CARLETON UNIVERSITY
SCHOOL OF COMPUTER SCIENCE

COMP 5306: Data Integration

Fall Term 2016

Instructor: Prof. Leopoldo Bertossi

Course web page: `http://www.scs.carleton.ca/~bertossi/DI/index.html`

Lectures: Tuesdays and Thursdays, 10:05-11:25. 378 UC.

Office hour: Room 5125A. Tuesdays 15:00 - 16:30.

Area: T, A

Prerequisite: A first course on database systems at the undergraduate level.

Evaluation: Assignments (40%), paper presentation (30%), final test (during last lecture time slot) (30%).

Description of the course:

This course offers an excellent opportunity to learn about foundational, classical and key subjects of databases, and data management in general, through applications to data integration and the interoperability of different kinds of data managements systems and data repositories.

Data integration is one of the main areas in databases, and one of increasing importance. We are frequently confronted to the problem combining data that reside in a growing number of different, independent and heterogeneous data sources. In order to use those data through applications, data have to be collected from those sources and integrated.

In this course several problems, concepts and techniques will be considered. They have to do with physical (or materialized) and virtual data integration. Emphasis will be placed on data models, architectures, query processing, meta-information, and consistency issues.

In the context of virtual data integration, logic-based techniques for specification of data sources and query answering will be presented and investigated. Connections will be established with other relevant and related areas, among others: Schema Mappings, Metadata Management, Data Exchange, Peer-to-Peer Data Exchange, Query Answering using Views, Query Containment, Semantic Web and Ontologies, Ontology-Based Data Access (OBDA).

This is a foundational and research oriented course on data integration and related areas of databases. A scientific, fundamental, and logical approach is followed. The course is based on state-of-the-art research material.

Table of contents (tentative): 0. Review of relational databases. 1. Introduction. 2. Introduction to data warehousing. 3. Logic-based query languages. 4. Introduction to virtual data integration. 5. Concepts and techniques for virtual data integration. 6. Consistency issues in data integration systems. 7. Query containment and query answering using views. 8. Data exchange. 9. Semantic P2P data exchange systems. 10. Multidimensional databases. 11. Ontologies and databases.

Reading: Course lectures notes will be posted on the course web page after every lecture (careful reading of them is mandatory). Additional, auxiliary material (books, papers, web resources) can be found at the course web page.

There are a few recent books and monographs that cover some of the themes of the course. They are listed as auxiliary recommended reading. The first three can be downloaded through the Carleton University Library.

1. M. Arenas, P. Barcelo, L. Libkin, and F. Murlak. *Relational and XML Data Exchange*. Synthesis Lectures on Data Management, Morgan & Claypool Publishers, 2010.

2. M. Genesereth. *Data Integration: The Relational Logic Approach*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2010.

3. S. Greco, C. Molinaro, and F. Spezzano. *Incomplete Data and Data Dependencies in Relational Databases*. Synthesis Lectures on Data Management, Morgan & Claypool Publishers, 2012.

4. A. Doan, A. Halevy, and Z. Ives. *Principles of Data Integration*. Morgan Kaufmann, 2012.

5. M. Arenas, P. Barcelo, L. Libkin, and F. Murlak. *Foundations of Data Exchange*. Cambridge University Press, 2014.

6. S. Abiteboul. I. Manolescu, Ph. Rigaux, M-Ch. Rousset, P. Sennellart. *Web Data Management*. Cambridge University Press, 2012.