Nội dung

* Bảng băm trong thư viện chuẩn của C++
* Tự cài đặt bảng băm dùng phương pháp đánh địa chỉ đóng - separate chaining

==================

Bảng băm trong thư viện chuẩn của C++

Unordered\_set và unordered\_map: thời gian trung bình cỡ O(1), tuy nhiên trong trường hợp tồi nhất có thể tới O(n)

Ví dụ 1. Nhập vào 1 danh sách số nguyên gồm n phần tử từ bàn phím và in ra lại các phần tử KHÔNG bị trùng trong danh sách

Dùng unordered\_set

|  |
| --- |
| #include <bits/stdc++.h>  using namespace std;    int main()  {  // declaring set for storing string data-type  unordered\_set <int> mySet;    printf("Nhap vao n phan tu:\n");  int n,x;  printf("n= ");  scanf("%d",&n);  for(int i=0; i<n;i++)  {  printf("Nhap vao so tiep theo: ");  scanf("%d",&x);  mySet.insert(x);  }  cout << "Cac phan tu KHONG trung trong day vua nhap : \n";  unordered\_set<int> :: iterator itr;  for (itr = mySet.begin(); itr != mySet.end(); itr++)  cout << (\*itr) << endl;  } |

Thời gian cỡ O(n)

Ví dụ 2. Tạo từ điển các từ nhập từ bàn phím và tra nghĩa

Unordered\_map

|  |
| --- |
| #include <iostream>  #include <unordered\_map>  using namespace std;    int main()  {  // Declaring umap to be of <string, int> type  // key will be of string type and mapped value will  // be of int type  unordered\_map<string, string> mydict;  char word[30], mean[50];  printf("Moi nhap vao tu moi: \n");  printf("Tu: ");  cin>>word;  // do bo dem ban phim  //cin.clear();  cin.ignore(numeric\_limits<streamsize>::max(), '\n');    cout<<"Nghia: ";  cin.getline(mean,50);  mydict[word]=mean;    // Traversing an unordered map  for (auto x : mydict)  cout << x.first << " : " << x.second << endl;    unordered\_map<string, string>:: iterator p;  p = mydict.find("hello");  if (p == mydict.end())  cout <<" word not found\n\n";  // If key found then iterator to that key is returned  else  cout << "Found " << p->first << " : "<< p->second<<" \n\n";    } |

Ví dụ 3. Nhập vào n số nguyên, in ra màn hình tần số xuất hiện của các từ

Dùng unordered\_map

|  |
| --- |
| #include <iostream>  #include <unordered\_map>  using namespace std;    int main()  {  unordered\_map<int, int> mylist;  int n, x;  cout<<"So luong so can nhap: ";  cin>>n;  printf("Nhap vao cac so: ");  unordered\_map<int, int>:: iterator p;  for(int i=0; i<n; i++)  {  cin >>x;  p = mylist.find(x);  if(p!=mylist.end()) // neu co  p->second++;  else  mylist[x]=1;  }  cout<<"danh sach cac gia tri va tan so tuong ung"<<endl;    for (auto x : mylist)  cout << x.first << " : " << x.second << endl;    return 0;  } |

**Bài tập 1**. Xây dựng ứng dụng thống kê tần số xuất hiện của các từ trong 1 đoạn văn tiếng anh

Đoan văn có thể là

Most candidates who have nabbed Trump's endorsement have refused to declare war on McConnell, who remains a powerful fundraiser and influential party figure in his own right, while those who are still angling for the former President's support have also stopped short of staking out opposition to the powerful senator from Kentucky. Trump's ineffective attempt, thus far, to challenge the top Senate Republican has forced him to temper his criteria as he aims to be a kingmaker in this year's midterm elections, according to multiple people close to the former President who spoke to CNN on the condition of anonymity.

Dùng unordered\_map

unordered\_map<string, int> wordlist;

Với mỗi từ tách được ta sẽ kiểm tra xem nó có trong wordlist hay không

|  |
| --- |
| p = mydict.find("hello");  if (p == mydict.end())  cout <<" word not found\n\n";  // If key found then iterator to that key is returned  else  cout << "Found " << p->first << " : "<< p->second<<" \n\n"; |

Nếu có xuất hiện thì tăng giá trị p->second thêm 1

Ngược lại, nếu chưa có thì thêm vào với giá trị 1 : mydict[word]=1;

Mở rộng:

* Không phân biệt chữ HOA/thường
* Thống kê thêm từ có tần số lớn nhất và nhỏ nhất và từ dài nhất/ngắn nhất

**Bài tập 2**. Xây dựng ứng dụng check chính tả cho văn bản nhập vào từ bàn phím

Đầu vào sẽ là 1 file từ điển tiếng anh được nạp sẵn

Tải về từ URL sau : <https://raw.githubusercontent.com/eneko/data-repository/master/data/words.txt>

hãy cho người dùng nhập vào 1 đoạn văn tiếng Anh từ bàn phím, sau đó check chính tả với các từ không viết Hoa có trong đoạn văn đó

In ra màn hình những từ có khả năng bị sai chính tả (không có trong từ điển)

VD. đầu vào người dùng nhập là

**Thi is an esample!**

thì từ có thể sai chính tả là **esample**

Chú ý:

* loại bỏ hết các dấu câu ; : , . ! ~ ? - trong đoạn văn do người dùng nhập vào trước khi check chính tả
* Để check chính tả hãy nạp từ điển vào **unordered\_set**

Chương trình cần lặp đi lặp lại cho tới khi người dùng không muốn check chính tả nữa