



COMP3220

Document Processing and the Semantic Web

Session 1, Weekday attendance, North Ryde 2020

Department of Computing

Contents

<u>General Information</u>	2
<u>Learning Outcomes</u>	2
<u>General Assessment Information</u>	3
<u>Assessment Tasks</u>	3
<u>Delivery and Resources</u>	3
<u>Unit Schedule</u>	5
<u>Policies and Procedures</u>	5
<u>Assessment Standards</u>	7

Disclaimer

Macquarie University has taken all reasonable measures to ensure the information in this publication is accurate and up-to-date. However, the information may change or become out-dated as a result of change in University policies, procedures or rules. The University reserves the right to make changes to any information in this publication without notice. Users of this publication are advised to check the website version of this publication [or the relevant faculty or department] before acting on any information in this publication.

General Information

Unit convenor and teaching staff

Rolf Schwitter

rolf.schwitter@mq.edu.au

Contact via Email

4 Research Park Drive; Office 359

By appointment

Diego Molla-Aliod

diego.molla-aliod@mq.edu.au

Contact via Email

4 Research Park Drive; Office 358

By appointment

Credit points

10

Prerequisites

130cp at 1000 level or above including COMP2110 or COMP249 or COMP2200 or COMP257

Corequisites

Co-badged status

Unit description

This unit explores the issues involved in building natural language processing (NLP) applications that operate on large bodies of real text such as are found on the world wide web. In this unit we discuss some core methods and tools for dealing with data on the web; in particular machine learning platforms widely used in industry. The unit also explores some recent developments of the web, such as emerging semantic web technologies and the corresponding standards promoted by the Word Wide Web Consortium (W3C). Application areas covered include web search, sentiment analysis, and information extraction.

Important Academic Dates

Information about important academic dates including deadlines for withdrawing from units are available at <https://www.mq.edu.au/study/calendar-of-dates>

Learning Outcomes

On successful completion of this unit, you will be able to:

ULO1: Explain the main techniques that are used to develop and implement intelligent

document processing applications.

ULO2: Describe the functionality of the key components in document processing architectures.

ULO3: Implement text processing applications using a programming language.

ULO4: Apply web technology to document processing.

Assessment Tasks

Coronavirus (COVID-19) Update

Assessment details are no longer provided here as a result of changes due to the Coronavirus (COVID-19) pandemic.

Students should consult [iLearn](#) for revised unit information.

[Find out more about the Coronavirus \(COVID-19\) and potential impacts on staff and students](#)

General Assessment Information

The assessment of this unit consists of three assignments and a final exam. You will submit the solutions to the three assignments via iLearn by the due date. The final examination is a closed book examination, and will be taken in person during the exam period.

Late Submission

No extensions will be granted without an approved application for [Special Consideration](#). There will be a deduction of 10% of the total available marks made from the total awarded mark for each 24 hour period or part thereof that the submission of the assignment is late. For example, 25 hours late in submission for an assignment worth 10 marks – 20% penalty or 2 marks deducted from the total. No submission will be accepted after solutions have been posted.

Supplementary Exam

If you receive [Special Consideration](#) for the final exam, a supplementary exam will be scheduled after the normal exam period, following the release of marks. By making a special consideration application for the final exam you are declaring yourself available for a resit during the supplementary examination period and will not be eligible for a second special consideration approval based on pre-existing commitments. Please ensure you are familiar with the policy prior to submitting an application. Approved applicants will receive an individual notification one week prior to the exam with the exact date and time of their supplementary examination.

Delivery and Resources

Coronavirus (COVID-19) Update

Any references to on-campus delivery below may no longer be relevant due to COVID-19.

Please check here for updated delivery information: https://ask.mq.edu.au/account/pub/display/unit_status

Required and Recommended Texts

Most of the contents of the unit will be based on the following two books:

- Steven Bird, Ewan Klein, Edward Loper. *Natural Language Processing -- Analyzing Text with Python and the Natural Language Toolkit*. Online at <http://www.nltk.org/book>.
- F. Chollet (2017). *Deep Learning with Python*. Manning Publications. Available in the library.

Additional material will be made available during the semester, in conjunction with the lecture notes. See the unit schedule for a listing of the most relevant reading for each week.

Technology Used and Required

The following software is used in COMP3220:

1. Anaconda for Python 3.7
2. NLTK (bundled with Anaconda)
3. Python SciKit-Learn (bundled with Anaconda)
4. gensim (can be installed using Anaconda)
5. spaCy (can be installed using Anaconda)
6. Keras (can be installed using Anaconda)
7. Tensorflow (can be installed using Anaconda)
8. XML Copy Editor
9. BaseX (XML Database Engine)
10. Saxon (XSLT and XQuery Processor)
11. rdflib (can be installed using Anaconda)
12. Protege (Ontology Editor)

This software is installed in the labs; you should also ensure that you have working copies of all the above on your own machine. Note that many packages come in various versions; to avoid potential incompatibilities, you should install versions as close as possible to those used in the labs.

Unit Web Page

Note that the majority of the unit materials is publicly available while some material requires you to log in to [iLearn](#) to access it.

The unit will make extensive use of discussion boards hosted within [iLearn](#). Please post questions there, they will be monitored by the staff on the unit.

Unit Schedule

Coronavirus (COVID-19) Update

The unit schedule/topics and any references to on-campus delivery below may no longer be relevant due to COVID-19. Please consult [iLearn](#) for latest details, and check here for updated delivery information: https://ask.mq.edu.au/account/pub/display/unit_status

Week	Topic	Reading
1	NLP Systems + Text Processing in Python	NLTK Ch 1
2	Information Retrieval	Manning et al. (2008)
3	Text Classification	NLTK Ch 6
4	Deep Learning for Text	Chollet, Ch. 2 & 3
5	Processing Text Sequences	Chollet, Ch. 6
6	Advanced Usage of Deep Learning for Text	Chollet, Ch. 8.1
7	Semi-structured Data	XSLT Tutorial at W3School
	<i>Recess</i>	
8	RDF, RDF Schema and SPARQL	RDF Primer SPARQL
9	Linked Data	DBpedia
10	Ontologies	Kroetzsch et al (2012) OWL Primer
11	Rule Languages	RIF Primer
12	Semantic Web Applications and Recent Trends	
13	Revision	

Policies and Procedures

Macquarie University policies and procedures are accessible from [Policy Central \(https://staff.mq.edu.au/work/strategy-planning-and-governance/university-policies-and-procedures/policy-central\)](https://staff.mq.edu.au/work/strategy-planning-and-governance/university-policies-and-procedures/policy-central). Students should be aware of the following policies in particular with regard to Learning and

Teaching:

- [Academic Appeals Policy](#)
- [Academic Integrity Policy](#)
- [Academic Progression Policy](#)
- [Assessment Policy](#)
- [Fitness to Practice Procedure](#)
- [Grade Appeal Policy](#)
- [Complaint Management Procedure for Students and Members of the Public](#)
- [Special Consideration Policy](#) (**Note:** *The Special Consideration Policy is effective from 4 December 2017 and replaces the Disruption to Studies Policy.*)

Students seeking more policy resources can visit the [Student Policy Gateway](https://students.mq.edu.au/support/study/student-policy-gateway) (<https://students.mq.edu.au/support/study/student-policy-gateway>). It is your one-stop-shop for the key policies you need to know about throughout your undergraduate student journey.

If you would like to see all the policies relevant to Learning and Teaching visit [Policy Central](http://staff.mq.edu.au/work/strategy-planning-and-governance/university-policies-and-procedures/policy-central) (<http://staff.mq.edu.au/work/strategy-planning-and-governance/university-policies-and-procedures/policy-central>).

Student Code of Conduct

Macquarie University students have a responsibility to be familiar with the Student Code of Conduct: <https://students.mq.edu.au/study/getting-started/student-conduct>

Results

Results published on platform other than [eStudent](#), (eg. iLearn, Coursera etc.) or released directly by your Unit Convenor, are not confirmed as they are subject to final approval by the University. Once approved, final results will be sent to your student email address and will be made available in [eStudent](#). For more information visit ask.mq.edu.au or if you are a Global MBA student contact globalmba.support@mq.edu.au

Student Support

Macquarie University provides a range of support services for students. For details, visit <http://students.mq.edu.au/support/>

Learning Skills

Learning Skills (mq.edu.au/learningskills) provides academic writing resources and study strategies to help you improve your marks and take control of your study.

- [Getting help with your assignment](#)
- [Workshops](#)
- [StudyWise](#)
- [Academic Integrity Module](#)

The Library provides online and face to face support to help you find and use relevant information resources.

- [Subject and Research Guides](#)
- [Ask a Librarian](#)

Student Services and Support

Students with a disability are encouraged to contact the [Disability Service](#) who can provide appropriate help with any issues that arise during their studies.

Student Enquiries

For all student enquiries, visit Student Connect at ask.mq.edu.au

If you are a Global MBA student contact globalmba.support@mq.edu.au

IT Help

For help with University computer systems and technology, visit http://www.mq.edu.au/about_us/offices_and_units/information_technology/help/.

When using the University's IT, you must adhere to the [Acceptable Use of IT Resources Policy](#). The policy applies to all who connect to the MQ network including students.

Assessment Standards

COMP3220 will be assessed and graded according to the University assessment and grading policies.

The following general standards of achievement will be used to assess each of the assessment tasks with respect to the letter grades.

Grade	Range	Description
HD	85-100	Provides consistent evidence of deep and critical understanding in relation to the learning outcomes. There is substantial originality, insight or creativity in identifying, generating and communicating competing arguments, perspectives or problem solving approaches; critical evaluation of problems, their solutions and their implications; creativity in application as appropriate to the course/program.
D	75-84	Provides evidence of integration and evaluation of critical ideas, principles and theories, distinctive insight and ability in applying relevant skills and concepts in relation to learning outcomes. There is demonstration of frequent originality or creativity in defining and analysing issues or problems and providing solutions; and the use of means of communication appropriate to the course/program and the audience.
CR	65-74	Provides evidence of learning that goes beyond replication of content knowledge or skills relevant to the learning outcomes. There is demonstration of substantial understanding of fundamental concepts in the field of study and the ability to apply these concepts in a variety of contexts; convincing argumentation with appropriate coherent justification; communication of ideas fluently and clearly in terms of the conventions of the course/program.
P	50-64	Provides sufficient evidence of the achievement of learning outcomes. There is demonstration of understanding and application of fundamental concepts of the course/program; routine argumentation with acceptable justification; communication of information and ideas adequately in terms of the conventions of the course/program. The learning attainment is considered satisfactory or adequate or competent or capable in relation to the specified outcomes.

Grade	Range	Description
F	0-49	Does not provide evidence of attainment of learning outcomes. There is missing or partial or superficial or faulty understanding and application of the fundamental concepts in the field of study; missing, undeveloped, inappropriate or confusing argumentation; incomplete, confusing or lacking communication of ideas in ways that give little attention to the conventions of the course/program.

Assessment Process

These assessment standards will be used to give a numeric mark to each assessment submission during marking. The mark will correspond to an appropriate letter grade when relevantly weighted. The final mark for the unit will be calculated by combining the marks for all assessment tasks according to the percentage weightings shown in the assessment summary.