

Assignment 4: Flight Ticket Management System Report

Experimental Observation of Hash Function Efficiency

by Swostik Pati

Hash Code 1: Cycle Shift Hash Code

Code:

```
int m = key.length();
unsigned int h = 0;
for (int i = 0; i < m; i++)
{
    h = (h << 5) | (h >> 27);
    h += (unsigned int)key[i];
}

return h % this->capacity;
```

Results:

1K records:

```
>count_collisions
-----
The total number of collisions are: 194
```

10K records:

```
>count_collisions
-----
The total number of collisions are: 7588
```

100K records:

```
>count_collisions
-----
The total number of collisions are: 96807
```

Average number of collisions per data set:

= (194+7588+96807) / 3

= 34863 collisions per dataset

Hash Code 2: Polynomial Hash Code

Code:

```
unsigned int sum = 0;
unsigned short int a = 2;
for (int i = key.length() - 1; i >= 0; i--)
{
    sum = (sum * a) + (unsigned int)key[i];
}
int32_t result = sum % this->capacity;

return result;
```

Results:

1K records:

```
>count_collisions
-----
The total number of collisions are: 459
```

10K records:

```
>count_collisions
-----
The total number of collisions are: 9040
```

100K records:

```
>count_collisions
-----
The total number of collisions are: 98295
```

Average number of collisions per data set:

= (459+9040+98295) / 3

= 35931.3 collisions per dataset

Hash Code 3: Summation Hash Code

Code:

```
unsigned int sum = 0;
for (int i = 0; i < key.length(); i++)
{
    sum += (unsigned int)key[i];
}
int32_t result = sum % this->capacity;
return result
```

Results:

1K records:

```
>count_collisions
-----
The total number of collisions are: 577
```

10K records:

```
>count_collisions
-----
The total number of collisions are: 9330
```

100K records:

```
>count_collisions
-----
The total number of collisions are: 98586
```

Average number of collisions per data set:

= (577+9330+98586) / 3

= 36164.3 collisions per dataset

Final Observation:

Based on the experimental results, we conclude that the cycle shift hashcodes causes the minimum average number of collisions per dataset (34863 collisions per data set on average) and therefore it is the most efficient implementation for the purpose of our project.