FYP Progress Check Form (Midterm)

Student Name & ID: Fan Yuang & 1155124596

Project title: A Machine Learning Approach to Building Index Structures

Instructions: This form is for each student to report his or her progress up to the midterm. Please complete the form using Microsoft Word and add rows if necessary. Completion percentage refers to the percentage of the original plan completed. If the percentage is less than 100%, explanation must be provided. Please report your progress accurately and evidence should be provided to the supervisors for them to verify.

|  |  |  |  |
| --- | --- | --- | --- |
| Task | Actual achievements | Completion  Percentage | Explanation |
| Apply different methods on 1-d data and compare their time efficiency on some synthetic data | I have implemented four methods which are AVL tree, binary search, trick method (apply a linear search for the data which obeys some special distributions) and hash table on uniform distribution and normal distribution. Besides, I have got some test results which are stored in MongoDB and some conclusions after analysis. Lastly, I have let the repository for the project open source on Github. | 75% | I have tested three aspects which are setting up time, memory occupation and querying time of these kinds of index/querying method. After analysis of the test result stored in database, I have some simple conclusions. In specific, AVL tree takes a lot of time to set up and its size is big. However, its querying time is slow on normal distribution and fast on unform distribution. Besides, binary search and trick method have the same index structure (sorted array), and trick method is faster than binary search. In addition., hash table’s querying time and setting up time are both short. Its size is bigger than the sorted array but is much smaller than AVL tree. |