Tools & Resources

- [Tools & Resources](#)

Our has been publishing a number of free/open source tools, components, frameworks, and resources for NLP and Semantic Computing.
toc_collapse=0; Table of Contents

1. Frameworks and Architectures

1.1. Semantic Assistants

The *Semantic Assistants* architecture provides for easy integration of NLP into (desktop) clients using W3C Web Services and ontologies. The server-side part integrates the GATE framework and allows to publish any existing pipeline as an NLP Web service and a plugin for the OpenOffice.org word processor provides for executing semantic services.

2. Corpora

2.1. Durm Corpus

A single book from a historic encyclopedia of architecture, written in German (in various formats).

3. Text Mining Systems

3.1. Open Mutation Miner (OMM)

The Open Mutation Miner (OMM) system provides a number of advanced text mining components for mutation mining from full-text research papers, including the detection of various forms of mutation mentions, protein properties, organism detection, impact mentions, and the relations between them.

4. NLP Components

4.1. LODeXporter

The LODeXporter is a GATE component that can export NLP annotations directly to a triplestore, with configurable vocabularies, for use in LOD applications.

4.2. Rhetector

Rhetector is a GATE plugin for the automatic detection of Rhetorical Entities (REs) in scientific literature. Rhetorical Entities are spans of text (sentences, passages, sections, etc.) in a document, where authors convey their findings, like Claims or Arguments, to the readers.

4.3. LODtagger

The LODtagger is a GATE component that allows linking entities in a document to their corresponding resources on the Linked Open Data (LOD) cloud. LODtagger relies on external tools to perform the actual content tagging and hides the complexity of communicating with LOD taggers, such as DBpedia Spotlight, from the perspective of pipeline developers.

4.4. OrganismTagger

The OrganismTagger is a hybrid rule-based/machine-learning system that extracts organism mentions from the biomedical literature, normalizes them to their scientific name, and provides grounding to the NCBI Taxonomy database.

4.5. OwlExporter

A GATE component for easy ontology population from text.

4.6. Predicate-Argument Extractor

A GATE component that extract predicate-argument structures (subject, predicate, object triples) in a common format from the output of different parsers (RASP, Minipar, Stanford, SUPPLE).

4.7. Durm German Lemmatizer

The *Durm* self-learning, context-aware lemmatizer for German nouns.

4.8. Reported Speech Tagger

Our *reported speech* tagger for English newspaper articles.

4.9. MuNPEx NP Chunker

The multi-lingual noun phrase (NP) chunker MuNPEx for GATE.

5. Other Tools and Resources

5.1. The Javadoc NLP Corpus Generation Doclet

A doclet for Javadoc that allow to generate a corpus from Java source code optimized for NLP processing of source code comments.

6. Support

For questions, comments, etc., please visit the Tools & Resources Forum.

Source URL (retrieved on *2025-12-13 06:05*): https://www.semanticsoftware.info/tools-and-resources