Text as Data Methods in R – Applications for Automated Analyses of News Content

Information about instructor:
Valerie Hase, M.Sc., M.A.
v.hase@ikmz.uzh.ch, Raum AND-3-26

Information about seminar:
Part of the Master “Internet & Society” (Research Areas in Internet & Society) takes place via in-person weekly meetings (Tuesdays, 10:15-12:00, HS2021) in room BIN-1-D.25

Link to course in UZH course catalogue
Link to course in OLAT
Link to corresponding bookdown tutorial

Summary of seminar
Automated content analyses, i.e., (semi-)automated analysis of text, are becoming an increasingly important method in communication science. In this course, students will learn how to apply different methods of automated content analysis using R and conduct their own analyses within a small project. Please note that interested participants should have a basic interest in working with the programming language R as well as in learning new quantitative methods. However, previous knowledge of R is not required.

Learning goals
- Students will learn basic knowledge of the programing language R and will be able to conduct their own analyses by writing corresponding functions.
- Students will learn basic knowledge of the method “automated content analysis” and will be able to conduct their own automated content analysis by writing corresponding functions.
- Students will be able to apply this knowledge to independently conduct automated content analyses using data on news media coverage. This means that given predefined research questions/hypotheses, they develop a suitable operationalization via R, conduct descriptive analyses/statistical tests to answer these questions/test hypotheses, and critically interpret results.
- Students will be able to critically discuss current German- and English-language research on the topic of automated content analyses, summarize its strengths and weaknesses, and evaluate it.
- Students will be able to engage in factual and competent discussions. They are able to communicate their acquired knowledge alone and in groups in a comprehensible and structured manner, both orally and in writing. They are also able to work in groups, resolve potential conflicts, and organize themselves.
Workload & Assessment

This is a 6 ECTS seminar which equals 180 hours of work.

Please know that the time spent in in-person meetings in class is only a fraction of the actual workload needed to pass the class. The bulk of the required work needs to be done between sessions.

Workload

<table>
<thead>
<tr>
<th>Activity</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attendance in seminar sessions</td>
<td>26h</td>
</tr>
<tr>
<td>Readings</td>
<td>24h</td>
</tr>
<tr>
<td>Working through R tutorials &amp; doing related assignments</td>
<td>70h</td>
</tr>
<tr>
<td>Seminar paper</td>
<td>60h</td>
</tr>
</tbody>
</table>

Grading

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignments in R (group or individual work)</td>
<td>30%</td>
</tr>
<tr>
<td>Seminar paper (individual work)</td>
<td>70%</td>
</tr>
</tbody>
</table>
Assignments in R (group or individual work)

During the semester, you will receive a total of three exercises in R which you can work on either in groups or individually. You will upload your solutions via OLAT. In the script, please specify the name of all students who worked on the corresponding task. The grade for each task will make up 10% of your final grade, leading to all tasks contributing a total of 30% to your final grade.

Task in R (I)
Please submit until: 22.10.2021 (folder closes as midnight) via OLAT, folder: “Task in R (1)”.

Task in R (II)
Please submit until: 19.11.2021 (folder closes as midnight) via OLAT, folder: “Task in R (2)”.

Task in R (III)
Please submit until: 17.12.2021 (folder closes as midnight) via OLAT, folder: “Task in R (3)”.

Seminar paper (individual work)

Since this seminar is a bit more practice- and programming oriented, you will do a seminar paper in a slightly different form than your usual term paper. In short, the seminar paper will consist of two parts:

(1) A theoretical part where you need to answer a set of pre-defined questions on automated content analyses using relevant literature (for instance by describing a specific method and discussing its (dis-)advantages

(2) An empirical part where you need answer a set of pre-defined questions with a given data set and write-up your results using both literature in the field and R.

Seminar paper
Please submit until: 07.01.2022 (folder closes as midnight) via OLAT, folder: “Seminar paper”.
Overview Seminar “Text as Data Methods in R” (HS2021)

<table>
<thead>
<tr>
<th>Date</th>
<th>Phase</th>
<th>Session</th>
<th>Work by students</th>
<th>Input by instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>21.09.2021</td>
<td>Introduction + organizational information</td>
<td>Session 1: Introduction</td>
<td></td>
<td>Basic information on the seminar; expectations; structure and assessment</td>
</tr>
<tr>
<td>28.09.2021</td>
<td>Introduction to R</td>
<td>Session 2: Introduction to R (I)</td>
<td>Work through necessary tutorials for corresponding session before in-person meeting</td>
<td>Summary of tutorials &amp; moderating Q&amp;A</td>
</tr>
<tr>
<td>05.10.2021</td>
<td>Session 3: Introduction to R (II)</td>
<td></td>
<td>Work through necessary tutorials for corresponding session before in-person meeting</td>
<td>Summary of tutorials &amp; moderating Q&amp;A</td>
</tr>
<tr>
<td>12.10.2021</td>
<td>Session 4: Introduction to R (III)</td>
<td></td>
<td>Work through necessary tutorials for corresponding session before in-person meeting</td>
<td>Summary of tutorials &amp; moderating Q&amp;A</td>
</tr>
<tr>
<td>19.10.2021</td>
<td>Session 5: Introduction to R (IV)</td>
<td></td>
<td></td>
<td>Assignment deadline: Task in R (I) until 22.10.2021</td>
</tr>
<tr>
<td>Date</td>
<td>Activity</td>
<td>Description</td>
<td>Assignment deadline</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------------------</td>
<td></td>
</tr>
<tr>
<td>16.11.2021</td>
<td>Automated Content Analysis in R</td>
<td>No in-person meeting (session used for working through next tutorials)</td>
<td>Task in R (II) until 19.11.2021</td>
<td></td>
</tr>
<tr>
<td>23.11.2021</td>
<td>Session 9: Introduction to automated content analysis with R (I)</td>
<td>Work through necessary tutorials for corresponding session before in-person meeting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30.11.2021</td>
<td>Session 10: Introduction to automated content analysis with R (II)</td>
<td>Work through necessary tutorials for corresponding session before in-person meeting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>07.12.2021</td>
<td>Session 11: Introduction to automated content analysis with R (III)</td>
<td>Work through necessary tutorials for corresponding session before in-person meeting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.12.2021</td>
<td>Session 12: Introduction to automated content analysis with R (IV)</td>
<td>Work through necessary tutorials for corresponding session before in-person meeting</td>
<td>Task in R (III) until 17.12.2021</td>
<td></td>
</tr>
<tr>
<td>21.12.2021</td>
<td>Wrap-Up &amp; information on seminar paper</td>
<td>Assignment deadline: Seminar paper until 07.01.2022</td>
<td>Wrap-Up of seminar, further information on seminar paper</td>
<td></td>
</tr>
</tbody>
</table>
Necessary readings

**Session 6 (26.10.2021): Theoretical Introduction: Automated Content Analysis (I)**


**Session 7 (02.11.2021): Theoretical Introduction: Automated Content Analysis (II)**


**Session 8 (09.11.2021): Theoretical Introduction: Automated Content Analysis (III)**


Further readings


Other tutorials: Introduction to R

Grolemund G., & Wickham, H.: R for Data Science. https://r4ds.had.co.nz/index.html
Haim, M. wegweisR. https://wegweisr.haim.it/

Other tutorials: Automated Content Analysis in R

van Atteveldt, W. Text Analysis in R workshop at University of Vienna http://vanatteveldt.com/vienna-r-text-analysis
Bail, C. Day 3: Automated Text Analysis. https://compsocialscience.github.io/summer-institute/curriculum#day_3
Watanabe, K., & Müller, S. Quanteda Tutorials. https://tutorials.quanteda.io/
Wiedemann, Gregor; Nickler, Andreas (2017): Hands-on: a five day text mining course for humanists and social scientists in R. Proceedings of the 1st Workshop Teaching NLP for Digital Humanities (Teach4DH@GSCL 2017), Berlin. https://tm4ss.github.io/docs/