

EC-504

Advanced Data Structures

Topic: Nearest State/County Finder

Team Information

Hitanshi Jain (U95073268)
Renyu Jiang (U18427594)
Shuyan Zhang(U09736276)
Yuxiang Wan

Abstract

The aim of this project is to input the latitude, longitude, and K (number of nearest reference points required) from the user and then output the K nearest reference points as an output. It also finds the state and county for the input referenced point, considering the 5 nearest points. To implement this, we have used two data structures Geohash and Red Black Trees, and compared their results for diverse locations. We have implemented the code in C language and used the US counties dataset to test our results.

Instructions to Run code

Copy files to your computer then compile and run it. After software outputs text, input the location (Latitude, Longitude) Then software would do automatic calculation and output nearest counties and their distances in descending order.

There are two folders with two methods for different locations, but we have combined them into "project_main.cpp", just use make command with its makefile and run it.

We have preprocessed counties data and uploaded it to the data folder, if you want to preprocess it, please change the file path in preprocess code.

In addition, our program needs database files in the "data" folder, don't forget it before running the code.

All compilation must be done by C++ 11 or later version.

if you have question about running code, project_main code or Geohash code, please contact renyu26@bu.edu

if you have question about R-B Tree code, please contact woodywan@bu.edu

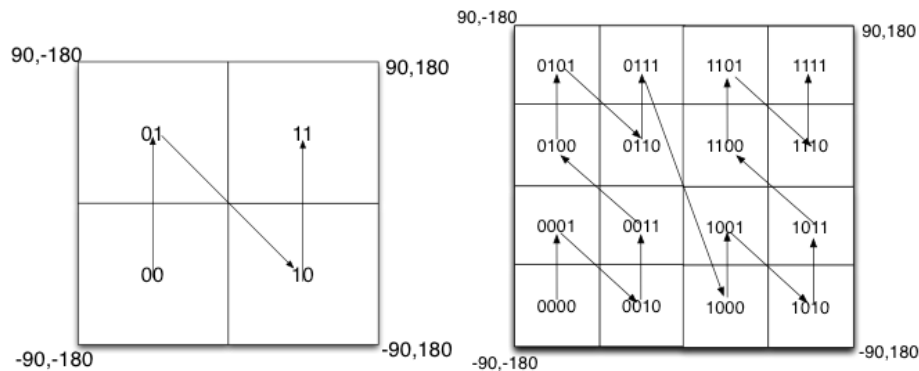
Methods Used

1. Geohash

Geohash is a public-domain geocode system that encodes a geographic location into a short string of letters and digits. It is a hierarchical spatial data structure that subdivides space into buckets of grid shape, which is one of the many applications of what is known as a Z-order curve, and generally space-filling curves.

Geohash Process:

1. Encode latitude and longitude
2. Combine encoded latitude and longitude, longitude in even digits, latitude in odd digits, creating a new location binary code
3. Base32 encodes the new location binary code
4. Perform prefix matching in order from long to short to find adjacent counties



5. Increase the prefix by one bit to re-screen the results of the preliminary search
6. Calculate the distances of the remaining counties and sort them

We used Geohash to find the nearest counties of local locations with $O(\log(\log(n)))$ time complexity.

Geohash Code Screenshots:

County Search:

```

while(temp.size()<target_num){
    temp.clear(); prefix_len--;
    prefix=your_location.substr(0,prefix_len);
    for(int i=0;i<data.size();i++){
        if(startsWith(data[i][1],prefix)){
            temp.push_back(data[i]);}}

```

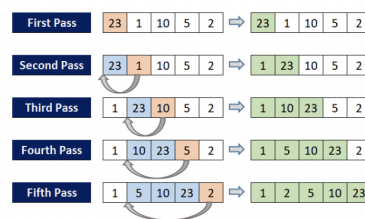


Majority State Finder:

```

int count=0;
for(auto & i : result_sorted){
    state_vote[i[3]]+=1;
    if (count++>target_num-2){
        break;
    }
}

```



Then, use **insertion sort** to find the K nearest counties

2. Red-Black Trees

In Red-Black Tree the insert/delete operations are $O(\log(N))$ and re-balancing rotation is an $O(1)$ operation, making the Red-Black tree more efficient in this aspect of the re-balancing stage. The R-B tree method is for oversea locations including locations far from North America like Hawaii and Guam. The average time complexity of the finding process will be $O((N)\log(N) + N)$ in the case of Red-black trees.

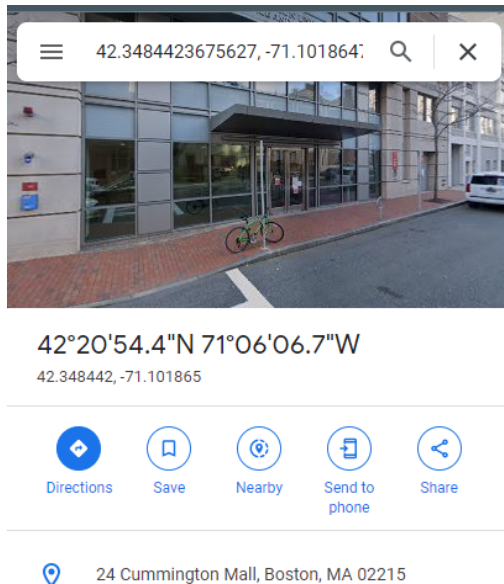
Results

Our software can find the nearest counties at a very fast speed (Avg 2.4 ms) no matter what location you input. For overseas locations, the speed would be a little slow, but it is still under 15 ms, which is significantly faster than other algorithms.

- Demo 1 (Domestic location):

Geohash Result

Test Location: BU LSE in
Massachusetts
Latitude: 42.3484423675627
Longitude: -71.10186472089339



```
loading data...
Input your latitude[-90,90]:
42.3484423675627
Input your longitude[-180,180]:
-71.10186472089339
Your location: [42.3484,-71.1019]
Geohash code: drt2yj31
10 counties found:
County      State      Distance(m)
Suffolk     MA         2980.423773
Norfolk     MA         22785.266053
Middlesex   MA         28299.651438
Essex       MA         38186.257006
Plymouth    MA         50300.613534
Bristol     MA         61372.760957
Providence  RI         66112.581848
Worcester   MA         66296.677576
Rockingham  NH         71176.775554
Bristol     RI         71854.254055
State majority: MA
Time: 0.0024714
```

Process finished with exit code 0

R-B Tree Result

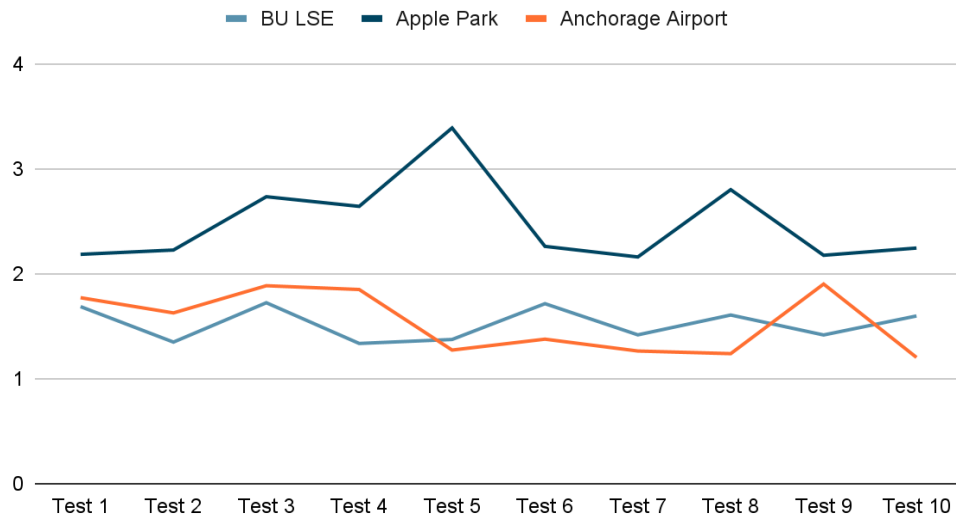
a

```
Enter the latitude: 42.3484
Enter the longitude: -71.1019
Enter the number of the closest city: 10
Suffolk, Massachusetts: 2977.69
Norfolk, Massachusetts: 22760.7
Middlesex, Massachusetts: 28276.1
Essex, Massachusetts: 38159.7
Plymouth, Massachusetts: 50255.8
Bristol, Massachusetts: 61316.6
Providence, Rhode Island: 66051.7
Worcester, Massachusetts: 66238.3
Rockingham, New Hampshire: 71121.8
Bristol, Rhode Island: 71788.9
MassachusettsTime difference (sec) = 0.014771s
```

The screenshots above show that Geohash is significantly faster as compared to R-B Trees for domestic US location.

Performance Analysis for Domestic Locations

Average Running Time (ms)

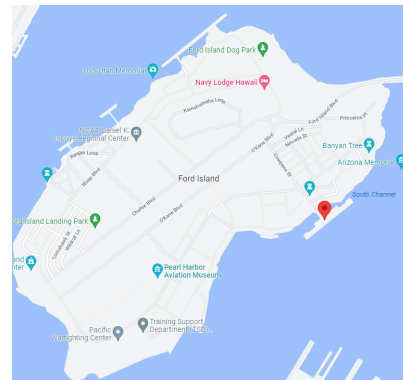


● Demo 2 (Overseas location):

Test Location: Pearl Harbor in Hawaii

Latitude: 21.362374050539042

Longitude: -157.95365360883602



Geohash Result

R-B Tree Result

loading data...

Input your latitude[-90,90]:

Input your longitude[-180,180]:

Your location: [21.3624,-157.954]

Geohash code: 87z9t8vs

10 counties found:

County	State	Distance(m)
Honolulu	HI	21986.463764
Kalawao	HI	106581.978728
Maui	HI	154688.237473
Kauai	HI	185946.594733
Hawaii	HI	321031.356393
Aleutians East	AK	3800202.018586
Mendocino	CA	3845356.428935
Marin	CA	3847675.050218
Sonoma	CA	3850764.910489
San Mateo	CA	3858208.515309

State majority: HI

Time: 2.06724

```

Enter the reference point:
Enter the latitude: 21.3624
Enter the longitude: -157.954
Enter the number of the closest city: 10
Honolulu, Hawaii: 21962.6
Kalawao, Hawaii: 106478
Maui, Hawaii: 154603
Kauai, Hawaii: 185719
Hawaii, Hawaii: 320752
Aleutians East, Alaska: 3.79702e+06
Mendocino, California: 3.84212e+06
Marin, California: 3.84452e+06
Sonoma, California: 3.84756e+06
San Mateo, California: 3.85501e+06
Hawaii
Time difference (sec) = 0.015433s
    
```

Process finished with exit code 0

The screenshots above show that R-B Trees is significantly faster as compared to Geohash for an overseas location.

References

1. <https://en.wikipedia.org/wiki/Geohash>
2. https://blog.csdn.net/weixin_41519463/article/details/88999339