CM3010 Databases and Advanced Data Techniques

Module description

Databases and data services form the basis of many of the systems we use today. Each element of the process of collecting, modelling, storing, filtering and processing data can raise technical, conceptual, methodological and ethical challenges that must be engaged with in a critical way. This module builds on concepts and skill relating to data structures, data handling and web technologies from previous models, providing advanced level knowledge and techniques for working with data in a computational environment.

In this module, you will develop how you think about and work with data. You will learn how to use relational databases, as well as alternative approaches (NoSQL, XML and Linked Data). You will explore the ethics of gathering, processing and sharing data, and look at issues of security. You will also learn about rich data sources such as text, audio and video and strategies for working with them.

Module goals and objectives

Upon successful completion of this module, you will be able to:

- Design and explain an appropriate model for data using the entity/relationship and relational model;
- Implement a database in an SQL RDBMS;
- Write SQL queries using indexes and joins to read and write data from a database;
- Access the full functionality of an SQL database from a variety of programming languages, and build a simple web application to read and write data;
- Understand and explain the principles, forms and commands for NoSQL databases;
- Understand and explain different paradigms for data modelling, with particular focus on the Semantic Web and Linked Data
- Discuss critical issues relating to the ethics of working with data and data security and apply this knowledge to particular situations
- Understand the challenges of working with text and media and explain how systems for working with this data overcome these challenges

Textbook and Readings

Specific essential readings for each week from the following list are included in the Readings page for each week. There is no required single textbook for this module, although we will draw

heavily on the study guide written by the course author, David Lewis, and on C. J. Date. Database Design and relational Theory. O'Reilly 2012.

Additional readings will come from academic writing from the early pioneers up to recent research.

Module outline

The module consists of ten topics that focus on key areas of the fundamentals of computer science.

Topic 1.	**Key concepts:** • Sources of data • Data and knowledge structures • Data sharing **Learning outcomes:** • Find, describe, evaluate data sources • Understand the forms data comes in • Evaluate data-related access and reuse rights
Topic 2.	**Key concepts:** • Relational theory • The Entity/Relationship model • SQL **Learning outcomes:** • Understand core concepts of relational theory • Design a database using E/R diagrams • Create and explore an SQL database
Topic 3.	**Key concepts:** • Database integrity • Transactions and ACID • Security and users **Learning outcomes:**

	• Understand the risk of repeated information and the role of normalization in managing that risk • Understand, explain and plan transactions • Control database access through security policies
Topic 4.	**Key concepts:** • Server/client interactions for RDBMS • Web Templating • Simple data visualisation **Learning outcomes:** • Connect to and interact with an SQL RDBMS from a range of clients • Build a simple web application reading and writing data from a database server • Build a simple data visualiser based on data retrieved from a server
Topic 5.	**Key concepts:** • Database Indexes • Denormalization • Query optimization **Learning outcomes:** • Analyse a database query for efficiency • Assess and implement ways to improve speed or reliability
Topic 6.	**Key concepts:** • Distributed database models • MongoDB • Map/Reduce **Learning outcomes:** • Assess a projects needs for NoSQL approaches

	• Understand and explain the principles of document and key/value databases • Have a basic understanding of using MongoDB and Hadoop
Topic 7.	**Key concepts:** • XML databases • XML Schemata • XPath, XSLT and XQuery **Learning outcomes:** • Read and understand XML and XML schemas • Model data in XML • Use XSLT to transform an XML document • Understand and evaluate XML as a database technology
Topic 8.	**Key concepts:** • Linked Data and RDF • Web ontologies • Triplestores and SPARQL **Learning outcomes:** • Understand and design simple Linked Data systems • Find, combine and extend web ontologies • Query public triplestores using SPARQL
Topic 9.	**Key concepts:** • Multimedia databases • Feature spaces • Information Retrieval **Learning outcomes:** • Understand and explain approximate matching

	• Understand the principles of features and their extraction • Understand and explain notions of distance in vector space • Understand and evaluate retrieval metrics
Topic 10.	**Key concepts:** • Data models and database models • Preparing and cleaning data • Data ethics **Learning outcomes:** • Use your knowledge to choose appropriate approaches to modelling and implementation • Understand approaches to data preparation • Discuss and act on ethical and legal considerations for gathering, using and sharing data

Activities of this module

The module is comprised of the following elements

- **Lecture videos**
  In each topic the concepts you need to know will be presented through a collection of short video lectures. You may stream these videos for playback within the browser by clicking on their titles or download the videos
- **Practice Quizzes**
  Throughout this module, you will find many short quizzes. These are intended for you to assess and consolidate your understanding of topics and techniques immediately after they are introduced or used. You are advised to take these quizzes directly after exploring the preceding item.
  None of the quizzes in this module in any way contribute toward your final mark for the module.
- **Ungraded lab exercises**
  These are practical exercises using a special virtual environment from within your browser. You will be able to build and interact with a database from within the environment. You are strongly encouraged to explore these labs well, since practice is

very important for fluency with databases. In some cases, you will also be able to save a read-only link to the lab that you can share with fellow students and tutors.
Several of the labs are followed by discussion prompts for sharing what you have found out.

- **Peer Reviewed Assignments**
  Some activities benefit from being shared and critically appraised, and the act of evaluating the work of others can be useful practice for teamwork and constructive self-criticism. For some tasks, you will read a short description of the assignment and submit your work. It will then be graded by your peers according to clear criteria, and you will evaluate the work of others.
  Peer assignment grades do not contribute towards your final mark for the module.
- **Discussion prompts**
  This module includes many discussion prompts. Each prompt provides a space for you to respond. After responding, you can see and comment on your peers' responses. All prompts and responses are also accessible from the general discussion forum and the module discussion forum.
- **Readings**
  Useful additional reading can be found throughout the module. These are good supplementary materials that will either expand on topics that we cover elsewhere or extend them.
- **Staff-Graded Assignment**
  Half-way through this module, you will submit an assignment consisting of a report and a link to a lab. The lab will contain a database and a simple web application for it. The whole process of modelling the data, and implementing the database and web application will be assessed. Clear instructions and information about assessment criteria will be provided.

How to pass this module

The module has two major assessments each worth 50% of your grade:

- Coursework: This consists entirely of the staff-graded assignment described above, which will be assessed half way through course
- Examination: Arrangements for this examination will be provided separately.

- The mark shown on the Coursera platform is your coursework mark. This is 50% of your grade, with the exam providing the remainder.

BSc Computer Science programme

Activity	Required?	Deadline week	Estimated time per module	% of final grade
Written, staff graded coursework	Yes	11	Approximately 25 hours	50%
Written examination	Yes	22	2	50%