

PhD Course – Introduction to Cultural Analytics

The PhD course “Introduction to Cultural Analytics” is given by the Centre for Digital Humanities Uppsala (CDHU) in the fall term 2022. It starts in week 36, runs on part time (50%), and ends in week 47. The main instructors on the course will be Karl Berglund, Marie Dubremetz, and Ekta Vats. Formally responsible (kursansvarig) for the course are Anna Foka, and Isto Huvila, Department of ALM.

The course is targeted towards PhD students in the humanities and social sciences who do not yet possess specific computational or technical skills, but who are interested to learn more for their future thesis work. Apart from some of the guest lectures, the course will be physical and take place at campus Engelska Parken, Uppsala University.

The course is module based, which means that you are free to choose how many modules you would like to take. Each module equals 1.5 ECTS, or 2 weeks at 50 %. Modules 1–2 are mandatory, while modules 3–6 are optional. To make the course equal 7.5 ECTS you thus take module 1–2 plus an additional three modules. For details on each module, please see the plan below.

Registration

If you are interested in taking the course, please check with your supervisors that it works and then send an email to karl.berglund@abm.uu.se, including your PhD topic, a short note of approval by your supervisors, and a sentence about why you are interested in the course.

The course will be offered on a first come, first served basis. PhDs at Uppsala University will have priority, thereafter PhDs at other universities in Sweden. There will be a maximum of 10 participants. When the course is completed you will receive a certificate of participation.

Main course goals

The aim of the course is to introduce methods for computational text analysis from a humanities and social sciences perspective. This covers and pays equal attention to providing:

- knowledge about computational text analysis methods and their relevance for humanities and social sciences tasks
- knowledge about machine learning and its basic concepts, as well as such methods' possibilities and limitations
- practical skills to employ computational methods for text analysis by using existing software, and by following and adapting basic programming scripts
- the ability to critically reflect upon the results derived from computational methods (regarding ethical, statistical, and empirical/material-oriented concerns)

- the ability to highlight epistemological concerns regarding computational and statistical methods from a humanities and social sciences perspective

Pedagogical idea

The general idea of the course is to combine practical hands-on tasks on technical (and statistical) methods with critical discussion and reflection on methodological concerns, limitations, and biases. Each module will typically consist of an introductory lecture, a discussion text seminar, and a practical lab. In some modules there will also be guest lectures by a leading expert. The hope is that the components feed into each other in the way that theory and reflection become concretised by practical work, and that practical work becomes meaningful and more critically aware through reflexive seminars, etc.

Examination

The course is examined: 1) through active participation in labs, seminars, and lectures; 2) through completion of lab assignments for each module; and 3) through a written final essay.

Practicalities

The workload for each module will be attending 1–2 lectures, some reading, attending a lab session and completing a lab report. In addition to the modules, a final essay is to be written.

The course will use a number of programs and modules—some of which can be hard or time-consuming to install for the beginner. To avoid all such trouble, during the course we will utilize a server solution, provided by CDHU, where everything the student needs (programs, libraries, dependencies, example corpora, tutorials) will be pre-installed.

Course program

Module 1: Introduction to Cultural Analytics and Python

- Wednesday 7 September 10.15–12: “Introduction to Cultural Analytics”, lecture by Karl Berglund, CDHU
- Thursday 8 September 10.15–11: “ Title TBA”, guest lecture by Benjamin Martin, Department of History of Science and Ideas, Uppsala University
- Tuesday 13 September 9.15–12: “Python Crash Course lab”, led by Karl Berglund and Marie Dubremetz
- Reading:
Melanie Walsh (2021), “Python Basics”, *Introduction to Cultural Analytics and Python*, <https://melaniewalsh.github.io/Intro-Cultural-Analytics/welcome.html>
Richard Jean So, (2017), “All Models Are Wrong”, *PMLA*, 132(3), 668–673, <https://doi.org/10.1632/pmla.2017.132.3.668>

Module 2: Data curation and analysis (with Pandas)

- Tuesday 20 September 10.15–12: “Working and thinking with tabular data”, lecture by Ekta Vats and Karl Berglund
- Thursday 22 September 9.15–12: “Data curation and analysis lab”, led by Ekta Vats and Karl Berglund. Main programs: Pandas, NumPy, Matplotlib.
- Reading:
Melanie Walsh (2021), “Data Analysis (Pandas)”, *Introduction to Cultural Analytics and Python*, <https://melaniewalsh.github.io/Intro-Cultural-Analytics/welcome.html>

Module 3: Data collection (web scraping, APIs, social media)

- Tuesday 27 September 10.15–12: “Gathering data through web scraping and APIs”, lecture by Marie Dubremetz
- Wednesday 28 September 10.15–11: “Title TBA”, online guest lecture by Simon Lindgren, professor in Sociology, DIGSUM, Umeå University
- Thursday 29 September 9.15–12: “Data collection and web scraping lab”, led by Marie Dubremetz. Main programs: Webscraper.io, BeautifulSoup, Pandas, NumPy.
- Reading:
Melanie Walsh (2021), “Data Collection (Web Scraping, APIs, Social Media)”, *Introduction to Cultural Analytics and Python*, <https://melaniewalsh.github.io/Intro-Cultural-Analytics/welcome.html>
Simon Lindgren, (2018), “The Concept of ‘Data’ in Digital Research”, *The Sage Handbook of Qualitative Data Collection*, pp. 441–450), <https://dx.doi.org/10.4135/9781526416070>

Module 4: Natural language processing (NLP)

- Tuesday 11 October 10.15–12: “Basics of NLP, core concepts and applied methods”, lecture by Marie Dubremetz
- Wednesday 12 October 13.15–14: “Title TBA”, guest lecture by Beáta Megyesi
- Thursday 13 October 9.15–12: “Applied NLP lab”, led by Marie Dubremetz. Main programs: spaCy (+ solution for Swedish text).
- Reading:
Melanie Walsh (2021), “Part-of-Speech Tagging”, *Introduction to Cultural Analytics and Python*, <https://melaniewalsh.github.io/Intro-Cultural-Analytics/welcome.html>
Nina Tahmasebi & Simon Hengchen (2019), “The Strengths and Pitfalls of Large-Scale Text Mining for Literary Studies”, *Samlaren* 140, pp. 198–227, <http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-406938>

Module 5: Computational text analysis

- Tuesday 25 October 10.15–12: “Computational text analysis”, lecture by Karl Berglund

- Thursday 27 October 9.15–12: “Computational text analysis lab”, led by Marie Dubremetz and Karl Berglund. Main programs: spaCy, Scikit-Learn, pandas, Matplotlib.
- Reading:
 - Melanie Walsh (2021), “Text Analysis”, *Introduction to Cultural Analytics and Python*, <https://melaniewalsh.github.io/Intro-Cultural-Analytics/welcome.html>
 - Jo Guldi, 2018, “Critical Search: A Procedure for Guided Reading in Large-Scale Textual Corpora”, *Journal of Cultural Analytics* 3 (1), <https://doi.org/10.22148/16.030>
 - Katherine Bode (2020), “Why You Can’t Model Away Bias”, *Modern Language Quarterly* 81 (1), 95–124, <https://doi.org/10.1215/00267929-7933102>

Module 6: Applied machine learning (ML)

- Tuesday 8 November 10.15–12: “Basics of ML, core concepts, applied methods”, lecture by Ekta Vats
- Wednesday 9 November 10.15–11: “Topic Modelling for the humanities and social sciences from a statistical perspective”, guest lecture by Måns Magnusson, Department of Statistics, Uppsala University
- Thursday 10 November 9.15–12: “Applied machine learning lab”, led by Ekta Vats
Main programs: Scikit-Learn, pandas, MALLET/Little MALLET wrapper, Matplotlib.
- Reading:
 - Melanie Walsh (2021), “Topic Modeling”, *Introduction to Cultural Analytics and Python*, <https://melaniewalsh.github.io/Intro-Cultural-Analytics/welcome.html>
 - Ted Underwood (2020), “Machine Learning and Human Perspective”, *PMLA*, 135(1), 92–109, <https://doi:10.1632/pmla.2020.135.1.92>

Final essay

A written case study that relates to the students’ PhD projects, where some cultural analytics method(s) introduced in the course are used to pose and answer a humanities or social sciences research question. The requested length of the essay will reflect the number of modules taken.