

EECS 317 – Data Management and Information Retrieval**Fall 2014****Syllabus****Class meets:** Mon./Wed./Fri. 2:00-2:50PM at G21, Annenberg Hall.**Instructor:** Goce Trajcevski (g-trajcevski@northwestern.edu)**Office Hours:** Mon./Wed. 3-4PM (or, by appointment) at L360, Tech**TAs:** Bing Zhang and Mas-ud Hussain**TAs Contact:**bing@u.northwestern.edu (Office hours: Tue. 1-2PM, Thu. 11AM-noon (L580))mmas-ud.hussain@u.northwestern.edu (Thu. 3-4PM; Fri. 3:30-4:30PM)**Graduate Assistant:** Boda Du (bodadu2013@u.northwestern.edu)

Course description: The main goal of this course is to expose the students to the fundamentals of the Data Management and Database Systems in general and, more specifically, to Database Design and Querying, providing a balance between the breadth and the depth.

Breadth: We will present many abstractions and concepts that are relevant for a domain-expert to know about databases (e.g., their role in enterprise-systems, their capabilities/use, their potential in decision-making and data mining), as well as information retrieval systems. However, due to the timing (and other) constraints, we cannot explore all of them in details. The rationale' for the selection of the topics is that, once the course has been completed, the students should feel comfortable discussing certain system-oriented aspects of the overall database management with the folks that are in charge of the actual implementation/maintenance/support details.

Depth: Some parts of the course (e.g., ER-design; Database Design; the Quality of the Database Design; Database Querying and SQL) will be allocated larger portions of the class. The rationale' behind this is three-fold: (1) Given the background of the targeted audience, these parts of the overall database theory and systems-aspect can be "tackled" at a greater depth; (2) These parts of the overall database design/querying are the ones that have closest "semantic-ties" with a given application domain; (3) The students will gain a level of independence/maturity to rely upon when starting a job (and/or internship) where databases play significant role in the enterprise.

Hence, after completing the course – in addition to the capabilities to comfortably design and query databases, along with exploiting their potential in the overall business process management (e.g., using data mining tools) – the students should have a solid understanding of some aspects of the data management that will enable you to communicate with many of the other kinds of professionals involved in the database issues (e.g., DB administrators; DB application-programmers) in an "almost peer-to-peer" manner.

I. Required text: "*A First Course in Database Systems*", by J. Ullman and J. Widom (Prentice Hall)

II. Reference text and/or other materials:

- a. "*Database Management Systems*", by R. Ramakrishnan and J. Gehrke (McGraw Hill);
- b. Several handouts, giving a bit deeper description of the material pertaining to some of the topics covered, will be posted on the Blackboard. Also, a couple of copies of Ramakrishnan & Gehrke's book will be placed under "Reserved" in the Engineering Library.

III. Tentative Course Outline

Topic #1	<ul style="list-style-type: none"> - Introduction/Motivation - Logic and Computing - “FF” course in Computer Organization - DBMS (centralized and Distributed)
Topic #2	<ul style="list-style-type: none"> - Essentials of the ER-diagrams (basic elements; limitations) - Modeling real-world scenarios with ER diagrams - Extended ER-diagrams (EER)
Topic #3	<ul style="list-style-type: none"> - Basics of Relational Databases and Representation of Data - From ER-diagrams to Relational Data Model - Integrity Constraints
Topic #4	<ul style="list-style-type: none"> - Quality of the Relational Database Design - Redundancy and Functional Dependencies - Normal Forms and Verification
Topic #5	<ul style="list-style-type: none"> - Manipulating Relational Data (Definition/Modification) - SQL - Other Relational Languages (QBE + Overview)
Potpourri Topics #1	<i>Information Retrieval:</i> Means to rank/select unstructured data (e.g., web search engines (Google)).
Potpourri Topics #2	<i>Data Mining and Warehousing/OLAP</i> a bit of a business intelligence used for decision-making.
Platforms/ Environments	<i>Mostly using Microsoft Access; Part of the SQL-related projects may be given in MySQL (if so, it will most likely be given in teams)</i>

IV. Grading

Your grades will be based on:

- 2-3 homeworks (~ 10%) Given during weeks #2, #3, (possibly, #7)
- 2 Quizzes (~15%) Given during weeks #4 and #8
- Project 1 (~10%) Given during week #4/5
- Project 2 (~10%) Given during week #6/7
- Project 3 (~10%) Given during week #8
- Midterm (~21%) Given during week #6
- Final Exam (~24%) Given during week #11

Notes: At this point, the weights of the assignments are approximate, however, on the Monday of the last week of classes, the final crisp distribution of the weights will be announced.

Closing Remarks (Awareness and Academic Responsibilities): Please be advised that it is each student’s individual responsibility to keep him/herself up-to-date with the announcements made in class, distributed via email, or otherwise posted. A particular assignment may be allowed to be done in teams, however, if an assignment is indicated as individual – although you are encouraged to always discuss high-level (e.g., design, solution) issues with your classmates – it is your responsibility to ensure that the

McCormick

Northwestern Engineering

Department of Electrical Engineering and Computer Science

work (e.g., entire coding) is done individually. The policies for cheating are well-defined. In addition, notwithstanding our willingness to be understanding for the students' commitments and time-constraints, please do not attempt to obtain an incomplete grade for the course, based solely on your poor performance – it is against the University regulations.

Welcome and good luck!!!