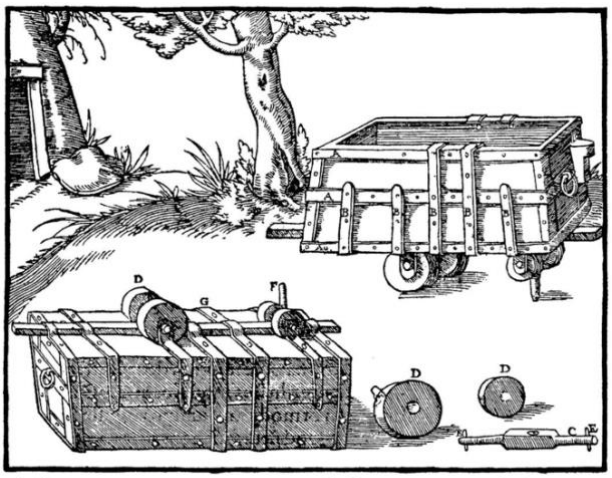
DIGI405 - Texts, Discourses & Data: the Humanities and Data Science

Course Guide
Semester 1, 2026



A—Rectangular iron bands on truck. B—Its iron straps. C—Iron axle.
D—Wooden rollers. E—Small iron keys. F—Large blunt iron pin.
G—Same truck upside down.

Leitnagel Hund (mining cart): 'Text mining' is a metaphor.

Course Description

This course examines computer-aided methods for analysing textual data, with consideration of the multiple ways that language reflects and shapes social meanings. Within this context, the course introduces concepts and methods for analysing natural language data and applies these through a series of practical lab classes.

The first part of the course focuses on discourse analysis methods drawn from corpus linguistics, as well as essential preprocessing steps used to prepare texts for a range of analytical purposes. In the second part of the course we explore supervised and unsupervised methods for identifying distinctive features of texts that allow us to group them according to attributes such as topic, sentiment, genre, or style.

Throughout the course we will examine the complexities of analysing texts, discourses and data at different scales. To avoid uncritically adopting the metaphor of 'text mining', we direct our attention to the wider methodological and interpretative issues involved and identify ethical and political dimensions to our analysis. We will read, analyse and discuss a broad selection of material, including political speeches, newspaper articles, fiction, and social media texts, and reflect on the knowledge claims that may be made when texts and discourses become data.

Learning Outcomes

Once you've successfully completed this course, you'll be able to:

- apply concepts and techniques to identify and explain patterns in text data, drawing on corpus linguistic approaches
- demonstrate concepts and techniques to model text data for information extraction and text classification
- explain how text data are connected to social, cultural, and linguistic structures and contexts
- identify ethical and political dimensions to the design and application of text analysis methods
- assess questions and claims made using text data as evidence and evaluate the alignment between research aims and methods
- appropriately link concepts from relevant research domains to interpret text analysis outputs
- develop written and verbal communication skills in applied text analysis.

Teaching Staff

The course coordinator is your main point of contact for the course. Contact details for the DIGI405 teaching team are available on AKO|LEARN. Please note that all course communications should be conducted via your UC student email account.

Dr Christopher Thomson
Kairuruku Akoranga | Course Coordinator
Office: Karl Popper 604
Email: christopher.thomson@canterbury.ac.nz
Office Hour: Tuesdays 3-4pm, Friday 1-2pm during term time.
Phone: +64 3 369 4646

Learn website

The main place to find information about classes and assessment is on our AKO|LEARN page. As well as information about labs and lectures, there are number of options for you to get support with your course work. Please read the Course Information page and Frequently Asked Questions to familiarize yourself with these options.

Class Times

Please see the Course Information Page (CIS) for class times and locations:
DIGI405-26S1 (C)
DIGI405-26S1 (D)
Please note these can be subject to change and it is recommended that you check before each class in the first couple of weeks of term.

Textbooks & Readings

There is no required textbook - all readings will be available through Learn. There are

required readings each week, which are an essential companion to lectures and labs. There will also often be optional readings that provide further context that extends material covered in class.

Schedule

This course will apply a 'flipped' learning model. Each week there will be video lectures, readings and other tasks to be completed before attending lectorials and labs. The lectorials will develop your understanding by working through examples, group activities and discussion. Each week there will be a 2-hour computer lab with hands-on exercises. Attendance at all classes is important for achieving well in the course.

Week	Topics	Assessment
1	Frequency and dispersion in text corpora	Corpus exploration assignment introduced.
2	Corpus analysis – keyness, collocation	
3	Discourse / Claims about text data	
4	Beyond words: annotating and analysing text data	
5	Text classification	Final project introduced.
6	Information extraction	Corpus exploration due, 5pm 25 March (30%)
Term break (3 weeks)		
7	Analysing topics	
8	~~In-class test – during lectorial time~~ Sentiment analysis	Written test (15%)
9	Genre classification	
10	Analysing style	
11	Writing the project report	
12	Project workshop / Course reflection	Draft of Final Project due 5pm 27 May.
Study week		Final Project due 5pm Wednesday 3 June. (35%) Learning Journal due 5pm Friday 29 May (20%) (Note: this is produced throughout the course – see Assessments)

Assessment

Task	Description	Due date
Learning Journal / Active participation (20%)	Each week you should make non-trivial contributions to active learning. Expectations for this will be set at the start of the course. 5% is awarded for weekly completion of journal entries (signed off in lab classes); the finished journal is worth 15%.	Weekly, then submitted when complete after week 12.
Corpus exploratory analysis (30%)	Analyse written texts using corpus linguistics techniques.	Due 5pm Wednesday 25 March.
~~In-class test~~ Written test (15%)	A closed-book written test completed during class time (or by arrangement for distance students). Includes multiple choice and short essay questions.	~~1pm, Tuesday 28 April (Week 8) – during normal lectorial time.~~ 7pm, Tuesday 28 April (Week 8). Room TBC.
Final project (35%)	Apply skills and concepts learned throughout the course to produce a final project on a selected problem or topic.	Draft report due 5pm Friday 29th May. Final report due 5pm Friday 5th June.

Your active participation assessment will sometimes build on tasks you complete **during lectorials and labs**. This applies to both on-campus and distance students. Make sure you are available for all lab sessions.

In order to be able to submit major assessments, you need to complete the **UC Academic Integrity Module**. This takes up to three hours to complete. A link to the module provided on the DIGI405 AKO|LEARN site. Many courses across the UC require completion of this module, and once you have completed it for one course, you count as having completed it for every other course. (I.e., you will not need to do it twice, and if you have already completed it for another course then you will not need to do it again for this one).

Workload

DIGI405 is a 15 point course and is designed to take about 150 hours to complete. The scheduled classes and regular preparation will take about 75 hours, with the remainder of the time devoted to completing your pieces of assessment. An approximate breakdown of the course work is as follows:

Lectures, lectorials and labs	40 hours
Reading, preparation and active participation	36 hours (12 x 3hrs per week)
Corpus exploratory analysis	25 hours
In-class test	10 hours
Final project	40 hours
Total	~150

JupyterHub

We provide a JupyterHub server for you to run code provided in labs and assessments in this course. We encourage you to use this to experiment with your own code, but please be aware of the following points.

- Please avoid generating large outputs or collecting very large amounts of data via web scraping. If you are unsure, please check with your course tutor or lecturer.
- Jupyterhub is provided for the duration of the course only. **Your access will end and your files will be removed from the server one week after the end of the examination period**. Please ensure you download any notebooks or files you wish to keep before this date. We recommend you download a copy of your important files regularly throughout the course.

Use of Generative AI

Generative AI includes generative AI tools like ChatGPT, but also any generative AI functionality built into software (e.g. Grammarly, Quillbot, word processing software, browser plugins, translation software).

As educators, we need to be sure we are assessing your work / thinking / learning / writing. In general, we expect your written assignments in this course to be **your own work.** We want to see how you describe, explain and present your ideas and results. For this reason, assignment marking criteria will focus on the demonstration of your own thinking, analysis, presentation of examples and evidence.

As students, we hope you want to make the most of your opportunities to develop knowledge and skills, including critical thinking and communication skills. You should also be aware that there are risks of generating false or ill-suited material if you use AI writing tools, translation software, or language/grammar tools backed by a language model. It is easy to use these tools to produce text material that sounds plausible but does not meet the criteria for an assignment.

Software branded as a tool to improve writing (e.g. Grammarly, Quillbot), with functionality to generate, substantially rewrite or augment text, potentially crosses a line into academic misconduct. Text generated, rewritten, or augmented by these tools is routinely flagged as AI and investigated.

If you are using software tools to translate from another language, it is likely that any translated output text may be flagged as AI. Please contact the course coordinator at the start of the course to discuss if you are planning to use translation software to help you study.

UC policy states that using generative AI to generate text or other content for assessments without permission is a breach of academic integrity and is considered academic misconduct. See the **Dishonest Practice** section for more information on academic misconduct processes and potential consequences.

Please be sure to read all assessment instructions carefully so you know whether use of generative AI tools is permitted or not. If permission is not expressly granted in assessment instructions, then use of these tools is not permitted. Ensure that you follow the guidance on usage, including how permitted usage should be acknowledged in your assignment. In some cases you may be required to provide draft writing or other proof you have written the assignment yourself.

There will be opportunities to learn about large language models and their applications during the course, as well as the limitations of generated text, problems of bias, and the potential to uncritically reproduce dominant ideas.

UC has an information page that provides general information about the use of generative AI.

If you are unsure about appropriate use of these tools, please ask the course coordinator.

AI Detection

We want to be confident that we are assessing your original work. We therefore may use AI detection on all written assessments, including any where AI tools are permitted.

We strongly recommend you document your writing process. You can use Microsoft Word and OneDrive to record the version history of your documents. You can then provide earlier drafts, if required, as evidence that your submitted assessment is your own work.

Dishonest Practice

Students found to be engaging in dishonest practices will be dealt with under the University Discipline Regulations. Penalties range from refusal to grade the assignment in question, to exclusion from the University. The Digital Humanities Programme understands 'dishonest practice' to include:

- Plagiarism: The presentation of any material without adequate acknowledgement of the source. This includes text, data, figures, computer code etc, on any medium (print or online). All sources should be adequately referenced, following accepted scholarly standards. Refer to the latest edition of the MLA Style Manual or the many available online resources for more information.
- Collusion: Working with another person or persons and presenting it as if it was your work alone. While collaboration and discussion is important in this course, students need to be careful about acknowledging everyone who has had a significant involvement in their work. Students are encouraged to discuss ideas and help each other develop them during the research stage of

their assignments, but it is not acceptable to submit identical work for assignments.

- ➔ Ghost Writing: Using another person to write an essay or assignment for you. This applies whether they are paid or not.
- ➔ Use of generative AI / text generation tools: Please read over the section in this course outline on **Use of Generative AI**. That section provides information on academic misconduct related to use of generative AI and course rules related to use of generative AI. If the course coordinator has good reason to suspect that a student has submitted work produced (or partially produced) using generative AI, they may ask for an explanation and/or evidence that the student did the required work. If the course coordinator is not satisfied with the explanation and/or evidence, they may withhold an assessment grade and delay a decision about academic misconduct action until they have had a chance to evaluate other work completed by the student.

Extensions and Late Submission

The Digital Humanities programme has a policy on extensions and late submission of assessed work, as follows:

- ➔ All assessed work must normally be submitted either on or before the due date.
- ➔ Extensions will likely be given if you are ill or cannot meet your deadline due to unforeseen circumstances. You must request an extension before the due date.
- ➔ For minor extensions (up to 3 days), please complete an extension application on Learn. If you think you may need to apply for a longer extension, please email the course coordinator directly.
- ➔ Your extension may not always be reflected in the due date visible on Learn, so if you have been given an extension please do not be concerned by this.
- ➔ Late work may not receive the same range or depth of feedback given to work submitted on time.
- ➔ Late work is generally accepted up to seven days after the due date, but will incur a part-grade penalty (i.e. from a B+ to a B) per day of lateness.

Disputes Procedures

Though rare, disputes between students and staff do occur. If a student feels they have valid cause for complaint or grievance on an academic matter, we advise the following step-by-step procedure. First, attempt to resolve it by discussion with the course coordinator. You may wish to make initial contact through and be supported by the class rep. If the previous step is not possible or is unsatisfactory, raise it with the Head of School of the School of Humanities. If the previous step is not possible or is unsatisfactory, raise it with the Associate Dean (Academic) in the Faculty of Arts.

Useful Links

Computer and technical support

Using Learn – student help
https://learn.canterbury.ac.nz/course/view.php?id=7726

Forum and email etiquette and best practices
https://web.archive.org/web/20160528134320/https://forums.adobe.com/thread/414764
https://web.archive.org/web/20140818123155/https://www.olympus.net/help-center/email-support/general-email/email-best-practices/email-best-practices-etiquette/

Information and Communication Technology Services
https://www.canterbury.ac.nz/its/

Academic support

Academic Skills Centre
https://www.canterbury.ac.nz/support/asc/

Citations and Referencing (UC Library)
https://www.canterbury.ac.nz/library/support/citations-and-referencing/

Key Dates and Information for Course Changes
https://www.canterbury.ac.nz/study/keydates/
https://www.canterbury.ac.nz/enrol/change/

Special Consideration
https://www.canterbury.ac.nz/study/special-consideration/

Academic Progress Reviews
https://www.canterbury.ac.nz/support/academic/progress-reviews/

Scholarships for students
https://www.canterbury.ac.nz/study/getting-started/scholarships

Personal support

Student Support
https://www.canterbury.ac.nz/support/

Student Counselling Services (Health Centre)
https://www.canterbury.ac.nz/healthcentre/counselling/

Student Life, Services and Communications
https://www.canterbury.ac.nz/life

Te Ratonga Whaikaha | Student Accessibility Service
https://www.canterbury.ac.nz/accessibility/

Māori Students
https://www.canterbury.ac.nz/support/akonga-maori/

Pasifika Students
https://www.canterbury.ac.nz/support/pasifika/

International Students
https://www.canterbury.ac.nz/international/

Appendix: Guidelines/Suggestions for Active Learning

Note: This Appendix reproduces portions of a course outline (https://courses.ucsd.edu/syllabi/FA16/877915.pdf) by Distinguished Professor Yến Lê Espiritu at University of California San Diego. We're using it with her permission! There are minor changes to replace US terminology with local equivalents and to tailor it to this course.

This course is designed to promote active participation, critical thinking, and intellectual advancement of *all* students. Toward this goal, I expect you to be active learners and to take responsibility to teach yourself and your classmates. This requires that you unlearn ways of learning that are passive, competitive, and uncollaborative. The American philosopher of education, John Dewey, believed that collaborative or cooperative learning, and the potential tensions that may arise in this interaction, are critical to the qualitative growth of individuals. In the same way, the Brazilian philosopher and educator Paulo Freire equates education with a people's triumph over the attempted "domestication" of their intellect.

Suggestions for Active Learning:

1) Attendance (should weigh very heavily).
2) Participate in class discussion on a regular basis.
3) Write and accumulate study notes on the readings.
4) Volunteer (in advance) to take responsibility (as an individual or as a group) to initiate discussion in class (discussion questions must be prepared in advance in consultation with the lecturer).
5) Come to class with written questions on scheduled readings / activities.
6) Write a brief review of an Academic Skills Centre workshop or resource and how it might apply to this course, and share it with the class (in consultation with the lecturer).
7) Meet with a study group on a regular basis (once a week, once every two weeks; each student needs to document meeting).
8) Visit your lecturer and/or tutors during office hours to discuss issues raised in class and in the readings; bring one or two classmates with you.
9) Contribute possible exam questions.
10) Watch a documentary film or read a text that explores how race, gender or sexuality is connected to topics studied in this course, then share your reaction and/or critical insights with your classmates and in a written commentary to the lecturer (one paragraph).
11) Offer novel learning resources or ideas to the class.
12) Offer reflections on your own personal experiences as they relate to topics studied in the course.
13) Notify the class of events on- or off-campus that address the issues discussed in class.
14) Attend a relevant community event and share your reaction with your classmates and in a written commentary to the lecturer (one paragraph).
15) Volunteer at a local organization; document your experience.
16) Any other creative way you can actively participate in class.

Some Criteria to Use to Evaluate Your Own Active Learning:

- Attendance record
- Degree to which I made conscious connections between different theories and explanations covered in class
- Degree to which I made myself available to classmates who needed help
- Degree to which I sought help from classmate or teaching staff when I needed help.
- The percentage of time I came to class having done the reading
- The amount of time I invested in reading, studying, and reflecting on the material
- The degree to which I engaged my classmates or teaching staff in thoughtful questions and comments (including in office hours)
- Degree to which I consciously and genuinely employed the participation methods above to advance my own and my classmates' learning and intellectual growth.
- Degree to which my motivation for following these suggestions was grounded in my desire to learn and to grow intellectually, to facilitate my classmates' learning experience, rather than in a desire to impress others.