


CMU SCS

15-826: Multimedia Databases and Data Mining

Christos Faloutsos
CMU
www.cs.cmu.edu/~christos




CMU SCS

Outline

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

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

15-826 Copyright: C. Faloutsos (2007) 2



CMU SCS

Problem

Given a large collection of (multimedia) records, find similar/interesting things, ie:

- Allow fast, approximate queries, and
- Find rules/patterns

15-826 Copyright: C. Faloutsos (2007) 3

CMU SCS

Sample queries

- Similarity search
 - Find pairs of branches with similar sales patterns
 - find medical cases similar to Smith's
 - Find pairs of sensor series that move in sync
 - Find shapes like a spark-plug
 - (nn: 'case based reasoning')

15-826 Copyright: C. Faloutsos (2007) 4

CMU SCS

Sample queries –cont'd

- Rule discovery
 - Clusters (of branches; of sensor data; ...)
 - Forecasting (total sales for next year?)
 - Outliers (eg., unexpected part failures; fraud detection)

15-826 Copyright: C. Faloutsos (2007) 5

CMU SCS

Outline

Goal: 'Find similar / interesting things'

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

15-826 Copyright: C. Faloutsos (2007) 6

CMU SCS

Detailed Outline

Intro to DB

- ➔ • Relational DBMS - what and why?
 - inserting, retrieving and summarizing data
 - views; security/privacy
 - (concurrency control and recovery)
- Object-Relational DBMS - what and why?

15-826 Copyright: C. Faloutsos (2007) 7

CMU SCS

What is the goal of rel. DBMSs

15-826 Copyright: C. Faloutsos (2007) 8

CMU SCS

What is the goal of rel. DBMSs

Electronic record-keeping:
Fast and convenient access to information.
 Eg.: students, taking classes, obtaining grades;

- find my gpa
- <and other ad-hoc queries>

15-826 Copyright: C. Faloutsos (2007) 9

CMU SCS

Why Databases?

15-826 Copyright: C. Faloutsos (2007) 10

CMU SCS

Why Databases?

- Flexibility
- data independence (can add new tables; new attributes)
- data sharing/concurrency control
- recovery

15-826 Copyright: C. Faloutsos (2007) 11

CMU SCS

Why NOT Databases?

15-826 Copyright: C. Faloutsos (2007) 12

CMU SCS

Why NOT Databases?

- Price
- additional expertise (SQL/DBA)
- over-kill for small data sets

15-826 Copyright: C. Faloutsos (2007) 13

CMU SCS

Main vendors/products

<p><u>Commercial</u></p> <ul style="list-style-type: none"> • Oracle • IBM/DB2 • MS SQL-server • Sybase • (MS Access, • ...) 	<p><u>Open source</u></p> <p>Postgres (UCB)</p> <p>mySQL, mSQL</p> <p>miniBase (Wisc)</p> <p>Predator (Cornell)</p> <p>(www.sigmod.org)</p>
--	---

15-826 Copyright: C. Faloutsos (2007) 14

CMU SCS

Detailed Outline

Intro to DB

- Relational DBMS - what and why?
 - ➡ – inserting, retrieving and summarizing data
 - views; security/privacy
 - (concurrency control and recovery)
- Object-Relational DBMS - what and why?

15-826 Copyright: C. Faloutsos (2007) 15

CMU SCS

How do DBs work?

```
%isql mydb /mydb
sql>create table student (
  ssn fixed;
  name char(20) );
```

student	
ssn	name

15-826 Copyright: C. Faloutsos (2007) 16

CMU SCS

How do DBs work?

```
sql>insert into student
  values (123, "Smith");
sql>select * from student;
```

student	
ssn	name
123	Smith

15-826 Copyright: C. Faloutsos (2007) 17

CMU SCS

How do DBs work?

```
sql>create table takes (
  ssn fixed,
  c-id char(5),
  grade fixed));
```

takes		
ssn	c-id	grade

15-826 Copyright: C. Faloutsos (2007) 18

CMU SCS

How do DBs work - cont'd

More than one tables - [joins](#)
 Eg., roster (names only) for 15-826

student	
ssn	name

takes		
ssn	c-id	grade

15-826 Copyright: C. Faloutsos (2007) 19

CMU SCS

How do DBs work - cont'd

```
sql> select name
      from student, takes
      where student.ssn = takes.ssn
      and takes.c-id = 15-826
```

15-826 Copyright: C. Faloutsos (2007) 20

CMU SCS

SQL-DML

General form:

```
select a1, a2, ... an
from r1, r2, ... rm
where P
[order by ...]
[group by ...]
[having ...]
```

15-826 Copyright: C. Faloutsos (2007) 21

CMU SCS

Aggregation

Find ssn and GPA for each student

student	
ssn	name

takes		
ssn	c-id	grade
123	603	4
123	412	3
234	603	3

15-826 Copyright: C. Faloutsos (2007) 22

CMU SCS

Aggregation

sql> select ssn, grade
from takes;

takes		
ssn		grade
123		4
123		3
234		3

15-826 Copyright: C. Faloutsos (2007) 23

CMU SCS

Aggregation

sql> select ssn, avg(grade)
from takes;

WRONG

15-826 Copyright: C. Faloutsos (2007) 24

CMU SCS

Aggregation

```
sql> select ssn, avg(grade)
      from takes
      group by ssn;
```

takes		
ssn	c-id	grade
123	603	4
123	412	3
234	603	3

ssn	avg(grade)
123	3.5
234	3

15-826 Copyright: C. Faloutsos (2007) 25

CMU SCS

Detailed Outline

Intro to DB

- Relational DBMS - what and why?
 - inserting, retrieving and summarizing data
 - ➡ - views; security/privacy
 - (concurrency control and recovery)
- Object-Relational DBMS - what and why?

15-826 Copyright: C. Faloutsos (2007) 26

CMU SCS

Views - what and why?

- suppose you ONLY want to see ssn and GPA (eg., in your data-warehouse)
- suppose secy is only allowed to see GPAs, but not individual grades
- -> VIEWS!

15-826 Copyright: C. Faloutsos (2007) 27

CMU SCS

Views

sql> create view fellowship as (
 select ssn, avg(grade)
 from takes group by ssn);

takes		
ssn	c-id	grade
123	603	4
123	412	3
234	603	3

ssn	avg(grade)
123	3.5
234	3

15-826 Copyright: C. Faloutsos (2007) 28

CMU SCS

Views

sql> create view fellowship as (
 select ssn, avg(grade)
 from takes group by ssn);

takes		
ssn	c-id	grade
123	603	4
123	412	3
234	603	3

ssn	avg(grade)
123	3.5
234	3

15-826 Copyright: C. Faloutsos (2007) 29

CMU SCS

Views

Views = 'virtual tables'

15-826 Copyright: C. Faloutsos (2007) 30

CMU SCS

Views

sql> select * from fellowship;

takes		
ssn	c-id	grade
123	603	4
123	412	3
234	603	3

ssn	avg(grade)
123	3.5
234	3

15-826 Copyright: C. Faloutsos (2007) 31

CMU SCS

Views

sql> grant select on fellowship to secy;

takes		
ssn	c-id	grade
123	603	4
123	412	3
234	603	3

ssn	avg(grade)
123	3.5
234	3

15-826 Copyright: C. Faloutsos (2007) 32

CMU SCS

Detailed Outline

Intro to DB

- Relational DBMS - what and why?
 - inserting, retrieving and summarizing data
 - views; security/privacy
 - ➔ – (concurrency control and recovery)
- Object-Relational DBMS - what and why?

15-826 Copyright: C. Faloutsos (2007) 33

CMU SCS

Detailed Outline

Intro to DB

- Relational DBMS - what and why?
 - inserting, retrieving and summarizing data
 - views; security/privacy
 - (concurrency control and recovery)
- ➔ • Object-Relational DBMS - what and why?

15-826 Copyright: C. Faloutsos (2007) 34

CMU SCS

Why more than RDBMSs?

- RDBMS: tuples, of numbers + strings
- What apps need only those?

15-826 Copyright: C. Faloutsos (2007) 35

CMU SCS

Why more than RDBMSs?

- RDBMS: tuples, of numbers + strings
- What apps need only those?
 - Banks
 - Airlines
 - Retailer stores
 - ...
- Q: Other apps, with more req's?

15-826 Copyright: C. Faloutsos (2007) 36

CMU SCS

Why more than RDBMS's

- Q: Other apps, with more req's?
- A:
 - text
 - multimedia; financial apps/forecasting
 - Geographic Inf. Sys.
 - CAD/CAM
 - Network management

15-826 Copyright: C. Faloutsos (2007) 37

CMU SCS

Ideally, we'd like to:

- create a new data type (eg., 'image', 'time-sequence')
- define functions on it (like (dist(im1, im2)))
- be able to ask queries like
select * from employee
where dist(employee.face, given-face) <= 10;

15-826 Copyright: C. Faloutsos (2007) 38

CMU SCS

OR DBMSs

traditional DBMS + attempts to provide

- user defined data types
- support for large / complex objects
- (inheritance - ISA hierarchies)

15-826 Copyright: C. Faloutsos (2007) 39

CMU SCS

SQL-3 proposed extensions

- complex types (sets, lists, multisets)
- inheritance (IS-A hierarchies)
- User Defined Functions (UDFs)

15-826 Copyright: C. Faloutsos (2007) 40

CMU SCS

Complex types sample syntax

eg,
create type MyDate (
 day **decimal**(2),
 month **char**(3),
 year **decimal** (4)
);

15-826 Copyright: C. Faloutsos (2007) 41

CMU SCS

BLOBs etc:

- Large objects, eg., video, images, 3d-MRI scans
- new data types: LOB (=Large Object)
 - BLOB: (up to 4Gb; binary: jpeg, mpeg, ...)
 - CLOB: (up to 2Gb; character: english text)
 - NCLOB:(.....; multi-byte characters)

15-826 Copyright: C. Faloutsos (2007) 42

CMU SCS sample syntax

Stored procedures

```

SQL> create or replace procedure del-st-rec
(s-id number) as
begin
  delete from student
  where s-id = ssn;
end del-st-rec;
SQL> execute del-st-rec ( 123 );

```

15-826 Copyright: C. Faloutsos (2007) 43

CMU SCS

Conclusions

- (relational) DBMSs: electronic record keepers
- customize them with **create table** commands
- ask SQL queries to retrieve info

15-826 Copyright: C. Faloutsos (2007) 44


CMU SCS

Conclusions cont'd

main advantages over flat files & scripts:

- logical + physical data independence (ie., flexibility of adding new attributes, new tables and indices)
- concurrency control and recovery for free


15-826 Copyright: C. Faloutsos (2007) 45

 CMU SCS

Conclusions cont'd

- OR-DBMS: user-defined data types (eg., images), and U.D. functions.

15-826 Copyright: C. Faloutsos (2007) 46

 CMU SCS

For more info:

- Microsoft Access: available on ANDREW clusters (PC)
- postgres:
 - www.cs.cmu.edu/~ipandis/15415/F06/HW/PostgreSQL_Readme.htm
- Ramakrishna + Gehrke, 3rd edition
- 15-415 web page, eg,
 - www.cs.cmu.edu/~ipandis/15415/F06/

15-826 Copyright: C. Faloutsos (2007) 47
