

CSC/DCSCI 2720: Data Structures

Assignment 1

Instructor: Shiraj Pokharel

Due : 10/27/2023 @ 10 PM ET
Late Submission deadline (with 25% penalty) 10/28/2023 @ 10 PM ET

Answer the below questions.

You must submit your responses as a SINGLE Jupyter Notebook, where each program is put in separate Jupyter Notebook cells within that SINGLE Jupyter Notebook. Do NOT submit Colab links.

Failure to comply with this simple requirement will result in a score of Zero. Please, be careful not to be assigned a Zero score this way.

Few Rules to be followed, else will receive a score of ZERO

- (1) Your submissions will work exactly as required.
- (2) Your files shall not be incomplete or worse corrupted such that the file does not compile at all. Make sure you submit a file that compiles
- (3) Your submission will show an output. Should you receive a Zero for no output shown do not bother to email me with "but the logic is perfect" !

Note that your program's output must **exactly** match the specs(design , style) given here for each problem to pass the instructor's test cases .

Design refers to how well your code is written (i.e. is it clear, efficient, and elegant), while *Style* refers to the readability of your code (commented, correct indentation, good variable names).

Context and Problem Statement: A web search engine returns you a list of web-links that match the set of words in your query. Each web-link is assigned an integer for identification purposes(page-ID) This is done by maintaining an "inverted index".

An inverted index, at a very elementary level, is a mapping that takes a word and returns to you a list of web-links which contains the word. That list would be a sorted(ranked) array of page-IDs where the web-link with the highest rank for the query is displayed on top. When your query to the web search engine contains multiple words, the search engine needs to find the sorted array for each word and then computes the intersection of these arrays. Given the number of web-pages existing in the world-wide-web, this work is very computationally challenging !

In this assignment you are required to write programs which takes as input **two sorted arrays** and returns a new array containing the elements found in both the sorted arrays. It is alright if the input arrays have duplicates but the returned array should be duplicate free !

Below is how the arrays are represented

ARRAY1[] = [1, 6, 9, 10, 11, 21]

Here length of ARRAY1 is m .

ARRAY2[] = [2, 6, 9, 11, 17, 21]

Here length of ARRAY2 is n .

The output array to be returned would be:

ARRAY[] = [6, 9, 11, 21]

ATTN : Please be reminded that you cannot use library functions to do any of the tasks required above. Doing so would straight up result in a score of Zero !

You will solve the problem in three ways:-

- (1) [40 points] Implement the function in such a way that your solution solves the problem with $O(mn)$ time complexity. $O(mn)$ is same as $O(m * n)$. This brute-force method suggested has a name called "loop-join" where you basically just traverse through the elements of one array comparing it to the elements of the other array.
- (2) [40 points] In a separate implementation, code up a solution in such a way that your solution solves the problem with $O(n \log(m))$ time complexity

or $O(m \log(n))$ time complexity. Here \log means to the base of 2. I'm sure you already know that the hint is to use Binary Search.

- (3) [20 points] In a separate implementation, code up a solution in such a way that your solution can run linearly with $O(m + n) \approx O(N)$ time complexity.

Very Very Important :

- (1) Your code should be well commented which explains all the steps you are performing to solve the problem. **A submission without code comments will immediately be deducted 15 points !**

- (2) As a comment in your code, please write your test-cases on how you would test your solution assumptions and hence your code.

A submission without test cases will immediately be deducted 15 points !
Example of cases to be tested for are like : What if the array input which is expected does not exist - that is , input is a null. How should your code handle such a situation ? Maybe output some message like "Null input case, so no output"? What if the length of the array is one ?....so on and so forth.

Please Remember : Although, written as comments - You will address your test cases in the form of code and not prose :)