

# Kursplan

för forskarkurs

**Digitala metoder för humaniora: text**  
**Digital Methods for the Humanities: Text**

**5.0 Högskolepoäng**  
**5.0 ECTS credits**

**Kurskod:** AR1DH02  
**Gäller från:** VT 2021  
**Institution** Institutionen för arkeologi och antikens kultur

## Förkunskapskrav och andra villkor för tillträde till programmet

Behörig är den som är antagen till utbildning på forskarnivå vid humanistisk eller annan fakultet vid Stockholms universitet, eller motsvarande vid annat lärosäte.

## Lärandemål

För godkänt resultat på kursen förväntas studenten kunna:

- beskriva och tillämpa metoder för insamling och rensning av textdata, vare sig den ursprungligen är i digitaliserad eller icke-digitaliserad form
- beskriva och tillämpa metoder för att konstruera, annotera, utvinna information ur och analysera textkorpusar
- kritiskt granska en datamängd eller korpus och utvärdera vilka forskningsfrågor det kan användas för

## Innehåll

Kursens mål är att introducera grundläggande färdigheter och kritiska problem i automatisk analys av text. Studenterna lär sig viktiga teoretiska begrepp och metoder och får praktisk erfarenhet av verktyg och tekniker, bland annat följande: Kvantitativa egenskaper hos språket. Frekvenser. Textbearbetning i Unix kommandoradsmiljö. Reguljära uttryck. Förekomster och samförekomster. Annotering med hjälp av manuella metoder och maskininlärningsmetoder. Övningar med verktyg för annotering, sökning, statistik och lingvistisk analys. Etik, integritet, upphovsrätt.

## Obligatoriska moment

Aktivt deltagande i seminarier och genomförande av samtliga skriftliga uppgifter.

## Examinationsformer

Kursen examineras genom en lärdagbok (learning journal) med högst 4000 ord, bestående av genomförda uppgifter och korta skriftliga reflektioner relaterade till studentens eget område.

## Arbetsform

Kursen genomförs i form av seminarier som blandar teoriinriktade föreläsningar och praktiska datorlaborationer. Alla kurstillfällen äger rum i ett datorrum.

Varje kurstillfälle (med undantag för de två sista) har ett visst tema, och använder ett eller flera digitala verktyg i enlighet med kursinnehållet. Inga tekniska färdigheter krävs från början. Studenter uppmuntras att arbeta tillsammans och vissa uppdrag kräver samarbete.

Kursen ges på engelska.

# Syllabus

for research course

**Digital Methods for the Humanities: Text**  
**Digitala metoder för humaniora: text**

**5.0 Higher Education**  
**Credits**  
**5.0 ECTS credits**

**Course code:** AR1DH02  
**Valid from:** Spring 2021  
**Department:** Department of Archaeology and Classical Studies

## Prerequisites and special admittance requirements

### Learning outcomes

In order to pass the course, students are expected to be able to:

- describe and apply methods for collection and cleaning of textual data, whether originally in digitized or non-digitized form
- describe and apply methods for constructing, annotating, mining and analysing text corpora
- critically review a dataset or corpus and assess the kinds of research questions it could be used to address.

### Content

The aim of the course is to introduce fundamental skills and critical issues in computational analytics of text. Students will learn key theoretical concepts and methods, and gain hands-on experience with tools and techniques, including the following: Quantitative properties of language. Frequencies. Text processing in the Unix command-line environment. Regular expressions. Occurrences and co-occurrences. Annotation using manual and machine learning methods. Practice with tools for annotation, search, statistics and linguistic analysis. Ethics, integrity, copyright.

### Mandatory exams

Active participation in seminars and completion of all written assignments.

### Forms of examination

The course will be examined through a learning journal of at most 4000 words, consisting of completed assignments and short written reflections relating to the student's own field.

### Form of instruction

The course is conducted in the form of seminars that mix theory-oriented lecturing and practical laboratory assignments using computers. All sessions take place in a computer room. Each session is devoted to a particular theme and (with exception of the last two) using one or more digital tools in accordance with the course contents. No technical skills are required at the outset. Students are encouraged to cooperate and some assignments require collaboration.

The course is given in English.