

**DSI TAP 1 – Workshop 1:  
Introduction to Text Analytics, Pre-processing Techniques, Collocation and Keyword Detection**

**Time:**

Friday, May 21, 2021 2pm – 5 pm (3 hours)

**Location:**

Zoom (Link will be shared upon registration)

*With your registration you agree to a planned recording of the Zoom Session.*

**Registration:**

Please send an e-mail to [tcc@cl.uzh.ch](mailto:tcc@cl.uzh.ch) to register and receive the link to the workshop.

Please provide the following information:

- Name
- Department
- Your computer's operating system (e.g. Mac/Windows/Linux)
- If you have any previous programming experience
- DSI membership status

*Please note that we might need to restrict the number of participants. We will give priority to DSI members and the temporal order of registrations that we receive.*

**Topics:**

- Introduction to Natural Language Processing and Text Analytics
- Pre-processing Techniques
- Collocation and Keyword Detection

**Software to be installed beforehand:**

- [R](#) with the following libraries: [readtext](#), [quantenda](#), [sentimentR](#)
- [Python](#) (>3.6) with the following libraries: [spaCy](#), [pandas](#)
- [AntConc](#)
- [LightSide](#)
- [Mallet](#)

**Note:**

- If you are not sure on how to install Python and R together with the respective libraries, best follow our **step-by-step guide** (shared with you upon registration).
- *If you have trouble installing any of the software, we will be available before the workshop at 1 pm for assistance*

**Detailed Schedule:**

13:00-14:00	Help with Installation if needed
14:00-15:00	Introduction, general overview, short introduction to programs and datasets used in the workshop
15:00-16:00	Introduction to R and to the libraries <i>readtext</i> and <i>quantenda</i> and <i>spacy</i>
16:00-17:00	Preprocessing and statistical analysis of a text corpus: Formatting, Encoding, Spelling Normalization, Concordancing, Collocations, Keywords

If you have no or very little experience in R, follow part I (“The Foundations”) of Gerold’s introduction to R and statistics before Workshop 1:

<https://dlf.uzh.ch/openbooks/statisticsforlinguists/>