Semantic Search for Biological Datasets: A Usability Study on Modes of Querying and Explaining Search Results

Felicitas Löffler[1], Fateme Shafiei[2], René Witte[3], Birgitta König-Ries[4], Friederike Klan[5]

Abstract: Dataset discovery is a frequent task in daily research practice, yet studies are missing that explore the usability of user interfaces (UI) in data portals. In particular, very few user studies exist that analyze whether particular elements in the user interface are useful for search tasks. We aim to address those needs for more specific usability evaluations in dataset search. In this work, we present a flexible semantic search over biological datasets with two user interfaces. The search result contains semantically related terms, such as synonyms or more specific terms, obtained from domain ontologies. We evaluated the system in a user study with 20 scholars. We focused on two components, the query input to explore a search in categories (entity types) in comparison to a single input field, and we analyzed textual highlightings in the returned datasets to study whether users are distracted by semantic information such as URIs. Our results show that users prefer interfaces with a single input field for search tasks they are not familiar with, and that users appreciate explanations with terminologies and URIs.

Keywords: Dataset Search; Semantic Search; Biodiversity; Life Sciences; User Interface; Usability

1 Introduction

Data-intensive research increasingly requires scientists to retrieve datasets in data portals. Scholars look for datasets to compare their own results with legacy data, to prevent repeating cost-intensive experiments or to merge multiple data sources into a new dataset, in order to explore novel hypotheses. Biodiversity research is one example of a data-intensive research area. It examines the variety of species and their genetic and ecological diversity [Lo10]. Multiple studies report on various obstacles users encounter while searching for datasets. Main problems are missing primary data and lacking information on data collection methods [Pa16], scattered data in different data repositories [Cu18], unsatisfactory user interfaces [Gr20] and insufficient metadata descriptions and unaligned terminologies [Lö21]. Semantic search approaches that are going beyond the classical keyword search can partially help

[1] Heinz Nixdorf Chair for Distributed Information Systems, Friedrich Schiller University, Jena, Germany, felicitas.loeffler@uni-jena.de
[2] Heinz Nixdorf Chair for Distributed Information Systems, Friedrich Schiller University, Jena, Germany
[3] Semantic Software Lab, Concordia University, Montréal, Canada
[4] Heinz Nixdorf Chair for Distributed Information Systems, Friedrich Schiller University, Jena, Germany
Michael-Stifel-Center for Data-Driven and Simulation Science, Jena, Germany
German Center for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Germany
[5] Data Acquisition and Mobilization Department, Institute of Data Science, German Aerospace Center (DLR), Jena, Germany

doi:10.18420/BTW2023-56

to overcome the current drawbacks: In particular, using standardized domain vocabularies and terminologies in the search process enhances sparse metadata and enables users to find more relevant data [LK16]. However, very few semantic search solutions for dataset search are publicly available. In this work, we present a semantic search over metadata files and domain vocabularies. We link entries in metadata files with concepts from relevant domain taxonomies and ontologies with respective URIs. Based on the linked vocabularies and its concepts, we also derive an entity type (entity category or type). Both information (URIs and entity types) are added to metadata as semantic annotations and are utilized in a semantic index with a URI-based retrieval model. In order to study which search input users prefer in dataset search, we implemented two user interfaces: a UI with a single input field and a second UI enabling a search over different semantic categories. In both interfaces, the search results are enhanced with text highlightings and additional information on demand, to support users in obtaining a quick overview with explanations of the returned datasets. We evaluated the system with 20 scholars working in the field of Life Sciences and Environmental Sciences. The participants performed various search tasks in a given time frame and answered questionnaires before and after each search task, for each user interface, and at the end of the session. Thus, our contribution is two-fold:

1. We present an architecture enabling a flexible, semantic dataset search; and
2. We provide a usability study on two main aspects in search: the query input and search result explanations, such as textual highlightings with additional information.

The structure of the paper is as follows: First, we present related work in Section 2, followed by the evaluation setup in Section 3, including information on our system architecture. The evaluation results are presented in Section 4.

2 Related Work

Due to the multitude of heterogenous data formats, research data, such as spread sheets, multimedia files or questionnaires, are described by metadata. This descriptive information on, e.g., author, title, abstract, collection time and data format, are utilized by data portals in search indexes [KCW18]. Very few usability studies for dataset search are publicly available. Therefore, besides a discussion on user studies in dataset retrieval in the Life Sciences, here we introduce semantic search approaches in the Life Sciences and user studies for semantic search systems.

User studies in dataset retrieval. [Di17] and [Ch18] are examples for user studies in the Datamed portal,[6] a federated approach providing datasets from numerous biomedical data repositories. A second data portal that conducted a user evaluation is DataONE[7] [Vo15],

[6] Datamed, `https://datamed.org/`

[7] DataONE, `https://www.dataone.org/`

a data portal providing environmental and biological datasets from various sources. The studies by [MM15] and [Ka20, Ka21] report results from user studies in the earth observation and oceanography domain (dealing with spatial and temporal data). Most of the studies collected qualitative data with less than 30 subjects. All studies focused on the evaluation of the usability and utilized various usability methods, such as thinking-aloud, questionnaires and interviews. Apart from [Ch18], all studies observe two main obstacles: (A) metadata quality [Vo15, Di17, Ka20, MM15], e.g., different spellings of variable names and (B) problems with the user interface, e.g., inconsistent information in title and description or a purely text-based presentation [Vo15, Ka20]. Users of the Datamed portal also suggested offering an enhanced search input, allowing them to search for domain specific topics, such as a search for phenotypes or genes [Di17]. Another source for guidelines and good practices in dataset search are the outcomes of the RDA Data Discovery Paradigms Interest Group.[8] Based on 79 data discovery use cases, heuristic evaluations and interviews with experts, they propose ten recommendations for enhancing dataset search and user experience [Wu19]. The main areas addressed in the RDA's suggestions are: *(i)* providing multiple search inputs in order to support different information needs, *(ii)* filtering options, and *(iii)* comprehensible search summaries on the search entry result page (SERP), by displaying dataset snippets that belong to a search query.

Semantic search systems in the Life Sciences. Semantic search is a "search beyond keywords" [BBH16]. Instead of matching query input and document content syntactically, the result set also contains semantically related content by exploiting additional knowledge to the search process. Data sources are either plain text, structured data from knowledge bases or a combination of both. Search approaches vary between *keyword search* (+ query expansion techniques), *structured search* by using specific query languages such as SPARQL and *full natural language questions* in question answering [BBH16]. Multiple semantic search approaches have emerged in the biomedical domain focusing on scientific articles from PubMed[9], e.g., [LLW15, Mu17, Al18, SPA18]. For dataset search, only very few semantic systems exist such as BioFid [Pa21] (German legacy collection data), Datamed [Ch18] (biomedical datasets) and GFBio (biological and environmental data) [Lö17]. Most systems expand the query input. Data discovery with named entities over semantic categories is only supported by [LLW15, SPA18, Al18]. The search summaries usually present snippets of the document or datasets, contain information on data sources and highlight the matched search terms. Explanations on semantic categories or matching entities [SPA18, Al18], full query information [Ch18] or additional information from external sources such as wikipedia [Pa21] are only partially available.

User studies on semantic search systems. [KB07] was one of the first studies exploring different query interfaces on a knowledge base with geographical information. This study also utilized a combination of search tasks and a System Usability Scale (SUS) questionnaire [Br96] to determine which kind of query input is useful for a search over knowledge

[8] RDA DDP IG, `https://rd-alliance.org/node/52248/outputs`

[9] PubMed, `https://pubmed.ncbi.nlm.nih.gov`

bases. The user interface with support for full natural language questions obtained the best success rate and resulted in the highest SUS score (75.73). The authors of [EWC12] utilized the same geographic knowledge base and analyzed two user types, experts and casual, for three different query inputs: natural language, form-based and graph-based. Their results show that casual users prefer the form-based query interface and experts favored the graph-based search. The study by [Ve16] also conducted an A/B test with 15 users and two semantic search systems over an RDF dataset with administrative and financial information of Norwegian companies. The participants performed four search tasks for each system and filled in questionnaires before and after each search task, after each system, and at the very end of the evaluation. Similar to the study of [EWC12], the form-based system was favored by casual users and the graph-based one by expert users.

The evaluation studies in dataset retrieval introduced above only collected qualitative data, through questionnaires and interviews. In contrast, the semantic search user studies followed the Text REtrieval Conference (TREC) guidelines for interactive information retrieval [Du05], but did not focus on dataset search. [Ch18] and [Lö17] are dataset search approaches with semantic enhancements based on query expansion. While [Ch18] extends the entered keywords with synonyms, scientific and common names obtained from the UMLS,[10] a knowledge base for the biomedical domain, [Lö17] utilizes the GFBio TS [Ka14]. Both systems do not offer searching for specific entity types and only partially support explanations on matched entities.

3 Evaluation Setup

Previous studies always examined user interfaces as a whole. In order to identify the impact of different UI choices more precisely, we decided to focus our study on two main parts of search: the query input and the search result summary. These two aspects are also top-ranked recommendations in the RDA guidelines [Wu19]. In the user study of [Di17], users mentioned the need for searching specific entity types. Our previous work [Lö21] also revealed that scientific information needs to differ in granularity. Search questions can be very specific, e.g., a search for a concrete species, or can have a broader scope, e.g., a search for datasets with organisms in water samples. Therefore, we implemented two user interfaces: the first one (Biodiv 1) provides a category-based input of search terms and the second interface (Biodiv 2) offers a classical single-input field and determines entity types automatically. The search results in both interfaces can contain datasets going beyond the entered query terms, to broaden the search on semantically related terms, like synonyms. Concerning the presentation of the search results, we followed the RDA recommendations for search summaries in dataset search. To ensure a consistent layout, we aligned the presentation of each dataset entry to existing data portals, such as GFBio[11] and Zenodo.[12] In a previous study [LK16], we determined that end users need more explanations

[10] UMLS, https://www.nlm.nih.gov/research/umls/index.html
[11] GFBio, https://www.gfbio.org
[12] Zenodo, https://zenodo.org/

when being faced with search results going beyond keywords. This motivated us to further focus our analysis of the search summary to the presentation of explanations, allowing us to study whether displaying terminologies and matching URIs confuse and distract users or help them. Through the biodiversity research projects we are involved in (e.g., iDiv,[13] NFDI4Biodiversity[14]), we know that most scholars in the Life and Environmental Sciences are casual users, with no expertise in semantic technologies. However, as FAIR data management with terminologies are increasingly forming the semantic backbone in academia, it is necessary to explore what additional information end users need from a semantic dataset search, in order to understand and assess the relevance of search results.

3.1 System Architecture

The overall architecture (Figure 1) is an extended version of a framework we introduced in previous work [Sh21]. According to [BBH16], our system represents a structured search over text and knowledge bases. A semantic search index (see paragraph below on semantic indexing) and a local terminology service with ontologies of the OBO Foundry initiative[15] form the back-end. In order to link entered keywords with concepts in ontologies, we utilize an updated version of the Semantic Assistants framework [WG08]. This new version provides several text mining pipelines from various sources in a Spring boot application and two micro services, accessible via REST services.[16] The first micro service offers pipelines of the text mining framework GATE.[17] Currently, three taggers are available: GATE ANNIE [Cu13], which extracts named entities, such as people, locations and organizations; the BiodivTagger [Lö20], focusing on the recognition of environmental terms, data parameters, materials, chemicals and processes; and the OrganismTagger [Na11] extracting species. A REST API in the middleware enables the front-end to communicate with the back-end applications.[18]

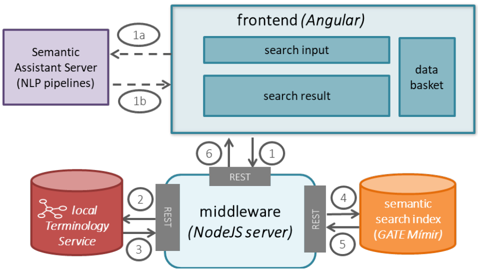

Fig. 1: Extended architecture and overall flow, based on the previous version introduced in [Sh21]

[13] iDiv, `https://www.idiv.de`
[14] NFDI4Biodiversity, `https://www.nfdi4biodiversity.org/de/`
[15] OBO Foundry, `https://obofoundry.org/`
[16] SA2.0, `https://github.com/fusion-jena/SA2.0`
[17] GATE, `https://gate.ac.uk/`
[18] [Dai:Si] semantic search, `https://github.com/fusion-jena/daisi-semantic-search`

Users can either enter keywords per category or they can type all search terms in one input field (1). In the latter case, the Semantic Assistants service is called to look for matching URIs and entity types in the entered search terms (1A–1B). Afterwards, the query terms are sent to the terminology service (2) and matching URIs are returned to the middleware (3). The obtained URIs and entity types are utilized to fill templates being sent to the search engine (4). The obtained datasets from the search index (5) are forwarded to the Angular application, and finally, the result is presented to the user (6).

Dataset corpus preparation and semantic indexing. We downloaded metadata files from GFBio[19] being relevant for the selected search tasks. The files were provided a in repository specific metadata format.[20] We manually annotated the search tasks with URIs from OBO Foundry ontologies to obtain descendants nodes, labels and alternate labels via SPARQL queries from our local terminology service. These additional terms were added to the original query terms. We downloaded only the top-100 datasets per query, to ensure a manageable corpus size. The obtained ~52.000 metadata files were semantically annotated with the OrganismTagger [Na11] and the BiodivTagger [Lö20]. As a result, key terms in the metadata were linked to matching entities and their types in domain specific ontologies, such as Organism, Environment, Data Parameter, Process, or Material. Using GATE's Semantic Enrichment Processing Resource, we added ancestor nodes to each entity as additional annotations (`subClassOf*` relations), facilitating a hierarchy search at run-time. SPARQL queries for hierarchy relations can become complex, and a call to a terminology service can take up to several minutes to come back. Therefore, we added these hierarchical relations to the metadata files in the pre-processing phase as 'broader' annotations. All metadata files were indexed with GATE Mímir [Cu13], a search engine that provides indexing of semantic annotations with entity types and features, like matching URIs.

User interfaces. We proposed several UIs in clickable paper prototypes, discussing them in a focus group with three biodiversity scholars. We finally selected two UIs for supporting scholars who are not experts in semantic web technologies. As a result, we decided to support keyword search in both user interfaces and to omit an entity-based search. The first UI (Biodiv 1, Figure 2 (left)) provides a category-based input. Here, users need to actively sort the search terms into given domain categories. The entered keywords are sent to the terminology service to find matching URIs. In the second UI (Biodiv 2, Figure 2 (right)), the query terms are additionally sent to the Semantic Assistants server to obtain the entity types. The retrieved URIs and entity types are then utilized to form the final query being sent to the search index. In case no matching URI can be found, we also consider the original keywords in the query (Listing 1).

```
((((``honeybee'' IN {Organism}) OR (``honeybee'' OVER {Organism}) OR (``honeybee'' AND
{Organism})) OR ({Organism broader=``http://purl.obolibrary.org/obo/NCBITaxon_7460''} OR
{Organism inst=``http://purl.obolibrary.org/obo/NCBITaxon_7460''})))
```

List. 1: Full query for the search 'honeybee' in Figure 2

[19] GFBio Metadata, `https://github.com/fusion-jena/GFBioMetadata`
[20] PAN-MD, `http://ws.pangaea.de/schemas/pangaea/MetaData.xsd`

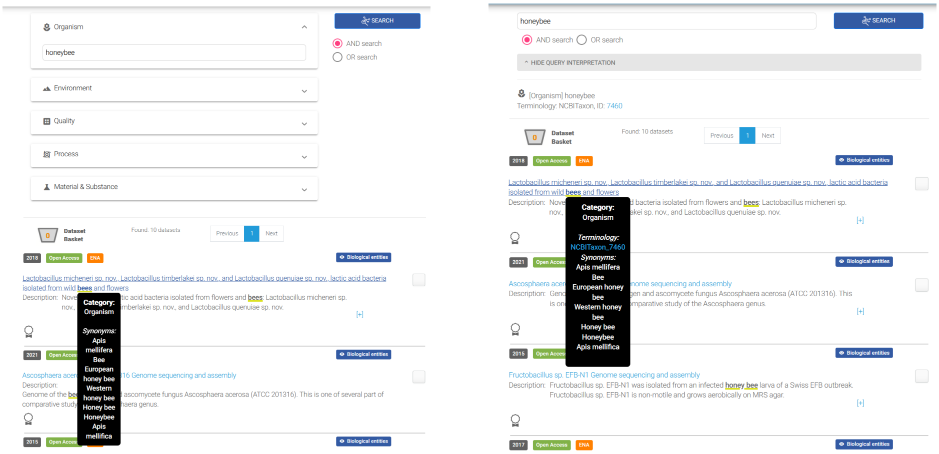

Fig. 2: Screenshots of Biodiv 1 and Biodiv 2, with a search for ‘honeybee’, returning alternate labels

We leverage search templates with the introduced ‘broader’ annotation to consider exact matches, as well as more specific terms, in the result set. The search results display also provides two types of highlighting: The original entered query terms and their related terms are highlighted in bold font (‘default highlightings’). If synonyms are available, they are displayed with a green underline. On demand (mouse-over) a separate dialog opens and displays alternate labels. In Biodiv 1, this dialog shows only the entity type of the highlighted term, whereas in Biodiv 2, the respective terminologies and URIs are presented additionally. A summarized version of the available synonyms is listed in an explanation tab. This tab also displays information on the search query. In Biodiv 1, users can view only a shortened search query, whereas in Biodiv 2, the full query to the search index is presented. Supplementary highlightings of further biological entities can be displayed with a ‘biological entities’ button. This function sends the textual information of a dataset to the Semantic Assistants service and returns annotations obtained from the taggers. These highlightings of data parameters, environmental terms, processes, materials and species are underlined in blue color.

3.2 Experimental Design

Our setup generally follows the TREC-9 guidelines.[21] The overall aim was to measure the usability of a semantic dataset search, but with a particular focus on the query input and provided explanations in the search summary. In total, 20 scholars with a research background in the Life Sciences and Environmental Sciences took part. Each participant performed eight search tasks in different orderings, e.g., *“What data exists in the repository*

[21] TREC Interactive Track, `https://trec.nist.gov/data/t9i/spec.html`

for bacteria in the groundwater?". For each user interface, every scholar carried out four tasks. In a maximum of five minutes, the users had to search for up to three datasets per task. When they considered a dataset as relevant, they added the dataset to the data basket, and at the end of the five minutes, the basket was downloaded. We conducted the evaluation in live sessions at the Friedrich Schiller University Jena (FSU Jena) or visited the participants at their working places, e.g., iDiv, Leipzig[22], Senckenberg, Frankfurt[23] and BGBM, Berlin.[24] The total evaluation time for each session was around 120 minutes. The search part took around 80 minutes, with ten minutes per task. The non-search part was around 40 minutes; 25 minutes were utilized for the post system questionnaires and five minutes for the exit questionnaire, with questions about a comparison between the two interfaces and some demographic questions. Before the start, we allotted 15 minutes for introductory explanations. For instance, we explained the purpose of the study and demonstrated some example searches to minimize the training effect [RC08].

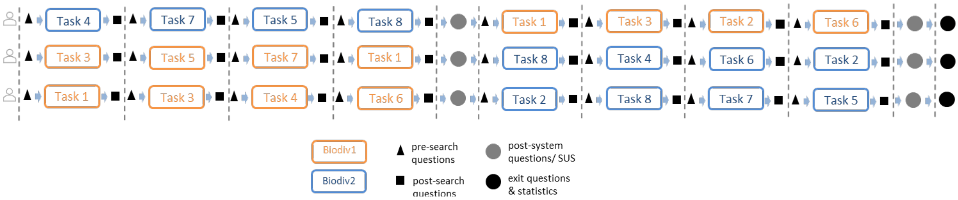

Fig. 3: Study flow for the first three users. Every user performed eight search tasks in different orderings. Half of the participants started with Biodiv 1, the other half started with Biodiv2.

Data collection and metric. Measuring the usability involves examining whether users are able to complete tasks in a given time (efficiency), how easy it is to work with the system (learnability), how many errors occur (issues), how easy it is to remember navigations and functions (memorability) and how satisfied users are overall with a system (satisfaction) [Ni93]. Apart from the memorability, we addressed all variables in our study. We evaluated the efficiency and effectiveness of the query input by means of user tasks (performance-based metrics). If users found three datasets in five minutes, we counted it as a complete task; if only one or two datasets could be found, it was considered as partially complete; and if no dataset was found, it counted as an uncompleted task. We measured the satisfaction, the learnability and collected usability issues with questionnaires (self-reported metrics) and through observation. Concerning the highlightings and explanations, we explored the comprehensibility, completeness and satisfaction through questionnaires. We also observed the participants and encouraged them to report occurring issues orally.

Search tasks and questionnaires. We selected eight search tasks from the question corpus we introduced in our previous work, to ensure that relevant categories from biodiversity research are contained [Lö21]. In order to better guide users through the evaluation, we

[22] iDiv, `https://www.idiv.de/`
[23] Senckenberg, `https://www.senckenberg.de`
[24] BGBM, `https://www.bgbm.org`

prepared a survey for each user.[25] The overall evaluation flow is presented in Figure 3. The survey provided questions before and after each task, according to the TREC guidelines, e.g., whether the participants expect certain content in the datasets, as well as questions on satisfaction, easy of use and ease of learnability afterwards. After four tasks (=each user interface), the scholars had to assess ten statements (a SUS questionnaire) on a five-point Likert scale to capture feedback on the usability. However, as we aimed to study the query input and explanation strategies, we adapted the classical SUS statements to our needs. The final SUS questionnaire contained two questions on the query input, three questions on the default highlightings, three questions concerning the biological entities highlighting and two questions on the provided query explanations. The final search tasks, the questionnaires and the original survey result files are available in Zenodo [Lö22].

4 Results

We compiled all results into one large CSV file and created a Jupyter notebook[26] to analyze the results. The code and the full results are available on GitHub.[27] Two-third of the 20 scholars search for datasets monthly or at least once in a year, and the other seven scholars use dataset search applications daily or weekly. The overall SUS scores for both interfaces resulted in values above 68 (Biodiv 1: 68, Biodiv 2: 71), which points to a good usability with respect to the two studied UI components in both interfaces. However, the dispersion in Biodiv 2 is large, because eleven users gave higher ratings for the interface with the single input field (Biodiv 2) than for the category-based search (Biodiv 1).

Search Input. With respect to the task success and task time, more scholars were able to retrieve three datasets, with fewer failure cases and in a shorter time, with Biodiv 2 (204ms) than with Biodiv 1 (225ms). However, the differences are small. Figure 4 reveals that for almost all tasks, the users were able to retrieve datasets. For two tasks, the success rate is low in both interfaces and the users complained that the returned datasets were not relevant or only partial information was relevant. This points to missing data in the metadata corpus. In addition, most scholars were not familiar with the search tasks (see the figures available on GitHub[28]). As the provided search tasks addressed a specific research question or a scientific information need, we did not expect that. However, this lack of specific knowledge impacted the results.

Concerning the learnability (see the figures on GitHub), the results of the questionnaires after each task and the respective statements in the SUS questionnaire reveal that it was easier for users to get started with Biodiv 2. But with respect to the ease of use, there is no

[25] Limesurvey, `https://www.limesurvey.org`
[26] Jupyter notebook, `https://jupyter.org/`
[27] Analysis, `https://github.com/fusion-jena/semantic-search-usability-analysis`
[28] Results for 20 users, `https://github.com/fusion-jena/semantic-search-usability-analysis/tree/main/analysis/results20`

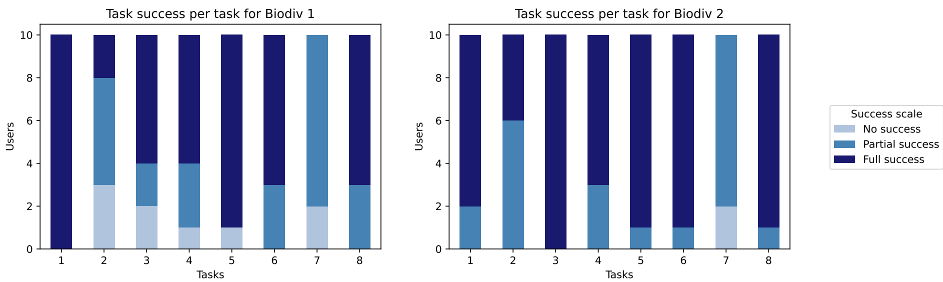

Fig. 4: Task success per task

difference between the two interfaces. Looking at the answers in the exit questionnaire paints the same picture. The pre-defined categories in Biodiv 1 were not as easy to understand, and it took some time for users to become familiar with the topical categorization of search terms. However, comments on Biodiv 1 reveal that one half of the users liked the category-based interface, because categories "helped to narrow down the search criteria and to get pertinent results." In contrast, the other half of scholars liked Biodiv 2, because it is "easier, as it goes straight forward without thinking about categories," and it "is more general and helped in searching for topics that I was unfamiliar with."

Highlightings and explanations. The users gave high ratings for the overall usefulness of default highlightings in both interfaces (Figure 5). However, in some result sets, the default highlightings were missing or too many terms were highlighted, which led to medium ratings. Overall, the provided information is sufficient and comprehensible. Concerning the highlighting of biological entities in the datasets (see figure on GitHub), we observed that only 50% of the participants utilized this function. Only when they had to give a rating on this function in the SUS questionnaire, they investigated it. Thus, this might have led to medium ratings with respect to the comprehensibility. However, the participants gave high ratings for the overall usefulness of this additional highlighting. For Biodiv 2, the users pointed out more than twice as often as for Biodiv 1 that they would need more information on the presented information of the biological entities. This also correlates with our observations. In some cases, the biological entity function took some time to deliver a result (as it was a request to the Semantic Assistant service calling various NLP pipelines) or not all expected terms were highlighted. In these cases, users looked up the terms in Google[29] or Wikipedia[30] (which was permitted), to obtain more information. With respect to the usage of the query explanation tab, we noticed less usage, too. Thus, the helpfulness obtained medium ratings (see figure on GitHub). Concerning the comprehensibility, users gave a little more preference to the simpler query explanations without URI information in Biodiv 1.

[29] Google, `https://www.google.de/`

[30] Wikipedia, `https://de.wikipedia.org`

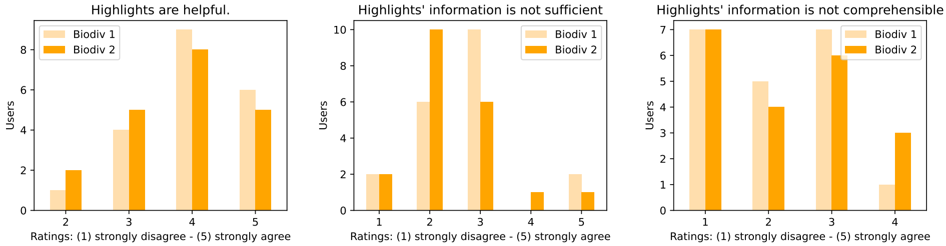

Fig. 5: SUS statements on the default highlighting of query terms in bold font

In comparision to other user studies [EWC12, Ve16], our results show that overall both interfaces are suitable for a semantic dataset search in biodiversity research. A user interface with a single input field and automatic category detection is more suitable when users have no expertise in the search topic. In order to get more insights on daily usage of the systems with the user's own search tasks, further long-term studies are needed. The users appreciated highlightings and explanations and were not confused about information on terminologies and URIs. Only with respect to query information, users preferred the simple explanations without information on the linked ontology. Concerning further improvements of the user interface, the participants suggested to integrate facets and filters to narrow down the results, e.g., to sort the data along metadata fields such as data repository or data type. This confirms the RDA recommendations [Wu19] that users need different entry points and filtering options in dataset search.

5 Conclusion

In this study, we proposed an improved user interface for a semantic search over datasets in the Life Sciences. We conducted a usability study of this novel system with two user interfaces, with a particular focus on query inputs and different explanation strategies. Our results reveal that both interfaces are suitable for semantic dataset search. For users that are not familiar with a search topic, a user interface with a single input field is slightly more efficient. Details about utilized ontologies and URIs are helpful and desired. More studies in other applied domains are needed to examine whether this outcome can be generalized. In addition, long-term studies would be benefical to study semantic search in daily usage.

Acknowledgement

We would like to thank all participants in this study for their time and valuable feedback.

Bibliography

[Al18] Allot, Alexis; Peng, Yifan; Wei, Chih-Hsuan; Lee, Kyubum; Phan, Lon; Lu, Zhiyong: LitVar: a semantic search engine for linking genomic variant data in PubMed and PMC. Nucleic Acids Research, 46(W1):W530–W536, 2018.

[BBH16] Bast, Hannah; Buchhold, Björn; Haussmann, Elmar: Semantic Search on Text and Knowledge Bases. Foundations and Trends® in Information Retrieval, 10(2-3):119–271, 2016.

[Br96] Brooke, John: SUS - A quick and dirty usability scale. In (Jordan, P.W.; Thomas, B.; McClelland, I.L.; Weerdmeester, B., eds): Usability Evaluation In Industry. CRC Press, chapter SUS - A quick and dirty usability scale, 1996.

[Ch18] Chen, Xiaoling; Gururaj, Anupama E; Ozyurt, Burak; Liu, Ruiling; Soysal, Ergin; Cohen, Trevor; Tiryaki, Firat; Li, Yueling; Zong, Nansu; Jiang, Min; Rogith, Deevakar; Salimi, Mandana; Kim, Hyeon-eui; Rocca-Serra, Philippe; Gonzalez-Beltran, Alejandra; Farcas, Claudiu; Johnson, Todd; Margolis, Ron; Alter, George; Sansone, Susanna-Assunta; Fore, Ian M; Ohno-Machado, Lucila; Grethe, Jeffrey S; Xu, Hua: DataMed – an open source discovery index for finding biomedical datasets. Journal of the American Medical Informatics Association, 25(3):300–308, 2018.

[Cu13] Cunningham, Hamish; Tablan, Valentin; Roberts, Angus; Bontcheva, Kalina: Getting More Out of Biomedical Documents with GATE's Full Lifecycle Open Source Text Analytics. PLoS Computational Biology, 9(2):e1002854–e1002854, 2013.

[Cu18] Culina, Antica; Baglioni, Miriam; Crowther, Tom W.; Visser, Marcel E.; Woutersen-Windhouwer, Saskia; Manghi, Paolo: Navigating the unfolding open data landscape in ecology and evolution. Nature Ecology & Evolution, 2(3):420–426, 2018.

[Di17] Dixit, Ram; Rogith, Deevakar; Narayana, Vidya; Salimi, Mandana; Gururaj, Anupama; Ohno-Machado, Lucila; Xu, Hua; Johnson, Todd R: User needs analysis and usability assessment of DataMed – a biomedical data discovery index. Journal of the American Medical Informatics Association, 25(3):337–344, 2017.

[Du05] Dumais, Susan: The Interactive TREC Track: Putting the User Into Search. MIT Press, 2005.

[EWC12] Elbedweihy, Khadija; Wrigley, Stuart N.; Ciravegna, Fabio: Evaluating Semantic Search Query Approaches with Expert and Casual Users. In (Cudré-Mauroux, Philippe; Heflin, Jeff; Sirin, Evren; Tudorache, Tania; Euzenat, Jérôme; Hauswirth, Manfred; Parreira, Josiane Xavier; Hendler, Jim; Schreiber, Guus; Bernstein, Abraham; Blomqvist, Eva, eds): The Semantic Web – ISWC 2012. Springer Berlin Heidelberg, pp. 274–286, 2012.

[Gr20] Gregory, Kathleen; Groth, Paul; Scharnhorst, Andrea; Wyatt, Sally: Lost or Found? Discovering Data Needed for Research. Harvard Data Science Review, 4 2020. https://doi.org/10.1162/99608f92.e38165eb.

[Ka14] Karam, Naouel.; Fichtmüller, David.; Gleisberg, Maren.; Becker, Florian.; Tolksdorf, Robert.; Müller-Birn, Claudia.; Paschke, Adrian.; Güntsch, Anton. (Eds.): , The Terminology Service of the German Federation for Biological Data (GFBio) - Service of semantic technologies in scientific environments. http://terminologies.gfbio.org/ - [Accessed 2023-01-22], 2014.

[Ka20] Kalantari, Mohsen; Syahrudin, Syahrudin; Rajabifard, Abbas; Subagyo, Hardi; Hubbard, Hannah: Spatial Metadata Usability Evaluation. ISPRS International Journal of Geo-Information, 9(7), 2020.

[Ka21] Kalantari, Mohsen; Syahrudin, Syahrudin; Rajabifard, Abbas; Hubbard, Hannah: Synchronising Spatial Metadata Records and Interfaces to Improve the Usability of Metadata Systems. ISPRS International Journal of Geo-Information, 10(6), 2021.

[KB07] Kaufmann, Esther; Bernstein, Abraham: How Useful Are Natural Language Interfaces to the Semantic Web for Casual End-Users? In (Aberer, Karl; Choi, Key-Sun; Noy, Natasha; Allemang, Dean; Lee, Kyung-Il; Nixon, Lyndon; Golbeck, Jennifer; Mika, Peter; Maynard, Diana; Mizoguchi, Riichiro; Schreiber, Guus; Cudré-Mauroux, Philippe, eds): The Semantic Web. Springer Berlin Heidelberg, pp. 281–294, 2007.

[KCW18] Khalsa, SiriJodha; Cotroneo, Peter; Wu, Mingfang: A survey of current practices in data search services. Technical report, Research Data Alliance Data (RDA) Discovery Paradigms Interest Group, 2018.

[LK16] Löffler, Felicitas; Klan, Friederike: Does Term Expansion Matter for the Retrieval of Biodiversity Data? In (Martin, Michael; Cuquet, Martí; Folmer, Erwin, eds): Joint Proceedings of the Posters and Demos Track of the 12th International Conference on Semantic Systems - SEMANTiCS2016 and the 1st International Workshop on Semantic Change & Evolving Semantics (SuCCESS'16), co-located with the 12th International Conference on Semantic Systems (SEMANTiCS 2016). CEUR Workshop Proceedings, 2016.

[LLW15] Liu, Yifeng; Liang, Yongjie; Wishart, David: PolySearch2: a significantly improved text-mining system for discovering associations between human diseases, genes, drugs, metabolites, toxins and more. Nucleic acids research, 43:W535–42, 2015.

[Lo10] Loreau, Michel: Excellence in Ecology: Book 17, The Challenges of Biodiversity Science. International Ecology Institute, Oldendorf, Germany, 2010.

[Lö17] Löffler, Felicitas; Opasjumruskit, Kobkaew; Karam, Naouel; Fichtmüller, David; Schindler, Uwe; Klan, Friederike; Müller-Birn, Claudia; Diepenbroek, Michael: Honey Bee Versus Apis Mellifera: A Semantic Search for Biological Data. In (Blomqvist, Eva; Hose, Katja; Paulheim, Heiko; Ławrynowicz, Agnieszka; Ciravegna, Fabio; Hartig, Olaf, eds): The Semantic Web: ESWC 2017 Satellite Events: Portorož, Slovenia. Springer International Publishing, pp. 98–103, 2017.

[Lö20] Löffler, Felicitas; Abdelmageed, Nora; Babalou, Samira; Kaur, Pawandeep; König-Ries, Birgitta: Tag Me If You Can! Semantic Annotation of Biodiversity Metadata with the QEMP Corpus and the BiodivTagger. In: Proceedings of The 12th Language Resources and Evaluation Conference. European Language Resources Association, pp. 4557–4564, 2020.

[Lö21] Löffler, Felicitas; Wesp, Valentin; König-Ries, Birgitta; Klan, Friederike: Dataset search in biodiversity research: Do metadata in data repositories reflect scholarly information needs? PLOS ONE, 16(3):1–36, 2021.

[Lö22] Löffler, Felicitas; Shafiei, Fateme; Witte, René; König-Ries, Birgitta; Klan, Friederike: , [Dataset] Supplementary material for a usability evaluation of a semantic search for biological datasets, `https://doi.org/10.5281/zenodo.7388037`, 2022.

[MM15] Megler, Veronica M.; Maier, David: Are Data Sets Like Documents?: Evaluating Similarity-Based Ranked Search over Scientific Data. TKDE: Transactions on Knowledge and Data Engineering, 27(1), 2015.

[Mu17] Mueller, Bernd; Poley, Christoph; Pössel, Jana; Hagelstein, Alexandra; Gübitz, Thomas: LIVIVO - the Vertical Search Engine for Life Sciences. Datenbank-Spektrum, 17(1):29–34, 2017.

[Na11] Naderi, Nona; Kappler, Thomas; Baker, Christopher J. O.; Witte, René: OrganismTagger: detection, normalization and grounding of organism entities in biomedical documents. Bioinformatics, 27(19):2721–2729, 2011.

[Ni93] Nielsen, Jakob: Chapter 6 - Usability Testing. In (NIELSEN, JAKOB, ed.): Usability Engineering, pp. 165–206. Morgan Kaufmann, 1993.

[Pa16] Parker, Timothy H.; Forstmeier, Wolfgang; Koricheva, Julia; Fidler, Fiona; Hadfield, Jarrod D.; Chee, Yung En; Kelly, Clint D.; Gurevitch, Jessica; Nakagawa, Shinichi: Transparency in Ecology and Evolution: Real Problems, Real Solutions. Trends in Ecology & Evolution, 31(9):711 – 719, 2016.

[Pa21] Pachzelt, Adrian; Kasperek, Gerwin; Lücking, Andy; Abrami, Giuseppe; Driller, Christine: Semantic Search in Legacy Biodiversity Literature: Integrating data from different data infrastructures. Biodiversity Information Science and Standards, 5:e74251, 2021.

[RC08] Rubin, Jeffrey; Chisnell, Dana: Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests, 2nd Edition. Wiley, 2008.

[Sh21] Shafiei, Fateme; Löffler, Felicitas; Thiel, Sven; Opasjumruskit, Kobkaew; Grabiger, Denis; Rauh, Pauline; König-Ries, Birgitta: [Dai:Si] - A Modular Dataset Retrieval Framework with a Semantic Search for Biological Data. In (Sanfilippo, Emilio M.; Kutz, Oliver; Troquard, Nicolas; Hahmann, Torsten; Masolo, Claudio; Hoehndorf, Robert; Vita, Randi; Algergawy, Alsayed; Karam, Naouel; Klan, Friederike; Michel, Franck; Rosati, Ilaria, eds): S4BioDiv 2021: 3rd International Workshop on Semantics for Biodiversity, held at JOWO 2021: Episode VII The BolzanoSummer of Knowledge, September 11–18, 2021, Bolzano, Italy. 2021.

[SPA18] Soto, Axel J; Przybyła, Piotr; Ananiadou, Sophia: Thalia: semantic search engine for biomedical abstracts. Bioinformatics, 35(10):1799–1801, 2018.

[Ve16] Vega-Gorgojo, Guillermo; Slaughter, Laura; Giese, Martin; Heggestøyl, Simen; Soylu, Ahmet; Waaler, Arild: Visual query interfaces for semantic datasets: An evaluation study. Journal of Web Semantics, 39:81–96, 2016.

[Vo15] Volentine, Rachel; Owens, Amber; Tenopir, Carol; Frame, Mike: Usability Testing to Improve Research Data Services. Qualitative and Quantitative Methods in Libraries, 4(1):59–68, 2015.

[WG08] Witte, René; Gitzinger, Thomas: Semantic Assistants – User-Centric Natural Language Processing Services for Desktop Clients. In (Domingue, John; Anutariya, Chutiporn, eds): The Semantic Web. Springer Berlin Heidelberg, pp. 360–374, 2008.

[Wu19] Wu, Mingfang; Psomopoulos, Fotis; Khalsa, Siri Jodha; de Waard, Anita: Data Discovery Paradigms: User Requirements and Recommendations for Data Repositories. Data Science Journal, 18(1):3, 2019.