Nearest state/country finder

• • •

EC504 TOPIC
JINGYI ZHANG, XUZHE YU ,Cynthia Young

Project Purpose

The main goal of this project is to create a system, which can efficiently find the nearest state or county of the given input location.

In this system, users are asked to input the latitude and longitude of targeted location.The system's data will be loaded from the official Geographic Committee website. We want this project to implement a nearest neighbor search algorithm to quickly respond to user queries.

Project mission

1 Data Loading:
- The system needs to load lots of reference points into a data structure.
- The reference points include information such as state, county, latitude, and longitude.

2 Nearest Neighbor Search:

- The system should use an efficient nearest neighbor search algorithm.(KD-Tree/ Quadtree/ Ball Tree/R-Tree/Red-Black Trees/Geohash)
- Users will input a latitude and longitude, and the system will return the nearest K reference points (K ranging from 1 to 10).
- The distance computation between two points should use the equirectangular approximation formula.

Project Implementation

3. Distance Computation:

- The distance between two points is calculated using the equirectangular approximation formula, which involves latitude and longitude.

  $x = (\lambda_2 - \lambda_1) * \cos((\varphi_1 + \varphi_2)/2);$
  $y = (\varphi_2 - \varphi_1);$
  $Distance = \sqrt{x*x + y*y} * R;$
  where φ is latitude, λ is longitude, R is earth's radius

4. State and County Identification:

- Once the nearest K reference points are found, the system should determine the state and county of the searched point.
- This is done by performing a majority voting among the K nearest points.

Technologies and Algorithms

Data Structures

- K-D Trees or R-Trees for efficient nearest neighbor searching.
- Hash tables for quick data retrieval.

Distance Calculation Algorithms:

- Equirectangular approximation method for calculating distances.

Majority Voting Algorithm:

- For determining the most likely state/county based on the nearest points.