![University of St.Gallen]

Course and Examination Fact Sheet: Spring Semester 2024

7,375 | 8,380: Quantitative Text Analysis

ECTS credits: 4

Overview examination/s
(binding regulations see below)
decentral - Portfolio, --, -- (100%)
Examination time: Term time

Attached courses
Timetable -- Language -- Lecturer
8,380,1.00 Quantitative Text Analysis -- English -- Gava Roy

Course information

Course prerequisites

Previous exposure to data analysis and a first experience with R (or any other programming language) is recommended.

Learning objectives

- Research design with "Text-as-Data"
- Collection and management of textual data
- Text wrangling and pre-processing
- Text classification (with manual coding, dictionaries, supervised/unsupervised machine learning and Large Language Models)

Course content

This course focuses on extracting content and producing inferences from texts by means of quantitative analysis. The course takes a practical approach focusing on applications and examples to introduce students to working with text as data. The goal is to provide participants with the necessary skills to critically assess and apply different methods and techniques in their own research projects. The first part of the course covers the collection and processing of political, legal and social media texts, as well as research design considerations. The second part of the course introduces various methods to quantitatively extract meaning from texts, ranging from traditional content analysis "by hand" to machine learning. Recent applications of quantitative text analysis in the social sciences will be critically discussed throughout the course.

All the presentations, examples and exercises will be based on the programming language R.

The course will cover the following topics:

- General introduction to quantitative text analysis in the social sciences
- Introduction to R for quantitative text analysis
- Acquiring text from APIs and social media
- Manual content analysis
- Text wrangling and pre-processing
- Text classification with supervised and unsupervised machine learning

Course structure and indications of the learning and teaching design

The course will be conducted in weekly two-hour sessions. The classes are structured as lectures with live-coding sessions.

University of St.Gallen

Course literature

A complete syllabus will be provided at the beginning of the semester. All readings will be made available on Canvas. *Indicative* of the literature and topics covered in this course are:

Grimmer, J., Roberts M., & Stewart, B. (2022). *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton; Oxford: Princeton University Press.

Hvitfeldt, E., & Silge, J. (2021). *Supervised machine learning for text analysis in R*. Boca Raton: CRC Press.

Silge, J., & Robinson, D. (2017). *Text Mining with R: A Tidy Approach*. Sebastopol: O'Reilly.

Welbers, K., Van Atteveldt, W., & Benoit, K. (2017). Text Analysis in R. *Communication Methods and Measures*, *11*(4), 245-265.

Wilkerson, J., & Casas, A. (2017). Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges. *Annual Review of Political Science*, *20*(1), 529-544.

Additional course information

--

Examination information

Examination sub part/s

1. Examination sub part (1/1)

Examination modalities

Examination type	Portfolio
Responsible for organisation	decentral
Examination form	Mixed form
Examination mode	--
Time of examination	Term time
Examination execution	--
Examination location	--
Grading type	--
Weighting	100%
Duration	--

Examination languages
Question language: English
Answer language: English

Remark
--

Examination-aid rule
no regulation necessary

- For written examinations at home (term paper), courses without credits, etc., no specific rules for examination aids are required.

- The regulations of the University of St. Gallen and the rules of academic work (sources and aids must always be identified) are applicable in a subsidiary fashion.

- All written work must be accompanied by a declaration of authorship.

Supplementary aids
--

Examination content

- Portfolio: Text-as-data research paper (100%)

Students (in small groups) address an original empirical research question with text-as-data by means of the different methods and techniques covered in the course. The portfolio consists of six components delivered throghout the semester:

- (1) Oral presentation outlining the proposed research design and potential data sources (10%, in groups, individual grade);
- (2) Oral presentation on data description and pre-processing (10%, in groups, individual grade);
- (3) Final written report with empirical analysis and corresponding computer code (60%, group grade);
- (4) Oral presentation of final report (10%, in groups, individual grade);
- (5) Written feedback for peers (10%, individual grade);
- (6) Reflection report on the learning journey and teamwork experience (not graded, individual).

Examination relevant literature

The relevant literature will be provided on Canvas at the start of the course.

Please note

Please note that only this fact sheet and the examination schedule published at the time of bidding are binding and takes precedence over other information, such as information on StudyNet (Canvas), on lecturers' websites and information in lectures etc.

Any references and links to third-party content within the fact sheet are only of a supplementary, informative nature and lie outside the area of responsibility of the University of St.Gallen.

Documents and materials are only relevant for central examinations if they are available by the end of the lecture period (CW21) at the latest. In the case of centrally organised mid-term examinations, the documents and materials up to CW 13 are relevant for testing.

Binding nature of the fact sheets:

- Course information as well as examination date (organised centrally/decentrally) and form of examination: from bidding start in CW 04 (Thursday, 25 January 2024);
- Examination information (supplementary aids, examination contents, examination literature) for decentralised examinations: in CW 12 (Monday, 18 March 2024);
- Examination information (supplementary aids, examination contents, examination literature) for centrally organised mid-term examinations: in CW 13 (Monday, 25 March 2024);
- Examination information (regulations on aids, examination contents, examination literature) for centrally organised examinations: Starting with de-registration period in CW 15 (Monday, 08 April 2024).