Assignment 2
Posted Date: Mar 8, 2023
Submission Due: Mar 27, 2023 (11:59 pm)
Late assignments will not be accepted and will result in a 0 on the assignment

---

Objective: This assignment covers two learning objectives (lo).

- ***lo#1:*** Perform research on data management and security – To achieve this task, you need to read and understand a given published research material and write a summary of your understandings.
- ***lo#2:*** Implementation of Data Extraction and Processing Engine for unstructured data + Use of NoSQL database – To achieve this task, you need to access a data source, and use API specification to extract data from the given data source. This can be done by a custom-built program written in Java, and then perform some filtration/ transformation on the data when data points enter the system. The final processed data should be kept in a document database
  [**Note**: Your code must be written by you, and will be checked for academic integrity]

Plagiarism Policy:

- This assignment is an individual task. Collaboration of any type amounts to a violation of the academic integrity policy and will be reported to the AIO.
- Content cannot be copied verbatim from any source(s). Please understand the concept and write in your own words. In addition, cite the actual source. Failing to do so will be considered as plagiarism and/or cheating.
- The Dalhousie Academic Integrity policy applies to all material submitted as part of this course. Please understand the policy, which is available at:
  https://www.dal.ca/dept/university_secretariat/academic-integrity.html

Assignment Rubric - based on the discussion board rubric (McKinney, 2018)

	Excellent (25%)	Proficient (15%)	Marginal (5%)	Unacceptable (0%)
Completeness including Citation	All required tasks are completed	Submission highlights tasks completion. However, missed some tasks in between, which created a disconnection	Some tasks are completed, which are disjoint in nature.	Incorrect and irrelevant
Correctness	All parts of the given tasks are correct	Most of the given tasks are correct However, some portions need minor modifications	Most of the given tasks are incorrect. The submission requires major modifications.	Incorrect and unacceptable
Novelty	The submission contains novel contribution in key segments, which is a clear indication of application knowledge	The submission lacks novel contributions. There are some evidences of novelty, however, it is not significant	The submission does not contain novel contributions. However, there is an evidence of some effort	There is no novelty
Clarity	The written or graphical materials, and developed applications provide a clear picture of the concept, and highlights the clarity	The written or graphical materials, and developed applications do not show clear picture of the concept. There is room for improvement	The written or graphical materials, and developed applications fail to prove the clarity. Background knowledge is needed	Failed to prove the clarity. Need proper background knowledge to perform the tasks

Citation: McKinney, B. (2018). The impact of program-wide discussion board grading rubrics on students' and faculty satisfaction. Online Learning, 22(2), 289-299.

Explanation of the rubric: Suppose you received different grades in Clarity for the 2 problems
Problem #1: 25% in clarity
Problem #2: 15% in clarity
Then your overall grade for the clarity will be avg of (25+15) % = 20%

Problem 1: Perform a systematic literature review and provide summary. Objective is lo#1

1) Read the given paper and write summary (Maximum 2 pages in your own words - Do not copy any content verbatim).
2) In addition, explore** and report, if the paper has any scope of improvements in terms of technical details or concepts.
   ** Due to time constraint, and limitation of formal research background, you do not need to explore in details the scientific equations or parameters presented in the paper.
   Visit and Login to https://dal.ca.libguides.com/ieee and search the following paper.
3) Do not forget to provide citation in proper format for the given paper and for any other supporting materials that you may use as reference.

G. Vonitsanos, E. Dritsas, A. Kanavos, P. Mylonas and S. Sioutas, "Security and Privacy Solutions associated with NoSQL Data Stores," *2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA, 2020),* pp. 1-5, doi: 10.1109/SMAP49528.2020.9248442.
Paper Link: https://ieeexplore-ieee-org.ezproxy.library.dal.ca/document/9248442

Problem 2: A light-weight data extracting, processing, transformation engine designed using Java programming language (no 3rd party libraries allowed). Objective is lo#2

Follow the given requirements strictly.

Req #	Task	Checklist
1	Use a standard Java IDE to develop your application. Any JDK version is acceptable	
2	While writing the Java code, follow JavaDocs specification for commenting styles, such as @param, @return etc.	
3	Write/Draw (using any tool), the design principles that you will be using or have used in your application program development/execution. Check SOLID design online	
4	Your application should be console based (no GUI needed)	
5	Visit the news API https://newsapi.org/ Create a developer account Once it is done, you need to write 3 programs (include the source code in gitlab), and each program should trigger the next Code A → Code B → Code C	
6	**Code A: Extraction Engine Implementation:** Write a well-formed program using Java to extract data from NewsAPI. (Do not use any online program codes or scripts, which is not part of the official API documentation and specification.) Search keywords (These are Not case sensitive)– "Canada", "University", "Dalhousie", "Halifax", "Canada Education", "Moncton", "hockey", "Fredericton", "celebration"	
7	You need to include an appropriate pseudocode of your data extraction program in the PDF file.	
8	**Code B: Data-Processing Engine Implementation:** This program is initiated with the Extraction Engine and waits for any incoming data. Once the raw data is captured using the API, this program writes the news contents, and the titles to files. **Note**: This is an automated process, there is no manual file creation or	

	writing process involved. In addition, each file should contain only 5 news articles or less.	
9	**Code C: Transformation Engine Implementation:** This program should automatically clean and transform the data stored in the files, and then upload each record to new MongodB database **myMongoNews.** You can create a single collection or multiple collections depending on your design. •For cleaning and transformation-Remove special characters, URLs, emoticons etc. •Write your own regular expression logic. You cannot use libraries such as, jsoup, JTidy You need to include a flowchart of the transformation engine in the pdf file.	

Submission Guidelines:

1. All written reports, images, code etc. must be added in a folder, and compress it with **.ZIP** format only.
2. If not mentioned by TAs, then please rename the .zip file with your B00xxxxx_FnameLname_A2
3. Submit your Java code in gitlab. Your TA must have provided guidelines for that. If not, please ask the TA.
4. You must include Test Cases (at least 3 – manual testing of functionality or validation testing) for the developed application and provide necessary screenshots as evidence of testing. Note: This is not Junit test.
5. Check the next point "Suggestions" for quality improvement and time management.

Suggestions:

Better Quality: To obtain good grades, you should follow the points given below:

- Try to understand the assignment requirement and follow all the steps required.
- Do not miss adding citations. If you write a single sentence taking the idea from somewhere else, then give credit to the author. Therefore, provide citation for any report you write, or any code you implement
- When you add citation, make sure to add it in a standard format and uniform format. E.g. if I refer 3 sources for writing a report, then I must cite the 3 sources in same format. One source in MLA, two sources in APA citation format will be a mismatch. Therefore, follow any one standard citation format
- Make sure to provide inline citations within report, and programming code
- Any image/picture/flowchart/diagram you add, make sure to provide a caption and a number for that image. It should be placed at the bottom of the image. E.g. "**Fig 1: Weekly time management chart for CSCI 5408**"
- Any table you add, must have a number and caption. This should be added on top of the table. E.g. "**Tab1: Table highlights the requirements in a ordered format**"

Time Management: Follow proper time management to reduce stress, and last-minute preparations. I am suggesting you follow the pie chart, which will require you to spend 5 hours in a week outside the classroom time for this course.

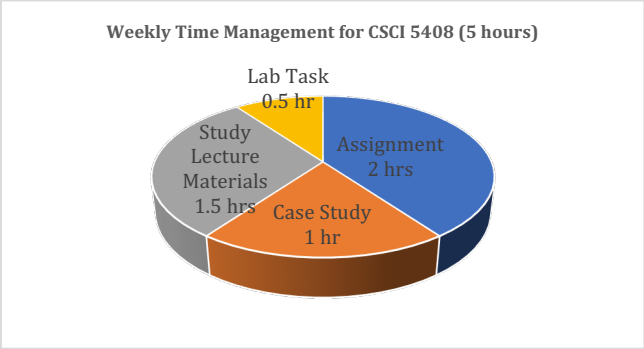

Fig 1: Weekly time management chart for CSCI 5408