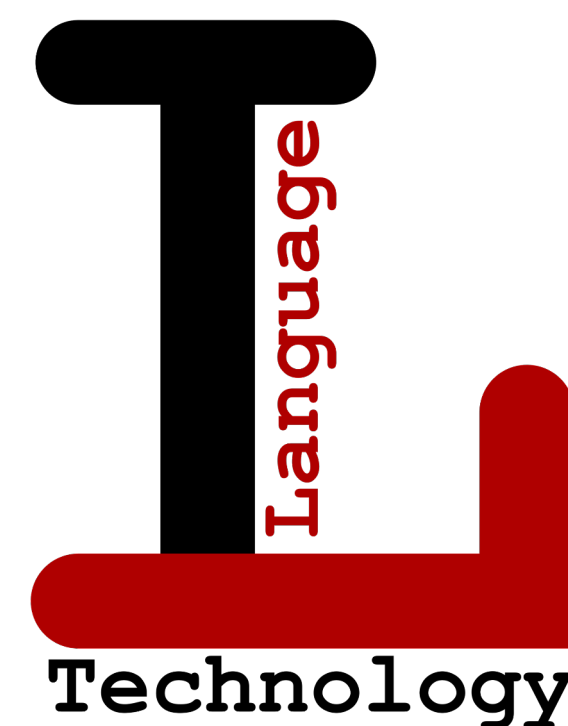**Universität Hamburg**
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR MATHEMATIK, INFORMATIK UND NATURWISSENSCHAFTEN

INSTITUT
FÜR EMPIRISCHE KULTURWISSENSCHAFT

Language Technology

The D-WISE Tool Suite:
Multi-Modal Machine-Learning-Powered Tools Supporting and Enhancing Digital Discourse Analysis

Florian Schneider[†], Tim Fischer[†], Fynn Petersen-Frey[†], Isabel Eiser[‡], Gertraud Koch[‡], Chris Biemann[†]
[†]Language Technology Group, Department of Informatics [‡]Anthropological Studies in Culture and History
{florian.schneider-1, tim.fischer, fynn.petersen-frey, isabel.eiser, gertraud.koch, chris.biemann}@uni-hamburg.de

ABSTRACT

- The D-WISE Tool Suite (DWTS) is a novel working environment for digital qualitative discourse analysis in the Digital Humanities.
- The DWTS addresses challenges introduced by large amounts of heterogeneous, unstructured, and multi-modal data in which current discourses are encoded.
- The DWTS provides meaningful insights by leveraging current NLP and CV technologies.
- The DWTS is an outcome of D-WISE, a collaborative and interdisciplinary research project between Humanities and Computer Science.

CENTRAL FEATURES

1) Support for multi-modal data including text, image, audio, and video documents.
2) ML-powered preprocessing pipelines for automatic annotations and semantic embeddings.
3) Lexical and multi-modal semantic search and filtering of large document collections.
4) Manual span, bounding box, time-span, and frame annotations.
5) Functionality to document and reflect the research process.

TECHNICAL DETAILS

Multi-Modal Embeddings
Sentence Transformers* (Text), CLIP-ViT* (Image), Whisper* (Audio), X-CLIP* (Video)

Efficient Retrieval and Storage
ElasticSearch*, FAISS*, PostgreSQL

Client-Server Application
React frontend, FastAPI backend

Preprocessing Pipeline
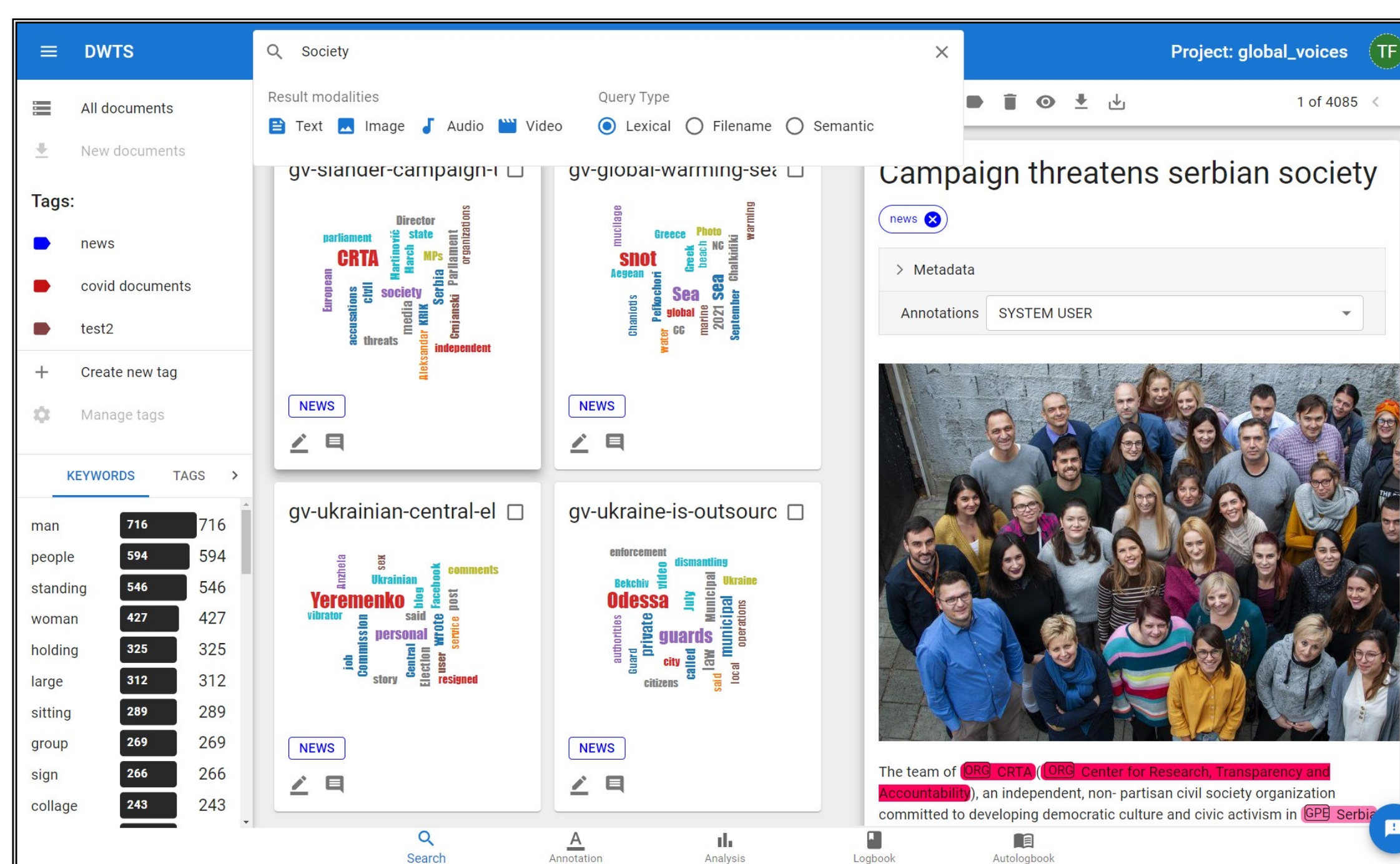Dockerized Celery workers

Automatic Information Enrichment
spaCy* (NER), DETR* (Object Detection), BLIP2* (Image Captioning), Whisper* (Transcriptions)
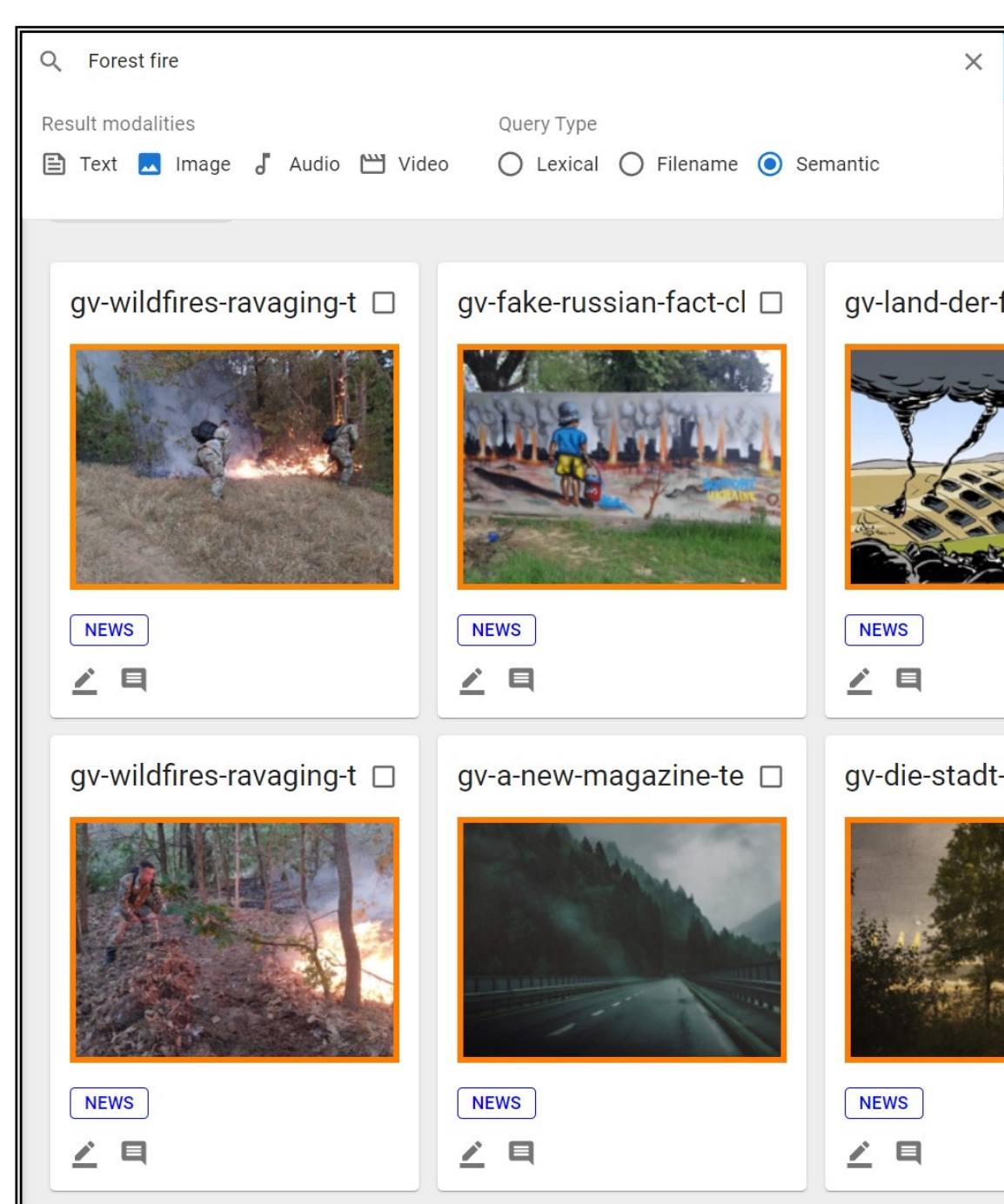
Document Import
Apache Tika (PDF, DOCX), Scrapy + Readability.JS (HTML, URL)
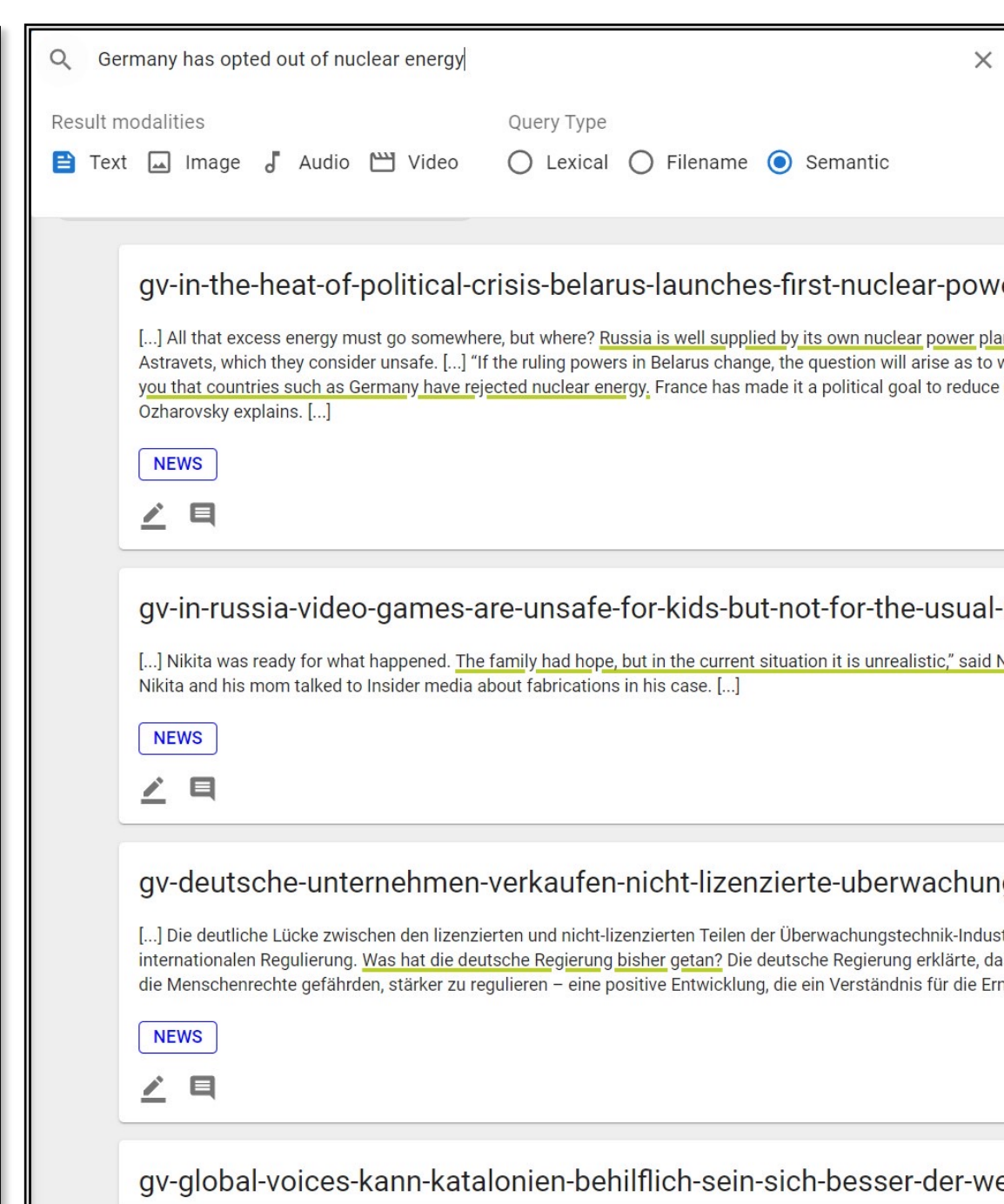
* Please find the respective reference in our paper

Search View

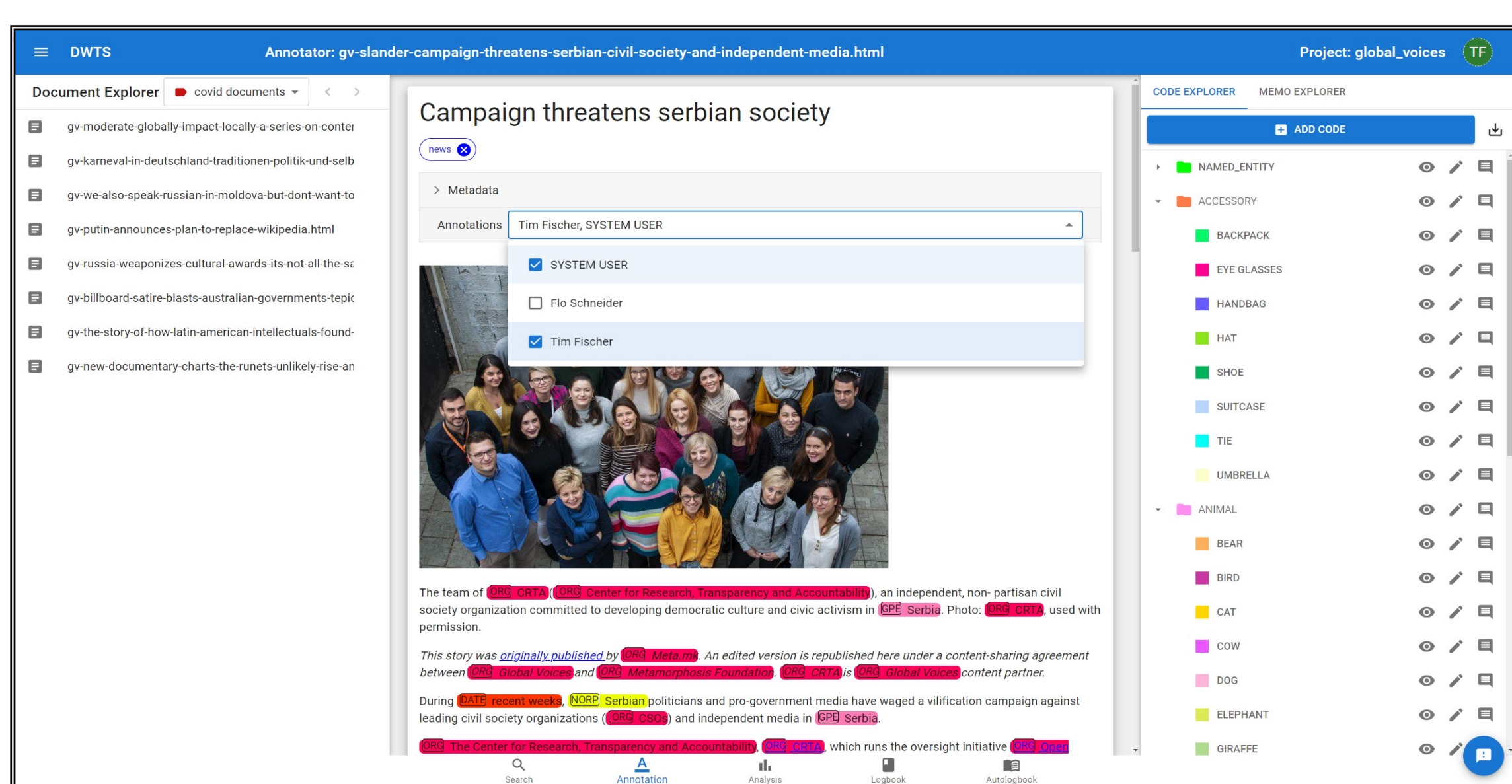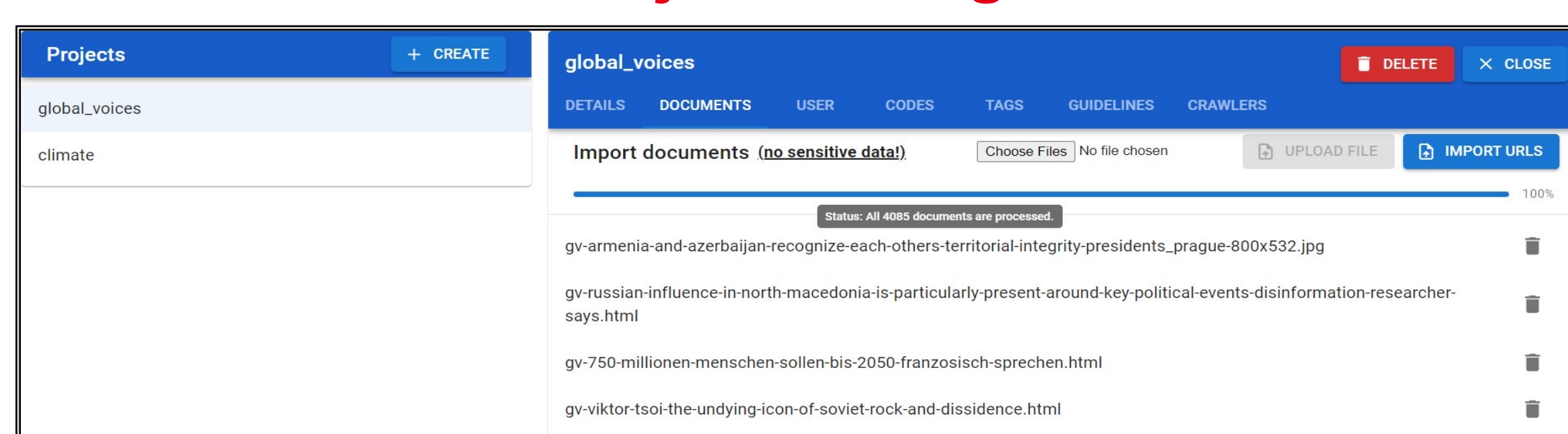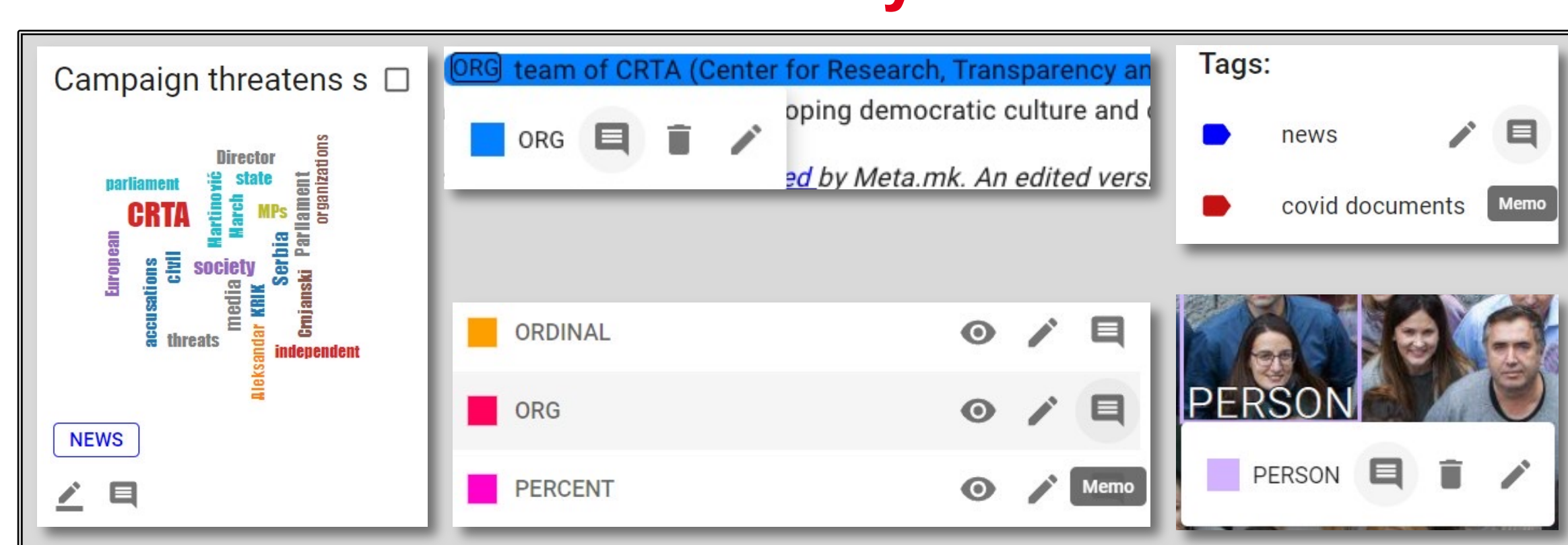### T2I Semantic Search
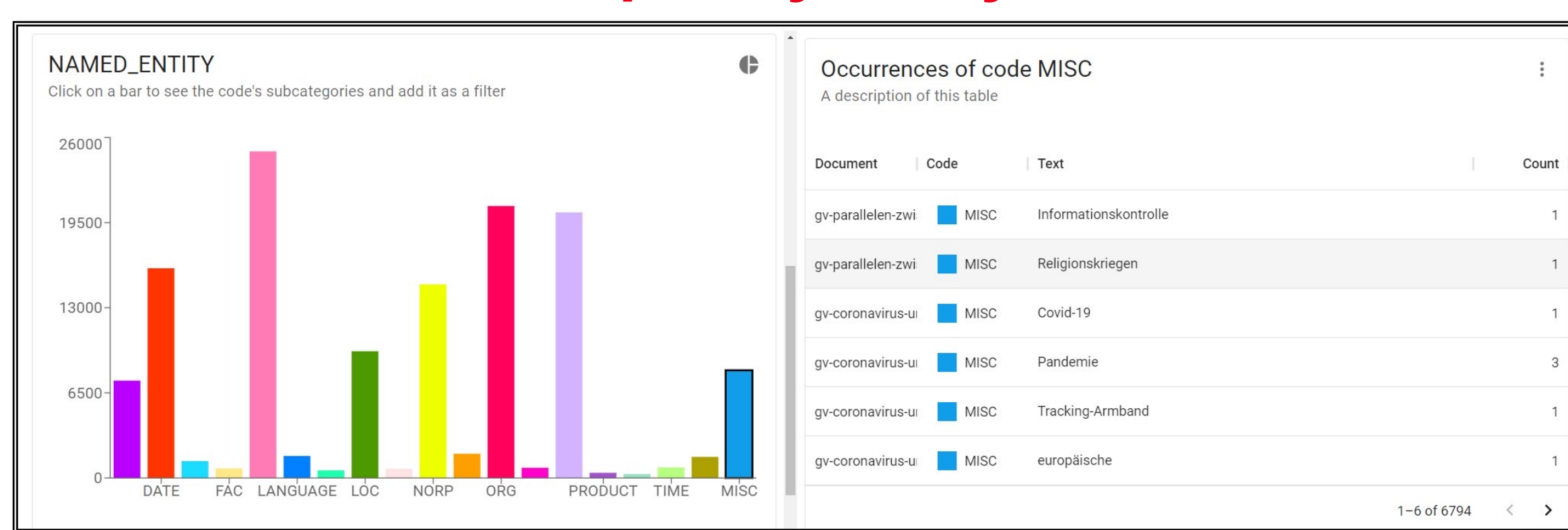

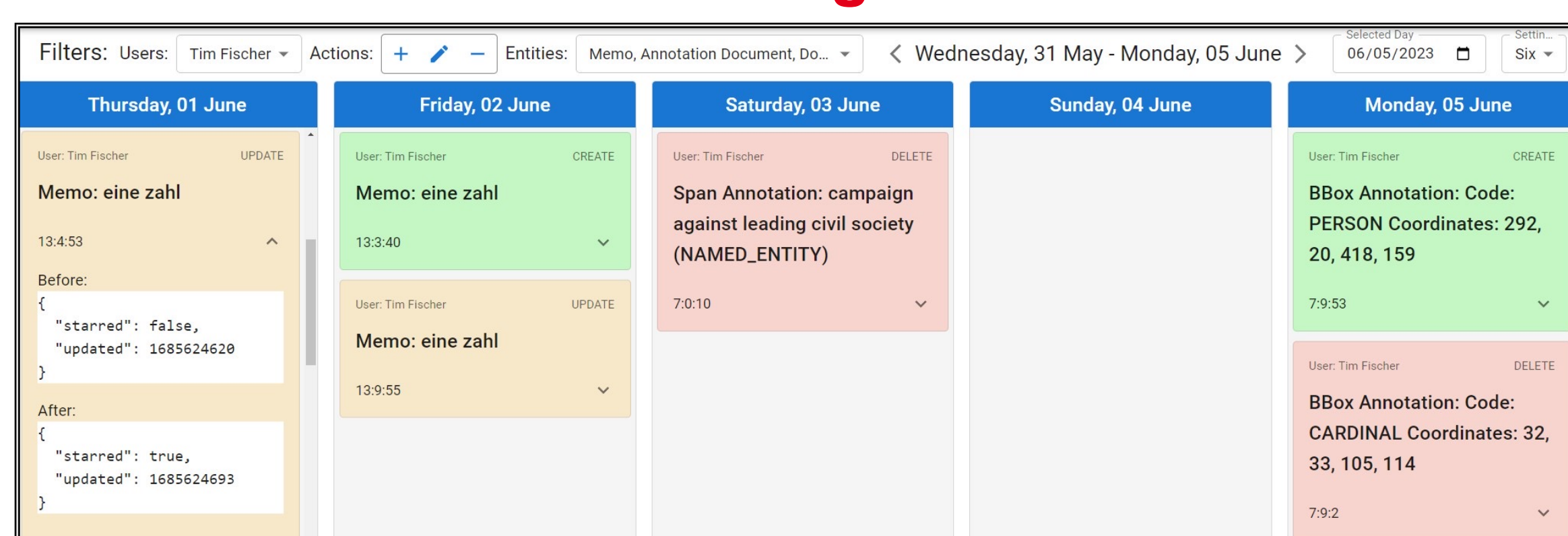### T2T Semantic Search

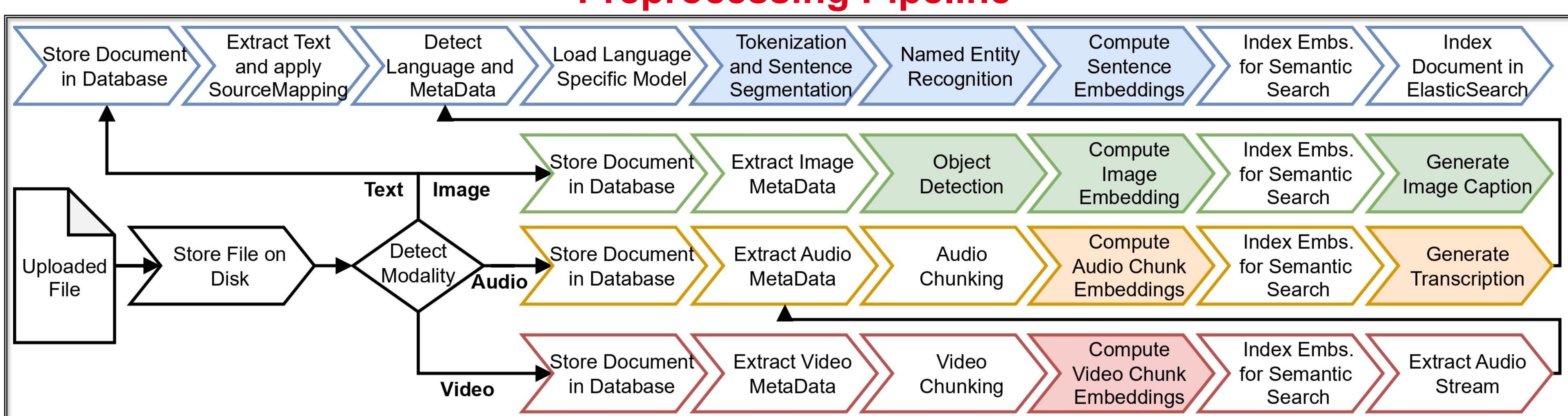### Annotation View

Project Settings

Memos Everywhere

Coding On-The-Fly

Code Creation

Frequency Analysis

Image Annotation

Export

Auto-Logbook

Preprocessing Pipeline

Try It Out!
shorturl.at /cgK39