

# Queue Implementation Report

## Introduction

The purpose of implementing a queue in our document analyzer system is to manage multiple tasks concurrently. This report outlines how queues are implemented in the `[upload.py]` script to handle PDF and text document processing and integrate these tasks seamlessly into our API.

### 1. Queue Implementation for PDF Analysis

To handle the processing of PDF files without stalling user interactions, we employ a queue mechanism. This is achieved by:

**Creating a Task Queue:** When a PDF file is uploaded, the file path and task details are enqueued in a `task_queue`. This queue holds all the pending tasks that need to be processed.

**Starting a Processing Thread:** A new thread is initiated specifically for processing each PDF file. This thread is responsible for extracting text from the PDF using the `pdfplumber` library. Once the text extraction is complete, the extracted text is enqueued in a `result_queue`, ready for the next stage of processing.

```
if filename.lower().endswith('.pdf'):
    # 启动一个线程来处理PDF
    # Starts a thread to process the PDF
    threading.Thread(target=process_pdf, args=(file_path, task_queue)).start()
elif filename.lower().endswith('.txt'):
    threading.Thread(target=process_text, args=(file_path, task_queue)).start()
```

### 2. Queue Implementation for NLP Analysis

Following the text extraction, the next step involves analyzing the text using Natural Language Processing (NLP):

**Retrieving Text from Task Queue:** The main thread waits to retrieve the extracted text from the `task_queue`. This blocking call ensures that the thread does not proceed until text is available.

**Starting NLP Analysis Thread:** Once text is retrieved, an NLP analysis thread is launched to process this text. The analysis involves calling a custom function `analyze_document` which performs the necessary computation and analysis.

```
text = task_queue.get() # 等待PDF处理完成
# Waits for the PDF processing to complete

# 启动另一个线程进行NLP分析
# Starts another thread for NLP analysis
threading.Thread(target=process_nlp, args=(text, result_queue)).start()
```

### 3. Integration into the API

The final step involves integrating this asynchronous processing mechanism into our API:

**Waiting for Analysis Results:** The API endpoint waits for the NLP analysis to complete by retrieving results from the result\_queue.

**Response Preparation:** Once the analysis is complete, the results are formatted into a JSON response, which is then returned to the client.

**File Cleanup:** Post-processing, temporary files are cleaned up to free resources and maintain the system's integrity.

```
analysis_result = result_queue.get() # 等待NLP分析完成
# Waits for the NLP analysis to complete

# 保存结果到文件
# Saves the results to a file
save_analysis_results(analysis_result, filename) # 调用函数保存结果 # Calls function to save results

# 清理操作: 删除临时文件
# Cleanup operation: delete the temporary file
os.remove(file_path)

return jsonify(analysis_result)
```

## Conclusion

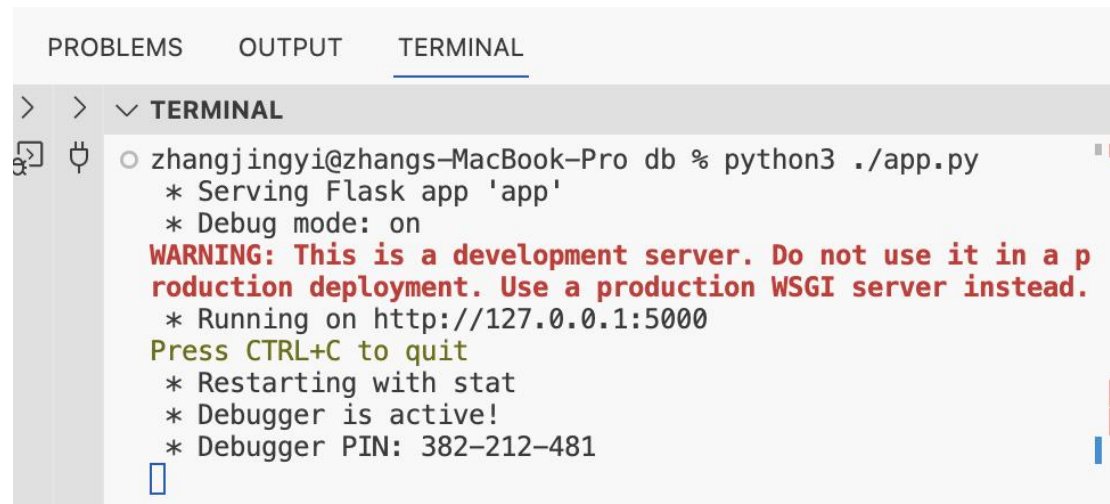
By using queues and threading in our document processing system, we manage to keep the API responsive and capable of handling multiple simultaneous requests efficiently. This setup ensures that our system can scale and handle heavy loads, especially when processing large documents or when multiple users are interacting with the system concurrently.

I will also use some screenshot to demonstrate you how two users use the nowclient.py at same time to upload documents.

## Usage Demonstration

Step 0: **download all code in main branch**

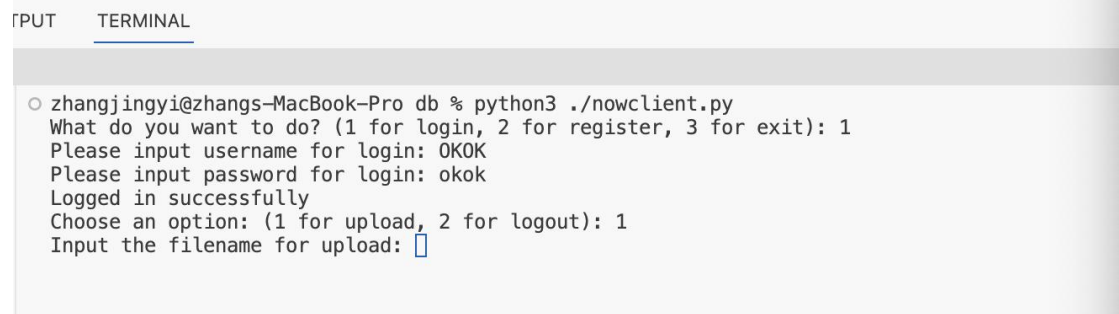
Step 1 open [\[app.py\]](#) : input `python3 ./app.py` in **terminal A** to run the app (the back end).



```
PROBLEMS  OUTPUT  TERMINAL
> >  TERMINAL
○ zhangjingyi@zhangs-MacBook-Pro db % python3 ./app.py
  * Serving Flask app 'app'
  * Debug mode: on
  WARNING: This is a development server. Do not use it in a p
  roduction deployment. Use a production WSGI server instead.
  * Running on http://127.0.0.1:5000
  Press CTRL+C to quit
  * Restarting with stat
  * Debugger is active!
  * Debugger PIN: 382-212-481
  □
```

Step 2 client 1 usage : input `python3 ./nowclient.py` in **terminal B** to run the app for client 1 (the front end).

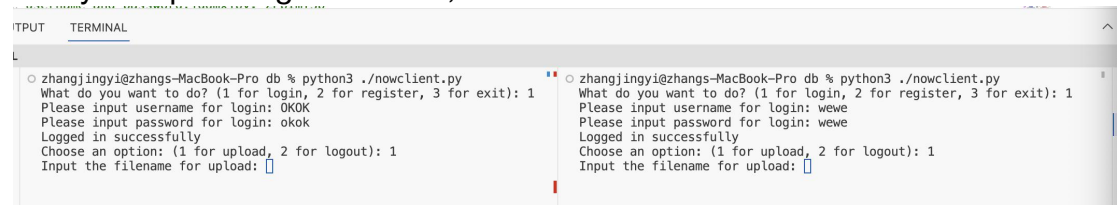
Then input 1 to login, then input username and password, then input 1 to be ready for uploading document.



```
INPUT  TERMINAL
○ zhangjingyi@zhangs-MacBook-Pro db % python3 ./nowclient.py
  What do you want to do? (1 for login, 2 for register, 3 for exit): 1
  Please input username for login: OKOK
  Please input password for login: okok
  Logged in successfully
  Choose an option: (1 for upload, 2 for logout): 1
  Input the filename for upload: □
```

Step 3 client 2 usage : input `python3 ./nowclient.py` in **terminal C** to run the app for client 2 (the front end).

Then input 1 to login, then input username and password, then input 1 to be ready for uploading document, similar with client 1 does.



```
INPUT  TERMINAL
○ zhangjingyi@zhangs-MacBook-Pro db % python3 ./nowclient.py
  What do you want to do? (1 for login, 2 for register, 3 for exit): 1
  Please input username for login: OKOK
  Please input password for login: okok
  Logged in successfully
  Choose an option: (1 for upload, 2 for logout): 1
  Input the filename for upload: □

○ zhangjingyi@zhangs-MacBook-Pro db % python3 ./nowclient.py
  What do you want to do? (1 for login, 2 for register, 3 for exit): 1
  Please input username for login: wewe
  Please input password for login: wewe
  Logged in successfully
  Choose an option: (1 for upload, 2 for logout): 1
  Input the filename for upload: □
```

Step 4 upload document for both clients: I upload [grade.pdf] for client 1, and upload [resume.pdf] for client2, so we see both analyzing report can be formed. The analyzing report is long , you can see [result\_of\_grade.txt] and [result\_of\_example.txt] since these documents print out the analyzing result of these two documents. You can use other pdf and txt documents to test.

```
ITPUT  TERMINAL
L
o zhangjingyi@zhangs-MacBook-Pro db % python3 ./nowclient.py
What do you want to do? (1 for login, 2 for register, 3 for exit): 1
Please input username for login: OKOK
Please input password for login: okok
Logged in successfully
Choose an option: (1 for upload, 2 for logout): 1
Input the filename for upload: grade.pdf
Document analyzed successfully with the following result:
Entities:
12/23/23, 10:04 AM (TIME); Boston University Office of the University Registrar
STUDENT GRADE REPORT FOR (ORG); December 23, (DATE); Jingyi Zhang (PERSON); 1st year (DATE); ENG EC504 (PRODUCT); ADV DATA (ORG); 14.8 4.0 (CARDINAL); ENG (ORG); 16.0 4.0 (CARDINAL); ENG EC602 (PRODUCT); 16.0 4.0 (CARDINAL); ENG EK690 A1 CAREER LAB MS (PRODUCT); F 0.0 1.0 (PRODUCT); PDP (ORG); 0.0 1.0 (CARDINAL); 46.8 (CARDINAL); Total Credit Hours (EVENT); 14.0 (CARDINAL); Semester Grade Point (ORG); 3.60 (CARDINAL); Cum Grade Point (ORG); 3.60 (CARDINAL); December 23, 2023 (DATE); Boston University Grading System (ORG); Explanation Grade per Credit Hour (ORG); 4.0 (CARDINAL); 3.7 (CARDINAL); 3.3 (CARDINAL); 2.7 (CARDINAL); 2.3 (CARDINAL); 2.0 (CARDINAL); NC Not Applicable (WORK_OF_ART); Fx Not Applicable Fail (PRODUCT); J Not Applicable Registration (WORK_OF_ART); Grade (ORG); 1/1 (CARDINAL)
Keywords:
AM, STUDENT, FALL, Name, 1st, year, Hours, A-, B+, B, C+, C, D, Credit, Credit, degree, Credit, acceptable, degree, P, P, course, course, Additional, work, same, course, Applicable, Applicable, Unresolved, https://www.bu.edu/link/bin/uiscgi_studentlink.pl/HREF=https://www.bu.edu/link/bin/uiscgi_studentlink.pl/1703343829?app=path=grades_print.pl&StudentKe
Sentiment Polarity: 0.07727272727272726
Summary:
12/23/23, 10:04 AM Grades
Boston University Office of the University Registrar
STUDENT GRADE REPORT FOR FALL 2023
AS OF December 23, 2023
Name: Jingyi Zhang College of Registration: ENG
ID Number: U26-57-8499 Class Year: Graduate 1st year
calculating the nearest state from the given input location.
BikeTrafficPythonPredictionTool PurdueUniversity, May2022-August2022
[] Designedanddevelopedapredictiontoolusingpythonwhichhelpsinforecastingbiketrafficduringdifferent
timeperiodsandrecommendedpositionstoinstallsensorsforrealtimesafetymonitoring..
[] Developed the prediction model using multivariate linear regression model by analyzing relationships between daily weather forecast data (5 attributes including temperature, (low/high temperature and precipitation and bicycle traffic using sklearn.
[] Implemented R square algorithm on calculating positions with predicted maximum traffic flow for installing sensors.
ISD-OCDDAILYDETAILApplication PurdueUniversity, Jan2021-May2021
[] OCDDailyDetailisanappdesignedforISD(IndianaschoolfortheDeaf)forteacheersgivingparentsprompt feedbackforstudentsontheirengagementandactivitiesatschool.
[] Enhancedtheiosfrontendfeaturesonformatting, displayingandaestheticsincludingusingCandSwift.
[] DesignedandImplementedtheloginauthenticationfeatureusingC++.
[] Implementedtheapplocalizationfeaturewhichtranslates the application between English and Spanish using C++ and java.
TikTokReminder FengChiaUniversitySep2019-Dec2019
[] Designedanandroidappwhichremindpeoplewhenshouldtheydowhat.
OtherProjects: Beautiful"handwriting"machine, GUARDTOURSUYSYSTEM, LASERHARPPROJECT
SKILLS
Languages/Tools: C/C++, Java, Python, SQL, Git, AWS, AndroidStudio, Matlab, PCB design, webdesign
Other: Machine learning, Critical Thinking , Data Science, Data visualization, Computer Engineering, System design, Economics, Algorithm, dataanalysis, Microsoftoffice, statistics, computer science.
Choose an option: (1 for upload, 2 for logout): 2
Logging out successfully...
o zhangjingyi@zhangs-MacBook-Pro db %
```

Step 5: input 2 and 3 to exit, or upload other document you want to test.

```
T  TERMINAL
Total Credit Hours 14.0
Semester Grade Point Index 3.60
Cum Grade Point Average 3.60 (As of December 23, 2023)
Boston University Grading System
Letter Honor Points
Explanation
Grade per Credit Hour
A 4.0 Excellent
A- 3.7
B+ 3.3
B 3.0 Good
B- 2.7
C+ 2.3
C 2.0 Satisfactory
C- 1.7
D 1.0 Low Pass
F 0.0 Fail; No Credit
AU Not Applicable Audit; No Credit Earned
CR Not Applicable Credit; Acceptable for degree
NC Not Applicable No Credit; Not acceptable for degree
H Not Applicable Honors
P Not Applicable Pass with Credit
P* Not Applicable Pass, P/F course
F* Not Applicable Fail, P/F course
I Not Applicable Incomplete; Additional work required
J Not Applicable Registration in the same or continuing course required
W Not Applicable Withdrew
MG Not Applicable Missing Grade; Grade not assigned
X Not Applicable Unresolved
https://www.bu.edu/link/bin/uiscgi_studentlink.pl/HREF=https://www.bu.edu/link/bin/uiscgi_studentlink.pl/1703343829?app=path=grades_print.pl&StudentKe= 1/1.
Choose an option: (1 for upload, 2 for logout): 2
Logging out successfully...
o zhangjingyi@zhangs-MacBook-Pro db %
calculating the nearest state from the given input location.
BikeTrafficPythonPredictionTool PurdueUniversity, May2022-August2022
[] Designedanddevelopedapredictiontoolusingpythonwhichhelpsinforecastingbiketrafficduringdifferent
timeperiodsandrecommendedpositionstoinstallsensorsforrealtimesafetymonitoring..
[] Developed the prediction model using multivariate linear regression model by analyzing relationships between daily weather forecast data (5 attributes including temperature, (low/high temperature and precipitation and bicycle traffic using sklearn.
[] Implemented R square algorithm on calculating positions with predicted maximum traffic flow for installing sensors.
ISD-OCDDAILYDETAILApplication PurdueUniversity, Jan2021-May2021
[] OCDDailyDetailisanappdesignedforISD(IndianaschoolfortheDeaf)forteacheersgivingparentsprompt feedbackforstudentsontheirengagementandactivitiesatschool.
[] Enhancedtheiosfrontendfeaturesonformatting, displayingandaestheticsincludingusingCandSwift.
[] DesignedandImplementedtheloginauthenticationfeatureusingC++.
[] Implementedtheapplocalizationfeaturewhichtranslates the application between English and Spanish using C++ and java.
TikTokReminder FengChiaUniversitySep2019-Dec2019
[] Designedanandroidappwhichremindpeoplewhenshouldtheydowhat.
OtherProjects: Beautiful"handwriting"machine, GUARDTOURSUYSYSTEM, LASERHARPPROJECT
SKILLS
Languages/Tools: C/C++, Java, Python, SQL, Git, AWS, AndroidStudio, Matlab, PCB design, webdesign
Other: Machine learning, Critical Thinking , Data Science, Data visualization, Computer Engineering, System design, Economics, Algorithm, dataanalysis, Microsoftoffice, statistics, computer science.
Choose an option: (1 for upload, 2 for logout): 2
Logging out successfully...
What do you want to do? (1 for login, 2 for register, 3 for exit): 3
Exiting...
o zhangjingyi@zhangs-MacBook-Pro db %
```