The Information School of the University of Washington

Social Computing

INFO/CSE 100, Spring 2005

Fluency in Information Technology

http://www.cs.washington.edu/100

Readings and References

- Reading
  - » *Fluency with Information Technology*
    - Chapters 12

Communicating over IT

- Synchronous communication
  - » Instant messaging
  - » Internet Relay Chat (IRC)
- Asynchronous communication
  - » Email
  - » Bulletin Boards
  - » UseNet news
  - » Blogs
  - » SMS

Problems with Text Communication

- Conveying emotion
  - » Emoticons :-), :D, ^--^
- Too much _emphasis_ ???
- Pace
- Ambiguity
  - » Sarcasm?!?
- Flame Wars
  - » Revenge of the Inconsolable Responder

Email Netiquette

- Only discuss 1 topic at a time
- Use a descriptive subject line
- Limit size and type of attachments
- Don't forward SPAM
- Use vacation messages (automated replies)
- Avoid mass mailing (use group aliases)
- Answer your email from the most current to the least current

Internet Netiquette

- Moderation
  - » And administrative or authoritative person who listens and/or approves communication
- http://www.dtcc.edu/cs/rfc1855.html
  - » Email
  - » Usenet
  - » Chat

Passwords

- Passwords are used to limit computer or software access
- Should be changed on a periodic basis (every 90 days at the UW)
- Forgotten passwords?!?
  - » As the administrator to reset it for you
- Select password topic areas
- Encode password with alternative characters

Intellectual Property

- Software licenses
  - » use
  - » shareware
  - » freeware
- Copyright gives the owner the right to:
  - Make a copy of the work
  - Use for a derivate work
  - Distribute or publish
  - Publicly perform/display

Why Study Databases?

- Some of us want to compute, but all of us want information …
  - Much of the archived information is in tables
  - Databases enhance applications, e.g. Web
  - Once you know how to create databases, you can use them to personal advantage
  - Databases introduce interesting ideas

The Internet Movie Database

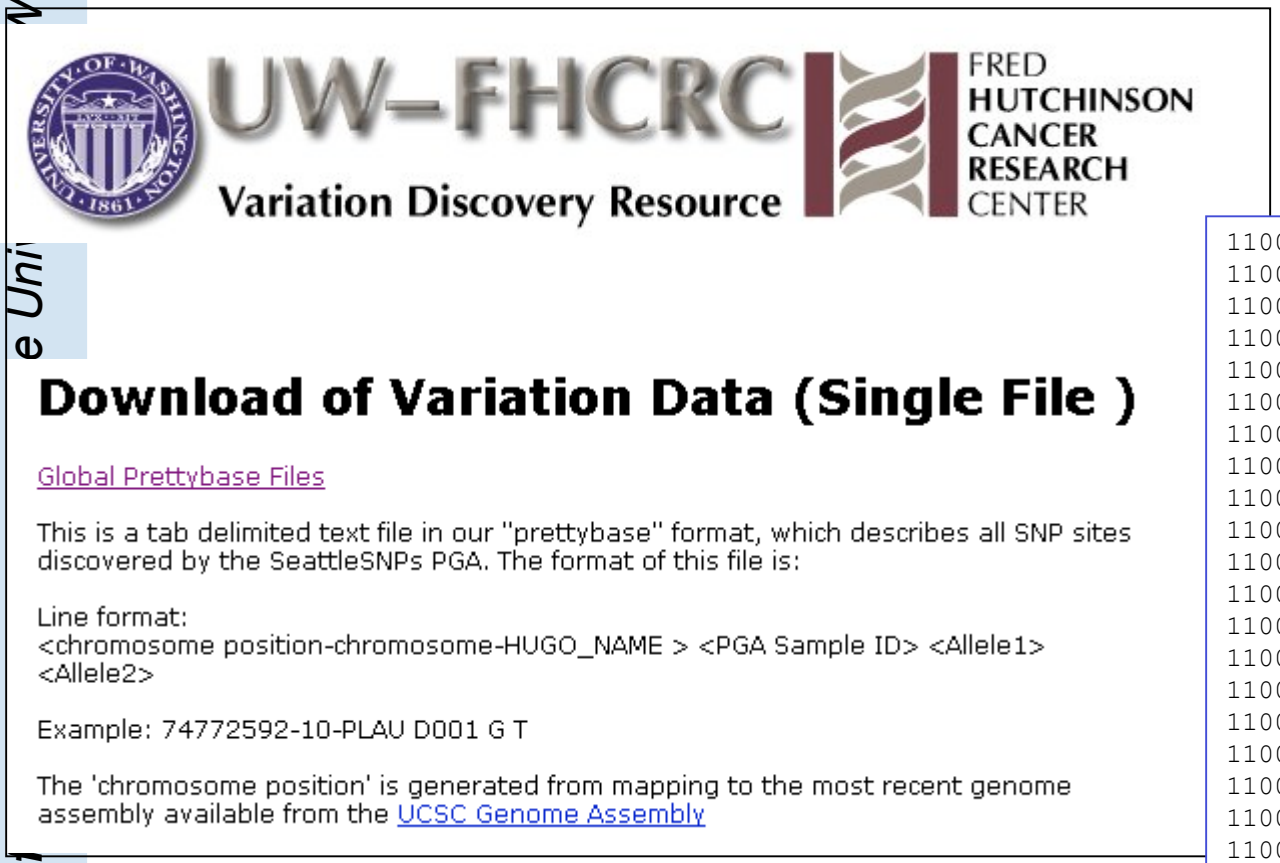
Visited by over 20 million movie lovers each month!

Welcome to the Internet Movie Database, the biggest, best, most award-winning movie site on the planet.

How to organize the data?

- Before relational databases (the kind we study) there were only "flat files"
  - » Structural information is difficult to express
  - » All processing of information is "special cased"
    - custom programs are needed
  - » Information repeated; difficult to combine
  - » Changes in format of one file means all programs that ever process that file must be changed
    - eg, adding ZIP codes

tab-delimited file example




```
1100322-IL3RA-X      D001    N  N
1100322-IL3RA-X      D002    G  G
1100322-IL3RA-X      D003    G  G
1100322-IL3RA-X      D004    G  G
1100322-IL3RA-X      D005    G  G
1100322-IL3RA-X      D006    G  G
1100322-IL3RA-X      D007    G  G
1100322-IL3RA-X      D008    G  G
1100322-IL3RA-X      D009    A  G
1100322-IL3RA-X      D010    N  N
1100322-IL3RA-X      D011    N  N
1100322-IL3RA-X      D012    N  N
1100322-IL3RA-X      D013    G  G
1100322-IL3RA-X      D014    A  G
1100322-IL3RA-X      D015    N  N
1100322-IL3RA-X      D016    N  N
1100322-IL3RA-X      D033    A  G
1100322-IL3RA-X      D034    A  G
1100322-IL3RA-X      D035    G  G
1100322-IL3RA-X      D036    A  G
1100322-IL3RA-X      D037    A  A
1100322-IL3RA-X      D038    G  G
1100322-IL3RA-X      D039    G  G
1100322-IL3RA-X      D040    G  G
...
```

Unix termcap example

```
# FILE FORMAT:
#
# The version you are looking at may be in any of three formats: master
# (terminfo with OT capabilities), stock terminfo, or termcap.  You can
    tell
# which by the format given in the header above.
#
# The master format is accepted and generated by the terminfo tools in the
# ncurses suite; it differs from stock (System V-compatible) terminfo only
# in that it admits a group of capabilities (prefixed `OT') equivalent to
# various obsolete termcap capabilities.
...
# ANSI capabilities are broken up into pieces, so that a terminal
# implementing some ANSI subset can use many of them.
ansi+local1:\
        :do=\E[B:le=\E[D:nd=\E[C:up=\E[A:
ansi+local:\
        :DO=\E[%dB:LE=\E[%dD:RI=\E[%dC:UP=\E[%dA:tc=ansi+local1:
ansi+tabs:\
        :bt=\E[Z:ct=\E[2g:st=\EH:ta=^I:
ansi+inittabs:\
        :it#8:tc=ansi+tabs:
```

The Information School of the University of Washington

Library example

notice the redundancy

ISBN	Title	AuID	AuName	AuPhone	PubID	PubName	PubPhone	Price
1-1111-1111-1	C++	4	Roman	444-444-4444	1	Big House	123-456-7890	$29.95
0-99-999999-9	Emma	1	Austen	111-111-1111	1	Big House	123-456-7890	$20.00
0-91-335678-7	Fairie Queene	7	Spencer	777-777-7777	1	Big House	123-456-7890	$15.00
0-91-045678-5	Hamlet	5	Shakespeare	555-555-5555	2	Alpha Press	999-999-9999	$20.00
0-103-45678-9	Iliad	3	Homer	333-333-3333	1	Big House	123-456-7890	$25.00
0-12-345678-6	Jane Eyre	1	Austen	111-111-1111	3	Small House	714-000-0000	$49.00
0-99-777777-7	King Lear	5	Shakespeare	555-555-5555	2	Alpha Press	999-999-9999	$49.00
0-555-55555-9	Macbeth	5	Shakespeare	555-555-5555	2	Alpha Press	999-999-9999	$12.00
0-11-345678-9	Moby Dick	2	Melville	222-222-2222	3	Small House	714-000-0000	$49.00
0-12-333433-3	On Liberty	8	Mill	888-888-8888	1	Big House	123-456-7890	$25.00
0-321-32132-1	Balloon	13	Sleepy	321-321-1111	3	Small House	714-000-0000	$34.00
0-321-32132-1	Balloon	11	Snoopy	321-321-2222	3	Small House	714-000-0000	$34.00
0-321-32132-1	Balloon	12	Grumpy	321-321-0000	3	Small House	714-000-0000	$34.00
0-55-123456-9	Main Street	10	Jones	123-333-3333	3	Small House	714-000-0000	$22.95
0-55-123456-9	Main Street	9	Smith	123-222-2222	3	Small House	714-000-0000	$22.95
0-123-45678-0	Ulysses	6	Joyce	666-666-6666	2	Alpha Press	999-999-9999	$34.00
1-22-233700-0	Visual Basic	4	Roman	444-444-4444	1	Big House	123-456-7890	$25.00

from Access Database book, Steve Roman

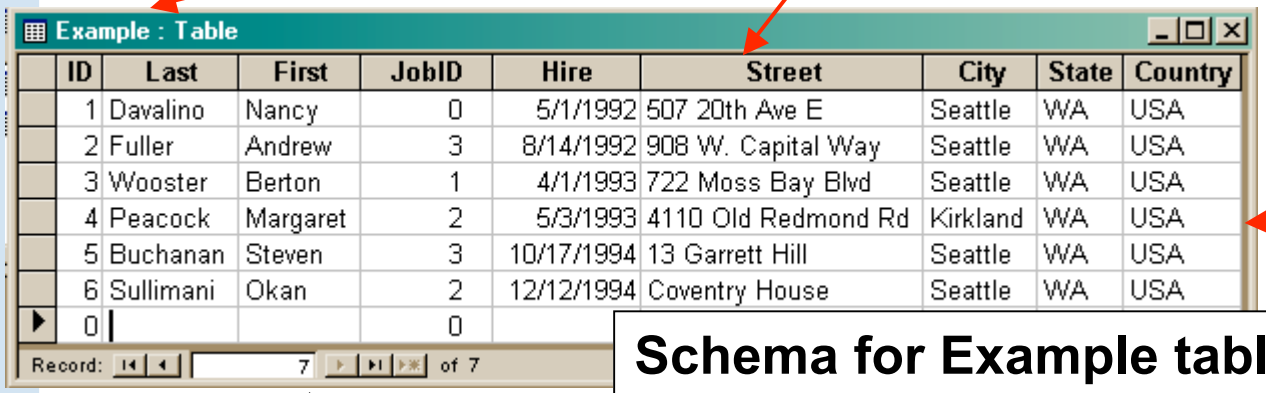
Relational Databases

- Information is stored in tables
  - » Tables store information about *entities*
  - » Entities have characteristics called *attributes*
  - » Each row in a table represents a single entity
    - Each row is a set of attribute values
    - Every row must be unique, identified by a key
  - » Relationships -- associations among the data values are stored

Table structure = schema
Table contents = instance

A Table in a Database

Tables have names, attributes, rows



Example : Table

ID	Last	First	JobID	Hire	Street	City	State	Country
1	Davalino	Nancy	0	5/1/1992	507 20th Ave E	Seattle	WA	USA
2	Fuller	Andrew	3	8/14/1992	908 W. Capital Way	Seattle	WA	USA
3	Wooster	Berton	1	4/1/1993	722 Moss Bay Blvd	Seattle	WA	USA
4	Peacock	Margaret	2	5/3/1993	4110 Old Redmond Rd	Kirkland	WA	USA
5	Buchanan	Steven	3	10/17/1994	13 Garrett Hill	Seattle	WA	USA
6	Sullimani	Okan	2	12/12/1994	Coventry House	Seattle	WA	USA
0			0					

Record: 7 of 7

instance

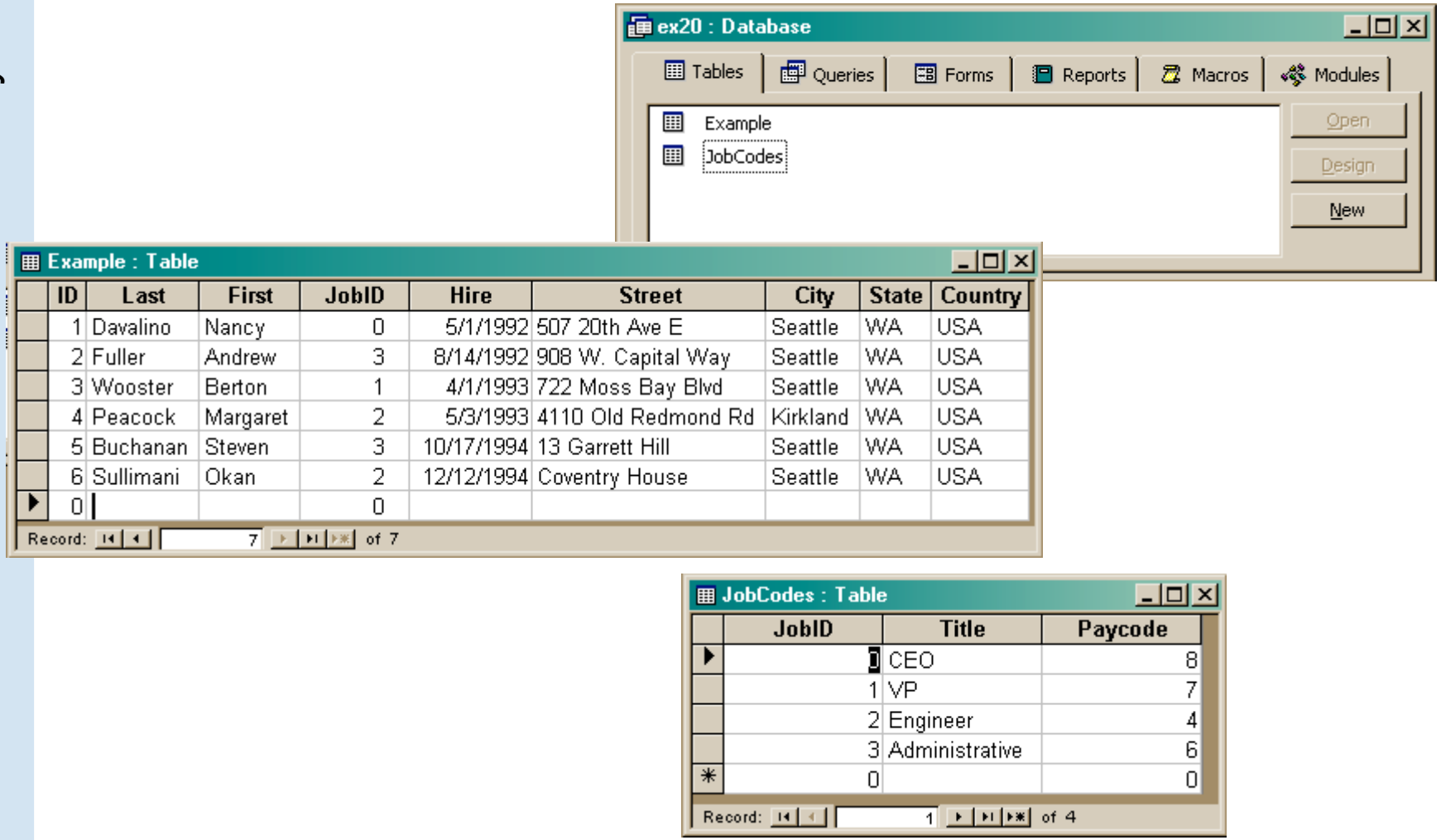
Schema for Example table:

ID	number	unique number(Key)
Last	text	person's last name
First	text	person's first name
JobCode	number	current position
Hire	date	first day on job
...		

schema

The Information School of the University of Washington

Two tables in a database



tion School of the University of Washington

Redundancy in a database is Very Bad

- Not every assembly of tables is a good database
- Repeating data is a bad idea
  - » Replicated data can differ in its different locations, e.g. multiple addresses can differ
    - Inconsistent data is worse than no data
  - » Keep a *single copy* of any data
    - if it is needed in multiple places, associate it with a key and store key rather than the data

Relationships between tables

Perms : Table

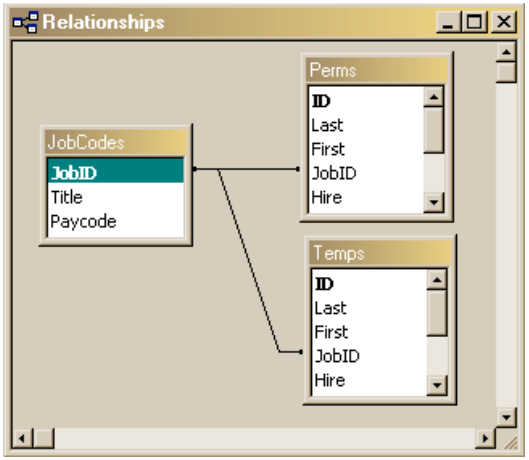
ID	Last	First	JobID	Hire	Street	City	State	Country
1	Davalino	Nancy	0	01-May-92	507 20th Ave E	Seattle	WA	USA
2	Fuller	Andrew	3	14-Aug-92	908 W. Capital Way	Seattle	WA	USA
3	Wooster	Berton	1	01-Apr-93	722 Moss Bay Blvd	Seattle	WA	USA
4	Peacock	Margaret	2	03-May-93	4110 Old Redmond Rd	Kirkland	WA	USA
5	Buchanan	Steven	3	17-Oct-94	13 Garrett Hill	Seattle	WA	USA
6	Sullimani	Okan	2	12-Dec-94	Coventry House	Seattle	WA	USA
0			0					

Record: 1 of 6

JobCodes : Table

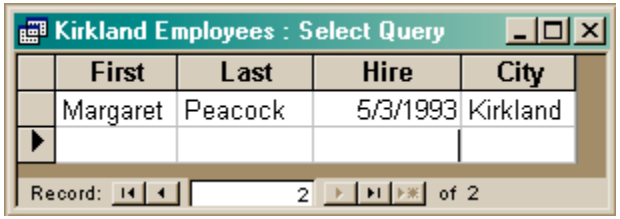
JobID	Title	Paycode
0	CEO	8
1	VP	7
2	Engineer	4
3	Administrative	6
0		0

Record: 1 of 4



"You can look it up"

- When looking for information, a single item might be the answer, but a table is more likely
  - » Which employees live in Kirkland?
    - Table of employees
  - » Who is taking INFO/CSE 100?
    - Table of students
  - » Whose mile run time ≤ 4:00?
    - Table of runners



Relational Algebra: Tables From Tables

There are five basic "algebraic" operations on tables:

- Select -- pick rows from a table
- Project -- pick columns from a table
- Union -- combine two tables w/like columns
- Difference -- remove one table from another
- Product -- create "all pairs" from two tables

From this basis, many more complicated operations can be built up