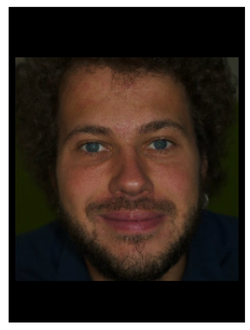
Essential Annotation Schema for Ecology (EASE)

Claas-Thido Pfaff[1], David Eichenberg[1], Birgitta König-Ries2, Mario Liebergesell[1], Christian Wirth[1]

1: Institut für Spezielle Botanik und Funktionelle Biodiversität, Universität Leipzig, Johannisallee 21, 04103 Leipzig,

2: Department of Mathematics and Computer Science, Universität Jena, Ernst-Abbe-Platz 2, 07743 Jena

Get The Poster



Introduction

Ecology has become a data intensive science over the last decade which often relies on the reuse of data in cross experimental analyses. However finding data which qualifies for being used in a specific analysis can be challenging. It requires good quality metadata as well as efficient search strategies. In ecology the most widely used search strategy is full text search which is known for being inaccurate. Faceted navigation provides a simple but efficient filter mechanism which can improve the results of a full text search.

Full text search

A keyword based full text search often fails to provide satisfactory results. This is related to the fact that it lacks an understanding of the semantic context and the meaning of a search query. For example it does not account for synonyms or homonyms nor does it know about closely related search terms. Thus it is likely to return not only irrelevant but also incomplete results (s. Figure 1).

Facet enhanced search

Facets provide a simple but efficient filter mechanism which can be combined with a full text search. The filter is based on metadata which provides information about a search object. The information is organized along overarching topics or dimensions which group related aspects (e.g. time, space). During a full text search such information can be used to provide selectable options which allow to clarify and refine the search query (s. Figure 2).

The facets of EASE

EASE provides 8 facets (s. Figure 3) forming large groups of specific information. They span axes in the search space of ecology and cover details like names of locations (e.g. location, country, continent), organisms and their taxonomy, parameterized biome descriptions (e.g. hemisphere, zonation, humidity), spheres as abstract location descriptors (e.g. abyssopelagic, crenon, a-horizon), methods and measured/manipulated variables (e.g. Producer Diversity, Radiation Quality), chemicals (Elements, Compounds, Biological functions) and processes (e.g. Adaption, Speciation). The time facet is unfolded as an example (s. Figure 3).

Conclusion

EASE is a framework based on two components being an XML schema and an ecological vocabulary. It provides the basis for an efficient annotation and faceted navigation discovery of data in ecology. The framework represents a synthesis work consolidating ideas which originate from widely used metadata standards in ecology (Ecological Metadata Language, Darwin Core, ABCD*), various vocabularies (e.g. CheBi*, ENVO*, Geonames, WRB*), text books, scientific literature, and expert knowledge. The framework is publicly available via GitHub: https://goo.gl/NJdygw

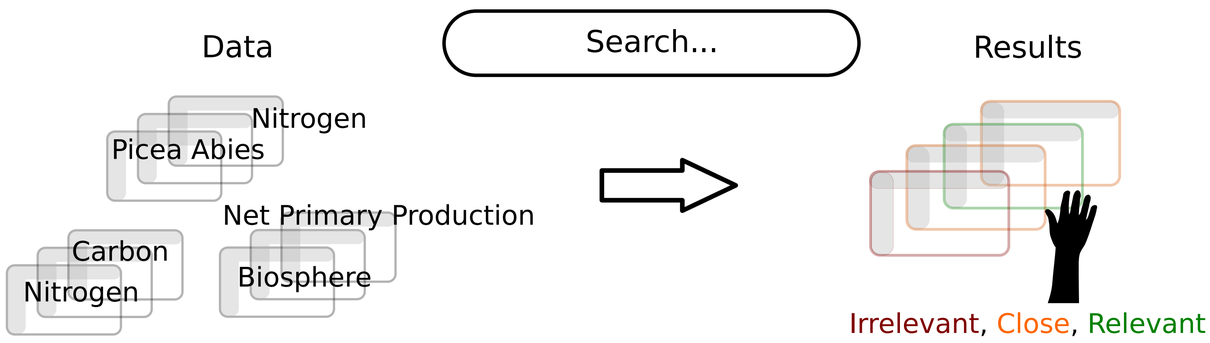


Figure 1: A keyword based full text search yields many but often not satisfactory results.

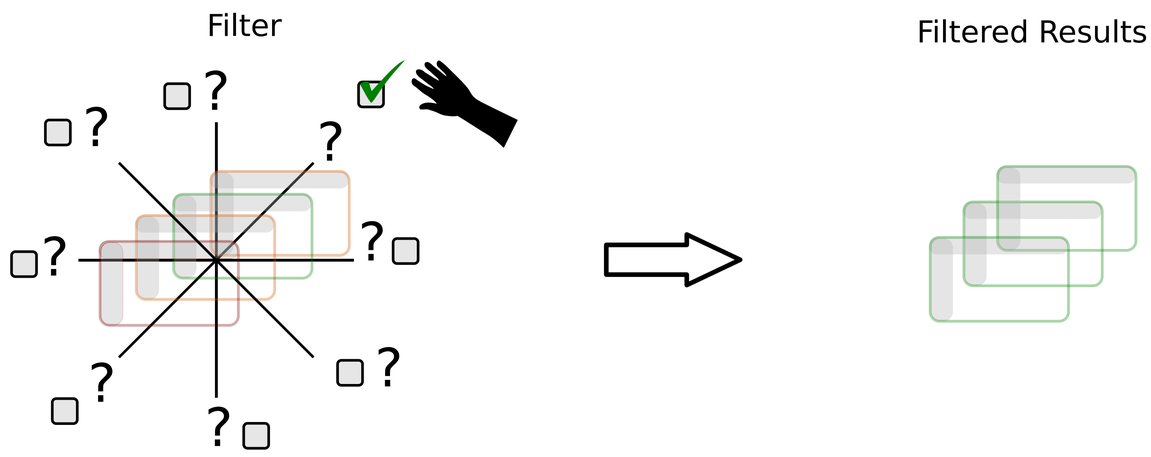


Figure 2: With facets the results of a full text search can be refined towards more relevance.

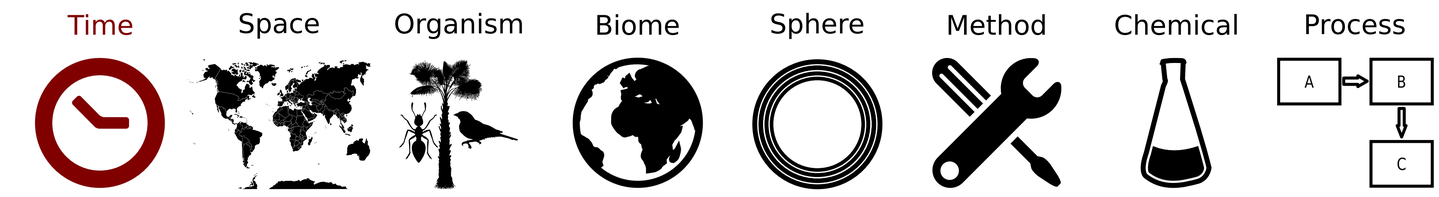


Figure 3: The facets which have been identified as most imporant categories for the classification of ecological search objects. Time is unfolded in figure 4.

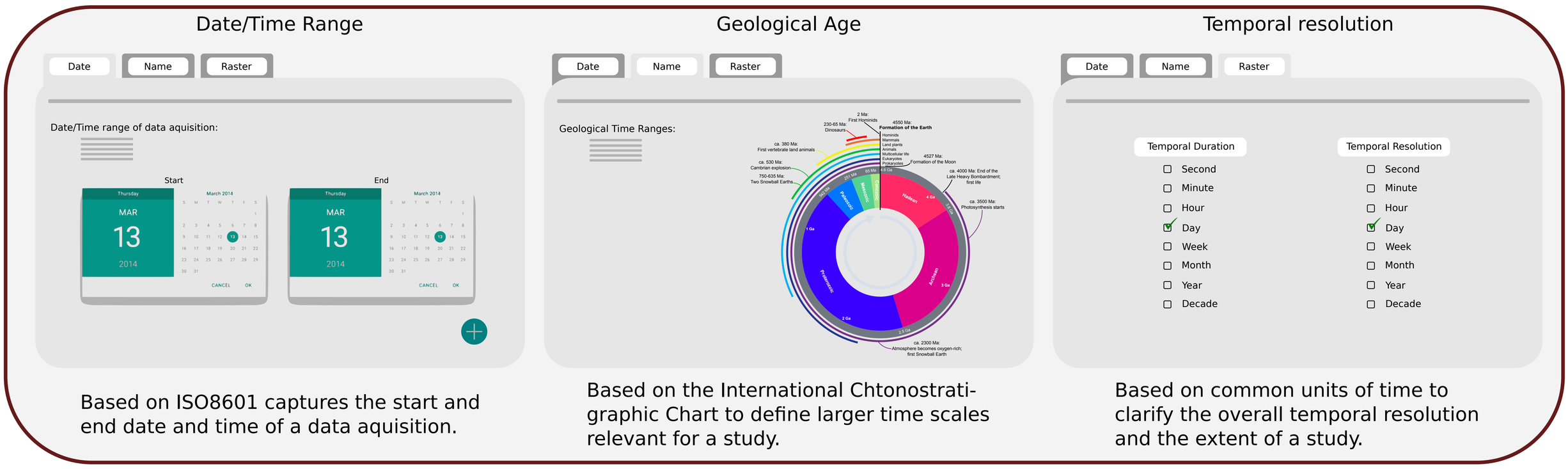


Figure 4: The time facet unfolded in form of mockups of a potential annotation application using EASE. Shown are the details which are captured and the sources which serve as a basis.

Acknowledgement: Thanks to all the people supporting the vocabulary through their expertise and discussions during the workshops: Alex Klein, Alexander Weinhold, Alexandra Weigelt, Alfred Radl, Anne Lang, Anton Güntsch, Bettina Ohse, Björn Reu, Claudia Guimaraes-Steinicke, Daniel Marra, David Fichtmüller, Harald Vacik, Hongmei Chen, Ingo Schöning, Ivaylo Kostadinov, Jens Kattge Karin Nadrowski, Katharina Grosser, Katherina Pietsch, Kristin Baber, Maren Gleisberg, Marion Schrumpf, Martin Freiberg, Naouel Karam, Nico Eisenhauer, Nicole Van Dam, Pelin Yilmaz, Pier Luigi Buttigieg, Robert Huber, Robert Tolksdorf, Rolf Engelmann, Ronny Richter, Sophia Ratcliffe, Steffen Bode.

Abbreviations: * ABCD: Access to Biological Collection Data, WRB: World Reference Base for Soil Resources, CheBi: Chemicals of Biological Interest, ENVO: Environment Ontology)

Printed by: Universitätsrechenzentrum Universität Leipzig

UNIVERSITÄT LEIPZIG