

# Text Technology (TT): SS 2020

Björn Rudzewitz and Johannes Dellert

Welcome to a new semester under very unusual circumstances! As every other course, we will likely have to organize and conduct the entire course remotely this year. We are just as new to this as everybody else, so please be patient in case things do not run as smoothly as we hope. We will reciprocate by handling attendance and deadlines more liberally than we otherwise would have. If you are suffering from circumstances that make things even more difficult than for most people (inability to travel, suboptimal internet connection, inadequate workplace, no childcare, illness), we would ask you to inform us as early as possible, and we will try to accommodate everyone in these difficult times.

Most importantly, the central hub for everything that happens in the course (including announcements about possible changes) is the **Moodle page**:

<https://moodle.zdv.uni-tuebingen.de/course/view.php?id=741>

**Room:** The course will be taught on Zoom (links will appear on Moodle) until further notice. We might return to Room 0.02 at some point, though.

**Time:** **Monday 16 c.t. - 18, Wednesday 16 c.t. - 18** (both for the Zoom sessions and live lectures in case things improve during the semester); the first session will take place on **April 22**.

## Lecturers:

Björn Rudzewitz ([bjorn.rudzewitz@uni-tuebingen.de](mailto:bjorn.rudzewitz@uni-tuebingen.de)),  
Johannes Dellert ([johannes.dellert@uni-tuebingen.de](mailto:johannes.dellert@uni-tuebingen.de))

**Office hours:** These should mostly be replaced by asking questions on Moodle forums, but individual appointments via Zoom are possible on request. It will generally also be possible to communicate directly via chat during the Zoom sessions with the one lecturer while the other is holding the lecture. This replaces the usual opportunity to drop by at the front desk shortly before or after class.

## Requirements for a 6 CP Proseminar :

- passing **homework assignments**: at least 60% of the points
- a **final exam**: This will probably require your physical presence in Tübingen at the end of the lecture period, but due to the very dynamic situation, we don't know anything yet about what the situation is going to be, and how this is going to be organized.
- **passing a small quiz** about each session (replaces attendance)
- attendance during Zoom sessions is recommended, but **not required**

**Grading :** The course grade depends entirely on the exam, but assignments will be graded as well in order to define the threshold for exam admission.

- exercises and quizzes count as a *Studienleistung*, they are a prerequisite for the exam
- the **final exam** is graded and counts as *Prüfungsleistung*

**Registration** : via the Moodle page (no password, just enroll)

**Access to Materials and Resources** :

Everything you need will be made available via Moodle. This will mostly be the lecture slides, but we are also going to point you to supplementary and/or alternative resources such as web tutorials and introductory videos produced by other people.

**Overview of Course Contents** :

<b>Block</b>	<b>Topics</b>
Introduction	definition of texts
Representing Text	binary and hexadecimal numbers binary representations historical encodings Unicode
Accessing Text	overview of legal issues general-purpose command line tools command line tools for handling text files regular expressions
Processing Text	data modeling and annotation basic principles of corpus linguistics using standard NLP tools on the command line
Structuring Text	TSV and TSV-based formats JSON XML XPath XML Schema
Presenting Text	HTML CSS Markdown LaTeX BibTeX
Wrap-up	recap session final exam