EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Philosophische Fakultät

Seminar für Sprachwissenschaft

Datenübernahme durch TALAR, das *Tübinger Archive of Language Resources*

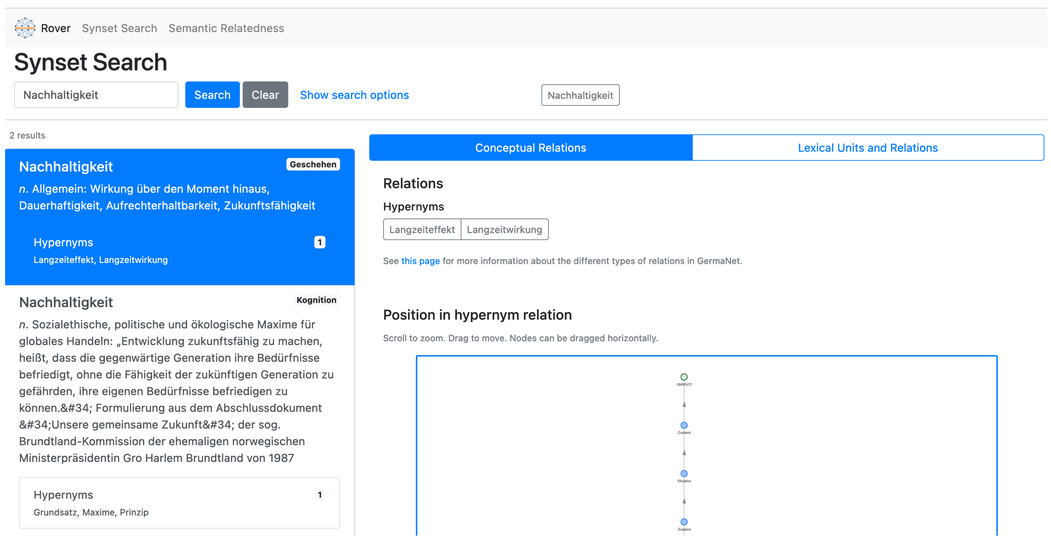
Vorhandene Daten in TALAR

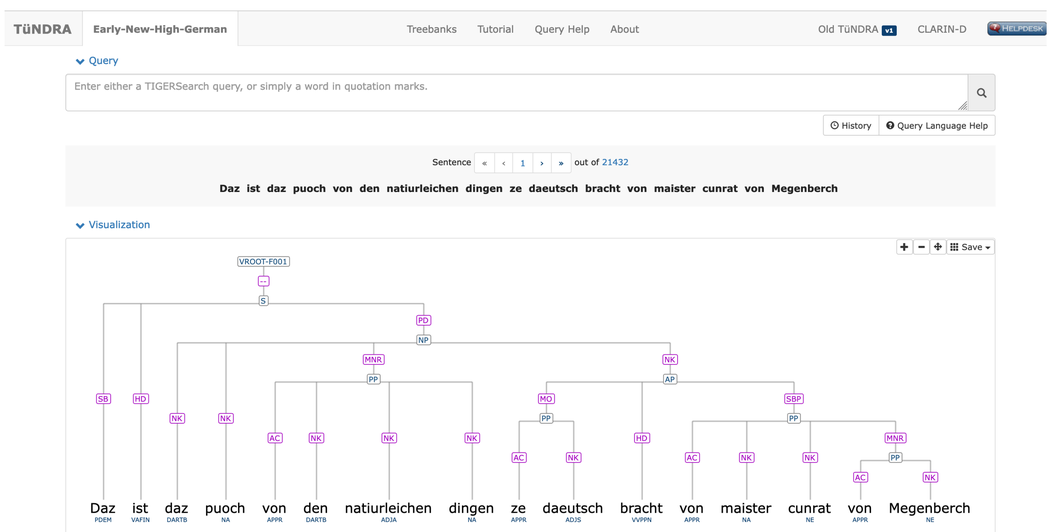
- Gepflegte Referenzdaten
  - Tübinger Baumbank-Collection
  - GermaNet
- Gehostete Daten
  - Daten des SFB 833/SFB 441
  - Daten des GRK 1808
- Daten externer Partner
  - SFB 638
  - Index Thomisticus Treebank
  - ...
- Community Daten
  - Universal Dependency Treebanks (UD)

Tübingen Archive of Language Resources (TALAR)

Spezialisierung	• Baumbanken/Treebanks • Wortnetze • Word-Embeddings
Modalität	• geschrieben • gesprochen
Akzeptierte Datenformate	• GermaNet XML-Format • CoNLL-U • Word-Embeddings • Weitere Formate nach Absprache
Ansprechperson(en)	• Thorsten Trippel • Claus Zinn

Suchmöglichkeiten in den gepflegten Daten





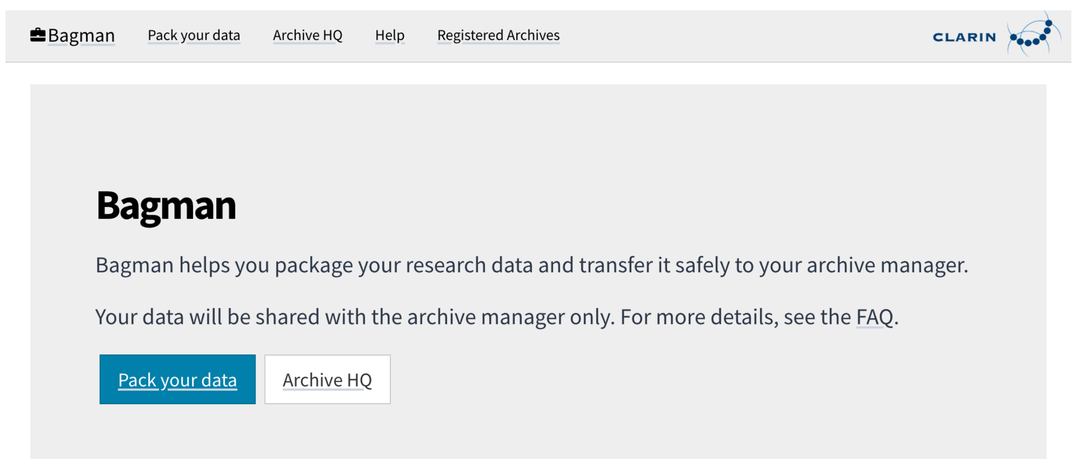
Daten Einbringen

Voraussetzung:

- Unterstützte Datentypen
  - Qualitätsgesichert
  - Auswahlprozess
- Datenüberlassungsvertrag
  - Muster siehe https://uni-tuebingen.de/de/134320
- Offene Lizenz: CC-BY 4.0 oder höher
- Metadaten nach ISO 24622
- Unterstützte Profile
  - *CourseProfile*
  - *ExperimentProfile*
  - *LexicalResourceProfile*
  - *ResourceBundle*
  - *SpeechCorpusProfile*
  - *TextCorpusProfile*
  - *ToolProfile*
  - *WebLichtWebService*

Initiale Datenübermittlung

- Übermittlung nach BagIT-Standard
- Werkzeugunterstützung: Bagman



F A I R

Findable

- Auffindbar mit persistentem Identifikator (PID)
  - Handle / DOI
- Nachweis über Forschungsdatensuchmaschinen
  - z. B. VLO
- Auffindbarkeit über Standardsuchmaschinen

Accessible

- Downloadmöglichkeit freier Daten
- Zugang zu vorhandenen zugangsbeschränkten Daten
- Metadaten frei zugänglich über technische Protokolle

Interoperable

- Je nach Datentyp
- Nutzbar in den vorhandenen Auswertungswerkzeugen
- Metadaten: standardkonform und maschinell interpretierbar

Reusable

- Nutzbar unter den Bedingungen der Lizenzen in eigener Software
- Verwendung in anderen Projektkontexten
- Klare Rechte an den Daten
- Vertrauenswürdige Aufbewahrung

Zertifizierung des Repositoriums

Zertifikat erteilt am 01. Juni 2023

CORE TRUST SEAL

Aktuell: turnusgemäße Erneuerung

CLARIN B-centre

Scott Martens (2013). **TüNDRA**: A Web Application for Treebank Search and Visualization. In: Proceedings of The Twelfth Workshop on Treebanks and Linguistic Theories (TLT12), Sofia, pp. 133—144.

Hinrichs, Marie and Lawrence, Richard and Hinrichs, Erhard (2020): Exploring and Visualizing Wordnet Data with GermaNet **Rover**. In Proceedings of the CLARIN Annual Conference 2020, pp. 32-36.

C. Zinn: **Bagman** - A Tool that Supports Researchers Archiving Their Data. Selected papers from the CLARIN Annual Conference 2021 (Virtual Event), Linköping University Electronic Press, vol. 189, pages 181-189, 2022.

Universität Tübingen · Philosophische Fakultät
Seminar für Sprachwissenschaft
Keplerstr. 2 · 72074 Tübingen
Telefon +49 7071 29-77352 · https://talar.sfb833.uni-tuebingen.de