

Introduction to Digital Text Analysis (A704153)

Course size *(nominal values; actual values may depend on programme)*

Credits 3.0

Study time 90 h

Course offerings and teaching methods in academic year 2023-2024

A (semester 2)

Dutch

Gent

lecture

independent work

Lecturers in academic year 2023-2024

Lefever, Els

LW22

lecturer-in-charge

Hoste, Veronique

LW22

co-lecturer

Offered in the following programmes in 2023-2024

	crdts	offering
Bachelor of Arts in Oriental Languages and Cultures(main subject Arabic and Islamic Studies)	3	A
Bachelor of Arts in Oriental Languages and Cultures(main subject China (China Track))	3	A
Bachelor of Arts in Oriental Languages and Cultures(main subject China (UGent Track))	3	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, English, French)	3	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, English, German)	3	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, English, Italian)	3	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, English, Russian)	3	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, English, Spanish)	3	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, English, Turkish)	3	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, French, German)	3	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, French, Italian)	3	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, French, Russian)	3	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, French, Spanish)	3	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, French, Turkish)	3	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, German, Italian)	3	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, German, Russian)	3	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, German, Spanish)	3	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, German, Turkish)	3	A
Bachelor of Arts in Oriental Languages and Cultures(main subject India)	3	A
Bachelor of Arts in Oriental Languages and Cultures(main subject Japan)	3	A
Bachelor of Arts in African Languages and Cultures	3	A
Bachelor of Arts in Moral Sciences	3	A
Bachelor of Arts in Philosophy	3	A
Micro-credential Introduction to Language Technology	3	A

(main subject Dutch, French, Spanish)		A
Preparatory Course Master of Arts in Interpreting: a combination of at least two languages, Master of Arts in Translation: a combination of at least two languages and Master of Arts in Multilingual Communication: a combination of at least two languages(main subject Dutch, French, Turkish)	3	A
Preparatory Course Master of Arts in Interpreting: a combination of at least two languages, Master of Arts in Translation: a combination of at least two languages and Master of Arts in Multilingual Communication: a combination of at least two languages(main subject Dutch, German, Italian)	3	A
Preparatory Course Master of Arts in Interpreting: a combination of at least two languages, Master of Arts in Translation: a combination of at least two languages and Master of Arts in Multilingual Communication: a combination of at least two languages(main subject Dutch, German, Russian)	3	A
Preparatory Course Master of Arts in Interpreting: a combination of at least two languages, Master of Arts in Translation: a combination of at least two languages and Master of Arts in Multilingual Communication: a combination of at least two languages(main subject Dutch, German, Spanish)	3	A
Preparatory Course Master of Arts in Interpreting: a combination of at least two languages, Master of Arts in Translation: a combination of at least two languages and Master of Arts in Multilingual Communication: a combination of at least two languages(main subject Dutch, German, Turkish)	3	A

Teaching languages

Dutch

Keywords

digital text analysis, digital humanities, digital literacy

Position of the course

The aim of this course is to:

- make the student aware of the importance of digital literacy, the competence that enables them to create, enrich, process and analyze digital resources using the computer.
- provide the student with a general overview of digital techniques and research methods within the research fields of the humanities, with a particular focus on digital text analysis (E.g., computational linguistics and sociolinguistics, natural language processing, corpus linguistics, network analysis).
- let the student reflect critically on the specific challenges of data-based research, with a specific focus on bias introduced by the selection of text corpora for research and building tools.
- introducing the student to concrete research projects and techniques within the digital humanities, again with a focus on digital text analysis (E.g., case studies within language and translation technology, history, literary studies, art history, etc.).

Contents

This course provides an overview of the most important digital techniques within the different research fields of the humanities, with a focus on digital text analysis.

The first part introduces the more general research domain of digital humanities, in which digital text analysis takes up a central place. A theoretical overview is provided of the main digital techniques and methods for automatic text analysis, starting with the collection, digitization and enrichment of resources, and the automatic search in large corpora or text collections. The introduction is followed by an overview of methods for natural language understanding, ranging from the more shallow lexical level to more complex semantic levels. We also elaborate on the problems inherent to processing natural language, such as ambiguity and world knowledge.

In addition to this theoretical foundation, much attention is also paid to critical reflection on the selection and use of resources for data-based research, the building and application of technology, the impact of machine learning systems and AI applications on society and the importance of a good understanding of the underlying algorithms for correctly applying and interpreting the output of these tools.

In the second part of the course, various use cases are discussed from the broader research domain of the digital humanities, again with a focus on digital text analysis. This way, an overview of state-of-the-art research and current trends is provided for some specific research domains (E.g., automatic text analysis for classical languages, the use of computational methods for the exploration and analysis of large text collections or contemporary digital media)

and one or more representative research projects are presented.
During the guided self-study, students read scientific articles to prepare specific course components, complete modules of a digital learning path and make exercises.

Initial competences

Average general knowledge that can be expected from someone who has completed secondary education.

Final competences

- 1 Having a basic knowledge and insight into the different digital techniques and research methods within the humanities, and more in particular methods for digital text analysis. [B.1.7.; assessed], [LWDC.6.1.; assessed]
- 2 Developing thinking and reasoning skills and the capacity for critical reflection on building and applying tools for digital text analysis [B.3.3.; assessed], [B.5.1.; assessed]
- 3 Research skills: scientific literature. [B.2.3.; assessed]
- 4 Reflect on the scientific and societal role of Digital Humanities research within the socio-cultural context in which it is conducted. [B.5.1.; assessed], [LWDC.6.2.; assessed]

Conditions for credit contract

Access to this course unit via a credit contract is unrestricted: the student takes into consideration the conditions mentioned in 'Starting Competences'

Conditions for exam contract

Access to this course unit via an exam contract is unrestricted

Teaching methods

Lecture, Independent work

Extra information on the teaching methods

Theoretical part: lectures, discussion of scientific research

Practical part: guided self-study by means of a digital learning path

Learning materials and price

Electronic learning platform: Course material (slides, scientific articles for self-study, digital learning path)

References

Rens Bod, *De Vergeten wetenschappen*, Prometheus, 2010.

Daniel Jurafsky & James Martin, "Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition", 2008. Updated Version December 2020.

Course content-related study coaching

Discussion forum

Assessment moments

end-of-term and continuous assessment

Examination methods in case of periodic assessment during the first examination period

Written assessment

Examination methods in case of periodic assessment during the second examination period

Written assessment

Examination methods in case of permanent assessment

Participation

Possibilities of retake in case of permanent assessment

examination during the second examination period is possible

Extra information on the examination methods

- Permanent evaluation: The students are assessed on the completion of digital learning paths and small assignments
- Periodic evaluation: written exam

Calculation of the examination mark

- Permanent evaluation: 10%
- Periodic evaluation: 90%

Facilities for Working Students

Possible exemption from educational activities requiring student attendance.

The examination cannot be rescheduled.

Possibility of feedback by e-mail or during consultation hour, restricted to answering specific questions.