Midterm Review

Midterm: November 3, 2008 in class – closed book

This review sheet is intended only as a study guide concerning the breadth of the exam. You are expected to know all the terminology presented as covered in class, the text and the required supplemental reading up to and including global alignment. (need not know Smith-Waterman and chaining).

Again, individual terms in this document are indicative of the breadth of the exam and do not represent an exhaustive list of the contents to be covered. Otherwise, the scope of the homework is indicative of exam contents. D.M.

Reading:
- Most of the database text, except the sections on SQL
- Papers

Material:
1. What is data modeling?
   a. Terminology
      i. Data model
      ii. Three schema model
         1. external schema
         2. conceptual schema
         3. internal schema
      iii. Entity, attribute, identifier, relation/association
      iv. Logical model, Physical Model, DDL
   ...

   b. What are the steps of a data modeling effort
      i. planning and analysis
      ii. conceptual design // logic without the details
      iii. logical design
      iv. physical design
      v. implementation

2. Basic Relational Database Concepts
   a. Schema(s)
   b. Content addressability
   c. Keys
      i. Candidate key
ii. Primary key
iii. Foreign key

3. Data Model Concepts
   a. Data modeling process
      i. Plan project
      ii. Determine requirements
      iii. Specify entities
      iv. … (see course slide handout)
   b. Data provenance
   c. Entity, attribute, relation/association
   e. Relations/associations properties
      i. Cardinality constraints
      ii. Aggregation
      iii. identifying relationships
      iv. Inheritance (subtypes)

4. Requirements analysis
   a. Why is it important?
   b. How does considering requirements change a database design compared to
      the simple use of databases as extensions of laboratory notebooks?
   c. What is the sequence of questions to be resolved in a requirements
      analysis?
      i. Who
      ii. What
      iii. Why
      iv. If they could do it now, how do they do it?

5. Sequence Alignment, Dynamic Programming Algorithms, (terminology and concepts
   only.  You will not be asked to complete a dynamic programming matrix)
   a. Terminology
      i. Principle of optimality
      ii. Indel
      iii. Substitution Matrix
      iv. PAM model
      vi. Backtrace
   b. Edit distance (global alignment)
   c. Local alignment (Smith-Waterman)
      i. Chaining