

Workshop Semantische Extraktion von Informationen

17.01.2023 10:00 – 13:30 Uhr

Moderation: Adrian Pachzelt



SENCKENBERG
world of biodiversity



Tagesordnung

1. Begrüßung
2. Keynote (Chris Biemann; Universität Hamburg) inkl. Diskussion
3. Umfrageauswertung der semantischen Technologien der FID
4. Kurzvorstellung der einzelnen FID
 - PAUSE (30 Minuten) --
5. Einsatzmöglichkeiten von semantischen Technologien in FID / Lessons Learned inkl. Diskussion
 - AutoSE: Automatisierung der Inhaltserschließung mit Machine-Learning-Methoden an der ZBW (Anna Kasprzik)
 - FID Pharmazie - Der Narrative Service (Hermann Kroll)
 - Anwendung von LOD-Technologien im Linguistik-Portal (Thorsten Fritze)
6. Schluss (geplant 13:30 Uhr)

Keynote

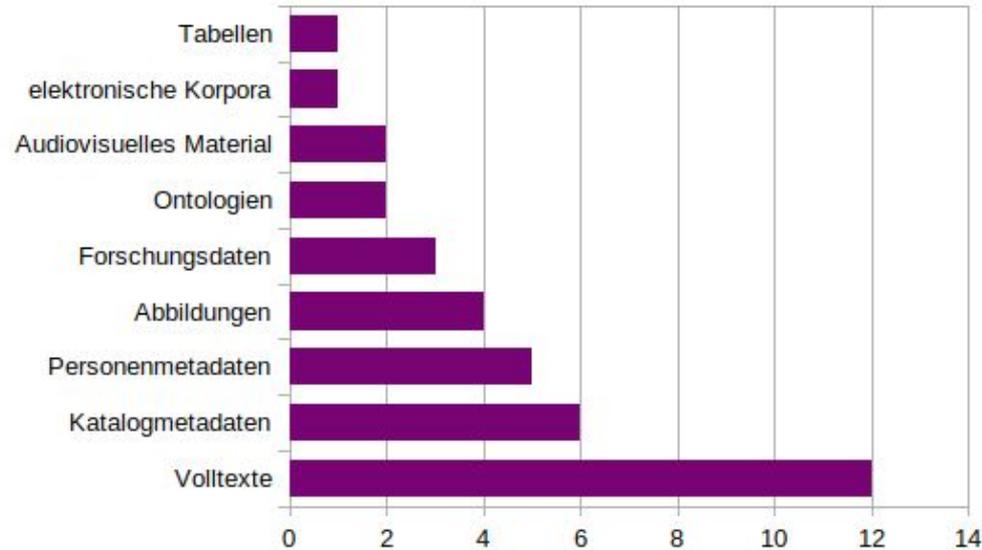


Prof. Chris Biemann
– Universität Hamburg

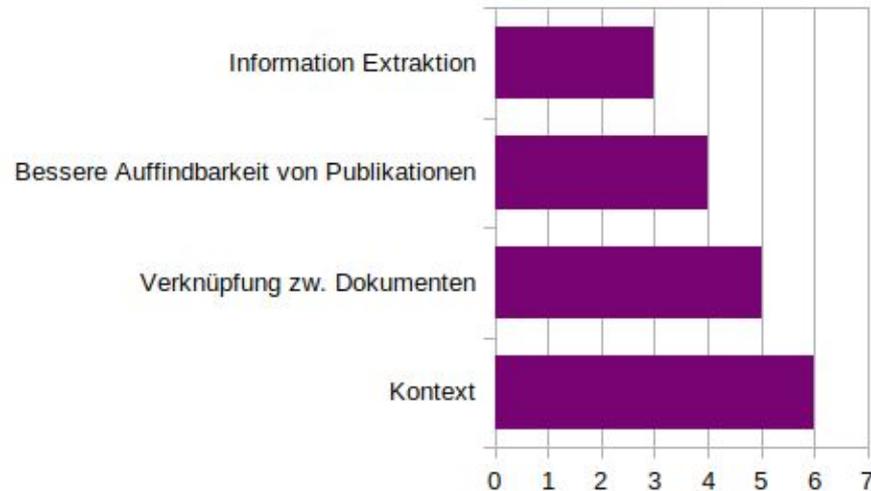
Professor für Sprachtechnologie

Umfrageauswertung der semantischen Technologien der FID

Welches Medium wird semantisch exploriert?



Für die Fachcommunity wichtige Daten aus der semantischen Extraktion?



Welche Software wird zur semantischen Erschließung genutzt?

Machine Learning Apps	Speichern & Abrufen	Management	Visualisierung
Annif	Apache Jena / Apache Tinkerpop	OpenRefine	SKOS Play
TextImager	GraphDB	TextAnnotator	
	Semantic MediaWiki	VocBench3	

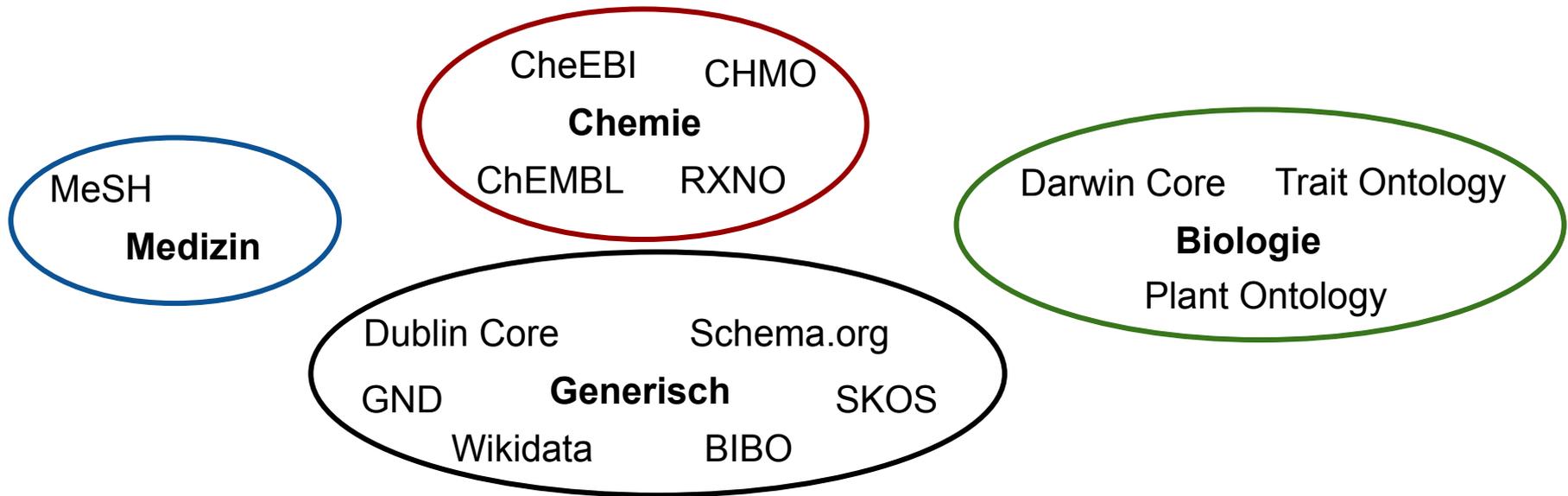
URLs zu allen Ressourcen finden sich im Anhang der Präsentation

Werden eigene Ontologien erstellt?

Vor allem Nachnutzung (10)

Vor allem eigene Ontologien (5)

Welche veröffentlichten Ontologien werden genutzt?



Wird Machine Learning im Projekt verwendet?

Ja (6)

Nein (6)

Wird Linked Open Data berücksichtigt?

Ja (12)

Nein (1)

Welche Daten-Download Formate werden angeboten?

- Datei-Download (CSV/TXT/JSON/PDF): 7
- Web-API (JSON/RDF): 8
- Sonstiges: SPARQL-Endpoint

Kurzvorstellungen der FID

FID adlr.link

FID Biodiversitätsforschung

FID Buch-, Bibliotheks- &
Informationswissenschaft

FID Darstellende Kunst

FID Erziehungswissenschaft &
Bildungsforschung

FID Jüdische Studien

FID Kriminologie

FID Lateinamerika, Karibik und Latino Studies

FID Linguistik

FID Materials Science

FID Move

FID Ost-, Ostmittel- & Südosteuropa

FID Pharmazie

FID Physik

FID Recht

FID Religionswissenschaften

FID Romanistik

FID Slawistik

FID Sozial- und Kulturanthropologie

FID Theologie

Heidelberg University Press (HeiUP)

Karlsruher Institut für Technologie (KIT)

Technische Informationsbibliothek (TIB)

ZBW - Leibniz Informationszentrum Wirtschaft

Pause

Einsatzmöglichkeiten von semantischen Technologien in FID

AutoSE: Automatisierung der Inhaltserschließung mit Machine-Learning-Methoden an der ZBW (Anna Kasprzik; ZBW)

FID Pharmazie - Narrative Service (Hermann Kroll; FID Pharmazie)

Anwendung von LOD-Technologien im Linguistik-Portal (Thorsten Fritze; FID Linguistik)

Diskussion

- Themenschwerpunkte in der Zukunft?
- Verbindung der FID Services via LOD gewünscht?
- Wünsche für weitere Formate?
 - Coding Club für technische Kommunikation der FID/Best Practice
- Wie kann auf administrativer Ebene mit dem Thema zusammengearbeitet werden?

Ressourcen

- Annif: <https://annif.org/>
- Apache Jena: <https://jena.apache.org/>
- Apache Tinkerpop: <https://tinkerpop.apache.org/>
- GraphDB: <https://www.ontotext.com/products/graphdb/>
- OpenRefine: <https://openrefine.org/>
- Semantic MediaWiki: https://www.semantic-mediawiki.org/wiki/Semantic_MediaWiki
- SKOS Play: <https://skos-play.sparna.fr/play/>
- VocBench3: <https://vocbench.op.europa.eu/#/Home>
- W3C Ontology Repositories: https://www.w3.org/wiki/Ontology_repositories

- [Homepage von Prof. Chris Biemann](#)

Semantische Extraktion von Informationen

- **Prämisse:** Automatisierter Prozess angewendet auf unstrukturierte Daten (Texte)
- **Aufgabenfelder:**
 - Herausarbeiten von “Entitäten” (z. B. Personen, Orte, Taxa u.ä.)
 - Zusammenhänge von Wörtern innerhalb eines Satzes aufzeigen
 - Fakten aus dem Text (auch satzübergreifend) extrahieren
 - Zusammenfassung eines Textes