

“Anwendung von LOD-Technologien im Linguistik-Portal”

Workshop zur Semantischen Extraktion von Informationen
2023

Thorsten Fritze

`t.fritze@ub.uni-frankfurt.de`

Fachinformationsdienst Linguistik

17. Januar 2023

Projektübersicht

Lin|gu|is|tik

PORTAL FÜR
SPRACHWISSENSCHAFT
www.linguistik.de



Universitätsbibliothek Johann Christian Senckenberg

- Heike Renner-Westermann
- Vanya Dimitrova
- Thorsten Fritze



Arbeitsgruppe Angewandte Computerlinguistik (2015-22)

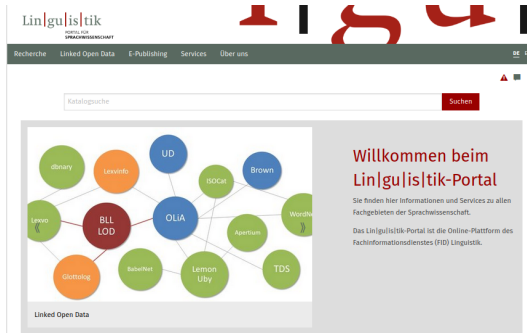
- Prof. Dr. Christian Chiarcos
- Frank Abromeit
- Christian Fäth

Die Rechte aller abgebildeter Logos liegen bei den jeweiligen Inhaber*innen. Eine unerlaubte Weiternutzung, Veränderung etc. ist nicht gestattet.

Zeittafel

- 1950–2015: Sondersammelgebiet 7.11
- 2012–2016: Virtuelle Fachbibliothek Linguistik (ViFa)
- Seit 2017: Fachinformationsdienst
- Seit 2015: Integration von LOD-Ressourcen
- Seit 2020: Umstellung des gesamten Linguistik-Portals auf LOD-Technologie

Linguistik-Portal



Alle Angebote des FID Linguistik werden im Linguistik-Portal gebündelt.

Angebote des FID Linguistik (Beispiele)

- Lizenzen für Sprachkorpora, Fachzeitschriften und Datenbanken
- Open-Access-Publizieren (OJS-Hosting, Linguistik-Repository)
- Katalogsuche
- Linkverzeichnisse
- Sammlung elektronischer Sprachressourcen

Merkmale des bisherigen Portals

- Katalogsuche: Traditionelle Bibliotheksressourcen (~62% Artikel)
- Linkverzeichnisse: Durch den FID kuratierte Webseiten
- Elektronische Sprachressourcen aus der LLOD-Cloud (Korpora, Elektronische Wörterbücher)
- Jeweils getrennte Datenbasis

LOD im FID Linguistik

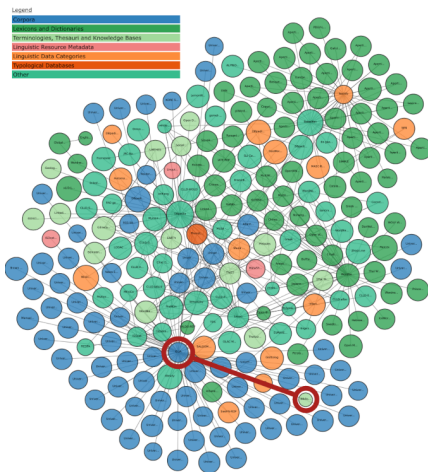
Drei Schwerpunkte im Bereich LOD

- 1 **Terminologie:** BLL THESAURUS & BLL ONTOLOGIE
- 2 **Datenmodell:** BIBFRAME im Linguistik-Portal
- 3 **Elektronische Sprachressourcen:** Annohub

Ausgangspunkt

Integration von Ressourcen aus der LLOD-Cloud in das Linguistik-Portal

Linguistic Linked Open Data Cloud



The Linguistic Linked Open Data Cloud from lod-cloud.net



Terminologie

BLL

- BLL: Bibliography of Linguistic Literature
- Linguistische Bibliografie, die an der UB JCS seit den frühen 1970ern erstellt wird
- 500.000+ bibliografische Nachweise
- Eigenständiger Schlagwortthesaurus mit 9.000+ Elementen in 5 Hauptkategorien (BLL THESAURUS)

Kernidee

Verknüpfung von BLL-Schlagwörtern mit anderen Vokabularen, die von Ressourcen in der LLOD-Cloud verwendet werden

Zielvokabularen:

- **OLiA** für morphosyntaktische Termini
- **lexvo** und **Glottolog** für Sprachbezeichner
- **Phoible** für phonologische Termini

BLL THESAURUS

Automatisch aus PICA-Klassifikationsnormdatensätzen
erzeugtes SKOS

```
:bll-133108015 a owl:Class , skos:Concept ;  
    rdfs:label "Kanauri"@en , "Kanauri"@de ;  
    rdfs:subClassOf :BLLConcept ;  
    skos:broader :bll-133122719 ;  
    skos:notation "02.13.22" ;  
    skos:prefLabel "Kanauri"@en , "Kanauri"@de .
```

BLL ONTOLOGIE

Intellektuell umgearbeitete Begriffe des BLL THESAURUS sowie weitere Nicht-BLL-Begriffe, die im Rahmen von Teilprojekten und Datenkonversionen benötigt werden

```
bllt:bll-133108015 a owl:Class ;  
  rdfs:label "Kanauri"@de, "Kanauri"@en ;  
  rdfs:subClassOf bllt:BLLConcept, bllt:WestHimalayish ;  
  skos:notation "02.13.22" ;  
  skos:prefLabel "Kanauri"@de, "Kanauri"@en .
```

Linkingmodell

Intellektuell erarbeitete Verknüpfungen zu anderen Vokabularen und Anbietern

```
:indiv_133108015 rdf:type bllo:bll-133108015 ,  
                    owl:NamedIndividual ;  
    lvont:iso639P3Code "kfk" ;  
    lvont:nearlySameAs glottolog:kinn1249 ;  
    owl:sameAs lexvoP3:kfk .
```


Datenmodell

Portal (bisher)

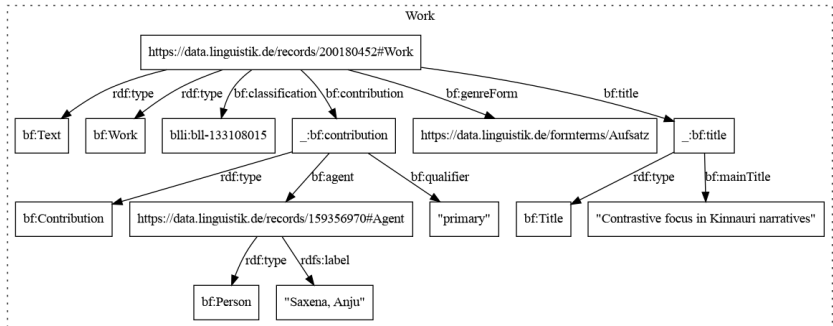
- Mit der bisherigen Technik konnten die Möglichkeiten der BLL ONTOLOGIE nicht ausgeschöpft werden
- Separate Datenpools für einige Teilbereiche des Portals
- Keine echten Normdatensätze; Beziehungen zwischen Ressourcen i.d.R. nur über reine Zeichenkettensuche

Ansatz: Neues Datenmodell

Einführung eines neues Datenmodells auf Basis von BIBFRAME
("Lingframe"-Anwendungsprofil)

- 1 Direktes Mapping von PICA+ nach BIBFRAME
- 2 Nachnutzung des LoC Marc21-nach-BIBFRAME-Mappings für Marc21-Ressourcen
- 3 Direktes Mapping des alten Datenformats nach BIBFRAME (übergangsweise)
- 4 Direkte Mappings für importierte Daten (perspektivisch)

Beispiel: BIBFRAME Work (1)



Anwendungsfälle

- Ontologische Suche über alle Portal-Ressourcen hinweg
- Einheitliche Suchbegriffe; Abbildung anderer Vokabularien auf die BLL ONTOLOGIE
- Geografische Suche auf Grundlage inferierter Informationen
- Fachspezifische Teilexporte von Metadaten für andere FID
- Thematische Module
- SPARQL-Endpoint

Planungen im Bereich LOD

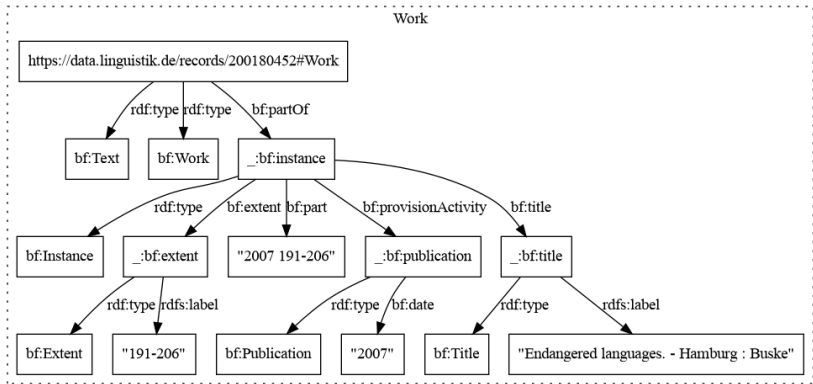
- Volltextsuche mit Named Entity Linking
- Verknüpfung mit META-Share-Ontologie
- Verknüpfung mit Wikidata
- Verknüpfung mit GND-Sachbegriffen

Vielen Dank für Ihre Aufmerksamkeit!

Weitere Informationen:

- <https://linguistik.de> (Linguistik-Portal)
- <https://data.linguistik.de> (BLL LOD Edition)
- <https://annohub.linguistik.de> (Annohub)
- <https://github.com/ubffm/lingframe> (Lingframe Anwendungsprofil)
- <https://www.linguistik.de/about/publications/> (Publikationsliste)

Beispiel: BIBFRAME Work (2)



Beispiel: BIBFRAME Instance

