How to Search for Biological Data?
A Comparison of User Interfaces in a Semantic Search

Felicitas Löffler[1], Friederike Klan[2], Birgitta König-Ries[1]

[1]*Heinz Nixdorf Chair for Distributed Information Systems, Friedrich Schiller University Jena, Germany*;

[2]*German Aerospace Center (DLR), Germany*;

Corresponding author(s) e-mail: felicitas.loeffler@uni-jena.de

ABSTRACT: Data discovery is a frequent task in a scholar's daily work. In biodiversity, data search is a particular challenge. Here, scholars have complex information needs such as the rich interplay of organisms and their environments that cannot be unambiguously expressed with a traditional keyword search, e.g., Does tree diversity reduce competition in a subtropical forest? Therefore, data repositories usually offer interfaces that enable users to browse datasets by a pre-determined set of categories or facets. Faceted search is a good compromise between cumbersome user interfaces for structured queries (e.g., using SPARQL) and natural language queries that are hard to interpret for machines. Thus, developers can specify relevant relationships between entities explicitly and users can filter search results by selecting proper categories. For the given query, appropriate categories could be *Organism* and *Habitat*.

However, there are two crucial design issues that have an impact on the effectiveness of category-based query interfaces: The choice of proper categories and the visual presentation of these categories in the query interface. In our work, we focus on the second aspect. We aim to develop two query interfaces: (a) a common one-box keyword search interface that analyzes the entered terms with respect to their categories automatically (b) a form-based query interface where users can enter their search keywords into a form with a query field per category. In both interfaces, the query keywords are matched against concepts in a knowledge base to make their semantics explicit. In case of a successful match the URI is used to obtain the labels of all sub-concepts to expand the query before sending it to the search engine. Retrieved results are displayed in a list. The aim of our system is not to answer the question completely but to support users in retrieving relevant datasets that give hints to answer a research question.

In our talk, we will introduce the two interfaces and invite the conference participants to give feedback. We are particularly interested in a discussion on the appropriateness of the suggested user interfaces. Do scholars prefer a form-based user interface or only a one-field search? What other functions might be helpful, e.g., providing more information about other relations and properties from the concept in the ontology? What kind of explanations might be helpful to understand why a certain result was returned?

KEYWORDS: user interfaces, semantic search, biological data, life sciences, biodiversity