

CME 2201 - Assignment 1

INVERTED INDEX BY USING HASH TABLES

Table 1. Performance matrix

Load Factor	Hash Function	Collision Handling	Collision Count	Indexing Time	Avg. Search Time	Min. Search Time	Max. Search Time
$\alpha=50\%$	SSF	LP	187096	67.832 seconds	1608445 nanoseconds	51 nanoseconds	8555 nanoseconds
		DH	177566	7.736 seconds	197000 nanoseconds	44 nanoseconds	3874 nanoseconds
	PAF	LP	483919	6.004 seconds	198828 nanoseconds	28 nanoseconds	25014 nanoseconds
		DH	483336	5.978 seconds	138281 nanoseconds	22 nanoseconds	7144 nanoseconds
$\alpha=80\%$	SSF	LP	187095	72.221 seconds	1659367 nanoseconds	52 nanoseconds	7501 nanoseconds
		DH	175489	8.12 seconds	188995 nanoseconds	44 nanoseconds	2605 nanoseconds
	PAF	LP	449459	6.393 seconds	166497 nanoseconds	21 nanoseconds	21553 nanoseconds
		DH	450600	5.911 seconds	143400 nanoseconds	22 nanoseconds	13062 nanoseconds

Prepared a table to help us understand which types of hashing are more efficient in the project. When we look at the table, the more the load factor increases, the more the hashing time increases but the number of collisions decreases. Other than that, there are minor changes in searching time. If we compare collision handling types, less collision occurs in DH and hashing happens faster. As a hash function, less collision occurs in SSF, but spends more

search time. Finally, hashing spends the least amount of time in the case of load factor 0.8, hash function type is PAF and collision handling type is DH;

Hüseyin Erdoğan

2017510031