This assignment is adopted from Dr. Lucet COSC 520 assignments.

a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

COSC 520
Assignment 1

The Login Checker Problem

The objective of the assignment is to: (1) review hash tables, (2) explore advanced data structures, and (3) numerically compare 2 algorithms.

The login checker problem is to quickly check that a login name has not been already taken so that all logins are unique. Assume you have millions of logins (aka usernames), and a user creates a new account. You need to ensure the new login is not a duplicate of an existing one. Using your existing knowledge of algorithms and data structures you can

1. store all logins in a list, and check existence using linear search;
2. store all logins in a sorted array, and check existence using binary search;
3. store logins in a hash table, and check existence by hashing.

The assignment explores 2 more advanced data structures that scale better than the above. Assume the number of logins is n=1 billion (you may use a smaller value of n depending on your computer memory and speed as long as you justify the value of n used to validate your approach).

- Read about bloom filters and look at examples.
- Read about Cuckoo filters and how they compare to Bloom filters.

Submission

To complete the assignment, submit the following:

1. A concise pdf file listing the time and space computational complexity of: linear search, binary search, hashing, Bloom filter, Cuckoo filter. Use appropriate parameters to describe the complexity, e.g., Bloom filter uses n (estimated number of elements/logins), m (bit array of m bits), and k (number of hash functions). You do not need to justify the complexity; only to explain each parameters used and provide the formula with appropriate references.
2. Plots showing the run time complexity for large enough data. You can generate strings dataset using synthesized functions. Aim for at least one million data. Then, run your

This assignment is adopted from Dr. Lucet COSC 520 assignments.

program with different implementations using linear search, binary search, hashing, Bloom filter, Cuckoo filter, and compare their run time in a plot. Include this analysis in your pdf file and explain if it supports your analysis. Remember to upload your dataset on the web and include a link to the dataset in the pdf file.

3. Python code comparing the hashing, Bloom filter, and Cuckoo filter. Add the link of your GitHub repository for this assignment to the pdf file. Your code should
    - Be well documented and clean.
    - Include unit tests.
    - follow common practises:
        - appropriate class/variable/method names (not too long and meaningful),
        - appropriate comments
        - comments for each method should indicate input, output, and a short explanation of the method
        - clear setup and running instructions.

Submit the PDF file to Canvas. Remember to add three items above (only the link of your GH) to the PDF file.

If you use online sources or GenAI to generate your code, first, make sure it is correct and executable. Second, make sure to include the resources you have used and/or mention that you used GenAI in your report. There is no penalty of using available code, the only drawback can be that you might not be engaged as writing the code yourself for your future references.

Grading rubric

	Weights	Subtotals
Report		10
Complexity	5	
References	5	

This assignment is adopted from Dr. Lucet COSC 520 assignments.

a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

Code static		40
naming	5	
comments	10	
running instructions	5	
code design	20	
Code execution		50
syntax-error free, runs	10	
unit tests run	10	
convincing demo	10	
performance	10	
comparison of the data structures (item 2 in the above list)	10	
Total	100	100

Bonus

1. Suggest an alternative competitive method; explain its data structure, how it works, and its complexity with appropriate reference [15% bonus]
2. Discuss your alternative method with the instructor, and if the instructor considers your method competitive, implement it and benchmark it vs. Bloom and Cuckoo [15% bonus]

This assignment is adopted from Dr. Lucet COSC 520 assignments.

a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

Submission deadline:

Feb 7, 2025, at 8 PM.