Spring 2007

CS 790-03: Information Integration & Analysis

Amit P. Sheth
Wright State University - Main Campus, amit.sheth@wright.edu

Follow this and additional works at: https://corescholar.libraries.wright.edu/cecs_syllabi
Part of the Computer Engineering Commons, and the Computer Sciences Commons

Repository Citation

This Syllabus is brought to you for free and open access by the College of Engineering & Computer Science at CORE Scholar. It has been accepted for inclusion in Computer Science & Engineering Syllabi by an authorized administrator of CORE Scholar. For more information, please contact corescholar@wwwlibraries.wright.edu, library-corescholar@wright.edu.
This course deals with the integration and analysis of heterogeneous data – a topic of significant research as well as practical importance. A typical large organization spends around 30% of all its IT effort and expense for data/information integration. Furthermore, organizations can greatly increase their competitiveness by exploiting the data they collect, and discovering trends and patterns leading to insights. High repositories of public data on the Web also provide opportunity to discovery undiscovered public knowledge.

This course will introduce the requirements/challenges and some approaches to integrate and analyze data. It will expose students to the following concept, topics, architectures, techniques and technologies:

- data, metadata, information, knowledge and ontologies
- unstructured, semi-structured, structured, multimodal, multimedia and sensor data
- syntax, structural/representational, and semantic aspects of data
- architectures: federated databases, mediator, information brokering
- integration and analysis of Web based information
- automatic information/metadata extraction (entity identification/recognition, disambiguation)
- relevant Web standards and technologies
- real world examples that have major research projects and commercial products

The class assumes basic knowledge in databases and Web technologies (e.g., search engines). It is an ideal first course among a series of advanced courses we will offer on Semantic Web and Services Sciences as part of our education and research components of the Advanced Data Management Program at Ohio's Wright Center of Innovation. See additional activities in this area at Kno.e.sis Center (http://knoesis.org) and dataOhio (http://www.daytaohio.com). The class will start with lectures and then move on to rigorous discussion of course material (most of which will be available on the Web) and end with a mini-project.

**Prerequisite:** Student should have taken at least one course related to data/information management or knowledge representation, such as: 605-4 (introduction to Database Management), 701-4 (database systems & Design), 705-4 (Data Mining), 711-4 (Knowledge-Based Systems in AI). Students with working knowledge of data management can get a waiver to the prerequisite from the faculty. Student must have working knowledge of programming in Java or C++.

Faculty: Prof. Amit Sheth, 367 Joshi Research Center, http://knoesis.org/~amit (amit.sheth@wright.edu, phone: 775-5203).

Type of material this course will cover is exemplified by:
- Multimedia Data Management: Using Metadata to Integrate and Apply Digital Media, Sheth and Klas (Eds).
- Reference reconciliation in complex information spaces, Dong, Halevy, Madhavan.