Introduction to Database Systems

References

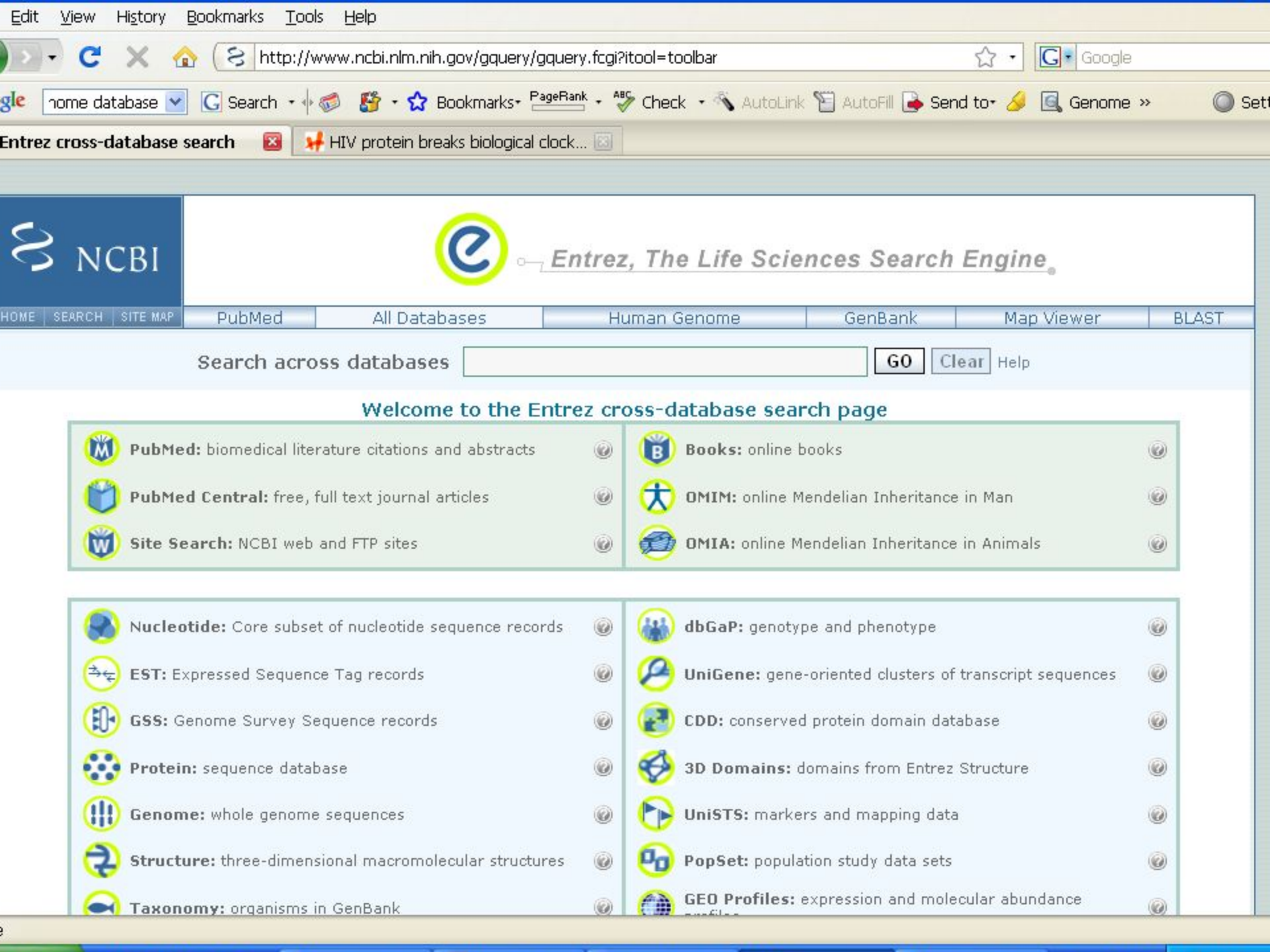
- A. Silberschatz, H. F. Korth, S Sudarshan,
  Database System Concepts, 5th Ed., McGrow Hill, 2005
  http://www.db-book.com
- Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom,
  Database Systems, The Complete Book, Prentice Hall, 2002
  http://www-db.stanford.edu/~ullman/dscb.html
- Class notes

Database Overview

- File Management vs Database Management (why do we need database?)
  - Advantages of Database systems: storage persistence, programming interface, transaction management
- Data Model (What is Data?)
- Database Language (How to manipulate data?)
- DBMS Architecture and Database System Components (How can you build a billion-dollar software, like Oracle? Or you can get it free, mySQL?)
- Users classification (What you can do and what you cannot do?)

Where are databases?

- You cannot avoid it and it's everywhere!
- You can say it actually makes the current society and your life work!
- Banking/Credit card /Social Security Info…
- Online shopping/booking…


Edit View History Bookmarks Tools Help
http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi?itool=toolbar
Google
nome database Search Bookmarks PageRank Check AutoLink AutoFill Send to Genome »
Entrez cross-database search
HIV protein breaks biological clock...
NCBI
Entrez, The Life Sciences Search Engine
HOME SEARCH SITE MAP
PubMed
All Databases
Human Genome
GenBank
Map Viewer
BLAST
Search across databases
GO
Clear
Help
Welcome to the Entrez cross-database search page
PubMed: biomedical literature citations and abstracts
PubMed Central: free, full text journal articles
Site Search: NCBI web and FTP sites
Books: online books
OMIM: online Mendelian Inheritance in Man
OMIA: online Mendelian Inheritance in Animals
Nucleotide: Core subset of nucleotide sequence records
EST: Expressed Sequence Tag records
GSS: Genome Survey Sequence records
Protein: sequence database
Genome: whole genome sequences
Structure: three-dimensional macromolecular structures
Taxonomy: organisms in GenBank
dbGaP: genotype and phenotype
UniGene: gene-oriented clusters of transcript sequences
CDD: conserved protein domain database
3D Domains: domains from Entrez Structure
UniSTS: markers and mapping data
PopSet: population study data sets
GEO Profiles: expression and molecular abundance

The DBMS Marketplace

- Relational DBMS companies – Oracle, Sybase – are among the largest software companies in the world.
- IBM offers its relational DB2 system. With IMS, a nonrelational system, IBM is by some accounts the largest DBMS vendor in the world.
- Microsoft offers SQL-Server, plus Microsoft Access for the cheap DBMS on the desktop, answered by “lite” systems from other competitors.
- OpenSource: mySQL, postgreSQL

Pre-Database Era: Stone Age of Data

- Imagine you want build an online shopping website
  - Maintain products/categories (price, picture, properties, …)
  - Customers accounts
- File is uninterpreted, unstructured collection of information
- File operations: delete, catalog, create, rename, open, close, read, write, find, …
- Access methods: Algorithms to implement operations along with internal file organization
- Examples: File of Customers, File of Products; Access method: implementation of a set of operations on those files

File Management System Problems

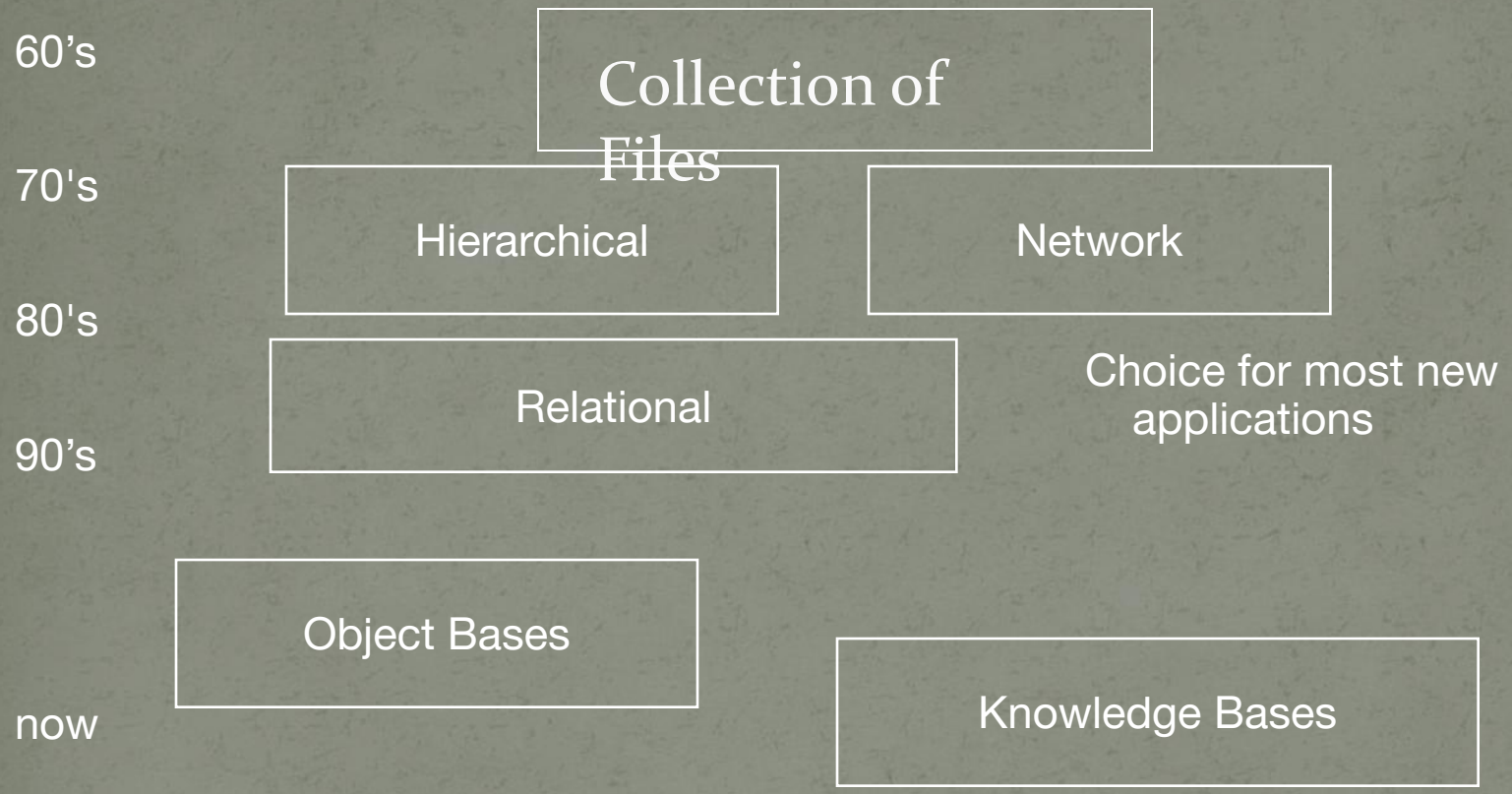
- Any question (access) on the data is a small program!!
- Data redundancy
- Data is not isolated from the access implementation (different format…)
- Multiple application (concurrent program) on the same file

Security Problems

- Allow access to the file only to the authorized personnel
- Ability to restrict access to parts of the record
- Ability to control operation usage by different users
- Protection from unauthorized use
- Protection from the derivation of unauthorized information

Data Integrity

- A database constraint is a logical constraint about the data expressed in a logical language.
  - **STUDENT.AGE >15**
  - ***If*** **(STUDENT.CLASS ==cs43005)** ***then*** **(STUDENT.PRIOR_CLASS ==cs31001)**
- Database is consistent if data at each time satisfies all integrity constraints.
- Input to any application is a set of consistent data. An application output is a set of consistent data.


60’s
Collection of Files
70's
Hierarchical
Network
80's
Relational
Choice for most new applications
90’s
Object Bases
Knowledge Bases
now

Advantages of Databases

- ***Persistent Storage*** – Database not only provides persistent storage but also efficient access to large amounts of data
- ***Programming Interface*** – Database allows users to access and modify data using powerful query language. It provides flexibility in data management
- ***Transaction Management*** – Database supports a concurrent access to the data

Early Database Applications

- Airline Reservation Systems – Data items are: single passenger reservations; Information about flights and airports; Information about ticket prices and tickets restrictions.
- Banking Systems – Data items are accounts, customers, loans, mortgages, balances, etc. Failures are not tolerable. Concurrent access must be provided
- Corporate Records – Data items are: sales, accounts, bill of materials records, employee and their dependents

Modern Database Applications

- Client – Server architecture
  - DBMS serves as a server and client queries are sent to servers
  - Where to locate servers
- Multimedia Applications
- Multidatabase Applications
- Data Warehouses
- It's everywhere!!

Data Administrator

- Coordinates all the activities of the database system; the database administrator has a good understanding of the enterprise's information resources and needs.
- Database administrator's duties include:
  - Schema definition
  - Storage structure and access method definition
  - Schema and physical organization modification
  - Granting user authority to access the database
  - Specifying integrity constraints
  - Acting as liaison with users
  - Monitoring performance and responding to changes in requirements

Database Users

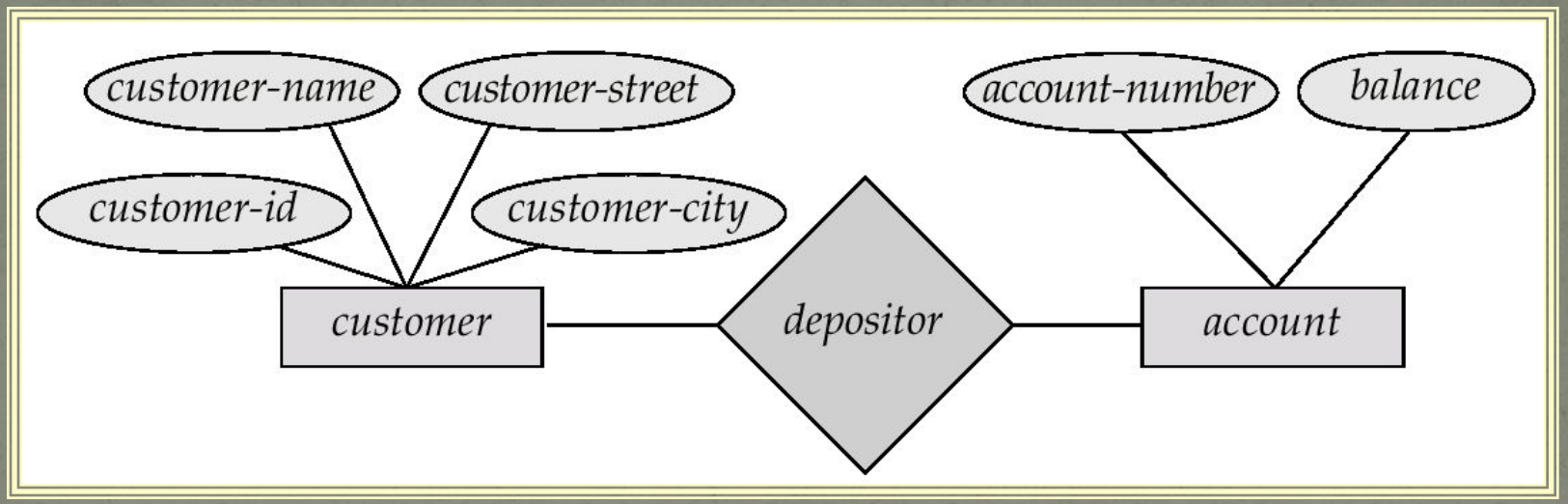
- Naïve – do not know about database too much, invoke application programs that are prepared already
- Application Programmers – know how to interact with the system but may not know how DBMS is designed
- Sophisticated users that know advanced use of the system and can use the system and packages on the top of the system
- DBMS system users – write specialized database applications that do not fit into the traditional data processing framework

A Little Design Methodology: Entity-Relationship Model

- The enterprise data can be described as a set of entities and a set of relationships between them.
- **Entity** – a data that pertains to, or describes some component of the enterprise
- Each entity is characterized by a set of **attributes**
- **Relationship** – describes an interconnection between different entities
- **Entity Set** – a set of entities that are characterized by the same entity definition
- **Relationship Set** – a set of relationships of the same type

Entity-Relationship Model

Example of schema in the entity-relationship model



Thank you