

# CSC/DSCI 2720 Data Structures Assignment 1

**Due: 02/06/2024 @ 11:59 PM ET**

Answer the below questions.

You may use whatever IDEs / editors you like, but you must submit your responses on iCollege as .py files. 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 us 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) list 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 list for each word and then computes the intersection of these lists. 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** lists and returns a new list containing the elements found in both the sorted lists. It is alright if the input lists have duplicates, but the returned lists should be duplicate free!

Below are how the lists are represented:

LST1 = [1, 5, 6, 6, 9, 9, 9, 11, 11, 21]. Here length of LST1 is m.

LST2 = [6, 6, 9, 11, 21, 21, 21]. Here length of LST2 is n.

List to be returned would be LST = [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 list comparing it to the elements of the other list.

(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 the form of sentences, as a comment in your code (at the bottom of your Solution2), you are required to suggest how can Solution2 be improved by leveraging the fact that both the lists are already sorted. Suggest a solution so that your suggested solution can run linearly with  $O(m + n)$  time complexity. Your suggestion should be no more than 5 sentences :)

**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 list input which is expected does not exist - that is, input is a null (empty list). 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 list 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 :)