Spreadsheets and Databases

Instructor: Dmitri A. Gusev

Spring 2007

CSC 120.02: Introduction to Computer Science

Lecture 17, April 19, 2007
Basic Definitions

An *information system* is software that helps the user organize and analyze data. Information systems:
- electronic spreadsheets
- database management systems

A *spreadsheet* is a software application that allows the user 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.

A *database* is a collection of records stored in a computer in a systematic (structured) way, so that a computer program can consult it to answer *queries*.

The computer program used to manage and query a database is known as a *database management system (DBMS)*.
Spreadsheet Software (Microsoft Excel)

Formulas begin with =. They may contain values, references to cells, symbols of arithmetic operations (+, -, *, /), and calls of spreadsheet functions.

A spreadsheet function is a computation provided by the spreadsheet software that can be incorporated into formulas.

=\text{AVERAGE}(A1,B1,C1,D1,A2,B2,C2,D2)

A range is a rectangular block of cells specified by two endpoints (references to corner cells). Example: A1:D2

=\text{AVERAGE}(A1:D2)

Properties of relative and absolute references for copy-paste and fill (right and down): F3, $F3, F$3, $F$3 (discussed in class)

Other features: Fill series; dynamic recalculation; circular reference detection; formatting; search; data analysis; charts
A query is a request for information submitted to a database.

The database schema provides the logical structure of the data in the database, independent of how it is physically stored.

The relational model is a database model in which the data items and the relationships among them are organized into tables.

A table is a collection of database records. A record (a.k.a. database object, entity) is a collection of related fields. Each field (a.k.a. attribute) contains a single data value. The key field(s) uniquely identify a record in the table.
Structured Query Language (SQL)

The *Structured Query Language (SQL)* is a comprehensive relational database language for data management and queries. SQL is not case sensitive. Spaces are used as separators in a statement.

The basic *select* statement format:

```
select attribute-list from table-list where condition
```

Sample query:
```
select Title from Movie where Rating = 'R' order by ProductionCost
```
Modifying Database Content

The *insert* statement adds a new record to a table.

The *update* statement changes the values in one or more records of a table.

The *delete* statement removes all records from a table matching the specified condition.

delete from Movie where Title like 'Naked Gun%'
Database Design

*Entity-relationship (ER) modeling* is a popular technique for designing relational databases. An *ER diagram* captures record types, attributes, and relationships in a graphical form.

- Types of records (classes for the database objects) are shown in rectangles
- Fields (attributes) are shown in ovals
- Relationships are shown in diamonds

*Cardinality relationships:*
- one-to-one
- one-to-many
- many-to-many