

CIS6930 Spring 2009, Homework 3/4: Spatial Indexing Structures

In this homework, you are required to study a certain type of advanced index structure, in addition to the B+-tree and hash-based indexes we discussed in class, or an algorithm to process specific queries using such index structures. The homework is to be finished by reading research papers and other materials I provide as well as anything you can find from other sources. You are required to deliver an oral presentation describing your understanding of the index structure.

In studying the index and writing your slides for presentations, please keep the following questions in your mind:

- 1) What type(s) of data is the index good for?
- 2) What type(s) of queries is the index good for?
- 3) Can you describe the data organization in the index?
- 4) Can you describe how the queries can be processed using the index?
- 5) What is the time complexity of answering the above queries using the index?
- 6) What is the space complexity of the index?
- 7) Other extensions (e.g., variants, bulk loading, popular implementation, compression, ...) to the basic index structure that can further improve performance?

A successful presentation would clearly show your understanding of the above issues. The presentation will be graded by the whole class based on the quality of your slides, your presentation skills shown, and responses to questions (asked by the audience).

The presentation will be given in-class during the week of 04/13/2009-04/17/2009. Please register a time slot for your presentation as early as possible.

This homework is to be finished by groups of **two**. It is your responsibility to find yourself a partner. However, I can definitely help if you have trouble finding a group. As there are 15 students in the class, one person will either join another group or work alone. I will adjust my expectations and grading standards accordingly of such groups. Once you form a group, please pick up a topic to study from the list attached to the end of this document. The assignment of topics will be done in a first-come-first-serve manner. I want to see a variety of topics in your presentations therefore no two groups are allowed to present the same content.

The following is a list of topics with pointers to corresponding sources as starting points for your study:

1. The **R-tree** for indexing low-dimensional data

The original paper: <http://www.cse.usf.edu/~ytu/Literature/Rtree-SIGMOD84.pdf>

Some slides you can borrow from the Ramakrishnan book:

<http://www.cse.usf.edu/~ytu/teaching/CowbookSlides/Chapter28.ppt>

Two variants of the R-tree can be found in

<http://www.cse.usf.edu/~ytu/Literature/Rstar-SIGMOD90.pdf> and
<http://www.cse.usf.edu/~ytu/Literature/Rplus-VLDB87.pdf>

2. **kd-trees**

The original paper with in-memory operations:

<http://www.cse.usf.edu/~ytu/Literature/KDtree-CACM75.pdf>

Two variants for secondary storage:

<http://www.cse.usf.edu/~ytu/Literature/KDtree-SE79.pdf>

<http://www.cse.usf.edu/~ytu/Literature/KDB-SIGMOD81.pdf>

3. The **Quad/Oct-tree** (this can be done by a 3-person group)

The original paper: Quad Trees: A Data Structure for Retrieval on Composite Keys,

URL: <http://www.cse.usf.edu/~ytu/Literature/QTree-ACTAInfo74.pdf>

An implementation of the quad-tree: <http://www.cse.usf.edu/~ytu/Literature/QTree-Represent-CACM82.pdf>

Efficient tree construction: <http://www.cse.usf.edu/~ytu/Literature/Qtree-VLDBJ02.pdf>

Numerous pointers can be found from the Internet about quad trees.

4. The **Bitmap** Index

The original paper: <http://www.cse.usf.edu/~ytu/Literature/Bitmap-SIGMOD97.pdf>

An extension to the original paper: <http://www.cse.usf.edu/~ytu/Literature/Bitmap-SIGMOD98.pdf>

5. The **Grid File** index

The original paper: <http://www.cse.usf.edu/~ytu/Literature/GridFile-TODS84.pdf>

6. Hash-based incremental query processing in spatio-temporal databases.

Paper: <http://www.cse.usf.edu/~ytu/Literature/SINA-SIGMOD04.pdf>

7. Constructing well-separated pair decomposition points sets using binary trees.

Paper: <http://www.cse.usf.edu/~ytu/Literature/WSPD-JACM95.pdf>

8. Space filling curves (this is good for a single-person group)

A tutorial in Chapter 5 of *Searching multimedia databases by content*. This book can be checked out from the USF library.

A set of slides: <http://www.win.tue.nl/~hermanh/stack/dagstuh108-talk.pdf>

I also found that WIKI provides a good introduction to all of the above structures (and probably everything on earth). Therefore, I suggest you search WIKI using the name of the index as keyword to get a rough idea of what the index is about and how it is organized. In addition to that, the following paper is a survey that covers most of the index structures in the above list (excluding the bitmap index):

<http://www.cse.usf.edu/~ytu/Literature/SpatialAccess-CS98.pdf>