Syllabus for Advanced Databases, AKA Database Theory in Fall 2006.

Course Number: CS 535, Section: 01, CRN: 26630, Credits: 3
Prerequisites: CS 435.
Meetings: Monday, Wednesday, and Friday, 9:10-10:00AM @ **EPS 350**
Course Web page: http://www.cs.montana.edu/courses/535/

Lecturer: Dr. Rafal A. Angryk

- ◊ Office: 362 Engineering/Physical Sciences Building
- ◊ E-mail: angryk@cs.montana.edu
- ◊ Phone: (406) 994-4440 (My Office), 994- 4780 (Main CS Office)
- ◊ Fax: (406) 994-4376 (CS Office)
- ◊ Home Page: http://www.cs.montana.edu/angryk/
- ◊ Consultations on the class-related issues:
  - Right after each class at the classroom
  - Office hours:
    - Monday -1:00 – 2:00 PM
    - Wednesday, Friday 12:00 -1:00 PM
    - Knock if the door is not open during these hours
    - Let me know if the above hours are not convenient for you!
  - By appointment (Please, set it up ahead of time by e-mail or phone).

Literature for 1st part of the course:

~ S. Shekhar, S. Chawla, *Spatial Databases: A Tour*, ISBN: 013-017480-7, Publisher: Prentice Hall, 2003.

Additional readings (for 1st and 2nd part of the course):

~ J. Han, M. Kamber, *Data Mining: Concepts And Techniques*, 2nd Edition, ISBN: 1-55860-901-6, Publisher: Morgan Kaufmann Publishers, 2006 (only 1 or 2 chapters)

~ NOTE: Additional readings from scientific journals and conferences will be recommended on some of the discussed topics (links are going to be placed on the course website).

Other recommended books:

1. P. Rigaux, M.O. Scholl, A. Voisard, *Spatial Databases: With Application to GIS*, ISBN: 1558605886, Publisher: Morgan Kaufmann, 2001. **Really good GIS book!**
2. H. Garcia-Molina, J.D. Ullman, J. Widom, *Database Systems, The Complete Book*, ISBN: 0-13-031995-3, Publisher: Prentice Hall, 2002.
3. H. Garcia-Molina, J.D. Ullman, J. Widom, *Database System Implementation*, ISBN: 0-13-040264-8, Publisher: Prentice Hall, 2000.
4. R. Elmasri, S.B. Navathe, *Fundamentals of Database Systems*, 5th Edition, ISBN: 0321369572, Publisher: Addison-Wesley, 2006.
5. A. Silberschatz, H. Korth, S. Sudarshan, *Database System Concepts*, 5th Edition, ISBN: 0-07-295886-3, Publisher: McGraw-Hill, 2005.

Course description
In short, the CS 535 course will have four major components:

- ◊ **Part 1.** Spatial Databases
- ◊ **Part 2.** Data Warehousing, Dealing with Graph-data, Fuzzy database models
- ◊ **Part 3.** Graduate research project

NOTE: The last part will have the most significant influence on the final grade from the course.

The goal of this course is to present in detail a number of advanced concepts of database systems: spatial databases, multi-dimensional indexing techniques, advanced data modeling techniques, common data mining approaches, and a few emerging areas of database-related research. The lectures should provide graduate students with sufficient foundation to conduct their own, but supervised research in the field of databases. Students will gain hands on experience on the chosen aspect of databases through completion of graduate research project, which will be reviewed by their peers at the end of the semester.

In the first part of the course the following components of spatial databases will be introduced: (1) spatial concepts and their modeling based on the pictogram enhanced ERD, (2) spatial querying language, (3) query optimization, and (4) popular methods for multi-dimensional indexing. If time allows maybe we will also go through industry standard for spatial data types (i.e. OGIS types).

In the second part of the course selected aspects of new database technologies will be presented: (1) data warehousing and OLAP technology, (2) graph data storage, processing and retrieval, and (3) a few unconventional database models (e.g. Fuzzy Databases Models, and Deductive Databases).

Significant (third!) part of the course will be devoted to students' individual research, conducted during work on graduate projects, which are going to be developed in the second half of the semester. The first two parts of the course contain these aspects of advanced database technologies, which in the author's opinion, are the most interesting in the field of current database systems. During the third section of the course the students are encouraged to extend the presented material by individual studies and by the development of projects meeting their own research interests and the gathered database expertise. At the current moment, databases technologies are developed to such extent, that almost every computer scientist can find something related to his/her individual interests. The work on projects will be closely supervised by the author of the course, and regular attendance to the individual meeting is strongly advised (i.e. required).

Course outcomes
At the end of the course, students should be able to:

- ◊ Understand concepts of spatial databases
- ◊ Understand components of popular data warehousing methodologies
- ◊ Develop a scientific project that involves processing of large repositories of data (i.e. is data-intensive)
- ◊ Present his/her findings on the conducted research in the form of scientific publication

---

Contributions of all grades to the final course grade:

- ◊ Regular homeworks and active participation in the lectures – 10%
- ◊ Daily quizzes, tests and final exam – 25%
- ◊ Research project, technical reports, presentation, and paper – 65%

A grading scale for my all Graduate Courses is following:

>= 93% = A
92%-85% = B
84%-77% = C
<= 76% = F

Attendance: Lack of sufficient attendance (2 absences) during the meetings related to the final project development may result in a grade of "F" for the whole course.

Plagiarism: The persons who are caught plagiarizing a work will receive a final grade of "F" in the course. Names of the persons, who are involved in plagiarism, will be reported directly to the Dean of Students office (SUB 120, (406) 994-2826). **NOTE:** Taking ideas (either by direct quotation or by paraphrase) from print or electronic sources without using end-of-sentence parenthetical documentation is also called plagiarism. Plagiarism also occurs when students copy word-for-word a sentence or passage from a source without placing double quotation marks around those words. A good Internet site for learning how to document sources is **A Guide for Writing Research Papers** (URL: http://webster.commnet.edu/mla/index.shtml).

Research project, technical reports, presentation, and **paper:**
Each of the graduate students needs to find a research project, which is relevant to the areas discussed in the course. Only projects, which gain approval of the instructor, are acceptable. During individual weekly meetings in the second part of the semester, students are expected to provide technical reports on their research achievements. This is how the final technical report (30-50-100 or more pages…) is going to be created. The reports, as well as the developed applications, are going to be used in the preparation of students' final project presentation (40 min. long). 10 days before the end of the course, all students are obligated to prepare a publishable research paper in the IEEE format. Papers not exceeding 5 letter size pages with the affiliation of the author including email address should be submitted electronically to the instructor before the deadline. Manuscripts should be prepared in the standard 2-column IEEE format. They will be peer reviewed by other participants of the course, and then graded by the author of the course.

NOTE: My students are expected to use computers to prepare ALL assignments. The print should be in black rather than a barely readable gray. I will not accept work which I can have hard time to read. I will not accept work on paper ripped from a spiral notebook.