Wednesday, October 22, 2014

Project #3

Nearest State/County Finder

You are given a huge number of reference points in the US, (We will use data extracted from the official US Board on Geographic Names dataset as input (you can download this dataset from the blackboard course page under Project/Data file for Project 3)). Your system will first load the reference points in an efficient nearest neighbor search amenable data structure. Then it will allow users to query your data structure by entering a decimal latitude and a decimal longitude. You will be asked to return the nearest K reference points, where K is a number from 1 to 10. Also, you will be asked to find the state and county of the a point by computing a majority voting among the 5 nearest points.

For distance computation between two points use the "equirectangular approximation" (http://www.movable-type.co.uk/scripts/latlong.html), which can be defined as:

$$x = (\lambda 2-\lambda 1) * \cos((\phi 1+\phi 2)/2);$$
$$y = (\phi 2-\phi 1);$$
$$\text{Distance} = \sqrt{x*x + y*y} * R;$$

where ϕ is latitude, λ is longitude, R is earth's radius (mean radius = 6371km);

In particular, you are to perform the following two tasks:
- *loading the province, state, decimal latitude, decimal longitude data into your data structure*
- accepting and *responding to user queries efficiently and accurately*

For example, assuming that the reference points are as follows:

STATE_ALPHA	COUNTY_NAME	PRIM_LAT_DEC	PRIM_LONG_DEC
AR	Benton	36.4805825	-94.4580681
AZ	Apache	36.4611122	-109.4784394
AZ	Maricopa	33.2486547	-112.7735045
AZ	Graham	32.4709038	-109.9361853
…			

Given the query lat: 33.24, long: -112.75, the nearest reference point would be in AZ Maricopa with approximately 67.05 km distance.

You are free to use any data structure you like for storing the reference points. You must define the complexity of search in your system. Efficient data structure selection & usage will get higher points.

Minimum Requirements [55%]

- Command line user interface that lets loading of reference points, and querying of coordinates for state and county with a given *K* nearest neighbors.

Possible Extensions [15%]

- Develop a GUI which displays a map of USA and lets users click on various locations on the map instead of entering coordinates, and displays the state and province of the clicked location. [15%]

Comparative Features [30%]

Your algorithms and data structures will be compared against others in the class and ranked according to:

- Speed of search operation.