



15-826: Multimedia Databases and Data Mining

Lecture #4: Multi-key and
Spatial Access Methods - I
C. Faloutsos



Must-Read Material

- MM-Textbook, Chapter 4
- [Bentley75] J.L. Bentley: *Multidimensional Binary Search Trees Used for Associative Searching*, CACM, 18,9, Sept. 1975.
- Ramakrishnan+Gehrke, Chapter 28.1-3

15-826 Copyright: C. Faloutsos (2012) 2



Outline

Goal: 'Find **similar** / **interesting** things'

- Intro to DB
- ➔ • Indexing - similarity search
- Data Mining

15-826 Copyright: C. Faloutsos (2012) 3



CMU SCS

Indexing - Detailed outline

- primary key indexing
- ➔ • secondary key / multi-key indexing
- spatial access methods
- text
- ...

15-826 Copyright: C. Faloutsos (2012) 4



CMU SCS

Sec. key indexing

- attributes w/ duplicates (eg., EMPLOYEES, with 'job-code')
- Query types:
 - exact match
 - partial match
 - 'job-code'='PGM' and 'dept'='R&D'
 - range queries
 - 'job-code'='ADMIN' and salary < 50K

15-826 Copyright: C. Faloutsos (2012) 5



CMU SCS

Sec. key indexing

- Query types - cont'd
 - boolean
 - 'job-code'='ADMIN' or salary > 20K
 - nn
 - salary ~ 30K

15-826 Copyright: C. Faloutsos (2012) 6

CMU SCS

Solution?

15-826 Copyright: C. Faloutsos (2012) 7

CMU SCS

Solution?

- Inverted indices (usually, w/ B-trees)
- Q: how to handle duplicates?

salary-index

Name	Job-code	Salary	Dept
Smith	PGM	70	R&D
Jones	ADMIN	50	R&D
....			
Tomson	ENG	50	SALES

15-826 Copyright: C. Faloutsos (2012) 8

CMU SCS

Solution

- A#1: eg., with postings lists

salary-index

postings lists

Name	Job-code	Salary	Dept
Smith	PGM	70	R&D
Jones	ADMIN	50	R&D
....			
Tomson	ENG	50	SALES

15-826 Copyright: C. Faloutsos (2012) 9

Solution

- A#2: modify B-tree code, to handle dup's

salary-index

Name	Job-code	Salary	Dept
Smith	PGM	70	R&D
Jones	ADMIN	50	R&D
....			
Tomson	ENG	50	SALES

15-826 Copyright: C. Faloutsos (2012) 10

How to handle Boolean Queries?

- eg., 'sal=50 AND job-code=PGM'?

salary-index

Name	Job-code	Salary	Dept
Smith	PGM	70	R&D
Jones	ADMIN	50	R&D
....			
Tomson	ENG	50	SALES

15-826 Copyright: C. Faloutsos (2012) 11

How to handle Boolean Queries?

- from indices, find lists of qual. record-ids
- merge lists (or check real records)

salary-index

Name	Job-code	Salary	Dept
Smith	PGM	70	R&D
Jones	ADMIN	50	R&D
....			
Tomson	ENG	50	SALES

15-826 Copyright: C. Faloutsos (2012) 12



CMU SCS

Sec. key indexing

- easily solved in commercial DBMS:

```
create index sal-index on  
  EMPLOYEE (salary);  
select * from EMPLOYEE  
  where salary > 50 and  
         job-code = 'ADMIN'
```

15-826 Copyright: C. Faloutsos (2012) 13



CMU SCS

Sec. key indexing

- can create combined indices:

```
create index sj on EMPLOYEE  
  ( salary, job-code);
```

15-826 Copyright: C. Faloutsos (2012) 14



CMU SCS

Indexing - Detailed outline

- primary key indexing
- ➡ • secondary key / multi-key indexing
 - main memory: quad-trees
 - main memory: k-d-trees
- spatial access methods
- text
- ...

15-826 Copyright: C. Faloutsos (2012) 15

CMU SCS

Quad-trees

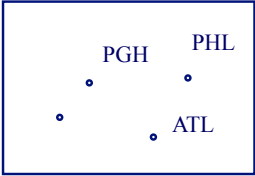
- problem: find cities within 100mi from Pittsburgh
- assumption: all fit in main memory
- Q: how to answer such queries quickly?

15-826 Copyright: C. Faloutsos (2012) 16

CMU SCS

Quad-trees

- A: recursive decomposition of space, e.g.:

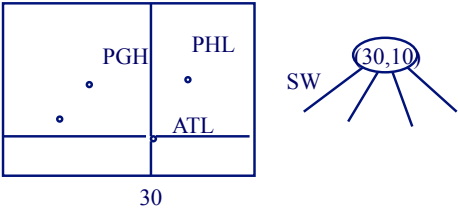


15-826 Copyright: C. Faloutsos (2012) 17

CMU SCS

Quad-trees

- A: recursive decomposition of space, e.g.:



15-826 Copyright: C. Faloutsos (2012) 18

Quad-trees

- A: recursive decomposition of space, e.g.:

15-826 Copyright: C. Faloutsos (2012) 19

Quad-trees - search?

- find cities with $(35 < x < 45, 15 < y < 25)$:

15-826 Copyright: C. Faloutsos (2012) 20

Quad-trees - search?

- find cities with $(35 < x < 45, 15 < y < 25)$:

15-826 Copyright: C. Faloutsos (2012) 21

CMU SCS

Quad-trees - search?

- pseudocode:


```

range-query( tree-ptr, range)
  if (tree-ptr == NULL) exit;
  if (tree-ptr->point within range){
    print tree-ptr->point}
  for each quadrant {
    if ( range intersects quadrant ) {
      range-query( tree-ptr->quadrant-ptr, range);
    }
  }
      
```

15-826 Copyright: C. Faloutsos (2012) 22

CMU SCS

Quad-trees - k-nn search?

- k-nearest neighbor algo - more complicated:
 - find ‘good’ neighbors and put them in a stack
 - go to the most promising quadrant, and update the stack of neighbors
 - until we hit the leaves

15-826 Copyright: C. Faloutsos (2012) 23

CMU SCS

Quad-trees - discussion

- great for 2- and 3-d spaces
- several variations, like fixed decomposition:

‘adaptive’

PGH •	PHL •
•	ATL •

‘fixed’

z-ordering (later)

PGH •	PHL •
•	ATL •

↖ middle

15-826 Copyright: C. Faloutsos (2012) 24



CMU SCS

Quad-trees - discussion

- but: unsuitable for higher-d spaces (why?)

15-826 Copyright: C. Faloutsos (2012) 25



CMU SCS

Quad-trees - discussion

- but: unsuitable for higher-d spaces (why?)
- A: 2^d pointers, per node!
- Q: how to solve this problem?
- A: k-d-trees!

15-826 Copyright: C. Faloutsos (2012) 26

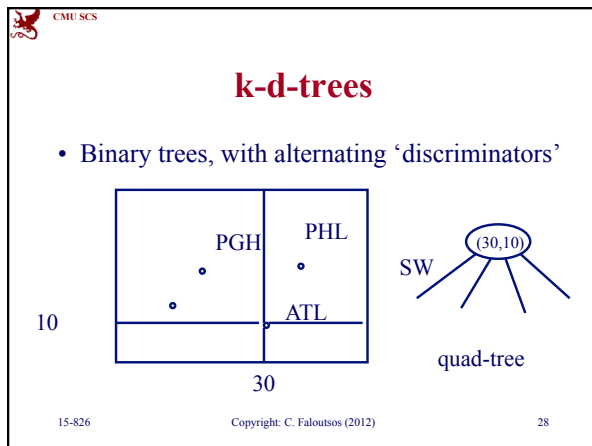


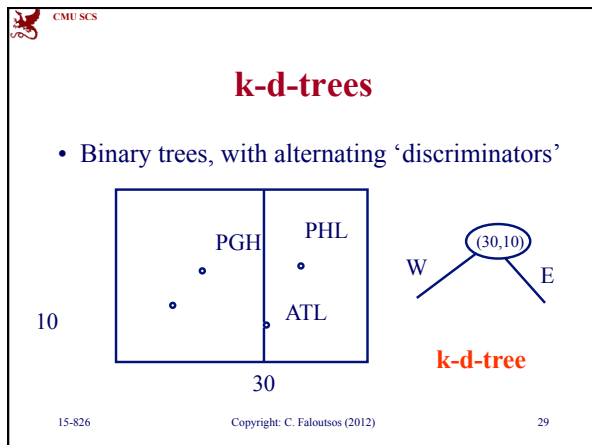
CMU SCS

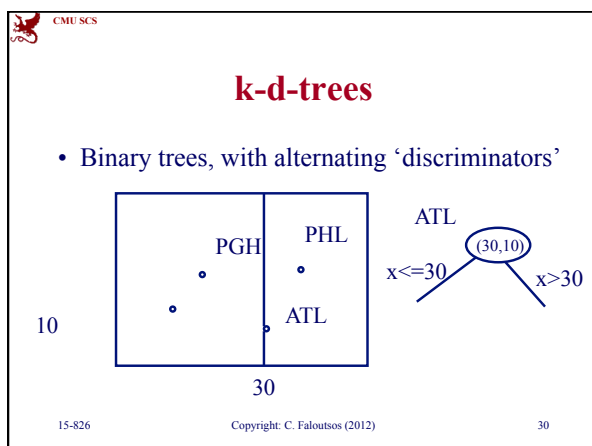
Indexing - Detailed outline

- primary key indexing
- secondary key / multi-key indexing
 - main memory: quad-trees
 - main memory: k-d-trees
- ➡ • spatial access methods
- text
- ...

15-826 Copyright: C. Faloutsos (2012) 27







CMU SCS

k-d-trees

- Binary trees, with alternating ‘discriminators’

15-826 Copyright: C. Faloutsos (2012) 31

CMU SCS

(Several demos/applets, e.g.)

- <http://donar.umiacs.umd.edu/quadtree/points/kdtree.html>

15-826 Copyright: C. Faloutsos (2012) 32

CMU SCS

Indexing - Detailed outline

- primary key indexing
- secondary key / multi-key indexing
 - main memory: quad-trees
 - main memory: k-d-trees
 - insertion; deletion
 - range query; k-nn query
- spatial access methods
- text
- ...

15-826 Copyright: C. Faloutsos (2012) 33

k-d-trees - insertion

- Binary trees, with alternating ‘discriminators’

15-826 Copyright: C. Faloutsos (2012) 34

k-d-trees - insertion

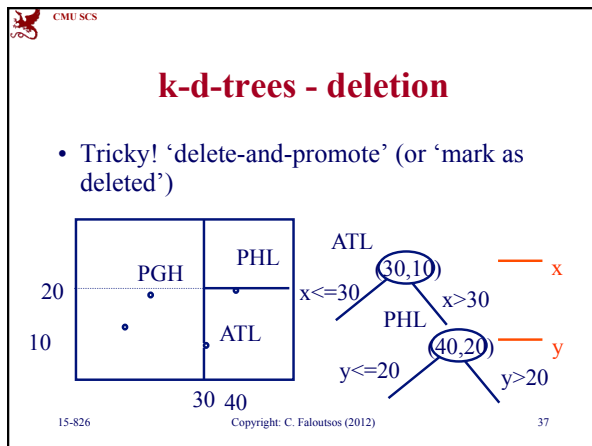
- discriminators: may cycle, or
- Q: which should we put first?

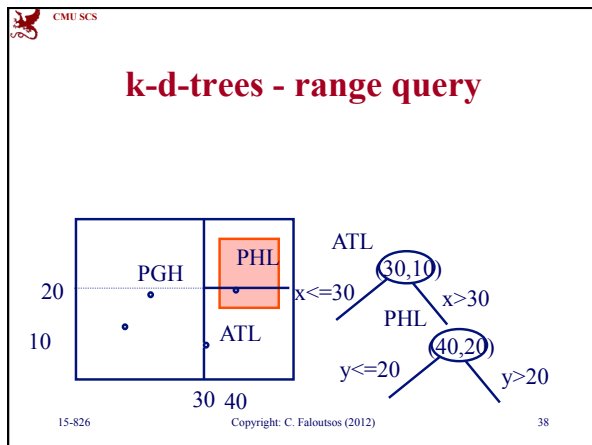
15-826 Copyright: C. Faloutsos (2012) 35

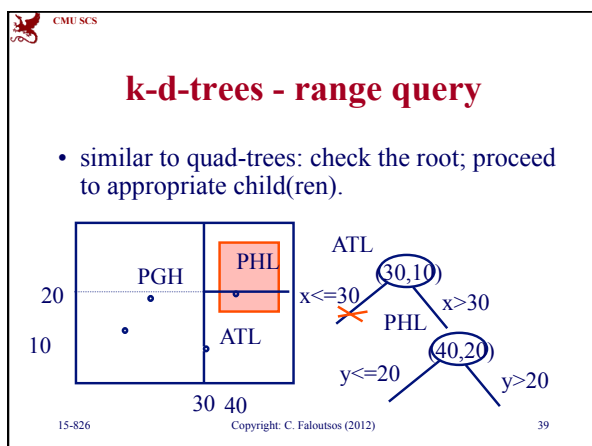
k-d-trees - deletion

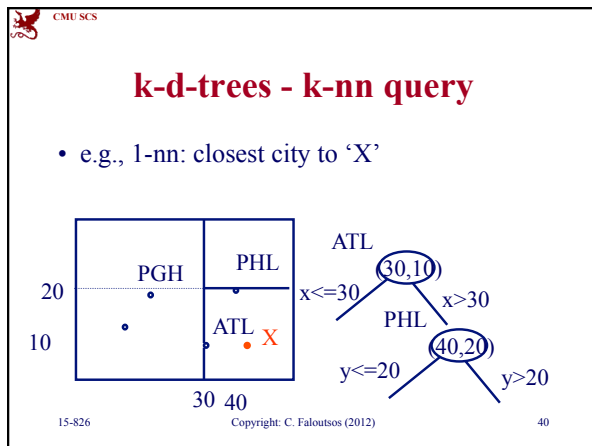
- How?

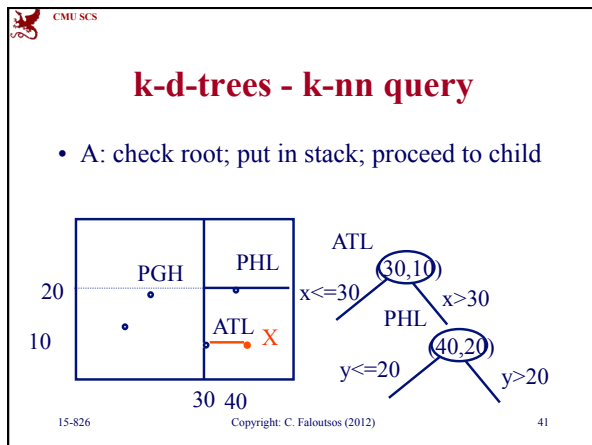
15-826 Copyright: C. Faloutsos (2012) 36

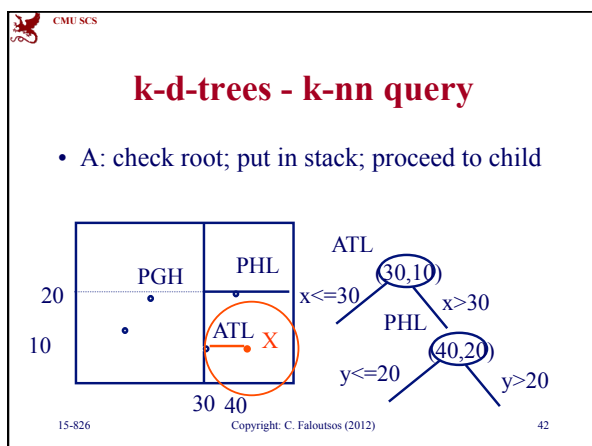














CMU SCS

Indexing - Detailed outline

- primary key indexing
- secondary key / multi-key indexing
 - main memory: quad-trees
 - main memory: k-d-trees
 - insertion; deletion
 - range query; k-nn query
 - discussion
- ➔ spatial access methods
- text

15-826 Copyright: C. Faloutsos (2012) 43



CMU SCS

k-d trees - discussion

- great for main memory & low 'd' ($\sim < 10$)
- Q: what about high-d?
- A:
- Q: what about disk
- A:

15-826 Copyright: C. Faloutsos (2012) 44



CMU SCS

k-d trees - discussion

- great for main memory & low 'd' ($\sim < 10$)
- Q: what about high-d?
- A: most attributes don't ever become discriminators
- Q: what about disk?
- A: Pagination problems, after ins./del. (solutions: next!)

15-826 Copyright: C. Faloutsos (2012) 45



CMU SCS

Conclusions

- sec. keys: B-tree indices (+ postings lists)
- multi-key, main memory methods:
 - quad-trees
 - k-d-trees

15-826 Copyright: C. Faloutsos (2012) 46



CMU SCS

References

- [Bentley75] J.L. Bentley: *Multidimensional Binary Search Trees Used for Associative Searching*, CACM, 18,9, Sept. 1975.
- [Finkel74] R.A. Finkel, J.L. Bentley: *Quadtrees: A data structure for retrieval on composite keys*, ACTA Informatica, 4,1, 1974
- Applet: eg., <http://donar.umiacs.umd.edu/quadtrees/points/kdtree.html>

15-826 Copyright: C. Faloutsos (2012) 47
