

2/18:

Progress:

1. Understand and discuss the lab instructions to make sure everyone understands the lab
2. Discuss how this lab can be assigned to different sub-tasks and assign the tasks

Commits on Feb 18, 2025

lab6 init

YiwenC23 committed 3 days ago

e8f4204

2/19:

Progress:

1. Everyone starts to work on the tasks that he/she is responsible for
2. Transparent discussion within the team for any questions that any team member encounters for his/her own part or connecting parts

Commits on Feb 19, 2025

skeleton of searching and extracting

farihashk02 committed 2 days ago

f3b426e

2/20:

Progress:

1. Continue working on the codes/tasks that each member is responsible for
2. Transparent discussion within the team for any questions that any team member encounters for his/her own part or connecting parts
3. Build up the high level code structure for most of the lab sections

Commits on Feb 20, 2025

Merge branch 'main' of https://github.com/YiwenC23/DSCI560-group_labs

HanluMaa committed yesterday

f9c17a9

Update a draft version for web_scraper.py

HanluMaa committed yesterday

499fb4b

Defined a function to preprocess images and a function to extract text from images.

YiwenC23 committed yesterday

169afaa

Defined a function to detect tables in images.

YiwenC23 committed yesterday

1dccc0

Defined pdf to images function.

YiwenC23 committed yesterday

5c94417

Add database config script.

YiwenC23 committed yesterday



















4fd44b1

2/21:

Progress:

1. Continue working on the codes/tasks that each member is responsible for
2. Start to finalize/fine-tune the code and discuss how to connect files together to wrap all things together
3. Transparent discussion within the team for any questions that any team member encounters for his/her own part or connecting parts

Commits on Feb 21, 2025

fix some errors in web_scraper.py HanluMaa committed 11 hours ago	60f55e9		
delete unnecessary files HanluMaa committed 11 hours ago	35bdd02		
extracting well data after searching farihashk02 committed 16 hours ago	a145016		
used csv to search web page farihashk02 committed 17 hours ago	73c51c7		
Update web_scraper.py HanluMaa committed 19 hours ago	3e0059d		
fix some errors in web_scrapper.py HanluMaa committed 19 hours ago	6cf0bd7		
Upload relevant information about the web scrapping part HanluMaa committed 19 hours ago	e82a436		
Update web_scraper.py HanluMaa committed 19 hours ago	a3044aa		
fix some details for web_scraper.py HanluMaa committed 19 hours ago	56acf81		

2/22:

Progress:

1. Finalize the code to secure its functionality
2. Work on the lab report and the demo video
3. Transparent discussion within the team for any questions that any team member encounters for his/her own part or connecting parts

Commits on Feb 22, 2025

Merge branch 'main' of https://github.com/YiwenC23/DSCI560-group_labs HanluMaa committed 1 minute ago	b71211f		
fix some details for web_scraper.py HanluMaa committed 1 minute ago	18701e8		
corrected input path issue. YiwenC23 committed 2 minutes ago	a15114c		
updated database info table YiwenC23 committed 3 minutes ago	d769802		
Finished all pdf extraction logic. YiwenC23 committed 1 hour ago	7683075		
Added several functions to ocr images. YiwenC23 committed 16 hours ago	ff8bf48		
Please also include the .gitignore file itself in your .gitignore file, and use the command <code>git rm --cached .gitignore</code> to untrack the file so it won't be pushed! YiwenC23 committed 17 hours ago	f4983ae		