COMP30018 Knowledge Technologies

Credit Points:	12.5

Level:	3 (Undergraduate)

Dates & Locations:	2015, Parkville This subject commences in the following study period/s: Semester 1, Parkville - Taught on campus. Semester 2, Parkville - Taught on campus.

Time Commitment:	Contact Hours: 36 hours, comprising of two 1-hour lectures and one 1-hour workshop per week Total Time Commitment: 170 hours

Prerequisites:	**One of the following:**

Subject	Study Period Commencement:	Credit Points
COMP20003 Algorithms and Data Structures	Semester 2	12.50
COMP90038 Algorithms and Complexity	Semester 1, Semester 2	12.50
COMP20007 Design of Algorithms	Semester 1	12.50

Corequisites:	None

Recommended Background Knowledge:	None

Non Allowed Subjects:	Students cannot enrol in and gain credit for this subject and:

Subject	Study Period Commencement:	Credit Points
COMP90049 Knowledge Technologies	Semester 1, Semester 2	12.50

433-352 Data on the Web

Core Participation Requirements:	For the purposes of considering request for Reasonable Adjustments under the Disability Standards for Education (Cwth 2005), and Student Support and Engagement Policy, academic requirements for this subject are articulated in the Subject Overview, Learning Outcomes, Assessment and Generic Skills sections of this entry. It is University policy to take all reasonable steps to minimise the impact of disability upon academic study, and reasonable adjustments will be made to enhance a student's participation in the University's programs. Students who feel their disability may impact on meeting the requirements of this subject are encouraged to discuss this matter with a Faculty Student Adviser and Student Equity and Disability Support: http://services.unimelb.edu.au/disability

Coordinator:	Assoc Prof Karin Verspoor, Prof Rao Kotagiri

Contact:	Email: **karin.verspoor@unimelb.edu.au (mailto:karin.verspoor@unimelb.edu.au)**

Subject Overview:	**AIMS** Much of the world's knowledge is stored in the form of unstructured data (e.g. text) or implicitly in structured data (e.g. relational databases). In this subject students will learn algorithms and data structures for extracting, retrieving and storing analysing explicit knowledge from various data sources, with a focus on the web.

The aim of this subject is to introduce students to knowledge technologies and to provide a foundational knowledge of data science. The subject will also give students exposure to what applied research is all about.

INDICATIVE CONTENT

Topics include: data encoding and markup, web crawling, regular expressions, document indexing, text retrieval, basic probability, clustering, pattern mining, Bayesian learning, instance-based learning, and prediction and approaches to evaluation of knowledge technologies.

Examples of projects that students may complete are:

  # A method for automatically predicting the geo-location of a Twitter user on the basis of their posts
  # An automatic method for tagging multilingual Wikipedia documents with Wikipedia categories
  # A search engine for Twitter data, which takes into account the time stamp of the query and documents
  # A search engine for web user forum data
  # A search engine servicing mixed monolingual queries (as in monolingual queries from a range of languages) over a large-scale document collection
  # Classification and prediction of some real world problems using machine learning techniques

Learning Outcomes:	**INTENDED LEARNING OUTCOMES (ILO)** Having completed this unit the student is expected to describe and apply the fundamentals of knowledge systems, including data acquisition and aggregation, knowledge extraction, text retrieval, machine learning and data mining. On completion of this subject the student is expected to: 1 Gain an understanding of a representative selection of knowledge technology techniques in both theoretical and applied contexts 2 Develop familiarity with component technologies used in commonly-deployed knowledge technology systems 3 Get a feel for what research is all about, especially relating to knowledge technology-related projects underway at The University of Melbourne
Assessment:	Project work during semester incorporating both programming and a report, requiring approximately 30 - 35 hours of work; one project due approximately mid-semester, and a second due in Week 11 or 12 (30%) A mid-semester test (10%) 2-hour end-of-semester examination (60%). Hurdle requirement: To pass the subject, students must obtain at least: 15/30 in project work And 35/70 in the mid-semester test and end-of-semester written examination combined. ILO 1 is addressed in the projects (applied) and the mid-semester test and final exam (theoretical). ILO 2 is addressed in the projects (through using a range of systems that are provided to students or that students experiment with themselves). ILO 3 is also addressed in the projects (which are generally themed around projects underway at the University, to give them a more applied feel).
Prescribed Texts:	TBA

Breadth Options:	This subject potentially can be taken as a breadth subject component for the following courses:
	
	  # **Bachelor of Arts (https://handbook.unimelb.edu.au/view/2015/B-ARTS)**
	  # **Bachelor of Commerce (https://handbook.unimelb.edu.au/view/2015/B-COM)**
	  # **Bachelor of Environments (https://handbook.unimelb.edu.au/view/2015/B-ENVS)**
	  # **Bachelor of Music (https://handbook.unimelb.edu.au/view/2015/B-MUS)**
	
	You should visit **learn more about breadth subjects (http://breadth.unimelb.edu.au/ breadth/info/index.html)** and read the breadth requirements for your degree, and should discuss your choice with your student adviser, before deciding on your subjects.

Fees Information:	Subject EFTSL, Level, Discipline & Census Date, http://enrolment.unimelb.edu.au/fees
Generic Skills:	On completion of this subject, students should have developed the following skills: # An ability to apply knowledge of basic science and engineering fundamentals

	# An ability to undertake problem identification, formulation and solution # The capacity to solve problems, including the collection and evaluation of information # The capacity for critical and independent thought and reflection # An expectation of the need to undertake lifelong learning, and the capacity to do so.
Notes:	**LEARNING AND TEACHING METHODS** The subject is delivered through a combination of lectures and tutorials. One feature of the subject is that the projects are designed to be relatively open-ended and broad enough that students have scope to get hands-on experience implementing the breadth of material covered in the subject, as well as building off the subject content in innovating their own methods/ researching related methods from the research literature and implementing them such methods themselves. **INDICATIVE KEY LEARNING RESOURCES** Students will have access to lecture slides, readings relating to the lecture materials (both from a textbook and conference/journal papers), tutorial worksheets with worked solutions for all numeric problems, and sample reports to use in writing the project reports. Students are permitted to do their programming in any language and any programming environment/OS, and may be given the option of working in a team (with suitably increased expectations on what they are required to do). In recent years, the projects have been hosted on Kaggle, supporting a live "scoreboard" for student systems, and giving the projects more of a real-world feel. **CAREERS / INDUSTRY LINKS** The knowledge technologies industry (encompassing machine learning, data science, natural language processing and information retrieval) has been growing rapidly over the past two decades, with key industry players including Google, Microsoft, Amazon, Facebook and Twitter. Google has sponsored a prize for the highest-achieving student in the subject each in recent years, underlining its interest in the subject material. There have been guest lecturers in the subject from organisations including Palantir Technologies and NICTA.
Related Majors/Minors/ Specialisations:	B-ENG Software Engineering stream Computer Science Computer Science Computer Science Master of Engineering (Software with Business) Master of Engineering (Software) Science-credited subjects - new generation B-SCI and B-ENG. Selective subjects for B-BMED