



Course and Examination Fact Sheet: Spring Semester 2022

7,375 | 8,380: Quantitative Text Analysis

ECTS credits: 4

Overview examination/s

(binding regulations see below)

Decentral - Group examination paper with presentation (all given the same grades) (50%)

Examination time: term time

Decentral - examination paper written at home (individual) (20%)

Examination time: term time

Decentral - examination paper written at home (individual) (20%)

Examination time: term time

Decentral - examination paper written at home (individual) (10%)

Examination time: term time

Attached courses

Timetable -- Language -- Lecturer

[8,380,1.00 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 recommended but not required.

Learning objectives

- Research design with "Text as Data"
- Collection of textual data
- Text wrangling and pre-processing
- Text classification (with manual coding, dictionaries and machine learning)

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 first part of 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 APIs and social media
- Manual content analysis
- Text wrangling and pre-processing
- Text classification with supervised and unsupervised machine learning



- Text similarity and reuse

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.

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. (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

In the case of the President's Board having to implement new directives due to the SARS-CoV-2 pandemic in SpS2022, the course information listed above will be changed as follows:

- The course is conducted online via the platform Zoom.

The examination information listed below would be changed as follows:

- There are no changes necessary to the examination information.

Examination information

Examination sub part/s

1. Examination sub part (1/4)

Examination time and form

Decentral - Group examination paper with presentation (all given the same grades) (50%)

Examination time: term time

Remark

--

Examination-aid rule

Term papers

Written work must be written without outside help according to the known citation standards, and a declaration of authorship must be attached, which is available as a template on the StudentWeb.



Documentation (quotations, bibliography, etc.) must be carried out universally and consistently according to the requirements of the chosen/specified citation standard such as e.g. APA or MLA.

The legal standard is recommended for legal work (cf. by way of example: FORSTMOSER, P., OGOREK R., SCHINDLER B., Juristisches Arbeiten: Eine Anleitung für Studierende (the latest edition in each case), or according to the recommendations of the Law School).

The reference sources of information (paraphrases, quotations, etc.) that has been taken over literally or in the sense of the original text must be integrated into the text in accordance with the requirements of the citation standard used. Informative and bibliographical notes must be included as footnotes (recommendations and standards e.g. in METZGER, C., Lern- und Arbeitsstrategien (latest edition)).

For all written work at the University of St.Gallen, the indication of page numbers is mandatory, regardless of the standard chosen. Where page numbers are missing in sources, the precise designation must be made differently: chapter or section title, section number, article, etc.

Supplementary aids

--

Examination languages

Question language: English

Answer language: English

2. Examination sub part (2/4)

Examination time and form

Decentral - examination paper written at home (individual) (20%)

Examination time: term time

Remark

--

Examination-aid rule

Term papers

Written work must be written without outside help according to the known citation standards, and a declaration of authorship must be attached, which is available as a template on the StudentWeb.

Documentation (quotations, bibliography, etc.) must be carried out universally and consistently according to the requirements of the chosen/specified citation standard such as e.g. APA or MLA.

The legal standard is recommended for legal work (cf. by way of example: FORSTMOSER, P., OGOREK R., SCHINDLER B., Juristisches Arbeiten: Eine Anleitung für Studierende (the latest edition in each case), or according to the recommendations of the Law School).

The reference sources of information (paraphrases, quotations, etc.) that has been taken over literally or in the sense of the original text must be integrated into the text in accordance with the requirements of the citation standard used. Informative and bibliographical notes must be included as footnotes (recommendations and standards e.g. in METZGER, C., Lern- und Arbeitsstrategien (latest edition)).

For all written work at the University of St.Gallen, the indication of page numbers is mandatory, regardless of the standard chosen. Where page numbers are missing in sources, the precise designation must be made differently: chapter or section title, section number, article, etc.

Supplementary aids

--

Examination languages

Question language: English

Answer language: English



3. Examination sub part (3/4)

Examination time and form

Decentral - examination paper written at home (individual) (20%)

Examination time: term time

Remark

--

Examination-aid rule

Term papers

Written work must be written without outside help according to the known citation standards, and a declaration of authorship must be attached, which is available as a template on the StudentWeb.

Documentation (quotations, bibliography, etc.) must be carried out universally and consistently according to the requirements of the chosen/specified citation standard such as e.g. APA or MLA.

The legal standard is recommended for legal work (cf. by way of example: FORSTMOSER, P., OGOREK R., SCHINDLER B., Juristisches Arbeiten: Eine Anleitung für Studierende (the latest edition in each case), or according to the recommendations of the Law School).

The reference sources of information (paraphrases, quotations, etc.) that has been taken over literally or in the sense of the original text must be integrated into the text in accordance with the requirements of the citation standard used. Informative and bibliographical notes must be included as footnotes (recommendations and standards e.g. in METZGER, C., Lern- und Arbeitsstrategien (latest edition)).

For all written work at the University of St.Gallen, the indication of page numbers is mandatory, regardless of the standard chosen. Where page numbers are missing in sources, the precise designation must be made differently: chapter or section title, section number, article, etc.

Supplementary aids

--

Examination languages

Question language: English

Answer language: English

4. Examination sub part (4/4)

Examination time and form

Decentral - examination paper written at home (individual) (10%)

Examination time: term time

Remark

--

Examination-aid rule

Peer review

Written work must be written without outside help according to the known citation standards, and a declaration of authorship must be attached, which is available as a template on the StudentWeb.

Documentation (quotations, bibliography, etc.) must be carried out universally and consistently according to the requirements of the chosen/specified citation standard such as e.g. APA or MLA.

The legal standard is recommended for legal work (cf. by way of example: FORSTMOSER, P., OGOREK R., SCHINDLER B., Juristisches Arbeiten: Eine Anleitung für Studierende (the latest edition in each case), or according to the recommendations of the Law School).

The reference sources of information (paraphrases, quotations, etc.) that has been taken over literally or in the sense of the



original text must be integrated into the text in accordance with the requirements of the citation standard used. Informative and bibliographical notes must be included as footnotes (recommendations and standards e.g. in METZGER, C., Lern- und Arbeitsstrategien (latest edition)).

For all written work at the University of St.Gallen, the indication of page numbers is mandatory, regardless of the standard chosen. Where page numbers are missing in sources, the precise designation must be made differently: chapter or section title, section number, article, etc.

Supplementary aids

--

Examination languages

Question language: English

Answer language: English

Examination content

- Term paper and presentation in groups (50%)

Students 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.

- Home assignments (2 x 20%)

Students solve (individually) two practical home assignments covering text wrangling and analysis with R and submit the corresponding computer code.

- Feedback to peers (10%)

Students review other students' work and provide (individually) written feedback to their colleagues.

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 12 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, 27 January 2022);
- Examination information (regulations on aids, examination contents, examination literature) for decentralised examinations: in CW 12 (Monday, 21 March 2022);
- Examination information (regulations on aids, examination contents, examination literature) for centrally organised mid-term examinations: in CW 12 (Monday, 21 March 2022);
- Examination information (regulations on aids, examination contents, examination literature) for centrally organised examinations: two weeks before the end of the registration period in CW 15 (Monday, 11 April 2022).



University of St.Gallen