FACULTY OF ENGINEERING
COMPUTER ENGINEERING DEPARTMENT
Tınaztepe Yerleşkesi, Buca-Kaynaklar, Dokuz Eylül Üniversitesi, İZMİR, TÜRKİYE
07.12.2022
CME 2201 Homework Report: Developing a Supermarket
Management
Prepared by: Kaan YILMAZ – İrem TEKİN

1- **Introduction**: Developing a simple supermarket management system. The program operates on text files that are kept in the retrieval of historical purchase records of customers. The goal is implementing a hashing algorithm in Java to index UUID (Universally Unique Identifier) for each customer and read these document, split it; customer ID, customer name, date, product name and index each word to the hash table according to rules. After insertion of all documents, the market manager gained access to the products purchased by the users, sorted according to their dates.

2- **Project Description**: To specify an index corresponding to given string UUID, firstly generating an integer hash code by using a special function. Then, resulting hash code must be converted to the range 0 to N-1 using a compression function, such as modulus operator (N is the size of hash table). There are two different functions and two different collision handlings,

2.1.1- Simple Summation Function (SSF):

Generate the hash code of a string *s* with the length *n* simply by the following formula:

$$h(s) = \sum_{k=0}^{n-1} ch_k$$

As an example:

SSF(“Ali”) = 1 (A) + 12 (l) + 9 (i) = 22

2.1.2- Polynomial Accumulation Function (PAF):

Generate the hash code by using the following polynomial:

$$h(s) = ch_0 * z^{n-1} + ch_1 * z^{n-2} + \ldots + ch_{n-2} * z^1 + ch_{n-1} * z^0$$

where ch_0 is the left most character of the string, characters are represented as numbers in 1-26 (case insensitive), and *n* is the length of the string. The constant *z* is usually a prime number. When the *z* value is chosen as 31, the string "car" has the following hash value:

$$h(\text{car}) = 3 * 31^2 + 1 * 31 + 18 * 1 = 2932$$

Horner rule is used for avoiding overflows.

2.2.1- Linear Probing (LP): Linear probing handles collisions by placing the collidingitem in the next available table cell.

2.2.2- Double Hashing (DH):

Double hashing uses a secondary hash function *d(k)* and handles collisions by placing an item in the first available cell of the series.

$$d(k) = q - k \bmod q$$

$$h_2(k) = (h(k) + j\, d(k)) \bmod N$$

where $q < N$ (table size), q is a prime, and $j = 0, 1, \dots, N - 1$.

3- Algorithms and Solution Strategies:

The hashed algorithm is mentioned on above. Unique UUIDs are stored in hashed key. The UUIDs hash value and values containing the customer are stored in the Customer class.

4- Performance Monitoring (Reading customer_1K.txt):

Load Factor	Hash Function	Collision Handling	Collision Count	Indexing Time	Avg. Search Time	Min. Search Time	Max. Search Time
α=50%	SSF	LP	500968	75962 ms	3152891 ns	750 ns	44200858 ns
		DH	499454	3291 ms	2952381 ns	650 ns	44100200 ns
	PAF	LP	467973	188 ms	1164631ns	600 ns	42984500 ns
		DH	467963	1012 ms	3354517 ns	600 ns	41892800 ns
α=80%	SSF	LP	510711	82064 ms	3178546 ns	750 ns	39520075 ns
		DH	499643	5017 ms	3803853 ns	600 ns	49539600 ns
	PAF	LP	478745	324 ms	1294617 ns	650 ns	29624200 ns
		DH	479000	1405 ms	2805911 ns	600 ns	36875100 ns

5-) Conclusion:

As we can see on the table that mentioned above. PAF & LP & 0.5 load factor is most efficient way to store data. PAF is distributing words clearly into the hashed table. If we distribute them again by DH they locate randomized into the table. But LP is increases index by one when there is a collision.

07.12.2022

Kaan Yılmaz – İrem Tekin
2020510072 2021510062