

# LAKEBASE SEMANTIC SERVICE

JAN MARTIN KEIL

Heinz Nixdorf Chair for Distributed Information Systems  
Friedrich Schiller University Jena



FRIEDRICH-SCHILLER-  
UNIVERSITÄT  
JENA

ICEI 2018 – 27/09/2018

# LAKEBASE SEMANTIC SERVICE

- Web service for data management platforms
  - Semantic annotation management
  - Semantic search
  - Ecological research data
- 

## ACKNOWLEDGMENTS

The development of the LakeBase Semantic Service was funded by DFG in the scope of the LakeBase project within the Scientific Library Services and Information Systems (LIS) program.

# SEMANTIC ANNOTATION

*There are eight jaguars in Germany.*<sup>1</sup>

- Is this about the animal or the car?
- How should a machine know?
- Annotation: Assign an identifier
  - Animal "jaguar":  
<http://www.wikidata.org/wiki/Q35694>
  - Car company "jaguar":  
<http://www.wikidata.org/wiki/Q30055>

1. <https://www.zootierliste.de/?klasse=1&ordnung=115&familie=11508&art=1120911>, 26/09/2018

# SEMANTIC SEARCH

*jaguar*

- We will find animals and cars
- Lets search for the identifier:  
<http://www.wikidata.org/wiki/Q35694> (animal "*jaguar*")
- What about searching for animals?  
<http://www.wikidata.org/wiki/Q729> ("*animal*")

# WORKFLOW



Data & Metadata



Data & Metadata Annotations Suggestion



Data & Metadata Annotations Adjustment



Query



Query Auto Completion &  
Annotations Suggestion



Query Annotations Adjustment



Search Results



# ANNOTATIONS SUGGESTION

*The data set contains the phosphorus and nitrate concentrations as well as the water temperatures measured mostly weekly or fortnightly in the main tributaries of the Saidenbach Reservoir in the period from 1975 to 2017.*

**nitrate** (100%)

[http://purl.obolibrary.org/obo/CHEBI\\_17632](http://purl.obolibrary.org/obo/CHEBI_17632)

**temperature** (98%)

<http://www.ontology-of-units-of-measure.org/resource/om-2/Temperature>

**temperate** (95%)

[http://purl.obolibrary.org/obo/ENVO\\_01000206](http://purl.obolibrary.org/obo/ENVO_01000206)

**Talsperre Saidenbach** (100%)

<http://www.wikidata.org/entity/Q895843>

## STORING ANNOTATIONS

- Usage of suggestion, preferring long terms
- Keyword fallback for unknown terms (ignoring stopwords)
- Optional adjustment of annotations
- Simultaneous annotation of multiple data objects

## DATA OBJECTS

- Packages of datasets
- Entire datasets
- Entire columns
- Entire rows
- Single cells
- Particular metadata values about one of these



# AUTO COMPLETION

- Provides completed term and concept

*Talsperre Sa*

**talsperre sander-lippe**

<http://www.wikidata.org/entity/Q1827503>

**talsperre saidenbach**

<http://www.wikidata.org/entity/Q895843>

# SEMANTIC SEARCH

- Usage of all suggested query annotations
- Keyword fallback for unknown terms (ignoring stopwords)
- Optional adjustment of query annotations to refine search
- Utilization of concept hierarchy to provide more results:
  - Search for *"artificial lake"* will include reservoirs

# FURTHER FEATURES

- Output Annotations
- Copy Annotation
- Delete Annotation
- Descriptions of Concepts

# SEMANTIC DATA SOURCES

- Java API for access of terminologies
- Functionality:
  - Label matching
  - Broader / narrower concepts
  - Concept description
  - Concept labels
  - ...

- Provided implementations:
  - OWL ontologies
  - SPARQL endpoints
  - [World Register of Marine Species \(WoRMS\)](#)
  - [GFBio Terminology Service](#)
- Easy addition of new terminologies

# APPLICATION

- IGB Freshwater Research and Environmental Database
- Free software (Apache 2.0 licensed)

## LAKEBASE SEMANTIC SERVICE

---

# QUESTIONS?

---

 [jan-martin.keil@uni-jena.de](mailto:jan-martin.keil@uni-jena.de)

 [@janmartinkeil](https://twitter.com/janmartinkeil)

 <https://github.com/fusion-jena/LakeBaseSemanticService>