Introduction to Informatics

Lecture 26:
Information Technology in the Real World

Databases
NO MORE LABS !!!
Exam Schedule

- 11595
  - Midterm
    - March 1\textsuperscript{st} (Thursday)
      - Regular Class time
  - Final Exam
    - May 3\textsuperscript{rd} (Thursday)
      - 7:15-9:15 p.m.
Readings until now

Lecture notes
- Posted online
  - [http://informatics.indiana.edu/rocha/i101](http://informatics.indiana.edu/rocha/i101)
    - *The Nature of Information*
    - *Technology*
    - *Modeling the World*

- [@ infoport](http://infoport.blogspot.com)
  - [http://infoport.blogspot.com](http://infoport.blogspot.com)

From course package
  - Chapters 1, 4 (pages 1-12)
  - Chapter 10 (pages 13-17)
- From Andy Clark's book *Natural-Born Cyborgs*
  - Chapters 2 and 6 (pages 19 - 67)
- From Irv Englander's book *The Architecture of Computer Hardware and Systems Software*
  - Chapter 3: Data Formats (pp. 70-86)
  - Chapter 2: Classical Logic (pp. 87-97)
  - Chapter 3: Classical Set Theory (pp. 98-103)
  - Chapters 1-3 (pages 105-129)
  - OPTIONAL: Chapter 4 (pages 131-136)
  - Chapter 13 (pages 147-155)
  - Chapter 5 (pages 141-144)
- Igor Aleksander, "Understanding Information Bit by Bit"
  - Pages 157-166
- Ellen Ullman, "Dining with Robots"
  - Pages 167-172
Assignment Situation

**Labs**

- **Past**
  - **Lab 1:** Blogs
    - Closed (Friday, January 19): Grades Posted
  - **Lab 2:** Basic HTML
    - Closed (Wednesday, January 31): Grades Posted
  - **Lab 3:** Advanced HTML: Cascading Style Sheets
    - Closed (Friday, February 2): Grades Posted
  - **Lab 4:** More HTML and CSS
    - Closed (Friday, February 9): Grades Posted
  - **Lab 5:** Introduction to Operating Systems: Unix
    - Closed (Friday, February 16): Grades Posted
  - **Lab 6:** More Unix and FTP
    - Closed (Friday, February 23): Grades Posted
  - **Lab 7:** Logic Gates
    - Closed (Friday, March 9): Grades Posted
  - **Lab 8:** Intro to Statistical Analysis using Excel
    - Closed (Friday, March 30): Grades Posted
  - **Lab 9:** Data analysis with Excel (linear regression)
    - Closed (Friday, April 6): Grades Posted
  - **Lab 10:** Simple programming in Excel and Measuring Uncertainty
    - April 12 and 13, Due April 20

**Assignments**

- **Individual**
  - **First installment**
    - Closed: February 9: Grades Posted
  - **Second Installment**
    - Past: March 2: Grades Posted
  - **Third Installment**
    - Past: Grades Posted
  - **Fourth Installment**
    - Presented April 10th, Due April 20th

- **Group**
  - **First Installment**
    - Past: March 9th, graded
  - **Second Installment**
    - Past: April 6th Graded
  - **Third Installment**
    - Presented Thursday, April 12; Due Friday, April 27
Step by step analysis of “dying” squares
- **4th Installment**
  - Presented: April 10th
  - Due: April 20th

Use inductive and deductive reasoning
- To uncover the algorithm in each quadrant
  - Build from inductive knowledge accumulated so far

<table>
<thead>
<tr>
<th>Q1</th>
<th>Q2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q3</th>
<th>Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

Cycles = 1
Summary of Black Box

- **Quadrant 1**
  - At the random initial state
    - All numbers have equal probability of being initially present
    - But the probability of changes are different
  - In Any State
    - Any number changes depending on its neighbors
    - It ‘gravitates’ towards the smallest number that it ‘sees’ most often.
    - Odd and Even numbers do not show different behavior

- What is the Algorithm?
Summary of Black Box

- **Quadrant 3**
  - At the random initial state
    - All numbers have equal probability of being initially present
    - But the probability of changes are different
  - In Any State
    - 0 can only change to 0
    - 5 can only change to 5 or 0
    - Even digits always change to even digits
    - Odd digits could change to any other digit

- What is the Algorithm?

<table>
<thead>
<tr>
<th>n(i)</th>
<th>p(i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
</tr>
</tbody>
</table>
Summary of Black Box

- Quadrant 2
  - At the random initial state
    - All numbers have equal probability of being initially present
    - But the probability of changes are different

- In Any State
  - 0 can only change to 0
  - 5 can only change to 5 or 0
  - Even digits always change to even digits
  - Odd digits could change to any other digit

- What is the Algorithm?

1. $0 \rightarrow 0$
2. $\{5\} \rightarrow \{0, 5\}$
3. $\{2, 4, 6, 8\} \rightarrow \{0, 2, 4, 6, 8\}$
4. $\{1, 3, 7, 9\} \rightarrow \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$
### Possible Operations Q2 and Q3

<table>
<thead>
<tr>
<th>Operator</th>
<th>Meaning</th>
<th>Excel</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>()</td>
<td>Brackets, grouping</td>
<td>()</td>
<td>y = (a + b) * (c + d)</td>
</tr>
<tr>
<td>*</td>
<td>Multiplication</td>
<td>*</td>
<td>i = j * k</td>
</tr>
<tr>
<td>+</td>
<td>Add</td>
<td>+</td>
<td>i = i + 1</td>
</tr>
<tr>
<td>-</td>
<td>Subtract</td>
<td>-</td>
<td>i = j - 3.2</td>
</tr>
<tr>
<td>/</td>
<td>Real division</td>
<td>/</td>
<td>i = 8 / 5 = 1.6</td>
</tr>
<tr>
<td>div</td>
<td>Integer division</td>
<td>Quotient (a,b)</td>
<td>i = 8 / 5 = 1</td>
</tr>
<tr>
<td>Mod, %</td>
<td>remainder</td>
<td>Mod (a, b)</td>
<td>i = 8 mod 5 = 3</td>
</tr>
<tr>
<td>ROUND</td>
<td>Rounds</td>
<td>ROUND (a, d)</td>
<td>i = ROUND(3.67, 0) = 4</td>
</tr>
<tr>
<td>INT</td>
<td>Integer Part</td>
<td>INT</td>
<td>i = INT(3.67) = 3</td>
</tr>
<tr>
<td>rand</td>
<td>Random number</td>
<td>Rand()</td>
<td>i = rand(n)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RandBetween(a,b)</td>
<td></td>
</tr>
</tbody>
</table>
Tip for Individual Assignment

- **Quadrant Q**
  - There are 100 cells in each 10x10 quadrant
    - C = 1...100
  - Each cell can take one of 10 colors
    - V(C) = 0..9
      - Is the value of the cell
      - This is the state cell C is in
  - Random initialization of quadrant Q at cycle 1
    - For c=1 to 100 do
      - V(C) ← randbetween(0,9) {random number 0 to 9}
    - EndFor
    - Cycle ← 1
  - Run for Number of cycles
    - n ← Input dialog
    - For k=1 to n do
      - Cycle ← cycle+1
      - {Pick random cell}
      - C ← randbetween(1,100)
      - {Update the value of the cell (NOT THE REAL THING)}
      - V(C) ← ((V(C) * randbetween(0,9)) div 2) - 5*X
    - EndFor
  - X may be a hidden variable
    - X ← ????
The Modeling Relation

Hertz’ Modeling Paradigm

- Organizing Data
  - After encoding
  - Modern Problems Require large storage capabilities
The Entity-Relationship Model

- Conceptual Data Model
  - A kind of “pseudocode” for models of data storage

What should we consider?
- What are the interesting entities and relationships in our model of reality?
- What information about these entities and relationships do we need to store?
- What are the reality constraints and rules that must hold?

Adapted from Yuqing Melanie Wu (I308: Information Representation)
Entities in Data Modeling

- Objects, people, places
  - Basically *a noun*: a discrete object
    - Choose a meaningful name
  - Represented by a rectangle
  - Attributes
    - Describe the properties of an entity

Adapted from Yuqing Melanie Wu
(1308: Information Representation)
Relationships in Data Modeling

- **Relationship:**
  - An association among two or more entities.
  - **Verbs**

```
Customer  Buy  Book
```

- **Attributes also describe relationships**

Adapted from Yuqing Melanie Wu
(I308: Information Representation)
Arity of Relationships

- The number of entities participate in a relationship
  - Binary, ternary, N-ary

Adapted from Yuqing Melanie Wu (1308: Information Representation)
ER Data Modeling Example

From Carol Brown
Try this at home...

How to represent the following?

- A book can have no more than 5 authors.
- A customer has to specify the shipping option.
- Each branch has only one manager.

**Book**
- Title
- ISBN
- Author1
- Author2
- Author3
- Author4
- Author5

**Customer**
- Name
- Address
- Manager

**Ship**
- Date/time
- Shipping option

**Branch**
- Name
- Address
- Manager

**Book**
The Relational Database Model

- Relational database management system (RDBMS)
  - Most popular commercial database type.
  - a data model based on **logic** and **set theory**.
- invented by Ted Codd in 1970
  - Oxford, IBM, U. Michigan, IBM
- System R
  - IBM's San Jose research center
  - Structured English Query Language ("SEQUEL")
    - Data Manipulation Language (DML)
  - SEQUEL was later condensed to SQL due to a trademark dispute
  - In 1979, Relational Software, Inc. (now Oracle Corporation) introduced the first commercially available implementation of SQL
The Relational Model

- All data are represented as mathematical relations
  - Represent the presence of association, interaction or interconnectedness between the elements of two or more sets.
  - A relation associates the elements of 2 or more sets
    - Set of books with sets of attributes (entities)
    - Set of purchases with sets of attributes (relationships)
- Tables store relations
The Relational Database Model

- A relational database is a collection of tables
  - 2-dimensional
- Each table has a unique name in the database.
- Tables define Relations
  - Columns (number of sets)
    - Attributes plus key (primary set)
  - Row (number of relation instances)
    - A table is a set of rows: tuples

### CDs

<table>
<thead>
<tr>
<th>ID</th>
<th>Title</th>
<th>Artist</th>
</tr>
</thead>
<tbody>
<tr>
<td>3592</td>
<td>Yes I am a Witch</td>
<td>Yoko Ono</td>
</tr>
<tr>
<td>2678</td>
<td>Big</td>
<td>Macy Gray</td>
</tr>
<tr>
<td>0623</td>
<td>Sound of Silver</td>
<td>LCD Soundsystem</td>
</tr>
<tr>
<td>0321</td>
<td>Welcome to Planet Sexor</td>
<td>Tiga</td>
</tr>
<tr>
<td>8854</td>
<td>Transparent Things</td>
<td>Fujiya &amp; Miyagi</td>
</tr>
</tbody>
</table>
### Example

**Relation (table)**

**Attributes (columns)**

**Customer**

<table>
<thead>
<tr>
<th>Phone</th>
<th>Name</th>
<th>Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>812-123-4567</td>
<td>Tom</td>
<td>408 3rd st. Bloomington, IN</td>
</tr>
<tr>
<td>812-304-2378</td>
<td>Bill</td>
<td>#113, Redbud Hall, Bloomington, IN</td>
</tr>
<tr>
<td>812-856-1190</td>
<td>Kate</td>
<td>1205, Maritime ct. Bloomington, IN</td>
</tr>
<tr>
<td>812-754-9567</td>
<td>Mary</td>
<td>#901 10th St. Bloomington, IN</td>
</tr>
<tr>
<td>317-897-4536</td>
<td>Pam</td>
<td>2400 Rd135, Greenwood, IN</td>
</tr>
<tr>
<td>812-906-2486</td>
<td>Jeff</td>
<td>#208 Union Ave. Bloomington, IN</td>
</tr>
</tbody>
</table>

**Degree** = 3  
**Cardinality** = 6  

From Yuqing Melanie Wu (L308: Information Representation)
Schema and Instance

- **Database schema**
  - Metadata or Model
  - The logical design of a database
    - E.g. using the *entity-relationship model*
      - *Entity* → *Table*
      - *Attribute* → *Columns*
      - *Relationship* → *Table*
  - Specifies names of tables/relations (*entities and relationships*), plus names and types of each column (*attributes*)

- **Database instance**
  - A snapshot of the data in the database at a given instant in time.

Adapted from Yuqing Melanie Wu (I308: Information Representation)
The identifying labels for the elements of the primary set of a table

- Every instance (row) in the database must have a distinct primary key
- Every instance in the database must have a particular (non-null) value for the primary key.

---

**Customer**

<table>
<thead>
<tr>
<th>Phone</th>
<th>Name</th>
<th>Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>812-123-4567</td>
<td>Tom</td>
<td>408 3rd st. Bloomington, IN</td>
</tr>
<tr>
<td>812-304-2378</td>
<td>Bill</td>
<td>#113, Redbud Hall, Bloomington, IN</td>
</tr>
<tr>
<td>.....</td>
<td>.....</td>
<td>.....</td>
</tr>
</tbody>
</table>

Adapted from Yuqing Melanie Wu (I308: Information Representation)
### Customer

<table>
<thead>
<tr>
<th>Phone</th>
<th>Name</th>
<th>Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>812-123-4567</td>
<td>Tom</td>
<td>......</td>
</tr>
<tr>
<td>812-304-2378</td>
<td>Bill</td>
<td>......</td>
</tr>
<tr>
<td>812-856-1190</td>
<td>Kate</td>
<td>......</td>
</tr>
</tbody>
</table>

### Book

<table>
<thead>
<tr>
<th>ISBN</th>
<th>Title</th>
<th>Publisher</th>
</tr>
</thead>
<tbody>
<tr>
<td>12345</td>
<td>Java</td>
<td>MIT press</td>
</tr>
<tr>
<td>49082</td>
<td>Snow White</td>
<td>......</td>
</tr>
<tr>
<td>72936</td>
<td>Honeymoon</td>
<td>......</td>
</tr>
</tbody>
</table>

### Sale

<table>
<thead>
<tr>
<th>ISBN</th>
<th>Phone</th>
<th>Price</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>12345</td>
<td>812-123-4567</td>
<td>$20</td>
<td>Feb 2, 05</td>
</tr>
<tr>
<td>49082</td>
<td>812-123-4567</td>
<td>$25</td>
<td>Dec 20, 04</td>
</tr>
<tr>
<td>12345</td>
<td>812-856-1190</td>
<td>$19</td>
<td>......</td>
</tr>
</tbody>
</table>

Primary keys in entities: ISBN, Phone, Name, Address

Primary keys in relationships: ISBN, Phone, time, price

Come from primary keys in entities

Adapted from Yuqing Melanie Wu (I308: Information Representation)
Example

From Yuqing Melanie Wu (I 308: Information Representation)
Structured Query Language (SQL)

- The most popular computer language used to create, modify and retrieve data from relational database management systems. (Wikipedia)

- Three subsets of SQL
  - Data Definition Language (DDL)
  - Data Manipulation Language (DML)
  - Data Control Language (DCL) (for authorization)
Data Definition Language

- Used to create, alter, and delete databases and tables.

- Statements
  - Create Table
    - CREATE TABLE table_name (column_name1 data_type primary key, column_name2 data_type);
  - Some other operations
    - “alter” and “drop”
Data Manipulation Language

- Used to retrieve, insert, delete and update data in a database

- Statements
  - Select
    - Selects rows (records) according to attribute criteria
      - E.g. Select CDs published in YEAR=x
  - Some other operations
    - “insert”, “update”, “delete”, and “truncate”
Select Statement

- **Select**
  - Selects rows (records) according to attribute criteria
    - E.g. papers published in YEAR=x
  - **SELECT** * FROM list-of-relations WHERE condition
    - **SELECT** * FROM CITATION_TABLE WHERE PUBLISHED_YEAR='1995';
  - * Denotes ALL
  - **SELECT** * FROM T;
    - Returns all elements of all the rows of the table T
Projection Operation

- **Project**
  - Extracts columns
  - E.g. projects a set of papers into a reduced set of attributes.

- **SELECT** \( C1, C7 \) FROM \( T; \)
Join Operation

- **Join**
  - Merges records that contain matching values for specified attributes
  - given a key value join records from both tables

- **SELECT * FROM employee, department;**
- **SELECT * FROM citation-table, author-table WHERE citation-table.MUID = author-table.MUID;**
AND YOUR AUTOGRAPH, PLEASE...

RIGHT

MAKE IT OUT TO “CYNTHIA MY NUMBER ONE FAN”

AND HERE'S YOUR CUSTOMER CARD!

THANKS

HAVE A NICE DAY, MRS. SLUG

THANKS

HAVE A SAFE TRIP BACK TO 1320 WEATHERWAX, CRESCENT; ENJOY A MEAL OF MACARONI AND CHEESE OR QUESADILLAS

READ A GOOD DETECTIVE NOVEL; HAVE A FRAGRANT BUBBLE BATH

WATCH “SURVIVOR” AND HIT THE SACK—BY THE WAY, YOUR HAIR COLOR GOES ON SALE NEXT WEDNESDAY

STILL TOO MUCH?

I SAY JUST STICK WITH THEIR NAMES. USING ANY MORE INFO TENDS TO GIVE CUSTOMERS THE HEEBIE JEEBIES
Group Assignment

Third Installment

Given any text such as the *library of babylon* or *Funes, the memorious*

- Create a **database model** and a **relational database instance** using *Microsoft Access* to store the data and conclusions from previous installments
  - Use the entity-relationship model
  - Examples of items that should appear
    - Title, author, language, publication date
    - Frequency/probability of each letter
    - Conditional probabilities for letters ‘e’ and ‘u’ (as produced in installment 2)
    - Positively and negatively dependent letters

- Use at least 4 texts

Due on April 27th, 2005

Upload to Oncourse
Next Class!

- Topics of next classes
  - Databases and SQL
  - Individual Assignment
  - Review
- Readings for Next week
  - @ infoport
  - course package
- No More Labs!!!!!!!!