WORD COUNT IN A FILE USING HASH TABLES

AIM:

To determine the number of words in a file and the repetition of each word usinh hash tables.

ABSTRACT:

The file is read word by word and hashed into a table of size 281(a prime number is used to avoid collisions) as default. The table size can be mentioned by the user too. If a same word is encountered again,the repetition count is incremented.

DESCRIPTION:

A class Hash is created with a string array to hash each word , an integer array to store the repetition count and an integer variable to keep track of the total number of words.

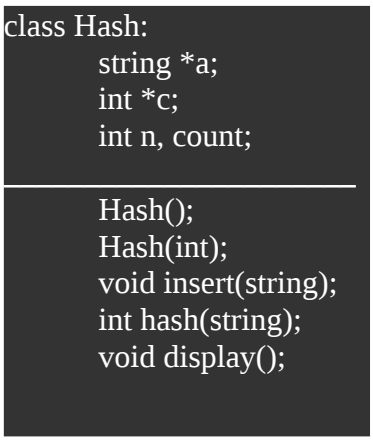
The methods of the class Hash are :

1.void insert(string);
2.int hash(string);
3.void display();

The insert() function inserts the string into the hash table and increments the count variable whenever a new word is inserted. The hash() function determines the index(hash value) at which the string needs to be inserted. The display() function displays the hash table along with number of occurences and the total word count.

The file is read using fstream. The file to be opened is received as input from the user. The file is read until the End-Of-File(EOF) marker is reached. Fullstops(.) and commas(,) are removed before hashing. After reading the whole file the display() function is called.

CLASS DIAGRAM:



SOURCE CODE:

```
#include<iostream>
#include<math.h>
#include<string.h>
#include<fstream>
#include<iomanip>
#include<algorithm>
using namespace std;
class Hash
{
        string *a;
        int *c, n;
        int count;
        public:
                Hash(){
                        count = 0;
                        n = 281;
                        c = new int[n];
                        a = new string[n];
                        for(int i =0;i<n;i++)
                                c[i] = 0;
                }
```

```
        Hash(int x){
            count = 0;
            n = x;
            c = new int[x];
            a = new string[x];
            for(int i=0;i<x;i++)
                c[i] = 0;
        }
        void insert(string);
        int hash(string);
        void display();
};
void Hash::insert(string x){
    int y = hash(x);
    a[y] = x;
    c[y] += 1;
}
int Hash::hash(string x){
    count++;
    int k = x.length();
    unsigned long long int s = 0;
    for(int i=0;i<k;i++){
        s+=(x[k-i-1]*pow(37,i));
    }
    return s%n;
}
void Hash::display(){
    for(int i=0;i<n;i++){
        if(c[i]!=0)
            cout<<setw(20)<<a[i]<<'\t'<<c[i]<<endl;
    }
    cout<<"Total word count:"<<count<<endl;
}
int main()
{
    string data;
```

```
        Hash a;
        char dat[25];
        cout<<"Enter the file name"<<endl;
        cin>>dat;
        ifstream infile;
        infile.open(dat);
        infile>>data;
        if(infile.fail())
                cout<<"Error in opening the file.Check the file name or enter
the entire path"<<endl;
        else{
                while(!infile.eof()){
                        a.insert(data);
                        infile>>data;
                        replace(data.begin(),data.end(),'.',' ');
                        replace(data.begin(),data.end(),',',' ');
                }
                a.display();
                infile.close();
        }
        return 0;
}
```

TESTCASES:

INPUT	EXPECTED OUTPUT	ACTUAL OUTPUT
gayathri@gayathri:~$./a.out Enter the file name ias.txt	gayathri@gayathri:~$ wc -w ias.txt 154 ias.txt	undergraduate 1 In 1 on 2 interest 1 years 1 like 1 with 1

		1
		Samuel 1
		very 1
		internship 2
		Google's 3
		I'm 2
		algorithm 1
		device 2
		the 8
		popular 1
		been 1
		in 3
		Computer 1
		able 1
		immense 1
		is 1
		capabilities 1
		lead 2
		I 3
		be 2
		studying 2
		computer 1
		various 1
		The 2
		engine 1
		e-commerce 1
		search 2
		Learning' 1
		recent 5
		a 5
		field 3
		course 1
		Since 1
		and 3
		play 1
		have 2

		papers 1 then 2 Gayathri R 1 sites 2 start 2 Similarly 1 algorithm 1 student 1 development 1 research 1 being 1 seen 1 algorithms 1 College 1 first 1 to 5 fascinated 1 taking 1 improve 1 me 1 after 1 This 2 deep 1 'Machine 1 currently 1 always 1 apply 1 learning 4 my 2 many 3 designed 1 developments 1 knowledge 2 look 1 own 1 who 1

		social 1 has 4 using 1 publish 2 also 1 was 2 now 1 learning 2 machine 6 of 5 Total word count:154
gayathri@gayathri:~$./a.out Enter the file name fi.txt	gayathri@gayathri:~$ wc -w fi.txt 16 fi.txt	programming 1 best 1 Have 1 the 2 is 2 sexy 1 learn 1 time 1 folks 1 new 1 to 1 fun 1 This 1 Programming 1 Total word count:16
gayathri@gayathri:~$./a.out Enter the file name test.txt	ERROR	Error in opening the file.Check the file name or enter the entire path
gayathri@gayathri:~$./a.out Enter the file name / home/gayathri/Others/test.txt	gayathri@gayathri:~/Others$ wc -w test.txt 4 test.txt	Atom 1 Inkscape 1 ownCloud 1 Wine 1 Total word count:4

RESULT:

Thus the total no. of words in a file no. of occurences of each word in the file has been determined and the table with hashed values has been displayed.

A PROJECT BY R.GAYATHRI, 2016103521
FOR PROGRAMMING AND DATA STRUCTURES II LAB
SUBMITTED TO DR.D.SHILOAH ELIZABETH.