Python/Go/Java Project Requirements

Project 1

Read a pdf file from a folder. Refer to the PDF file Chemistry Questions.pdf

Requirements

1. Store a PDF file in a folder called “/content”
2. Read PDF file from the folder
3. Write the content to a text file called “output.txt”
4. Store this file under the “/content” folder

Error Handling

1. Take care of case where folder is not available
2. Take care of case where PDF file is not present in the content folder
3. Take care of case where the output.txt file is not available

Project 2

Traverse through folder tree and filter pdf files

Requirements

1. Add sub-folders called “One”, “Two”, “Three” under the folder called “/content”
2. Add PDF files under each of the sub-folders
3. Load all PDF files under the sub-folders and load the PDF content
4. Write the content to a text file called “output.txt” under each sub-folder respectively

Error Handling

5. Take care of case where folder is not available
6. Take care of case where PDF file is not present in a sub-folder
7. Take care of case where the output.txt file is not available in a sub-folder

Project 3

Read content from a particular page

Requirements

1. Update project 1 and update the reading of content
2. Take a page number as an input from command prompt
3. Read content of the page number provided and write to the output file

Error Handling

1. Take care of case where folder is not available
2. Take care of case where PDF file is not present in a sub-folder
3. Take care of case where the output.txt file is not available in a sub-folder

Project 4

Read regular expression from a config file and extract content

Requirements

1. Update project 3
2. Add support for a configuration file
3. In the configuration file set a config with key “regex” and value some regular expression that will match a part of the content in the PDF
4. Update code to extract only the content matching the regular expression
5. Write to the output file

Error Handling

1. Take care of case where folder is not available
2. Take care of case where PDF file is not present in a sub-folder
3. Take care of case where the output.txt file is not available in a sub-folder
4. Take care of case where no configuration file is available
5. Take care of the case where configuration file does not have the regular expression

Project 5

Store extracted questions in mysql

Requirements

1. Update project 4 and add support for database
2. Create a database to store the following
    a. Subject Name
    b. Question Text
    c. Answer options
    d. Chapter name
3. Load a PDF containing questions
4. Extract each question as per a regular expression
5. Store each question in the database

Error Handling

1. Take care of case where database is not available
2. Take care of case where table is not available
3. Take care of any error handling in DB operations

Project 6

Load all questions from a chapter

Requirements

1. Update project 5 and add support for taking a chapter name as input in the command line
2. Load all questions from the input chapter
3. Print all questions on the console

Error Handling

1. Take care of case where empty string is provided as input from command line
2. Take care of case where there are no questions corresponding to the provided chapter name

Project 7

Load RSS content and then extract content from each link. Do this in multiple threads

Requirements

1. Load an RSS xml file (Format: https://www.w3schools.com/xml/xml_rss.asp)
2. Loop through each link

3. Extract content from each link and write to “output.txt”
4. Execute reading from multiple links in parallel

Error Handling

1. Take care of case where no RSS xml file is available
2. Take care of case where xml file is empty

Project 8

On top of Project 5, add more subjective questions, objective questions with multichoice options to the database from the command line.Use concept of inheritance.