

# Biology 2YY3: Practical Databases and Data Mining in Biology

G. Brian Golding

January 7, 2009

# Outline

- ▶ Introduction: Databases, Database systems, Data Mining
- ▶ Biological Databases - Lecture #1-#5
- ▶ Database systems - Lectures #6-#15
- ▶ Data Mining - Lectures #16-#24

# Outline

- ▶ Introduction: Databases, Database systems, Data Mining
- ▶ Biological Databases - Lecture #1-#5
  - ▶ Why databases
  - ▶ Popular existing databases
- ▶ Database systems - Lectures #6-#15
- ▶ Data Mining - Lectures #16-#24

# Outline

- ▶ Introduction: Databases, Database systems, Data Mining
- ▶ Biological Databases - Lecture #1-#5
- ▶ Database systems - Lectures #6-#15
  - ▶ PART 1 DATA MODELS
  - ▶ PART 2 RELATIONAL DATABASES
  - ▶ PART 3 OBJECT-BASED DATABASES AND XML
  - ▶ PART 4 DATA STORAGE AND QUERYING
  - ▶ PART 5 TRANSACTIONMANAGEMENT
  - ▶ PART 6 DATABASE SYSTEM ARCHITECTURE
  - ▶ PART 7 OTHER TOPICS
  - ▶ PART 8 CASE STUDIES
- ▶ Data Mining - Lectures #16-#24

# Outline

- ▶ Introduction: Databases, Database systems, Data Mining
- ▶ Biological Databases - Lecture #1-#5
- ▶ Database systems - Lectures #6-#15
- ▶ Data Mining - Lectures #16-#24
  - ▶ Visualization
  - ▶ Clustering
  - ▶ Modelling

# Texts

- Database Systems; Concepts, Management and Applications by  
A.C. Lorents and J.N. Morgan, 1998, Harcourt Brace &  
Company.
- First Course in Database Systems, A, 3/E Jeffrey D. Ullman 2008  
Prentice Hall.
- Databases Demystified by A. Opper, 2004, McGraw-Hill.
- Database System Concepts, 5/e A. Silberschatz, H.F. Korth, and  
S. Sudarshan, 2006, McGraw-Hill.
- A First Course in Database Systems, 3/e by J.D. Ullman and  
J. Widom, 2008, Prentice Hall.
- Learning MySQL By Seyed M.M. "Saied" Tahaghoghi, Hugh E.  
Williams November 2006, O'Reilly Publ.
- Introduction to Data Mining by Pang-Ning Tan, Michael Steinbach, and  
Vipin Kumar, 2006, Addison-Wesley
- DataMining; Concepts and Techniques by J. Han and M. Kamber,  
2001, Academic Press.
- Survey of Text Mining; Clustering, Classification and Retrieval edited by  
M.W. Berry, 2004, Springer Verlag.

# Calendar

---

Week of. . .	Monday	Wednesday	Thursday
January 5	Intro	Biodatabases	NA
January 12	Biodatabases	Biodatabases	cmd line
January 19	Biodatabases	Database systems	cmd line
January 26	Database systems	Database systems	mySQL
February 2	Database systems	Database systems	mySQL
February 9	Database systems	Database systems	mySQL
February 23	Database systems	Database systems	mySQL
March 2	Database systems	Data mining	mySQL
March 9	Data mining	Data mining	R
March 16	Data mining	Data mining	R
March 23	Data mining	Data mining	R
March 30	Data mining	Data mining	R
April 6	Data mining		

---