UNSW Course Outline

COMP6714 Information Retrieval and Web Search - 2023

Course Code : COMP6714
Year : 2023
Term : Term 3
Teaching Period : T3
Delivery Mode : in person
Delivery Format : Standard
Delivery Location : Kensington

General Course Information

Course Code : COMP6714
Year : 2023
Term : Term 3
Teaching Period : T3
Is a multi-term course? : No
Faculty : Faculty of Engineering
Academic Unit : School of Computer Science and Engineering
Delivery Mode : in person
Delivery Format : Standard
Delivery Location : Kensington
Campus : Sydney
Study Level : Undergraduate, Postgraduate
Units of Credit : 6

Useful Links

Handbook Class Timetable

Course Details & Outcomes

Course Description

Information retrieval (IR) is the process of retrieving relevant information by specifying a query to

an IR system. Web search is one form of an IR system that allows users to search information on the Internet based on a search engine. It involves computing a numeric score on how well each result matches the query and ranking the results according to this score.

This course aims to introduce the concepts, theories, and algorithmic issues important to Information Retrieval. If time allows, the course will also cover some recent topics and common practices. The course is composed of the following parts:

Information Retrieval:

1. Document modeling
2. Inverted index construction and compression
3. Vector space model and ranking methods
4. Probabilistic and language models
5. Evaluation methods
6. Relevance feedback and query expansion.

Web Search:

1. Web search engine architecture
2. Web crawling and indexing
3. Web structure and usage analytics.

The lecture materials will be complemented by a non-programming assignment and a programming project.

Course Aims

This course aims to introduce the concepts, theories, and algorithmic issues important to Information Retrieval, including its manifestation as Web Search. At the end of the course students will have a solid understanding of IR, which will give them the capability to conduct better searches in large document collections, or to build systems to do such searching.

The course is an advanced elective for both undergraduate and postgraduate students.

Course Learning Outcomes

Course Learning Outcomes
CLO1 : Analyze the whole process of information retrieval and search engines
CLO2 : Apply various document and retrieval models in information retrieval
CLO3 : Use various indexing and query processing techniques and their variants
CLO4 : Develop solutions for real problems using existing IR technologies
CLO5 : Discuss the past, present and future of information retrieval and search engine technologies

Course Learning Outcomes	Assessment Item
CLO1 : Analyze the whole process of information retrieval and search engines	• Assignment (non-programming) • Project (programming) • Final Exam
CLO2 : Apply various document and retrieval models in information retrieval	• Project (programming) • Final Exam
CLO3 : Use various indexing and query processing techniques and their variants	• Assignment (non-programming) • Project (programming) • Final Exam
CLO4 : Develop solutions for real problems using existing IR technologies	• Project (programming)
CLO5 : Discuss the past, present and future of information retrieval and search engine technologies	• Assignment (non-programming) • Final Exam

Learning and Teaching Technologies

Moodle - Learning Management System | Echo 360 | Blackboard Collaborate | WebCMS3

Other Professional Outcomes

https://www.unsw.edu.au/engineering/student-life/student-resources/program-design

Additional Course Information

The official prerequisite of this course is COMP9020 and COMP9024 for postgraduates; and (MATH1081 and (COMP1531 or COMP2041)) or (COMP1927 or COMP2521) for undergraduates. That is, we assume you have:

- experience with procedural programming, and an understanding of a range of data structures (e.g., trees, graphs, hash-tables) and algorithms (e.g., sorting, divide-and-conquer); and
- knowledge of discrete mathematics, including sets, logic, functions and relations, and graphs and trees.

Furthermore, at the start of this course students should be familiar with Python programming:

- produce correct programs in Python, i.e., coding, running, testing, etc.;
- produce readable code with clear documentation; and

- appreciate the use of abstraction in computing.

Assessments

Assessment Structure

Assessment Item	Weight	Relevant Dates
Assignment (non-programming) Assessment FormatIndividual	10%	Due DateWeek 5
Project (programming) Assessment FormatIndividual	40%	Due DateWeek 9
Final Exam Assessment FormatIndividual	50%	Start DateNot Applicable Due DateTBA during Exam Week

Assessment Details

Assignment (non-programming)

Assessment Overview

This is a warm-up, non-programming assignment for the course. It will be lightweight (students are expected to be able to finish the assignment in a few hours).

Assessment will be in a form of short questions or multiple choices. It will be marked semi-automatically (e.g., via Moodle) and feedbacks/comments will be provided via a LMS system such as Moodle. Overall feedbacks will be discussed in the lectures, and students may discuss with the tutors in consultation sessions for further assessment feedbacks.

Detailed Assessment Description

Details are on the course website.

Submission notes

Submitted via Moodle

Project (programming)

Assessment Overview

This is the programming project for the course and it will be relatively heavier weight than the non-programming assignment, since it involves utilizing various techniques that students have learnt from the course and put them together to form a solution (students are expected to be work on the project over a few weeks).

Assessment of assignments will be primarily based on how accurately they satisfy the requirements; this means that most of the marks will be based on automatic marking. However, we may also manually examine submitted assignments to determine (a) whether they are written with good style, (b) how closely they satisfied the requirements, if time allows.

Individual graded results with optional comments will be emailed to each student. Overall feedbacks will be discussed in the lectures, and students may discuss with the tutors in

consultation sessions for further assessment feedbacks.

Detailed Assessment Description

Details are on the course website.

Submission notes

Submitted via Give

Final Exam

Assessment Overview

The final exam will be a major assessment in this course and aims to test what students learned about data compression and search during the course of the semester. To pass this course, students are required to have satisfactory performance on the final exam even if they do very well on the assignments. In order to meet the hurdle requirement, students must score better than 40% on the final exam. Note that the hurdle will be enforced after any required scaling.

Detailed Assessment Description

Details are on the course website.

Hurdle rules

To pass this course, students are required to have satisfactory performance on the final exam even if they do very well on the assignments. In order to meet the hurdle requirement, students must score better than 40% on the final exam. Note that the hurdle will be enforced after any required scaling.

General Assessment Information

Grading Basis

Standard

Course Schedule

Teaching Week/Module	Activity Type	Content
Week 1 : 11 September - 15 September	Topic	Introduction, Boolean Retrieval
Week 2 : 18 September - 22 September	Topic	Preprocessing
Week 3 : 25 September - 29 September	Topic	Index Construction
Week 4 : 2 October - 6 October	Topic	Compression

Week 5 : 9 October - 13 October	Topic	Vector Space Model
Week 7 : 23 October - 27 October	Topic	Evaluation
Week 8 : 30 October - 3 November	Topic	Crawling
Week 9 : 6 November - 10 November	Topic	Link Analysis
Week 10 : 13 November - 17 November	Topic	Optional Topics; Course Revision

Attendance Requirements

Students are strongly encouraged to attend all classes and review lecture recordings.

General Schedule Information

The course schedule is an **approximate** guide to the sequence of topics in this course. It is subject to change as the term progresses.

Course Resources

Prescribed Resources

Lecture slides will be posted on the course website. These slides summarise the major contents and help you understand the materials when you read the textbook later. You definitely need to read the corresponding chapters in the textbook to gain a full understanding of the lectures.

- [MRS08] Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze, Introduction to Information Retrieval, Cambridge University Press. 2008.
- [CMS09] W. Bruce Croft, Donald Metzler, and Trevor Strohman, Search Engines: Information Retrieval in Practice. Pearson. 2009.

Recommended Resources

Reference books for this course are:

- [JM19] Dan Jurafsky, James H. Martin, Speech and Language Processing (3rd ed. draft). 2019.
- [BCC10] Stefan Buettcher, Charles L. A. Clarke, Gordon V. Cormack, Information Retrieval: Implementing and Evaluating Search Engines, The MIT Press. 2010.
- [BB99] Ricardo Baeza-Yates and Ber thier Ribeiro-Neto, Modern Information Retrieval, Addison Wesley. 1999.

Useful Resources:

- http://nlp.stanford.edu/IR-book/information-retrieval.html
- http://www.ir.uwaterloo.ca/book/

Course Evaluation and Development

This course is evaluated each session using MyExperience. Based on the MyExperience comments from the previous offering, in this term, we will try to bring forward some topics (such as evaluation) such that the project can cover a wider scope (rather than on boolean search).

Since this is the second time that I take over and run this course, the course will be maintained with a similar style and structure as before, with the exception of some topics will be brought forward. Therefore, your feedback is particularly important and will be considered to improve future offerings of this course.

Students are also encouraged to provide informal feedback during the term and let the lecturer know of any problems, as soon as they arise. Suggestions will be listened to very openly, positively, constructively, and thankfully, and every reasonable effort will be made to address them as soon as possible.

Staff Details

Position	Name	Email	Location	Phone	Availability	Equitable Learning Services Contact	Primary Contact
Convenor	Raymond Wong	ray.wong@unsw.edu.au				Yes	No
	COMP6714 Class Account	cs6714@cse.unsw.edu.au				No	Yes

Other Useful Information

Academic Information

I. Special consideration and supplementary assessment

If you have experienced an illness or misadventure beyond your control that will interfere with your assessment performance, you are eligible to apply for Special Consideration prior to, or within 3 working days of, submitting an assessment or sitting an exam.

Please note that UNSW has a Fit to Sit / Submit rule, which means that if you sit an exam or submit a piece of assessment, you are declaring yourself fit enough to do so and cannot later apply for Special Consideration.

For details of applying for Special Consideration and conditions for the award of supplementary assessment, please see the information on UNSW's Special Consideration page.

II. Administrative matters and links

All students are expected to read and be familiar with UNSW guidelines and polices. In particular,

students should be familiar with the following:

- Attendance
- UNSW Email Address
- Special Consideration
- Exams
- Approved Calculators
- Academic Honesty and Plagiarism
- Equitable Learning Services

III. Equity and diversity

Those students who have a disability that requires some adjustment in their teaching or learning environment are encouraged to discuss their study needs with the course convener prior to, or at the commencement of, their course, or with the Equity Officer (Disability) in the Equitable Learning Services. Issues to be discussed may include access to materials, signers or note-takers, the provision of services and additional exam and assessment arrangements. Early notification is essential to enable any necessary adjustments to be made.

Note: This course outline sets out the description of classes at the date the Course Outline is published. The nature of classes may change during the Term after the Course Outline is published. Moodle or your primary learning management system (LMS) should be consulted for the up-to-date class descriptions. If there is any inconsistency in the description of activities between the University timetable and the Course Outline/Moodle/LMS, the description in the Course Outline/ Moodle/LMS applies.

Academic Honesty and Plagarism

UNSW has an ongoing commitment to fostering a culture of learning informed by academic integrity. All UNSW students have a responsibility to adhere to this principle of academic integrity. Plagiarism undermines academic integrity and is not tolerated at UNSW. *Plagiarism at UNSW is defined as using the words or ideas of others and passing them off as your own.*

Plagiarism is a type of intellectual theft. It can take many forms, from deliberate cheating to accidentally copying from a source without acknowledgement. UNSW has produced a website with a wealth of resources to support students to understand and avoid plagiarism, visit: student.unsw.edu.au/plagiarism. The Learning Centre assists students with understanding academic integrity and how not to plagiarise. They also hold workshops and can help students one-on-one.

You are also reminded that careful time management is an important part of study and one of the identified causes of plagiarism is poor time management. Students should allow sufficient time for research, drafting and the proper referencing of sources in preparing all assessment tasks.

Repeated plagiarism (even in first year), plagiarism after first year, or serious instances, may also be investigated under the Student Misconduct Procedures. The penalties under the procedures can include a reduction in marks, failing a course or for the most serious matters (like plagiarism in an honours thesis or contract cheating) even suspension from the university. The Student Misconduct Procedures are available here:

www.gs.unsw.edu.au/policy/documents/studentmisconductprocedures.pdf

Submission of Assessment Tasks

Work submitted late without an approved extension by the course coordinator or delegated authority is subject to a late penalty of five percent (5%) of the maximum mark possible for that assessment item, per calendar day.

The late penalty is applied per calendar day (including weekends and public holidays) that the assessment is overdue. There is no pro-rata of the late penalty for submissions made part way through a day. This is for all assessments where a penalty applies.

Work submitted after five days (120 hours) will not be accepted and a mark of zero will be awarded for that assessment item.

For some assessment items, a late penalty may not be appropriate. These will be clearly indicated in the course outline, and such assessments will receive a mark of zero if not completed by the specified date. Examples include:

- Weekly online tests or laboratory work worth a small proportion of the subject mark;
- Exams, peer feedback and team evaluation surveys;
- Online quizzes where answers are released to students on completion;
- Professional assessment tasks, where the intention is to create an authentic assessment that has an absolute submission date; and,
- Pass/Fail assessment tasks.

Faculty-specific Information

Engineering Student Support Services – The Nucleus - enrolment, progression checks, clash requests, course issues or program-related queries

Engineering Industrial Training – Industrial training questions

UNSW Study Abroad – study abroad student enquiries (for inbound students)

UNSW Exchange – student exchange enquiries (for inbound students)

UNSW Future Students – potential student enquiries e.g. admissions, fees, programs, credit transfer

Phone

(+61 2) 9385 8500 – Nucleus Student Hub

(+61 2) 9385 7661 – Engineering Industrial Training

(+61 2) 9385 3179 – UNSW Study Abroad and UNSW Exchange (for inbound students)

School Contact Information

CSE Help! - on the Ground Floor of K17

- For assistance with coursework assessments.

The Nucleus Student Hub - https://nucleus.unsw.edu.au/en/contact-us

- Course enrolment queries.

Grievance Officer - grievance-officer@cse.unsw.edu.au

- If the course convenor gives an inadequate response to a query or when the courses convenor does not respond to a query about assessment.

Student Reps - stureps@cse.unsw.edu.au

- If some aspect of a course needs urgent improvement. (e.g. Nobody responding to forum queries, cannot understand the lecturer)

You should **never** contact any of the following people directly:

- Vice Chancellor

- Pro-vice Chancellor Education (PVCE)

- Head of School

- CSE administrative staff

- CSE teaching support staff

They will simply bounce the email to one of the above, thereby creating an unnecessary level of indirection and a delay in the response.