty Policy statement at the beginning, signed by all the team members.
 3. Following individual files from your project folder (not zipped):
 - a. Project's urls.py
 - b. App's urls.py
 - c. App's views.py
 - d. App's models.py (if created)
 - e. App's serializers.py (if created)
 - f. All the html files you created
- If the framework used is other than Django, you will upload the whole project (zipped) and other individual files as asked later.