Information Systems

Managing Information

- Information system—software that helps users organize and analyze data
- Types of IS desktop software include electronic spreadsheets and database management systems
  - Software tools that allow the user to organize, manage, and analyze data is various ways

What Is Information Systems?
- http://www.youtube.com/watch?v=fJYFpwbuDc&feature=related

Spreadsheets

- Software application that allows users to organize and analyze data using a grid of labeled cells
  - A cell can contain data or a formula that is used to calculate a value
  - Data stored in a cell can be text, numbers, or “special” data such as dates
  - Spreadsheet cells are referenced by their row and column designation

Spreadsheet Formulas

- The power of spreadsheets comes from the formulas that we can create and store in cells
  - When a formula is stored in a cell, the result of the formula is displayed in the cell
  - If the spreadsheet is set up correctly:
    - Tutors can be added or removed
    - Additional weeks of data can be added
    - Any stored data can be changed and the corresponding calculations would automatically be updated

Spreadsheet Formulas

- Formulas make use of basic arithmetic operations using standard keyboard symbols (+, -, *, and /)
  - Formulas always begin with the equal to (=) operator, i.e., $= G18+G19–D13$
- Spreadsheet functions—internal computations provided by the spreadsheet software that can be incorporated into formulas
  - Sum, Count, Today, etc.
- Range—a set of contiguous cells specified by the endpoints, i.e., C4..E4

Spreadsheet Functions

- SUM(val1, val2, ...) or SUM(range)
  - Sum of the specified set of values
- COUNT(val1, val2, ...) or COUNT(range)
  - Count of the number of cells that contain values
- MAX(val1, val2, ...) or MAX(range)
  - Sum of the specified set of values
- SINE(angle)
  - The sine of the specified angle

Spreadsheet Functions

- PI()
  - The value of PI
- STDEV(val1, val2, ...) or STDEV(range)
  - The standard deviation from the specified samples
- TODAY()
  - Sum of the specified set of values
- LEFT(text, numb_chars)
● Sum of the specified set of values

12 Spreadsheet Functions (Page 3)
  ● IF(test, true_val, false_val)
  ● If the test is true, it returns the true_val; otherwise, it returns the false_val
  ● ISBLANK(value)
  ● Returns true if the

13 Circular References
  ● Circular reference—a set of formulas that ultimately rely on each other

14 Spreadsheet Analysis (Page 1)
  ● Possible tasks a spreadsheet could perform:
    ● Track sales
    ● Analyze sport statistics
    ● Maintain student grades
    ● Keep a car maintenance log
    ● Record and summarize travel expenses
    ● Track project activities and schedules
    ● Plan stock purchases

15 Spreadsheet Analysis (Page 2)
  ● Spreadsheets are also useful because of their dynamic nature, which provides the powerful ability to do what-if analysis
    ● What if the number of attendees decreased by 10%?
    ● What if we increase the ticket price by $5?
    ● What if we could reduce the cost of materials by half?

16 Database
  ● A computerized method of storing data ...
    ● Maintenance:
      ● Adding new information
      ● Updating information
      ● Deleting information
    ● Reporting:
      ● Printed reports
      ● On-line retrieval (on-screen)

17 Database Examples
  ● Possible tasks a database management system could perform:
    ● Banking records including ATM's
    ● Order entry and billing
    ● Personnel and payroll
    ● Customer, client, contacts, etc.
    ● Inventory control

18 Traditional Files vs. Database
  ● Traditionally data files held data for single entities
    ● Lead to duplication of data as information was stored for multiple systems
    ● It was almost impossible to extract information that involved data from more than one system
    ● A database is a structure that can store information about various types of entities and their relationships

19 Database Management Systems (Page 1)
  ● Database—a structured set of data
  ● Database management system (DBMS)—a combination of software and data, made up of a physical database, a database engine, and a database schema
- Physical database—a collection of files that contain the data

20 Database Management Systems (Page 2)
- Database engine—software that supports access to and modification of the database contents
- Database schema—a specification of the logical structure of the data stored in the database
- Database query—a request to retrieve data from a database

22 The Relational Model (Page 1)
- Relational DBMS—a DBMS in which the data items and the relationships among them are organized into tables
  - Organized by the relationships among the tables
- Tables—a collection of records
  - Individual categories of data entities

The Relational Model (Page 2)
- Records (object, entity)—a collection of related fields that make up a single database entry
  - All of the data for one person or entity
  - Can be thought of as a row in a table
- Fields (attributes)—a single value or element in a database record
  - Can be thought of as a column within a table

25 Indexes
- An object that relates key values to records that contain those key values...
  - Can be maintained for any column or combination of columns
- Made up of key values and record numbers...
  - Record number is used by DBMS, not by the user who usually does not see them
- Most important index is the primary key

27 A Database Table
- Key (also known a primary key)—one or more fields of a database record that uniquely identify it among all other records in the table
- The schema for this part of the database can be expressed as follows:
  - Movie (MovieId: key, Title, Genre, Rating)

28 The Primary Key and Entity Integrity
- The primary key (or key) uniquely distinguishes each record from every other in a table
- No column or columns that make up the primary key may contain a null (blank) value
- Guarantees that each entity (record) will be unique (can be distinguished from another)

29 Relationships
- When there are records in more than one table that have a relationship to one other, how is that relationship established
  - For example how do we relate movies to customers?
- By a table, of course

30 Integrity Rules: Referential Integrity
- The foreign key is the column or columns in a table whose value is required to:
  - Match value of a primary key in another table (the table it is linking to)...
  - Or be null

32 Views
- An application program's or individual user's "picture" of a database
  - A saved object
- A view is never the actual data (physical database) as it exists on disk but...
- Rather, it is an object that provides a specific way of looking at the data in memory (logical database)
33 **Views: Advantages and Limitations**

- **Advantages:**
  - Views (using tools such as Structure Query Language or Query by Example) can greatly simplify the process of querying a database
  - *Sensitive* tables or columns can be omitted from a view—*hide* tables, queries, etc.

- **Disadvantage:**
  - Updating a database through a view can present problems especially when it involves a join

34 **Structured Query Language (Page 1)**

- SQL (pronounced see'-quel)
- A comprehensive relational database language for data manipulation and queries
- Represents any data as one or more tables
- Non-procedural language ...
  - Lets DBMS determine *how* the operation is executed
- Functionality includes:
  - Maintaining data (INSERT, UPDATE and DELETE)
  - Providing information from the database by retrieving data from one or multiple tables via SELECTs

35 **Structured Query Language (Page 2)**

- The basic statement for data retrieval in SQL is SELECT in the format:
  ```sql
  SELECT attribute-list
  FROM table-list
  WHERE condition
  ```
  - The SELECT clause limits the columns returned
    - *attribute-list* is the name of the field or fields
  - The FROM lists the table or tables
    - *table-list* is tables from which the fields are extracted

36 **Structured Query Language (Page 2)**

- The basic statement for data retrieval in SQL is SELECT in the format:
  ```sql
  SELECT attribute-list
  FROM table-list
  WHERE condition
  ```
  - The WHERE clause limits the rows returned
    - *condition* is a value restriction that controls which records are included in the query

37 **Structured Query Language (Page 4)**

- Example:
  ```sql
  SELECT Title
  FROM Movie
  WHERE Rating = 'PG'
  ```
  - “Title” is the name of the field (the *attribute-list*)
  - “Movie” is the name of the table from which the “Title” field is extracted (the *table-list*)
  - “Rating = 'PG'” is the value restriction that controls which records are included in the query (the *condition*)

- Result is a table containing all PG movies in table “Movie”

38 **SELECT Queries in SQL**

- SELECT Name, Address
  FROM Customer
- SELECT *
  FROM Movie
WHERE Genre LIKE 'action%'
SELECT *
FROM Movie
WHERE Rating = 'R' ORDER BY Title

Joining Columns from Multiple Tables
- Used to link matching attributes from more than one table
- Format:
  select columns
  from tables
  where primary_key = foreign_key;

Joining Columns from Multiple Tables
SELECT Name, CreditCardNumber, MovieID, DateRented
FROM Customer, Rental
WHERE Customer.CustomerID = Rental.CustomerID;
SELECT Name, Address, Title, DateDue
FROM Customer, Rental, Movie
WHERE Customer.CustomerID = Rental.CustomerID
AND Rental.MovieID = Movie.MovieID;

Modifying Database Content in SQL
- INSERT INTO Customer
  VALUES (9876, 'John Smith', '602 Greenbriar Court', '2938 3212 3402 0299')
- UPDATE Movie
  SET Genre = 'thriller drama'
  WHERE title = 'Unbreakable'
- DELETE from Movie
  WHERE Rating = 'R'

Database Design
- Entity-relationship (ER) modeling—a popular technique for designing relational databases
- ER Diagram—graphical representation of the ER model that captures:
  - Record types are shown in rectangles
  - Attributes (fields) are shown in ovals
  - Relationships are shown in diamonds (relationships may have their own attributes)

Database Design
- Cardinality constraints—help database designer convey the number of relationships that may exist at one time among entities in an ER diagram
  - One-to-one
  - One-to-many
  - Many-to-many

E-Commerce
- Electronic commerce—deals with the process of buying and selling products and services using the World Wide Web
- The development of user interaction with on-line applications, secure protocols, and secure funds transfers made e-commerce feasible and easy
- Biggest problems remaining for expansion include:
  - Some consumers still have misgivings about conducting business on-line
  - The need for still greater security

How to Create a Successful E-Commerce Web Site
- http://www.youtube.com/watch?v=lWW5TAVQtLw
Information Security

- Information security—the techniques and policies used to ensure proper access to data
- Three ways to look at the problem:
  - Confidentiality—ensuring that data is protected from unauthorized access
  - Integrity—ensuring that data can be modified only by appropriate mechanisms (hardware and software)
  - Availability—the degree to which authorized users can access information for legitimate purposes (information not useful if authorized users cannot get to it)

Risk analysis—determining nature and likelihood of the risks to key data
Planning for information analysis requires risk analysis
Goal is to minimize vulnerability to threats that put a system at the most risk

Cryptography

- Cryptography—the field of study related to encoded information (comes from Greek word for “secret writing”)
- Encryption—the process of converting plaintext into ciphertext
- Decryption—the process of converting ciphertext into plaintext

Encrypted information cannot be read
Decrypted/Encrypted information can be read

Cipher—an algorithm used to encrypt and decrypt text
Key—the set of parameters that guide a cipher
Neither is any good without the other

Substitution cipher—a cipher that substitutes one character with another
Caesar cipher—version of the substitution cipher that simply shifts the characters a certain number of positions in the alphabet
Transposition ciphers—a cipher that rearranges the order of existing characters in a message in a certain way (e.g., a “route cipher”)

Substitution Cipher

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
D E F G H I J K L M N O P Q R S T U V W X Y Z A B C
- Substitute the letters in the second row for the letters in the top row to encrypt a message
  - This specific example is a “Caesar cipher”
  - Key consists of number of characters shifted and direction
- Encrypt (COMPUTER) gives FRPSXWHU
  - Substitute the letters in the first row for the letters in the second row to decrypt a message
- Decrypt (FRPSXWHU) gives COMPUTER

Transposition Cipher

T O D A Y
+ I S + M
O N D A Y
- One version of a transposition cipher—write the letters in a row of five, using ‘+’ as a blank; encrypt by spiraling inward from top left corner moving counter clockwise
- Encrypting TODAY IS MONDAY gives T+ONDAYMYADOIS+
- Decrypt by recreating grid and reading letters across row
- The key is the dimension of the grid and the route used to encrypt the data
Cryptanalysis

- Cryptanalysis—process of decrypting a message without knowing cipher or key used to encrypt it
- Substitution and transposition ciphers are easy for modern computers to break
  - To protect information the use of more sophisticated schemes are needed

Public/Private Keys

- Public-key cryptography—an approach in which each user has two related keys, one public and one private
- One’s public key is distributed freely
- A person encrypts an outgoing message, using the receiver’s public key.
- Only the receiver’s private key can decrypt the message

Public/Private Keys

- Digital signature—data that is appended to a message, made from the message itself and the sender’s private key
  - To ensure the authenticity of the message
- Digital certificate—a representation of sender’s authenticated public key that is used to minimize malicious forgeries

Computer Security

- Authentication credentials—the first level of security in which users provide information to identify themselves for computer access
  - User knowledge—something the user knows, such as user name, password, PIN (users should follow certain guidelines and procedures when creating passwords)
  - Smart card—card with embedded memory chip used for identification
  - Biometrics—human characteristics such as fingerprints, retina or voice patterns

Malicious code—a computer program (commonly called a “virus”) that attempts to bypass appropriate authorization and/or perform unauthorized functions
  - Often transmitted across network or via removable media
  - Worm—stands alone, targets network resources
  - Trojan horse—disguised as benevolent resource
  - Virus—self-replicating
  - Logic bomb—set up to execute at a system event

Security attacks—an attack by “hackers” on the computer system itself
  - Password guessing—(obvious)
  - Phishing—attempts to trick users into revealing security information
  - Spoofing—malicious user who is masquerading as an authorized user
  - Back door—an unauthorized access to anyone who knows it exists (sometimes left as an emergency access for authorized user, i.e. the development programmer)

Buffer overflow—a program defect that might attempt to store more data than a buffer can accommodate causing a system crash
  - Leaves unauthorized user with heightened privileges
  - Denial-of-service—attack that prevents an authorized user from accessing the system
  - Man-in-the-middle—an attack on network in which communication is intercepted in an attempt to obtain key data