THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

BUSINESS SCHOOL

Course Outline 2017
INFOMGMT 390: SPECIAL TOPIC—DATA WRANGLING (15 POINTS)

Semester 2 (1175)

Course prescription

Organisations are increasingly adopting big data analysis, predictive analytics, social data mining, and deep machine learning to gain business intelligence and insight. The value of such technologies relies on having high-quality data, yet raw data is messy. Data wrangling is the often neglected, yet highly value-adding process of transforming raw data into a useful form for downstream consumption. This course provides students with a data wrangling toolbox to add value to data. Students will work with data sets from Instagram, Facebook, Twitter, IMDB, Statistics NZ, and other public sources.

Programme and course advice

Prerequisite: 15 points at Stage II in Information Management or Information Systems

Goals of the course

This course aims to:

- provide students with a digital data management toolbox that will be universally valuable for business data in all disciplines and industries;
- develop students' skills in analytical thinking, programmatic problem solving, and business communication;
- develop students' proficiency in industry-relevant software tools to make raw data valuable and meaningful for a business audience; and
- expose students to the digital data life-cycle model.

Learning outcomes

Students will use industry-relevant software to achieve all learning outcomes.
By the end of this course it is expected that students will be able to:

1. work with big data that is unstructured and non-relational;
2. recognise, read, comprehend, and manage structured data in various formats;
3. source XML and JSON data from online sources using REST APIs;
4. clean data in various formats;
5. profile data to verify its veracity and usefulness;
6. manipulate raw data from one format into another;
7. document and communicate data wrangling processes; and

8. add value to data by transforming it for downstream consumption.

Content outline

The general content outline for this course is:

Topic 1: Wrangling data in spreadsheets

Topic 2: Wrangling XML and JSON data from the web

Topic 3: Wrangling Big Data with MongoDB

Topic 4: Data cleaning and ETL tools

An hour-by-hour course schedule will be provided on Canvas.

Learning and teaching

The learning and teaching philosophy behind this course is that proficiency in practical skills reflects sound understanding of the content.

Students are expected to participate in two 2-hour classes each week. Classes are structured so that students are guided through practical exercises to develop their data wrangling skills. The class content is then reinforced by self-study exercises, instant-feedback assignments, and practical tests.

In addition to the weekly classes, students are expected to spend approximately another six hours per week on activities related to this course. These activities include reading, revising, practising, completing assignments, and preparing for assessments.

Teaching staff

Lecturer

Ron Tiong | OGGB Room 460 | r.tiong@auckland.ac.nz

Learning resources

This course will require students to use the following software/tools:

Excel 2016 (Some BI features are only available on the Windows version)
Notepad++ (Windows) or TextWrangler (OSX) or other text editor
Google Chrome DevTools
MongoDB and Robomongo (or a cloud-hosted option on Cloud9)
Visual Studio Code
Windows command prompt or OSX Terminal (bash)

Students may be required to use:

wget or curl
OpenRefine
Pentaho Data Integration

All the learning resources used in this course are available online for free online to students enrolled in this course. Links to resources will be provided in Canvas.

Assessment

Excel Assignment	20%
MongoDB Assignment	15%
Cleaning Assignment	15%
Excel Test	25%
MongoDB Test	25%
Total	100%

Pass requirement: Students are required to pass the tests to pass this course, i.e. their combined score for all tests is at least 25 out of 50.

Academic integrity: In attempting any assessment, students agree to adhere to all the principles and practices of academic honesty and integrity for the University of Auckland outlined here: https://www.auckland.ac.nz/en/about/learning-andteaching/policies-guidelines-and-procedures/academic-integrity-info-forstudents.html. The work that a student submits for marking must be the student's own work, reflecting his or her learning. A student's submitted work may be reviewed against electronic source material using computerised detection mechanisms. Any form of cheating, plagiarism, assistance in cheating, unfair collaboration, or other behaviour deemed to be academic misconduct will not be tolerated. Academic misconduct will be dealt with according to University's Student Academic Conduct Statute outlined here: https://cdn.auckland.ac.nz/assets/central/about/the-university/how-the-university-works/policy-and-administration/student-academic-conduct-statute.pdf.

The broad relationship between these assessments and the course learning outcomes is as follows:

Learning outcome	Excel Assignment	MongoDB Assignment	Cleaning Assignment	Excel Test	MongoDB Test
1		✓	✓		✓
2	✓	✓	✓	✓	✓
3			✓		
4	✓	✓	✓	✓	✓
5	✓	✓	✓	✓	✓
6	✓	✓	✓	✓	
7			✓	✓	✓
8	✓	✓	✓	✓	✓

Inclusive learning

Students are urged to privately discuss any impairment-related requirements with the course staff.

Student feedback

Student feedback is important to us and will be used to improve the course from semester to semester. This semester you may be asked to complete evaluations on

the teaching of the course. Please note that you do not have to wait until these evaluations are conducted in order to provide feedback. If there is something that you think we could improve then please let the course staff know as soon as possible.

In the event of an unexpected disruption

We undertake to maintain the continuity and standard of teaching and learning in all your courses throughout the year. If there are unexpected disruptions, the University has contingency plans to ensure that access to your course continues and your assessment is fair, and not compromised. Some adjustments may need to be made in emergencies. In the event of a disruption, the University and your course coordinators will make every effort to provide you with up-to-date information via Canvas and the university website.