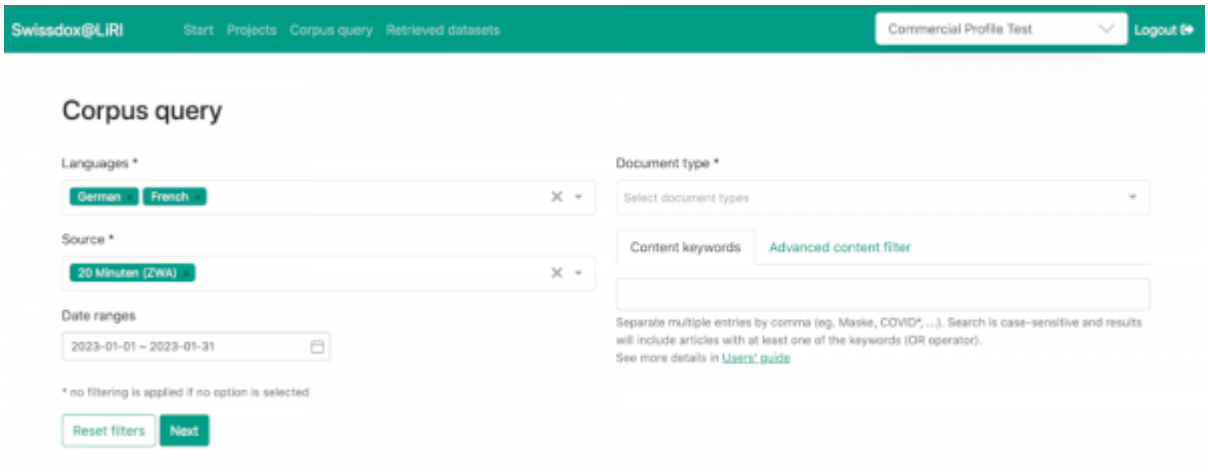
Corpora & Assistive Technology

The Language Technology group has expertise in handling various types of corpora. We are building tailor-made applications to explore large and structurally complex collections of language data. In particular, we are competent in:

- The design of databases to hold application-relevant data
- Generating interactive visualizations
- Efficiently querying large data collections (in particular corpora)
- Anonymisation of large data sets
- Data crawling/scraping and processing of web sources, batch download of documents
- Data extraction and conversion

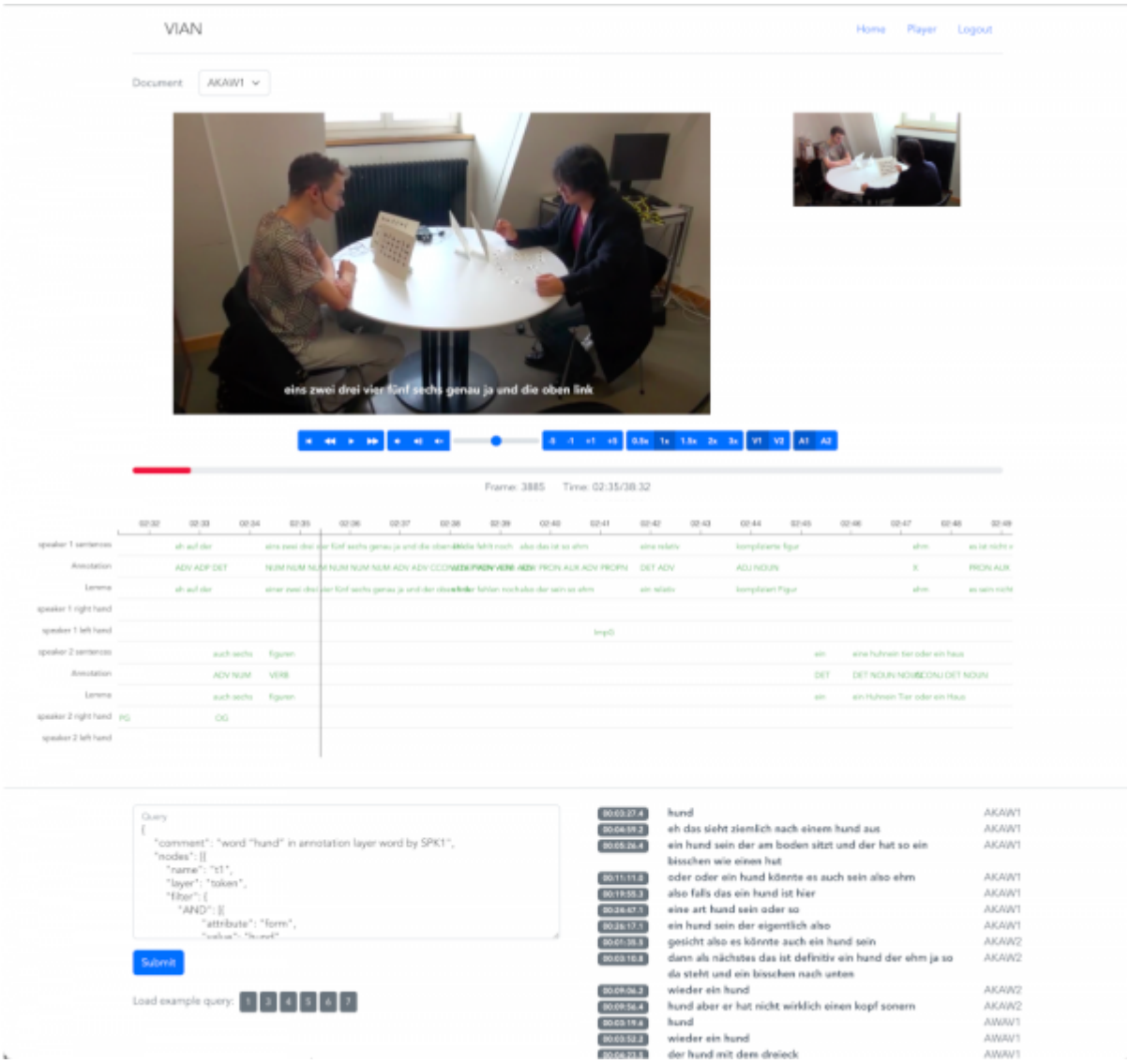
Examples of our work

Swissdox@LiRI – web based service for extraction of subcorpora from the Swiss media database Swissdox

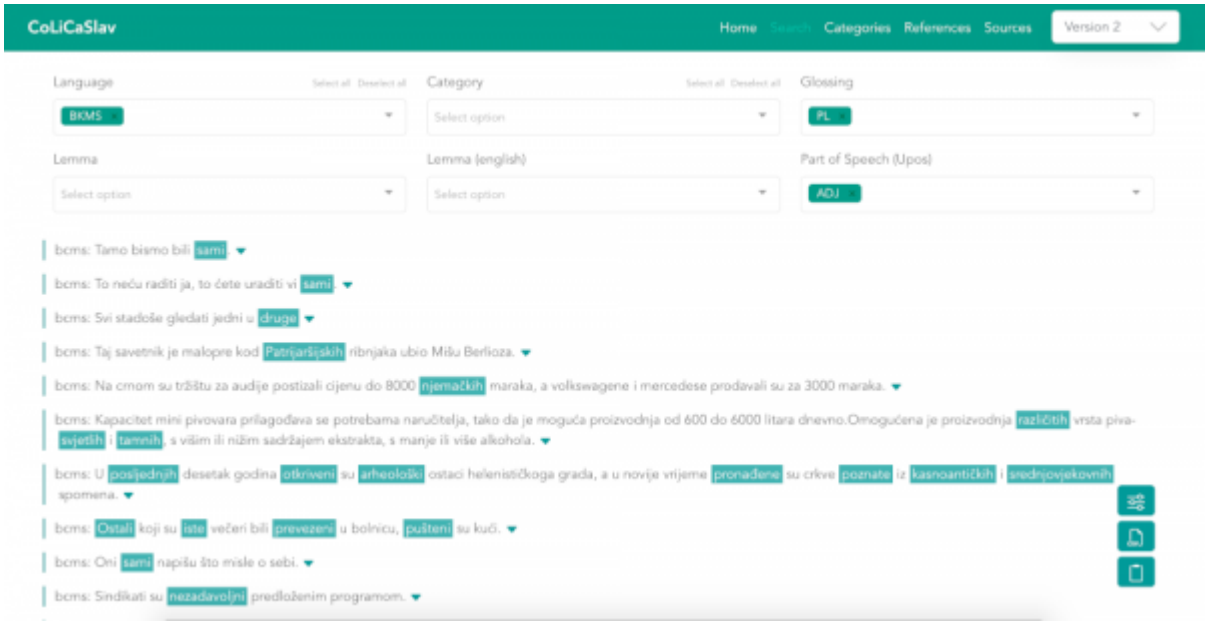


https://swissdox.linguistik.uzh.ch/

VIAN – web application for multimodal corpora; comprises of corpus querying interface, multimodal corpus viewer, video and audio player and timeline with time-aligned text and annotations



CoLiCaSlav – web corpus application used as an empirical basis for teaching and studying the principle categories and concepts relevant for the Slavic languages



https://lehrkorpus-slav.linguistik.uzh.ch/

Kollo – command line tool for extracting collocations from VERT formatted corpora