



How to Search for Biological Data?

A Comparison of User Interfaces in a Semantic Data Search

Felicitas Löffler¹, Friederike Klan², Birgitta König-Ries¹

*¹Heinz Nixdorf Chair for Distributed Information Systems
Department of Mathematics and Computer Science
Friedrich Schiller University Jena, Germany*

²German Aerospace Center (DLR) , Jena, Germany

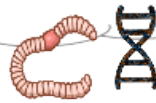
MOTIVATION

Do butterflies occur on calcareous grassland?



Ecologist

Is there genome data about nematodes?



Molecular Biologist

How does agriculture influence the groundwater?



Marine Ecologist

I'm interested in data about the frog distribution in Europe.



Zoologist

Scholars have complex information needs

Metadata Standards

EML

```
<eml><title>Tree measurements in a  
subtropical forest  
</title>  
<creator>Hans Meier</creator>  
<description> We analyzed [...] </description>  
<keywords>schima superba</keywords>  
<methods></methods>  
...  
</eml>
```

ABCD

```
<abcd:dataset>  
<abcd:TechnicalContact>Museum  
XYZ</abcd:TechnicalContact>  
<abcd:title>Herbarium </abcd:title>  
...  
<abcd:ScientificName>Helianthus  
annus</abcd:ScientificName>  
...  
</abcd:dataset>
```

What? Who? Where? When?

→ Metadata are focused on data and citation aspects



<https://www.gfbio.org/>

Search:



Show entries per page

Showing 1 to 10 of 5,507,820 entries

Prev

[Asmus, Harald; Hussel, Birgit; Kadel, Petra; Asmus, Ragnhild; Rick, Johannes J; Wiltshire, Karen Helen \(2018\): Clupeidae larvae in the S](#)

Data Center: PANGAEA: Data Publisher for Earth & Environmental Science

Parameters: Sample code/label; LATITUDE; LONGITUDE; Latitude 2; Longitude 2; Depth, bathymetric; DATE/TIME; Habitat; Mass; Clupeidae, larvae, tr

License/Rights: CC-BY: Creative Commons Attribution 3.0 Unported

[Data Description](#) - [Data Download](#)

[Asmus, Harald; Hussel, Birgit; Kadel, Petra; Asmus, Ragnhild; Rick, Johannes J; Wiltshire, Karen Helen \(2018\): Merlangius merlangus in](#)

Data Center: PANGAEA: Data Publisher for Earth & Environmental Science

Parameters: Sample code/label; LATITUDE; LONGITUDE; Latitude 2; Longitude 2; Depth, bathymetric; DATE/TIME; Habitat; Merlangius merlangus, tot

License/Rights: CC-BY: Creative Commons Attribution 3.0 Unported

[Data Description](#) - [Data Download](#)

[Asmus, Harald; Hussel, Birgit; Kadel, Petra; Asmus, Ragnhild; Rick, Johannes J; Wiltshire, Karen Helen \(2018\): Crangon crangon in the S](#)

Data Center: PANGAEA: Data Publisher for Earth & Environmental Science

Parameters: Sample code/label; LATITUDE; LONGITUDE; Latitude 2; Longitude 2; Depth, bathymetric; DATE/TIME; Habitat; Mass; Crangon crangon, tr

License/Rights: CC-BY: Creative Commons Attribution 3.0 Unported

[Data Description](#) - [Data Download](#)

I'm looking for
shrub species
in a subtropical
forest



search over organisms, processes or habitats is not possible

How to search for biological data?

- 1.) What information categories are required to determine the search context?
- 2.) How to enable data search over these categories?
- 3.) What kind of user interface would be appropriate considering the categories?

OUR APPROACH

Information Categories

A) Question Corpus with 169 search and research questions, e.g.:

Do *butterflies* occur on *calcareous grassland*?

What are *associated taxa*?

Information Categories

A) Question Corpus with 169 search and research questions, e.g.:

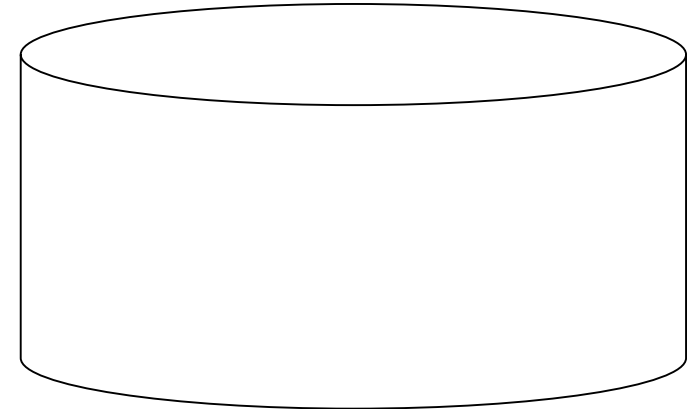
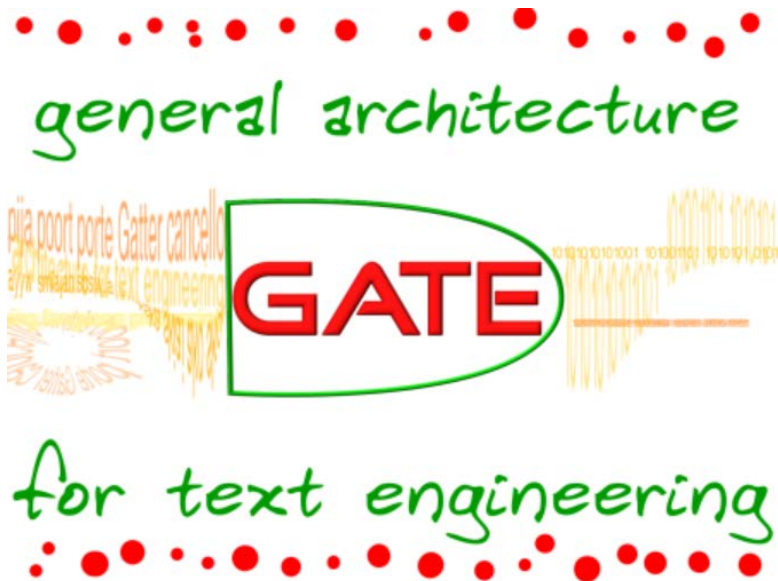
*Do **butterflies** occur on **calcareous grassland**?*

*What are **associated taxa**?*

B) 13 categories were identified, evaluation with 9 domain experts, finally 10 categories have been verified

***Organism**, **Environment**, Process, Material, Quality, Event, Person, Location, Time, Human Intervention*

Categories – Ontologies



Organism	Environment	Quality	Process	Material
OrganismTagger	ENVO (parts)	PATO, TO, FLOPO	ENVO (parts), GO (parts), UEBERON (parts)	ENVO (parts), ChemicalTagger

USER INTERFACES

User Interface I

- form-based - one search field per category
- SPARQL templates per category
- includes synonyms and more specific results

User Interface II

- NLP based - category is determined automatically
- SPARQL template over all categories for a first result
- includes synonyms and more specific results

Demo

User Interface I: <https://cloud.uni-jena.de/index.php/s/iibAyfwPcmGMjLo>

User Interface II: <https://cloud.uni-jena.de/index.php/s/mcCTiHqF8ZHy7Sj>

Which interface do you prefer?

*Please take part in a
short interactive survey!*

www.mentimeter.com

Code: 775281

Thank You! Questions?



@felicitasloe

@fusionUniJena



FUSION **F**unctionality **S**haring **I**n **O**pen **e**Nvironments

Heinz-Nixdorf Endowed Chair for Distributed Information Systems
Department of Mathematics and Computer Science - Friedrich Schiller University of Jena