Assignment 4 & 5 – Hashing

Dates of work: Aug 30, 2022 & Sep 6, 2022;

Submission of complete assignment (code + report of analyses): Sep 23, 2022;

Viva: Sep 24, 2022 & Sep 25, 2022; Total Points: 140

Questions:

1. Consider the following partial table of an ordered library catalogue:

Author_ID	Book_ID	Author_Name	Book
An_Ch_0103	Aest_AC_0103	Anjan Chatterjee	The Aesthetic Brain
An_Da_0104	Self_AD_0104	Antonio Damasio	Self Comes to Mind
Ca_Sa_0319	Anim_CS_0319	Carl Safina	What Animals Think
Jo_Ro_1018	Deat_JR_1018	Joanne K. Rowling	Deathly Hallows_Harry Potter
Jo_Ro_1018	Fant_JR_1018	Joanne K. Rowling	Fantastic Beasts and Where to Find Them
Jo_Ro_1018	Gobl_JR_1018	Joanne K. Rowling	Goblet of Fire_Harry Potter
Jo_Ro_1018	Phil_JR_1018	Joanne K. Rowling	Philosopher's Stone_Harry Potter
Jo_Ro_1018	Pris_JR_1018	Joanne K. Rowling	Prisoner of Azkaban_Harry Potter
La_Ch_1203	Mind_LC_1203	Lars Chittka	The Mind of a Bee
Ma_Mi_1313	Emot_MM_1313	Marvin Minsky	Emotion Machine
Ma_Mi_1313	Soci_MM_1313	Marvin Minsky	Society of Mind
Pe_Wo_1623	Aunt_PW_1623	Pelham G. Wodehouse	Aunts Aren't Gentlemen
Pe_Wo_1623	Wode_PW_1623	Pelham G. Wodehouse	Wodehouse at the Wicket
Vi_Ra_2218	Emer_VR_2218	Vilayanur Ramachandran	The Emerging Mind
Vi_Ra_2218	Phan_VR_2218	Vilayanur Ramachandran	Phantoms in the Brain

Note: Majority of searches in the catalogue involve Author name and/or Book name

[All codes must be written in C / C++]

Q1.	a. Choose a primary key for the given table. Reasons for doing so? [5] b. Design a hash function for what you choose as the primary key for the given table. Use the evaluated hash values to work on the following questions [5] c. Comment on the provided codes for book_id and author_id. Do you think these are sufficiently effective? [5]	[15]
Q2.	a. Use extendible hashing with bucket size 4 to design an effective access strategy for the above [5] b. Do you think a different bucket size would have been more effective? Experiment with at least one smaller and at least one larger bucket size to understand the scenario [5+5] c. Experiment using a different data structure, instead of a linear list, for a bucket [5]	[20]
Q3.	a. Choose a global bucket order (n) for the given table. What inspired your choice of 'n'? [5] b. Use linear hashing with a local bucket size of 4 and your chosen value of 'n' to design an effective strategy for the given scenario [5] c. Do you think a different value of 'n' would have been more effective? Experiment with at least one smaller and one larger value - keeping the bucket size unchanged - to understand the scenario [5+5] d. Do you think a different bucket size would be better? Experiment with at least one smaller and one larger value - keeping 'n' constant - to understand the scenario [5+5] e. Experiment using a different data structure, instead of a linear list, for a bucket [5]	[35]
Q4.	a. Choose a value for the number (n) of higher positioned bits that you would like to consider for the distributed hash tree. What inspired your choice of 'n'? [5] b. Use distributed hashing with a bucket size of 4 and your chosen value of 'n' to design an effective strategy for the given scenario [5] c. Do you think a different value of 'n' would have been more effective? Experiment with at least one smaller and one larger value - keeping the bucket size unchanged - to understand the scenario [5+5] d. Do you think a different bucket size would be better? Experiment with at least one smaller and one larger value - keeping 'n' constant - to understand the scenario [5+5] e. Experiment using a different data structure, instead of a linear list, for a bucket [5]	[35]
Q.5.	a. Comment on your experiments with all the hashing mechanisms [5]	[35]

	b. If you have to insert a new record ➔ Finding Muchness by Kobi Yamada – which - amongst the 3 mechanisms (bucket size 4) - was the fastest, and why? [10] c. If you have to retrieve a record ➔ What Animals Think by Carl Safina – which - amongst the 3 mechanisms (bucket size 4) - was the fastest, and why? [10] d. If you have to retrieve names of all books by Marvin Minsky, which technique (bucket size 4) was the fastest and why? [10]	

Assessment Rubric for Submitted Work - Evaluation per answer (Viva & Analyses):

Proper understanding of question and work done accordingly:	80 - 95%
Attempt to work beyond what has been asked, with in-depth understanding: (Definite contender for full score for question)	95 - 100%
Vague understanding, bursts of in-depth answers:	70 - 80%
Vague understanding, bursts of broad conceptual answers:	50 - 70%
Weird hash of work submitted, some understanding:	40 - 50%
No understanding, just work submitted: (Probable plagiarism)	0

Submission Rubric:

Within 2 days of Deadline:	No penalty
Within 5 days of Deadline:	30% penalty
Within 7 days of Deadline:	50% penalty
After 7 days of Deadline:	Will not be evaluated