CS6390/CS4390
Special Topics in Computer Science:
Web-based Data Integration
Spring 2019

Instructor: Natalia Villanueva Rosales , e-mail: nvillanuevarosales [at] utep.edu, office: CCSB 3.0508, phone: (915) 747-8643.
Class time: TThu, 1:30-2:50pm.
Location: CCSB G.0208.
Office hours:TThu 9:15-10:15am, noon-1:15pm or by appointment.

Course Description

This course introduces foundations and applications of integrating heterogeneous data from decoupled sources. Data integration approaches introduced in the class include the use of high-level data models (e.g., ontologies) with a focus on semantic web technologies and current research trends in data retrieval and processing. Hands-on activities are used to illustrate practical aspects of data integration.

This course is intended for senior and graduate students in science and engineering. There are no hard prerequisites for this class. However, computer science students are expected to have some experience in programming and databases. Students from other programs interested (or in need of) data science to publish and integrate resources (e.g. data, information, and methods) from their research are highly encouraged to register in this course.

Objectives

By the end of this course, students will be able to:

Manage large quantities of data generated from experiments, sensors and/or personal devices, a current challenge in every domain.

Understand challenges in data management and create data mo

dels.

Use and create

cyberinfrastructure

to publish, retrieve, integrate and exchange data with a focus on reproducible research.

Identify current trends in data science.

Improve your research, communication, teamwork and problem-solving skills by working on real-world problems.

Topics

The topics covered in this course include:

- Cyberinfrastructure platforms and applications.
- The quest for semantics.
- Semantic Web Languages and ontology engineering.
- Provenance-based trust.
- Current and future trends of cyber-infrastructure for scientific research.

Grading

1. Project 30% .
   Use cyberinfrastructure to access, integrate, query, and manipulate scientific data
2. Presentation and assignments 30% .
    You will present project's progress reports and topics related to your project. • Assignments will be related with your project and will be discussed in class 1-2 weeks before the deadline. You will present papers about data science.
3. Exam(s) and Quiz(zes) 40%.
   Written evaluation of your understanding of the topics reviewed.

Recommended readings
- Book: Pascal Hitzler, Markus Krotzsch, Sebastian Rudolph, <u>Foundations of Semantic Web Technologies. </u>Chapman & Hall/CRC, 2009. ISBN: 9781420090505.
- Protégé, a free open source ontology editor and knowledge base framework. <u>Download here. </u>An extensive Protégé OWL tutorial can be found <u>here.</u>
- Ian Horrock's Sem Tech tutorial for Description Logics. <u>Part 1 </u>and <u>Part 2</u>

Resources

Special Accommodations:
If you have a disability and need classroom accommodations, please contact the Center for Accommodations and Support Services (CASS) at 747-5148 or by email to cass@utep.edu, or visit their office located in UTEP Union East, Room 106. For additional information, please visit the CASS website at www.sa.utep.edu/cass. CASS' staff are the only individuals who can validate and if need be, authorize accommodations for students with disabilities.

Scholastic Dishonesty:
Any student who commits an act of scholastic dishonesty is subject to discipline. Scholastic dishonesty includes, but not limited to cheating, plagiarism, collusion, submission for credit of any work or materials that are attributable to another person. Cheating is copying from the test paper of another student. Communicating with another student during a test to be taken individually. Giving or seeking aid from another student during a test to be taken individually. Possession and/or use of unauthorized materials during tests (i.e. crib notes,   class notes, books, etc.). Substituting for another person to take a test. Falsifying research data, reports, academic work offered for credit. Plagiarism is using someone's work in your assignments without the proper citations. Submitting the same paper or assignment from a different course, without direct   permission of instructors. To avoid plagiarism, see:

http://sa.utep.edu/osccr/wpcontent/uploads/sites/8/2012/09/AvoidingPlagiarism.pdf. Collusion is unauthorized collaboration with another person in preparing academic assignments   NOTE: When in doubt on any of the above, please contact your instructor to check if you are following authorized procedure.