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
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

## Download of Variation Data (Single File)

**Global Prettybase Files**

This is a tab delimited text file in our "prettybase" format, which describes all SNP sites discovered by the SeattleSNPs PGA. The format of this file is:

Line format:

<chromosome position-chromosome-HUGO_NAME > <PGA Sample ID> <Allele1> <Allele2>

Example: 74772592-10-FLAU D001 G T

The 'chromosome position' is generated from mapping to the most recent genome assembly available from the [UCSC Genome Assembly](http://genome.ucsc.edu).

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Sample ID</th>
<th>Allele1</th>
<th>Allele2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1100322-IL3RA-X</td>
<td>D001</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D002</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D003</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D004</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D005</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D006</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D007</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D008</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D009</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D010</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D011</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D012</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D013</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D014</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D015</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D016</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D033</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D034</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D035</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D036</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D037</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D038</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D039</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>1100322-IL3RA-X</td>
<td>D040</td>
<td>G</td>
<td>G</td>
</tr>
</tbody>
</table>

...
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+local:\
   :do=\E[B:le=\E[D:nd=\E[C:up=\E[A:
ansi+local:\
   :DO=\E[%dB:LE=\E[%D:RI=\E[%dC:UP=\E[%dA:tc=ansi+local:
ansi+tabs:\
   :bt=\E[Z:ct=\E[2g:st=\E[H:ta=^I:
ansi+inittabs:\
   :it#8:tc=ansi+tabs:
Library example

<table>
<thead>
<tr>
<th>ISBN</th>
<th>Title</th>
<th>AuID</th>
<th>AuName</th>
<th>AuPhone</th>
<th>Pub ID</th>
<th>PubName</th>
<th>PubPhone</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1111-1111-1</td>
<td>C++</td>
<td>4</td>
<td>Roman</td>
<td>444-444-4444</td>
<td>1</td>
<td>Big House</td>
<td>123-456-7890</td>
<td>$29.95</td>
</tr>
<tr>
<td>0-99-999999-9</td>
<td>Emma</td>
<td>1</td>
<td>Austen</td>
<td>111-111-1111</td>
<td>1</td>
<td>Big House</td>
<td>123-456-7890</td>
<td>$20.00</td>
</tr>
<tr>
<td>0-91-335678-7</td>
<td>Fairie Queene</td>
<td>7</td>
<td>Spencer</td>
<td>777-777-7777</td>
<td>1</td>
<td>Big House</td>
<td>123-456-7890</td>
<td>$15.00</td>
</tr>
<tr>
<td>0-91-045678-3</td>
<td>Hamlet</td>
<td>3</td>
<td>Shakespeare</td>
<td>555-555-5555</td>
<td>2</td>
<td>Alpha Press</td>
<td>999-999-9999</td>
<td>$20.00</td>
</tr>
<tr>
<td>0-103-45678-9</td>
<td>Iliad</td>
<td>3</td>
<td>Homer</td>
<td>333-333-3333</td>
<td>1</td>
<td>Big House</td>
<td>123-456-7890</td>
<td>$25.00</td>
</tr>
<tr>
<td>0-12-345678-6</td>
<td>Jane Eyre</td>
<td>1</td>
<td>Austen</td>
<td>111-111-1111</td>
<td>3</td>
<td>Small House</td>
<td>714-000-0000</td>
<td>$49.00</td>
</tr>
<tr>
<td>0-99-777777-7</td>
<td>King Lear</td>
<td>5</td>
<td>Shakespeare</td>
<td>555-555-5555</td>
<td>2</td>
<td>Alpha Press</td>
<td>999-999-9999</td>
<td>$49.00</td>
</tr>
<tr>
<td>0-555-55555-9</td>
<td>Macbeth</td>
<td>5</td>
<td>Shakespeare</td>
<td>555-555-5555</td>
<td>2</td>
<td>Alpha Press</td>
<td>999-999-9999</td>
<td>$12.00</td>
</tr>
<tr>
<td>0-11-345678-9</td>
<td>Moby Dick</td>
<td>2</td>
<td>Melville</td>
<td>222-222-2222</td>
<td>3</td>
<td>Small House</td>
<td>714-000-0000</td>
<td>$49.00</td>
</tr>
<tr>
<td>0-12-333433-3</td>
<td>On Liberty</td>
<td>8</td>
<td>Mill</td>
<td>888-888-8888</td>
<td>1</td>
<td>Big House</td>
<td>123-456-7890</td>
<td>$25.00</td>
</tr>
<tr>
<td>0-321-32132-1</td>
<td>Balloon</td>
<td>13</td>
<td>Sleepy</td>
<td>321-321-1111</td>
<td>3</td>
<td>Small House</td>
<td>714-000-0000</td>
<td>$34.00</td>
</tr>
<tr>
<td>0-321-32132-1</td>
<td>Balloon</td>
<td>11</td>
<td>Snoopy</td>
<td>321-321-2222</td>
<td>3</td>
<td>Small House</td>
<td>714-000-0000</td>
<td>$34.00</td>
</tr>
<tr>
<td>0-321-32132-1</td>
<td>Balloon</td>
<td>12</td>
<td>Grumpy</td>
<td>321-321-0000</td>
<td>3</td>
<td>Small House</td>
<td>714-000-0000</td>
<td>$34.00</td>
</tr>
<tr>
<td>0-55-123456-9</td>
<td>MainStreet</td>
<td>10</td>
<td>Jones</td>
<td>123-333-3333</td>
<td>3</td>
<td>Small House</td>
<td>714-000-0000</td>
<td>$22.95</td>
</tr>
<tr>
<td>0-55-123456-9</td>
<td>MainStreet</td>
<td>9</td>
<td>Smith</td>
<td>122-222-2222</td>
<td>3</td>
<td>Small House</td>
<td>714-000-0000</td>
<td>$22.95</td>
</tr>
<tr>
<td>0-123-45678-0</td>
<td>Ulysses</td>
<td>6</td>
<td>Joyce</td>
<td>666-666-6666</td>
<td>2</td>
<td>Alpha Press</td>
<td>999-999-9999</td>
<td>$34.00</td>
</tr>
<tr>
<td>1-22-233700-0</td>
<td>Visual Basic</td>
<td>4</td>
<td>Roman</td>
<td>444-444-4444</td>
<td>1</td>
<td>Big House</td>
<td>123-456-7890</td>
<td>$25.00</td>
</tr>
</tbody>
</table>

from Access Database book, Steve Roman

notice the redundancy
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

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

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

...
Two tables in a database
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
“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 $\leq 4:00$?
    • 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