Syllabus (draft)
CSE 636 Data Integration
Dr. Jan Chomicki
Spring 2020

General information

• Registration number 23543
• Time: T R 9:30-10:50
• Location: Bell 138
• Textbooks:
• Prerequisites: At least one database course (graduate or undergraduate), some knowledge of logic and computational complexity.
• Office hours: TBD.
• Office: 338I Davis.

Course description

The availability of integrated data from multiple independent, heterogenous data sources is crucial for many applications. Data integration requires combining and matching information from different sources, and resolving a variety of discrepancies. XML is becoming a de facto data integration standard. Recently, Semantic Web techniques have been considered for data integration.

This course will survey selected issues arising in data integration. Of particular interest are the theoretical foundations of the area, and algorithms and software systems facilitating integration. The students in the class will be working on team projects (size: 3 people) involving research and/or programming, and will give class presentations about their projects. There will also be two exams and two homeworks.

The course team project can be used as an MS project (provided the course grade is at least B+). The projects involve building mashups, implementing some aspects of data integration, or working with real datasets.
Course outline

1. Datalog, negation, recursive query evaluation.
2. XML: data model, schemas, integrity constraints.
3. XML query languages: XPath, XQuery, query evaluation.
5. Schema matching and mapping.
6. Query containment and rewriting.
7. Data exchange, source-to-target dependencies.
8. Database incompleteness and inconsistency, nulls, consistent query answers.
9. Data quality and cleaning.
10. Data provenance.
11. Information extraction.

Policies

1. Grade distribution: 2 exams (40%), 2 homeworks (10%), team project (50%).
2. Grade allocation: A (at least 85%), B (at least 70%), C (at least 50%), F (below 50%).
3. Make-up policy: Make-ups have to be scheduled sufficiently in advance.
4. Late submission policy: The submissions are due at midnight on the due date. No late submissions are accepted. Any exceptional deadline extensions need to be applied for and approved before the deadline. Supporting documentation (for example, medical) should also be submitted at the same time.
5. Questions about the grading have to be raised within a week after the graded assignment is returned.
6. Attendance is mandatory. No late arrivals in class.
7. UB and CSE academic integrity policies: https://engineering.buffalo.edu/computer-science-engineering/information-for-students/policies/academic-integrity.html
8. Other relevant UB and CSE policies.
Accessibility resources

Students with physical or learning disabilities should register with the university Disability Services Office (http://www.student-affairs.buffalo.edu/ods/) in order to receive accommodation.