

CIS 349: Databases with Web Interfaces

Scott D. Anderson

Spring 2001

Instructor

Dr. Scott D. Anderson. My office hours are Monday and Thursday 9:50–11:00 and 4:10–5:30. You can also meet me by appointment, or just drop by if I'm not busy. Contact me by email at sanderso@wellesley.edu or by phone at 781-283-3249. My office is room E114 in the Science Center. My web page is <http://cs.wellesley.edu/~anderson/>.

Policies

My policies regarding academic dishonesty, attendance, late assignments, grading standards, and so forth are listed on my web page. You are expected to read them and understand them.

Description and Prerequisites

A study of the design of file systems and databases, including file organization and access methods, concepts of database management, and database querying using SQL. We will look at the entity-relationship model as a way of structuring data, and we will use relational algebra and relational calculus as a formal system for operating on data. We will investigate how databases are represented, including B+ trees and hash indexes. We will briefly discuss sorting methods for databases. Finally, we will create dynamic web documents driven by database entries. Prerequisite: CS 230

Purpose

This course is taught because one of the major uses of computers has always been to store and allow access to large amounts of data. This usage is only increasing thanks to the Internet, because companies can now conceive of everyone from sales representatives to upper level managers querying and entering data into massive databases. Customers even get into the act, with point-of-sale data entry, by bar code or swipe card. Thus, databases are a crucial aspect of modern computing. Furthermore, because databases are so often used in commercial web sites (think about Amazon.com) we can use this course as an opportunity to do some dynamic web pages, where the content of the page is generated from a database query.

Objectives

Upon successful completion of this course, the student should be able to:

- Describe basic file organizations, the distinction between physical and logical files, and concepts of file management (fields, records, buffers).
- Explain the strengths and weaknesses of different secondary storage devices (tape, hard disks, floppy disk, CD-ROM, DVD).
- Code file structures algorithms using the *C⁺⁺* language as a tool, particularly B+ trees and Linear Hash files.
- Evaluate trade-offs between various file structures and techniques for optimizing file access

(buffers, sorting, compression, indices, parallelism).

- Describe advanced file structures for optimizing access (multilevel indexes, B+ trees, ISAM, hashing).
- Write Entity-Relationship diagrams to capture the entities, relationships, and participation constraints in real-world situations.
- Convert ER diagrams to tables and build those tables in SQL's Data Definition Language (DDL)
- Give relational algebra and relational calculus expressions to define query results.
- Solve query problems using SQL's Data Manipulation Language (DML).
- Display an understanding of data management issues such as database creation, access, and update.
- Write simple Perl scripts, using the CGI and DBI modules, to allow database operations via a web-page interface.

Instruction

Instruction will consist of classroom lectures, discussions, exercises, and programming assignments.

The evaluation is weighted as follows:

Category	Weight
midterm	20
final exam	30
assignments	50

It is virtually impossible for you to learn the material in this course without some type of directed hands-on experience. There will be a minimum of four programming assignments for you to complete independently to reinforce the material covered in class. Much of our classroom discussion will be devoted to the concepts which these programs will use.

You will be held responsible on tests for the concepts illustrated. Written descriptions of all programming assignments will be provided as the semester progresses. The due date for each assignment will be announced in class, and the assignments must be submitted by midnight of the announced date in order for the assignment to be considered on time.

Textbook

Database Management Systems, by Raghu Ramakrishnan and Johannes Gehrke, 2000, WCB/McGraw-Hill, second edition, ISBN 0-07-232206-3.

Course Outline

The table below gives the *tentative* schedule for the course, including readings and due dates for assignments. Students are expected to come to class having read the material for the week.

Assignments (tentative):

1. Simple database construction and queries using MS Access.
2. Simple random-access file I/O to store a database as a “heap”
3. buffer/file manager using the minibase software
4. B+ tree implementation using the minibase software
5. database creating and querying using Oracle from the command line
6. database manipulation by Perl CGI programming, using the Perl DBI and CGI modules.

Each assignment may also include some homework problems based on the text.

Lecture		Topic	Reading	Due
1.	M 1/29	Introduction	Chapter 1	
2.	R 2/1	C++ and random access file I/O	handout	
3.	M 2/5	more on C++ and file I/O	handout	A1
4.	R 2/8	Storing Data: Disks and Files	Chapter 7	
5.	M 2/12	File Organizations and Indexes	Chapter 8	A2
6.	R 2/15	File Organizations and Indexes	Chapter 8	
7.	T 2/20	Tree-Structured Indexing	Chapter 9	
8.	R 2/22	Tree-Structured Indexing	Chapter 9	
9.	M 2/26	Entity-Relationship Model	Chapter 2	A3
10.	R 3/1	Entity-Relationship Model	Chapter 2	
11.	M 3/5	The Relational Model	Chapter 3	
12.	R 3/8	The Relational Model	Chapter 3	
13.	M 3/12	SQL: Queries, Programming, Triggers	Chapter 5	A4
14.	R 3/15	SQL: Queries, Programming, Triggers Spring Break	Chapter 5	midterm
15.	M 3/26	Internet Databases	Chapter 22	
16.	R 3/29	Perl, DBI, CGI, Oracle	handouts	
17.	M 4/2	Perl, DBI, CGI, Oracle	handouts	
18.	R 4/5	Relational Algebra and Calculus	Chapter 4	
19.	M 4/9	Relational Algebra and Calculus	Chapter 4	A5
20.	R 4/12	External Sorting	Chapter 11	
	M 4/16	no class (Patriot's day)		
21.	R 4/19	Evaluation of Relational Operators	Chapter 12	
22.	M 4/23	Evaluation of Relational Operators	Chapter 12	
23.	R 4/26	Query Optimization	Chapter 13	
24.	M 4/30	Schema Refinement, FDs, Normalization	Chapter 14	
25.	R 5/3	Transaction Management	Chapter 18	
26.	M 5/7	Review and Collapse		A6