Assignment 6 - CS618A

Sumedh Masulkar
sumedh@,11736

April 13, 2015

1 System Configuration

- Processor: Intel(R) Core(TM) i5-2450M CPU @ 2.50GHz
- Memory: 16 GB

2 Results

Following table shows time statistics for VA-file of 10^4 nodes in each tree and followed by 300 queries. All values are in milliseconds.

Table 1: Using uniform data and VA-File

	Minimum	Maximum	Mean	Standard Deviation
Search Query	18	23	20.76	0.51
Range Query	22	26	22.35	0.21
kNN Query	23	65	29.69	6.32

Table 2: Using uniform data and linear file

	Minimum	Maximum	Mean	Standard Deviation
Search Query	81	156	84.39	7.54
Range Query	82	114	84.13	4.33
kNN Query	82	141	84.92	7.51

Table 3: Using exponential data and VA file

	Minimum	Maximum	Mean	Standard Deviation
Search Query	11	18	11.76	1.17
Range Query	12	20	12.77	0.85
kNN Query	13	119	54.12	38.51

Table 4: Using exponential data and linear file

	Minimum	Maximum	Mean	Standard Deviation
Search Query	82	83	82.66	0.47
Range Query	82	87	83.27	0.48
kNN Query	83	102	84.17	2.9

3 Observations

- For VA File, we observe,

$$\text{kNN Query} > \text{Range Query} > \text{Point Search Query}.$$

- This can be justified as the number of output points is more in KNN which leads to more I/Os for file access.
- For linear files, all are identical since there is no additional cost for accessing any file for object. It only requires one scan of entire file for all queries.
- For VA file and linear file, we observe that VA file is better than linear file for all queries and data distribution. This can be explained as follows - Data points in VA file can be pruned based on lower bound on distance , whereas distance from all points has to be calculated in case of linear file. Also, VA file takes less memory, so more points can be brought to memory at once than in VA file.