



Course and Examination Fact Sheet: Autumn Semester 2018

7,375: Quantitative Text Analysis

ECTS credits: 4

Overview examination/s

(binding regulations see below)

Decentral - Group examination paper (all given the same grades) (80%)

Decentral - Presentation (in groups - all given the same grades) (20%)

Attached courses

Timetable -- Language -- Lecturer

[7,375,1.00 Quantitative Text Analysis](#) -- Englisch -- [Gava Roy](#)

Course information

Course prerequisites

Some exposure to quantitative data analysis and statistical software is recommended. Students should be comfortable with the idea of typing commands to interact with the computer. Previous experience with R or any other programming language is helpful but not required.

Course content

This course focuses on extracting content and producing inferences from texts by means of quantitative analysis. The course takes a practical approach in order to provide students with the necessary skills to critically assess and apply different methods and techniques in their own research projects. The course covers the collection and processing of political, legal, newspaper 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 the web, APIs and social media
- Manual content analysis
- Text wrangling and pre-processing
- Text classification with supervised and unsupervised machine learning

Course structure

Weekly two-hour sessions.

Course literature

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

Grimmer, J. (2015). We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together. *PS: Political Science & Politics*, 48(01), 80-83.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267-297.



Lucas, C., Nielsen, R. A., Roberts, M. E., Stewart, B. M., Storer, A., & Tingley, D. (2015). Computer-assisted text analysis for comparative politics. *Political Analysis*, 23(2), 254-277.

Munzert, S., Rubba, C., Meissner, P., & Nyhuis, D. (2015). *Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining*. Chichester: Wiley.

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 (1/2)

Examination time and form

Decentral - Group examination paper (all given the same grades) (80%)

Remark

--

Examination-aid rule

Term papers

- Term papers must be written without anyone else's help and in accordance with the known quotation standards, and they must contain a declaration of authorship.
- The documentation of sources (quotations, bibliography) has to be done throughout and consistently in accordance with the APA or MLA standards. The indications of the sources of information taken over verbatim or in paraphrase (quotations) must be integrated into the text in accordance with the precepts of the applicable quotation standard, while informative and bibliographical notes must be added as footnotes (recommendations and standards can be found, for example, in METZGER, C. (2017), *Lern- und Arbeitsstrategien* (12th ed., Cornelsen Schweiz).
- For any work written at the HSG, the indication of the page numbers both according to the MLA and the APA standard is never optional.
- Where there are no page numbers in sources, precise references must be provided in a different way: titles of chapters or sections, section numbers, acts, scenes, verses, etc.
- For papers in law, the legal standard is recommended (by way of example, cf. FORSTMOSER, P., OGOREK R. et SCHINDLER B. (2018), *Juristisches Arbeiten: Eine Anleitung für Studierende* (6. Auflage), Zürich: Schulthess, or the recommendations of the Law School).

Supplementary aids

--

Examination languages

Question language: English

Answer language: English

2. Examination sub part (2/2)

Examination time and form

Decentral - Presentation (in groups - all given the same grades) (20%)



Remark

--

Examination-aid rule

Practical examination

No examination-aid rule is necessary for such examination types. The rules and regulations of the University of St. Gallen apply in a subsidiary fashion.

Supplementary aids

--

Examination languages

Question language: English

Answer language: English

Examination content

Students are asked to address an empirical research question in small groups by means of the different methods and techniques covered in the course. The examination consists of two parts. First, students (in small groups) are expected to present their work in class and receive feedback to further develop their research designs and empirical analyses. Second, students (in small groups) are expected to hand in a paper with their empirical analyses and the corresponding computer code.

Examination relevant literature

The relevant literature will be provided on StudyNet (as of 25 September 2018).

Please note

We would like to point out to you that this fact sheet has absolute priority over other information such as StudyNet, faculty members' personal databases, information provided in lectures, etc. When will the fact sheets become binding?

- Information about courses and examination time (central/decentral and grading form): from the start of the bidding process on 23 August 2018
- Information about decentral examinations (examination-aid rule, examination content, examination relevant literature): after the 4th semester week on 15 October 2018
- Information about central examinations (examination-aid rule, examination content, examination relevant literature): from the start of the enrolment period for the examinations on 05 November 2018

Please look at the fact sheet once more after these deadlines have expired.