A Semantic Search Engine for the Exploration of Explainability Phenomena

Meisam Booshehri[1,2,*], Christian Kullik[1,2] and Philipp Cimiano[1,2]

[1]Semantic Computing Group, Faculty of Technology, Bielefeld University, Bielefeld, Germany
[2]Transregional Collaborative Research Center 318 "Constructing Explainability", Paderborn and Bielefeld, Germany

Abstract

In this paper, we present a novel semantic search engine designed to facilitate the exploration of explainability phenomena in various contexts. The search engine is built upon an ontology specifically developed to capture explanatory moves, enabling the semantical indexing and retrieval of related concepts. This ontology integrates multiple existing annotation schemas, mapping them into a unified framework that describes diverse explanatory behaviors. We detail the development and implementation of both the search engine and the ontology, outlining their architecture and the methodology behind the integration of annotation schemas. Additionally, we present the results of a usability evaluation that assesses the effectiveness and user experience of the search engine. Our findings provide insights into the potential of semantic search for advancing the study and understanding of explainability phenomena.

Keywords

eXplainable AI, Semantic Search, Explanations, Ontology design

1. Introduction

Recently, there has been a resurgence of interest in eXplainable Artificial Intelligence (XAI)—a subfield of AI focused on "mak[ing] a system's behavior intelligible and thus controllable by humans" [1]. Miller [2] argues that explanations are *social*, selective, contrastive, and causal rather than statistical, all converging on the point that explanations are *contexual*. He further argues that for making truly explainable AI systems that offer explanations, one must conform to these key desiderata. Yet, as surveyed by Alpsancar et al. [3], it becomes evident that the XAI community tends to abstract away from the social and political embeddedness of AI systems. Thommes [4] also discusses how failing to capture the diversity of contextual factors limits the utility of explainable systems for users– e.g., due to failing to consider the explanatory needs of users by being focused on an one-fits-all explanation. On this basis, Rohlfing et al. [1, 5] have revisited the notion of explainability, presenting *explanation as a social practice* where explainers and explainees *co-construct* an explanation. The explainer provides an explanation that takes the explainee's knowledge into account [6]. The explainee, on the other hand, attends to the explanation, evaluates whether it is suitable, and provides feedback to the explainer[6]. Accordingly, we would like to stress that the design of XAI systems can be informed by a better understanding of how humans explain to each other. We have therefore developed a semantic search engine which allows to explore human-human explanatory dialogues collected and manually annotated using annotation schemes compiled by the research projects within the scope of the Collaborative Research Center "Constructing Explainability" at the Universities of Bielefeld and Paderborn (TRR 318). We integrate these annotation schemes, mapping them into a unified framework, which describes diverse explanatory behaviors. In the rest of this paper, we provide an overview of the proposed system (Section 2), present how we integrate the data from multiple projects (Section 3), and report on a usability evaluation of the system (Section 4). Section 5 follows with concluding remarks.

SEMANTiCS'25: International Conference on Semantic Systems, September 3–5, 2025, Vienna, Austria
*Corresponding author.
mbooshehri@techfak.uni-bielefeld.de (M. Booshehri); ckullik@techfak.uni-bielefeld.de (C. Kullik); cimiano@cit-ec.uni-bielefeld.de (P. Cimiano)
0009-0008-4895-2699 (M. Booshehri); 0009-0001-7197-4801 (C. Kullik); 0000-0002-4771-441X (P. Cimiano)
© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

CEUR-WS.org/Vol-4064/PD-paper15.pdf

2. System Overview

Figure 1 shows the search engine interface with numbered sections; for example, number 1 is the **facetted search** panel, which allows users to narrow down results by applying multiple filters (or "facets"). Selecting a category (e.g., *INF:assert*) retrieves utterances annotated with it, as well as those utterances annotated with categories mapped to *INF:assert*—e.g., *A01:declarative* and *A02:assert* as close matches, and *A01:additional_info* and *A02:statement* as broader classes. However, if a user selects categories located lower in the taxonomy, the search results are further narrowed to those more specific categories. Figure 2 shows the software architecture of the search engine. It consists of two Docker containers: one for the application and another for the database. The application container includes a frontend built with *Bootstrap* and *JavaScript*, and a backend using an *express.js* server. Users interact with the frontend to submit queries, which the backend processes and translates into SPARQL queries. These SPARQL queries are sent to a Fuseki server housed in the second Docker container, where the RDF triples are stored. The Fuseki server executes the queries and returns the resulting annotation data. This data is then passed back through the backend to the frontend, where it is displayed to the user.

Use case: To illustrate the application of this framework, we present a simple use case from the perspective of target users. *Albert is a first year PhD student in linguistics. He is interested in understanding displays of understanding in human-human explanatory dialogues. He knows that there is this group of datasets created by TRR 318 that has some exploration interface. He looks at the tree of categories and makes an educated guess: completion could be related to display of understanding. He can now conveniently browse through utterances annotated with the category completion across multiple datasets [even though not in every dataset the category "completion" was used, but other categories known to be similar to completion]. He checks a couple of these utterances and builds the hypothesis that completion could actually be related to display of understanding.* In such a scenario the interface has helped Albert in several ways: (1) It helped him find out that these datasets are relevant for his research, and (2) that 'completion' is related to display of understanding. (3) He has looked at a couple of utterances and now understands better what kind of analysis he could do next. (4) The facetted search in the interface has helped him in building a hypothesis relevant to his research. An alternative would be the following: Albert has access to a database containing those datasets. Yet, doing all of this just using a query language would have been tedious, and would have required him to understand the structure of the data in the databases and how to write the necessary queries.

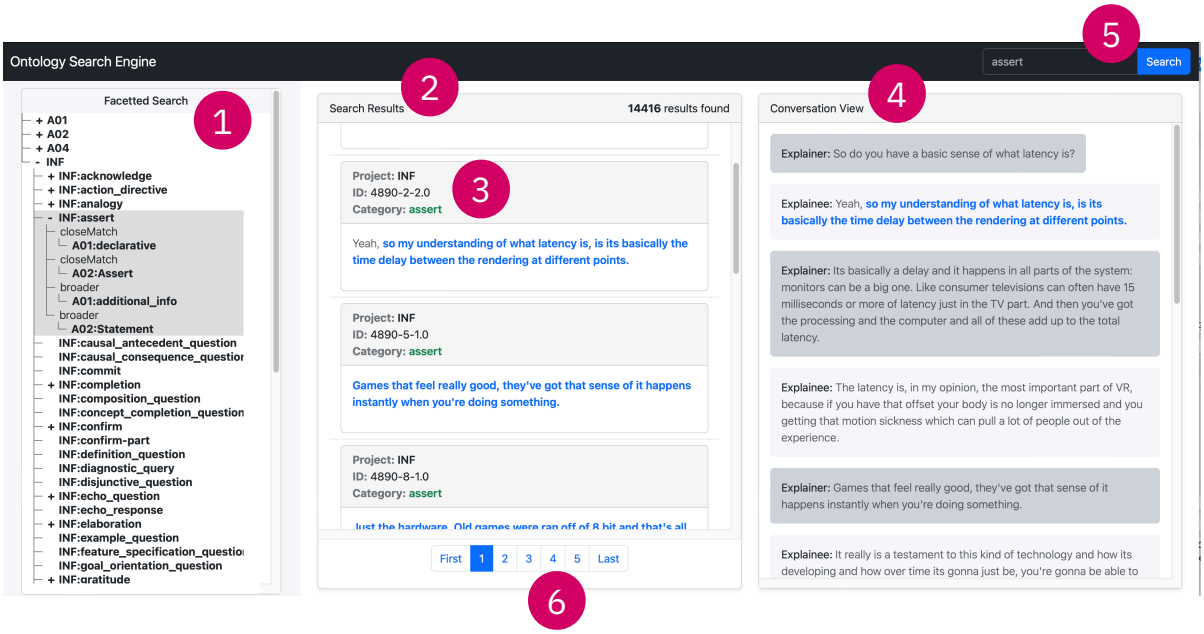


Figure 1: Search Engine Interface; Available at: https://ag-sc.techfak.uni-bielefeld.de/sems/

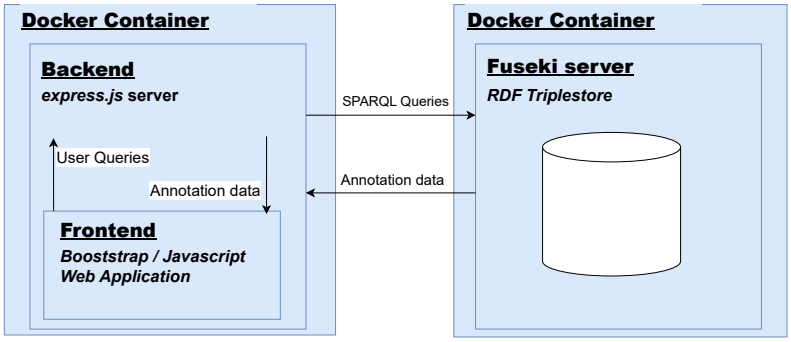


Figure 2: Search Engine Architecture

3. Data Integration

What we present here is part of an ongoing larger ontology development process working towards an ontology of explanations with applications to Social XAI (see Booshehri et al. [7, 6]). Here we discuss how we have provided a unified taxonomy of all existing annotation schemes used in the TRR 318 projects. First, in the INF project, we initiated with a lexicon of explanatory moves—speech act-like categories enacted by explainers and explainees in explanatory dialogues, such as *assert*, *confirm*, *acknowledge*, etc. The list of explanatory moves was compiled from the literature in fields of tutoring and computational linguistics, and was subsequently used to annotate various datasets (e.g., see [8]). Next, we held ontology developer meetings with all projects within TRR 318 that are developing an annotation schema or categorization (i.e., projects A01, A02, and A04). In these meetings, we discussed how existing categories from the individual annotation schemes could be mapped to the INF vocabulary. In this way, the INF vocabulary serves as an *interlingua* [9], to which the vocabularies of other projects are mapped. This interlingua also serves as the default entry point for exploration in the faceted search panel of the search engine. Each annotation scheme was treated as a separate namespace, and SKOS vocabulary was employed to establish semantic connections between categories, as in:

```
INF:assert skos:broader A01:additional_info ;
           skos:closeMatch a01_rhetorical:declarative
```

Table 1 summarizes the mapping statistics: the projects collectively define 122 categories, with 68.03% included in or mapped to the INF vocabulary.

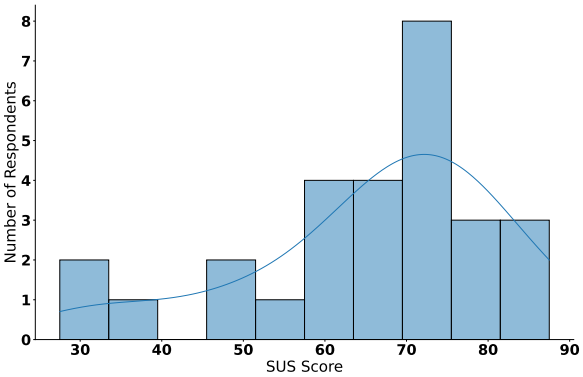
Project names	INF	A2	A1	A4	sum
Total number of categories in annotation schemes	55	28	32	7	122
No. of categories linked to INF	55	11	17	0	83
Percentage of categories linked to INF	100%	39.29%	53.13%	0%	**68.03%**

Table 1
Linking of annotation scheme concepts from different TRR 318 projects to the INF vocabulary.

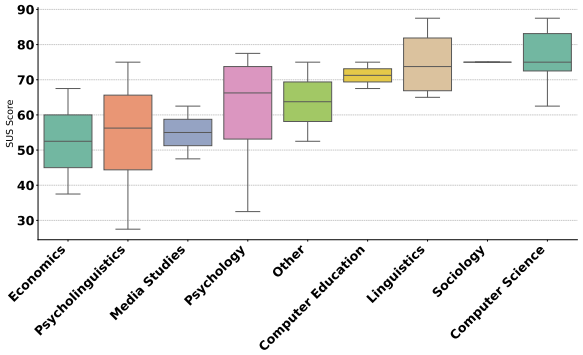
4. Evaluation

We evaluated the proposed system using a System Usability Scale (SUS) questionnaire [10]. A total of 28 participants from different backgrounds (see Figure 3c) participated in the usability study. The SUS scores from 28 participants showed a mean of 66.07 (SD = 15.58), with a median of 70.00, a 25th percentile of 61.88, and a 75th percentile of 75.00. The SUS scores are skewed towards the higher end of the scale (see Figure 3a), indicating that users generally perceive the software system as *acceptable* or of

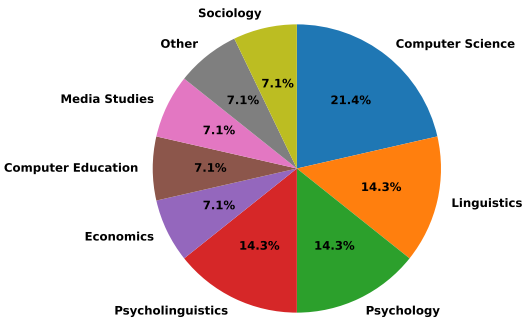
good quality (see Figure 3 in Bangor et al. [10] for the adjective rating scale). This reflects a positive user experience and suggests that the system is, overall, well-received. However, a small number of lower scores point to some usability challenges. These could be attributed to the interdisciplinary context of current study, which can lead to varying—and at times conflicting—expectations about how the system should look like or function.



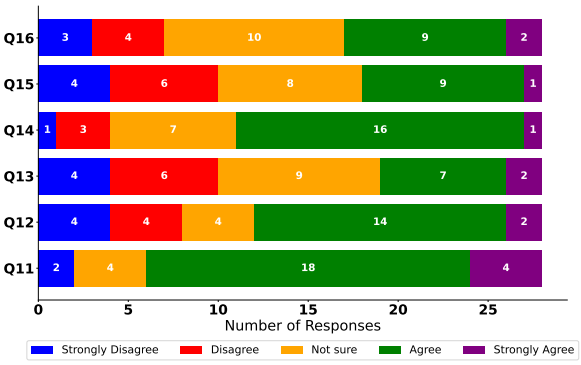
(a) Distribution of SUS scores



(b) Distribution of SUS scores by Research Area



(c) Distribution of Main Area of Research for Participants in usability evaluation



(d) Distribution of extended Survey Questions (Q11-Q16)

Figure 3: Distribution of SUS scores, the extended survey questions, and main areas of research

Category	Questions
Interface	**Q11**. The organization and display of search results are clear **Q12**. The browsing of data using the taxonomy is intuitive
Function of ontology as interlingua	**Q13**. The taxonomy facilitates the understanding of relationships between concepts/categories from different TRR projects **Q14**. The taxonomy or the search engine makes it easy to search for the data for the key phenomena across TRR projects
Use of Search Engine	**Q15**. I see how the exploration of data in the search engine could help me in new hypothesis formation **Q16**. The search engine helps me in better understanding of explanatory interactions at the micro level

Table 2
Further Survey Questions in addition to SUS Questions

In addition to the original 10 items of the standard SUS, we included six additional questions in the questionnaire (**Q11-Q16**; see Table 2), focusing on three specific aspects: the system interface, the function of the ontology as an interlingua, and the use of the search engine. Regarding the interface of the search engine (Figure 3d; **Q11-Q12**), most participants found the organization and display of search results clear, and about half considered browsing via the taxonomy intuitive. On the ontology's role as an interlingua (Figure 3d; **Q13-Q14**), responses were mixed: roughly one-third found the taxonomy helpful for understanding cross-project relationships, while others were either unsure or disagreed.

Still, a majority agreed it facilitates searching for data or key phenomena across TRR 318 projects. As for the system's potential to support research activities (Figure 3d; **Q15**-**Q16**), over 30% believed it could help form new hypotheses and improve understanding of explanatory interactions at micro-level, though uncertainty remained high.

5. Conclusion

We have developed a semantic search engine for exploring the behavior of interlocutors in human-human explanations, with the first experimental results showing an overall positive user experience. We hope this framework can support research activities in the realm of XAI by providing insights into the contextual factors of explanations. As future work, we plan to add more features (e.g., search for sequences of explanatory moves) to the search engine and enrich it with additional annotated data.

Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): TRR 318/1 2021 – 438445824, Project INF ("Toward a framework for assessing explanation quality"). We further thank Basil Ell for brainstorming ideas about framing the use-case scenario.

Declaration of Generative AI Use

During the preparation of this work, GPT-4.5 ChatGPT was used in order to: Paraphrase and reword. The authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] K. Rohlfing, P. Cimiano, I. Scharlau, et al., Explanation as a social practice: Toward a conceptual framework for the social design of ai systems, IEEE Transactions on Cognitive and Developmental Systems 13 (2021) 717–728. doi:10.1109/TCDS.2020.3044366.

[2] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, Artificial Intelligence 267 (2019) 1–38. doi:10.1016/j.artint.2018.07.007.

[3] S. Alpsancar, H. M. Buhl, T. Matzner, I. Scharlau, Explanation needs and ethical demands: unpacking the instrumental value of XAI, AI and Ethics 5 (2025) 3015–3033. URL: https://link.springer.com/10.1007/s43681-024-00622-3.

[4] K. Thommes, Evaluation principles, in: [5], 2025.

[5] K. Rohlfing, K. Främling, B. Lim, K. Thommes, S. Alpsancar (Eds.), Social Explainable AI, Communications of NII Shonan Meeting, Springer, Singapore, 2025.

[6] M. Booshehri, H. Buschmeier, P. Cimiano, A BFO-based ontological analysis of entities in Social XAI, in: Proc. of the 15th Int. Conf. on Formal Ontology in Information Systems, IOS, Catania, Italy, 2025.

[7] M. Booshehri, H. Buschmeier, P. Cimiano, Towards a BFO-based ontology of understanding in explanatory interactions, in: Proc. of the 4th Int. Workshop on Data Meets Applied Ontologies in Explainable AI (DAO-XAI), CEUR-EU, Santiago de Compostela, Spain, 2024. URL: https://ceur-ws.org/Vol-3833/paper3.pdf.

[8] M. Booshehri, H. Buschmeier, P. Cimiano, A model of factors contributing to the success of dialogical explanations, in: H. Hung, C. Oertel, M. Soleymani, T. Chaspari, H. Dibeklioglu, J. Shukla, K. P. Truong (Eds.), Proceedings of the 26th International Conference on Multimodal Interaction (ICMI), ACM, 2024, pp. 373–381. URL: https://doi.org/10.1145/3678957.3685744.

[9] E. Hovy, S. Nirenburg, Approximating an interlingua in a principled way, in: Proceedings of the Workshop on Speech and Natural Language, HLT '91, Association for Computational Linguistics, USA, 1992, p. 261–266. URL: https://doi.org/10.3115/1075527.1075588.
[10] A. Bangor, P. Kortum, J. Miller, Determining what individual sus scores mean: adding an adjective rating scale, J. Usability Studies 4 (2009) 114–123.