

Scalable Language Technologies Lab

Summer Term 2026

Foundations

- ❑ Transformers [\[exbert\]](#) [\[temir.org\]](#) [\[bert from scratch\]](#)
- ❑ CLIP and generative models [\[webis.de\]](#)

Exercise

- ❑ Introductory exercise from [Session 1](#)
- ❑ Download the AOL Query Log from the [web archive](#)
- ❑ Load the AOL Query Log into a [pandas DataFrame](#) and experiment with it:
 - What queries did user 711391 issue?
 - How many queries are navigational?
 - Identify search sessions based on time (same user with at most 30 min between queries)
 - Identify query reformulations
 - Filter out search sessions that are purely navigational
 - Other stuff you can come up with :)
- ❑ Do not use neural models or embedding models yet
- ❑ Prepare a 5–10 min presentation with your findings