

# Final Exam Review Sheet

CS329e Elements of Biological Data Models, Spring '07  
Professor Daniel Miranker  
4/6/07

**Exam Time:** Saturday, May 12, 2:00–5:00 PM, in the regular lecture room, Painter 3.14

This is a comprehensive exam. 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. Again, individual terms in this document are indicative of the breadth, they do not represent the actual content.

## Reading:

At this juncture you should have read the entire contents of the course reading packet. For the exam, pay special attention to:

- Brazma et. al. Minimum information about a microarray experiment (MIAME) - toward standards for microarray data. Nat Genet. 2001 Dec;29(4):365-71.
- The Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 32: D258-D261 (2004).

## Topics:

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
      - i. 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

If they could do *it* now, how do they do it?

5. Sequence Alignment, Dynamic Programming Algorithms, (~~be prepared to fill in a dynamic programming matrix by hand~~)
- a. Edit distance (global alignment)
  - b. Local alignment (Smith-Waterman)
  - c. Backtrace
  - d. Chaining
6. Rosetta Analysis
- a. Goals?
  - b. How is it done?
  - c. Gene fusion
  - ~~d. Use of GO in the evaluation of the method~~
7. MIAME, Gene Expression
- a. What are the aims of MIAME?
  - b. What are the six components of MIAME?
  - c. Data provenance
  - d. Controlled vocabulary
8. Ontologies
- a. What is an Ontology?
  - b. What purposes to they serve?
  - c. Why are they important?
  - d. GO (Gene Ontology)
    - i. What is the purpose of GO?
    - ii. How is it organized?
    - iii. How was GO implemented?
9. Database Indexing
- a. How does a B+ tree compare to a Binary tree?
  - b. Why are secondary indexes important?
  - ~~c. What is the heuristic for when to include a secondary index?~~

## 10. Programmatic Interfaces to Databases

- a. Why is it important for a database to have a programmatic interface? (compared to a simple web/browser interface)
- b. What is the structure of a three tier architecture?
- c. What are the software components of a JDBC interface?
  - i. Driver
  - ii. Connection
  - iii. Statement
  - iv. ResultSet
  - v. ...

// There will be no exam questions that require you to write any Java  
// This topic is good for high-level short answer questions. For example,  
// “How is security/authentication handled in JDBC?” Answer: When  
// opening a connection the program must supply a user name and password.

## 11. XML – exception to the above warning concerning scope. The following is precisely the limit of expected XML knowledge for the final. (Miranker’s personal policy, No homework on a subject → no exam questions, except basic definitions.)

- a. Basic definitions
  - i. Element
  - ii. Content
  - iii. Attribute
  - iv. ~~DTD XML schema~~
- b. Does XML solve semantic integration problems (U.S.dollar vs. Canadian dollars?)  
Not by itself. As a method for defining standards, such details are only unambiguous if the standard requires the details to be spelled out in the data transfer.
- c. Why is XML [really] valuable?
  - i. Human readable, expanding the who can use it.
  - ii. Looks like HTML and can be transferred using http and thus get through firewall more easily than proprietary binary representation.
- d. XML can be structured per a DTD or an XML schema. But even these definitions are optional.