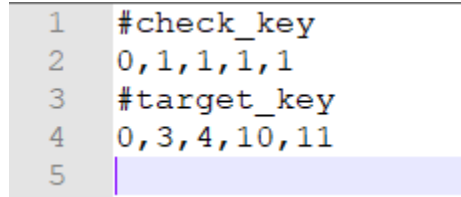


Java KD+-Tree Index User Guide

How To Build KD+-Tree Index File Using Your Datasets

- Set the KD+-Tree parameter in the file kdtree_parameter.txt

Figure 1 is a screenshot to explain how to set the parameter



```
1 #check_key
2 0,1,1,1,1
3 #target_key
4 0,3,4,10,11
5
```

Figure 1: kdtree_parameter.txt screenshot

Check Key is to help program know the value's type. It has 0, -1, 1.

0 means original value is Integer

-1, mean original data value is a time format string < “yyyy-MM-dd HH:mm:ss “ >

1 means original value is Float like longitude, latitude

Right now, this program cannot support long or double type value, and float value cannot be greater than 210000 since program will transfer float to Integer by multiplying 10000, new version will be updated to support.

Target Key is to help program pick up the fields value from original dataset to build tree.

For example, 0 is the 0th field value of original record. 3 is the 3rd field value of original record. The value order will set how to build the tree. For example, based on the figure 1, the build tree order is 0-3-4-10-11. And also, the target key's order must correspond as its value type of Check Key.

- Run KD+-Tree Index program

Type in “java -jar kdtree1.jar <datafile> Num_knode <kdtree_parameter >”

<datafile> is your original data set

Num_knode is the number of leaf node size which you want to set

<kdtree_parameter> is the kdtree_parameter set file

For example, “java -jar kdtree1.jar testdata1.csv 1 kdtree_parameter.txt”

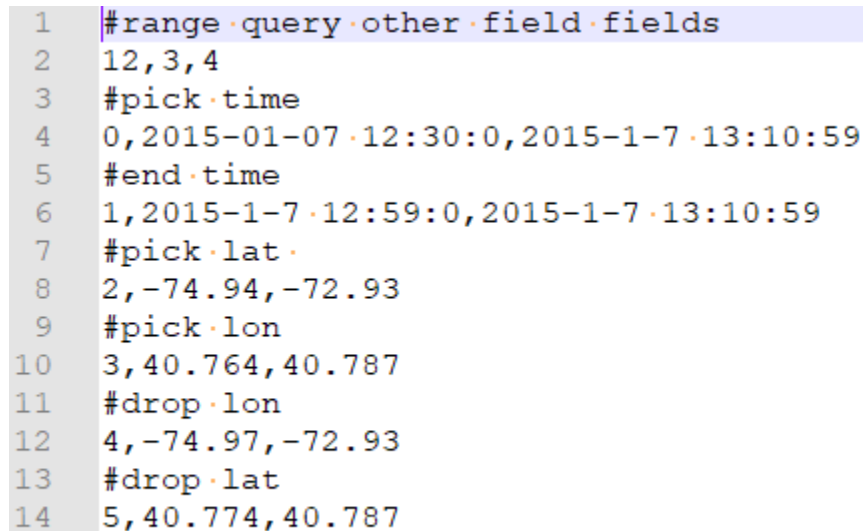
How To Run The Query To Search

- Set Query Parameter File

There are two kinds of different setting parameter files. One is for top k nearest search. Another one is for range search. The program can distinguish the parameter based on if the first line contains “@” or not.

Range query parameter setting

Figure 2 is a screenshot to explain how to set the parameter



```
1 #range.query.other.field.fields
2 12,3,4
3 #pick.time
4 0,2015-01-07.12:30:0,2015-1-7.13:10:59
5 #end.time
6 1,2015-1-7.12:59:0,2015-1-7.13:10:59
7 #pick.lat.
8 2,-74.94,-72.93
9 #pick.lon
10 3,40.764,40.787
11 #drop.lon
12 4,-74.97,-72.93
13 #drop.lat
14 5,40.774,40.787
```

Figure 2

- first line should not contain “@”, if the line is starting with “#”, it is a comment line.
- second line is to set which fields you want to output from original record. The default output is to output the key value of KD+-Tree but cannot output other fields of original record. So this line is to set the parameter that you want to output. For example, like figure 2, you want to output the 12th, 3rd, 4th fields of original record.
- From the third line which is not starting from “#”, it is to set the search range. The first number is stand for order of the target key which you used to build tree. For this example, my KD+-Tree’s target key order is (pick time, drop time, pick lon, pick lat, drop lon, drop lat) . So my pick time order is 0, then “0,2015-01-07 12:30:0,2015-1-7 13:10:59” is to set pick time’s min time is 2015-01-07 12:30:0,and pick time’s max time 2015-1-7 13:10:59. You don’t need set all the target key range, the default range is from Integer.MIN_VALUE to Integer.MAX_VALUE.

Top K nearest query parameter setting

Figure 3 is a screenshot to explain how to set the parameter

```
@query top k , x key, y key  
3,4  
#original record field key 5,6,7  
5,6,7  
#set x  
3,100.0,100.0  
#set y  
4,600.0,600.0  
#set other condition  
0,1000,1000
```

Figure 3

- first line should contain "@", which is to stand for this file is the top k query parameter file.
- 3,4 mean let program know which location's lon, lat you want to based on to do the top k search. 3,4 are the order key of my target key which I used to build tree. My target key value is (id, lon, lat).
- 5,6,7 are to set which fields you want to output from original record. The default output is to output the key value of KD+-Tree but cannot output other fields of original record. So this line is to set the parameter that you want to output. For example, like figure 3, you want to output the 5th, 6th, 7th fields of original record.
- From the forth line which is not starting from "#", it is to set the search condition. The first number is stand for order of the target key which you used to build tree. For this example, my KD+-Tree's target key order is (id, lon, lat). Because it is to search top K, so min and max are same.

- **How To Run Query Program**

Range query

The command is “java -jar Query.jar <datafile> <query_parameter >”

For example, “java -jar Query.jar testdata1.csv query_range_parameter.txt”

Top K query

The command is “java -jar Query.jar <datafile> <query_parameter > <k>”

For example, “java -jar Query.jar testdata1.csv query_topK_parameter.txt 10”

In the test folder, it has built KD+-Tree to run. Before run it, please read ReadMe.txt first.

For reproduce the benchmark test, please login in window and linux folder to read ReadME.txt.