Semantic Tool Hub: Towards A Sustainable Community-Driven Documentation of Semantic Web Tools

Achim Reiz[1][0000-0003-1446-9670], Fajar J. Ekaputra[2][0000-0003-4569-2496], and Nandana Mihindukulasooriya[3][0000-0003-1707-4842]

[1] Universität Rostock, Rostock, Germany
[2] WU Vienna, Vienna, Austria
[3] IBM Research AI, New York, USA

Abstract. The semantic web community has developed and still is developing a tremendous number of tools and software. While the activity underlines the continuous importance of the field and the commitment of its members, it also poses a challenge, especially for people entering the field. Identifying the right tools for one's use case is increasingly difficult. A lot of software is no longer actively maintained, and going through all publications and source repositories to find the software with the proper set of functions is tedious. In this demo, we propose a workflow and an initial Wikidata-based toolkit to support knowledge engineers and developers in documenting and finding the right tools. We categorized existing tools into a pre-defined taxonomy and integrated them with GitHub metadata about their recent developments when applicable. The condensed information with the new taxonomy is integrated into Wikidata, ready for further use.

Keywords: Semantic Web, Knowledge Graph, Software, Tools

1 Introduction

From 2010 to 2023, more than 2,400 papers were presented and published alone at the Extended and International Semantic Web Conferences (e.g., ESWC and ISWC). Moreover, there are many more additional journals, conferences, and workshops (e.g., SWJ, KEOD, LDAC). We develop countless knowledge graph construction, querying, and storing approaches, often supported by or implemented in software. While that is, in some respect, a sign of a healthy research community, it also poses a challenge, especially for people entering the field.

Currently, documenting and searching for the right Semantic Web (SW) software for a given practical problem is tedious: (1.) There is no common repository for documenting SW software (for example, similar to LOV for vocabularies), and relevant information is scattered throughout the various research outlets. (2.) There is no standardized semantics and taxonomy to describe the SW tools. Additionally, many tools and frameworks cover more than one element of the knowledge graph (KG) development toolchain. For example, Apache Jena is mainly regarded as a Java framework. However, it also integrates a triple store with a reasoning engine and allows the validation of incoming data for conformance based on SHACL shapes.

Many SW researchers and practitioners acknowledge these challenges and try to address them in many forms. The Awesome Semantic Web initiative[4] provides a GitHub repository to collect and report SW tools. The page contains more than 100 software tools related to SW. However, it contains many outdated software and does not provide information beyond categories and descriptions. A recent technical report from the OntoCommons [1] collects a collection of tools metadata, including description, homepage, code repository, documentation page, and related publications from more than 60 tools based on a survey. The report, however, is only available as a PDF and not as a machine-readable resource. Existing Knowledge Graphs, such as Wikidata and DBPedia, contain information about traditional and popular software, such as Protégé [5]. However, these pages typically are missing for newer or less popular software –albeit potentially of similar or higher importance – such as Chowlk [2] for ontology creation or Widoco [3] for ontology documentation. Furthermore, there is currently no dedicated visualization page to render their sub-graphs on SW software.

The Semantic Tool Hub targets to ease these challenges: it aims to bring the scattered knowledge on the SW software tools into Wikidata –an open and community-driven Knowledge Graph– according to a predefined taxonomy representing the semantic artifact development process.

The data is further enriched with metadata from their GitHub repositories (if applicable) to identify recent activity. The Semantic Tool Hub is meant to strengthen the application of FAIR principles [4] for semantic web research by making the tools **f**indable, **a**ccessible at a centralized, open location, **i**nteroperable through standardized semantic web protocols and the Wikidata vocabulary and **r**eusable by stating the license and source code, if applicable.

In the rest of the paper, we will describe the proposed methodology and our initial toolkit to support the documentation and retrieval process of the SW software centered around Wikidata.

2 The Semantic Tool Hub

The main idea behind the Semantic Tool Hub is to develop a Wikidata-based solution for sustainable community-driven documentation and retrieval of SW tools. In this work, we defined our scope only to include tools targeting RDF-based technologies. To this end, we propose a workflow (cf. Fig. 1) consisting of the following steps:

[4] https://github.com/semantalytics/awesome-semantic-web

[5] https://dbpedia.org/page/Protege_(software); https://www.wikidata.org/wiki/Q2066865

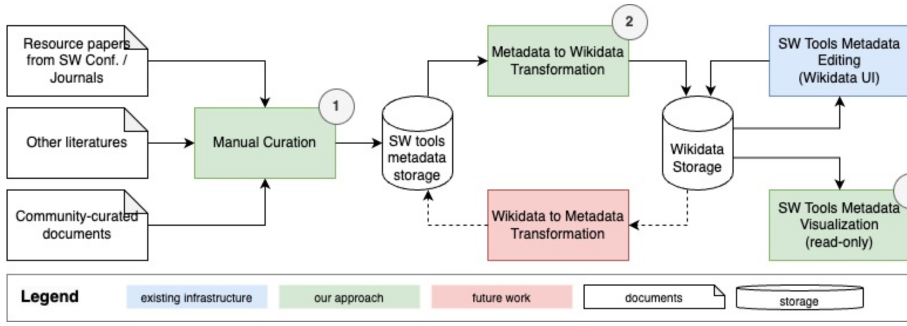

Fig. 1. The Semantic Tool Hub Workflow

Taxonomy Development. At the beginning of our research, we realized a strong need for a solid categorization of the tool. To this end, we combine the existing categorization tool from OntoCommons report [1] and the semantic web lifecycle [5] to develop our SW software taxonomy, shown in Fig. 2. The taxonomy is modeled in Wikidata and available to the full extent in GitHub [6].

Step 1. Manual Data Curation. Next, we collected and manually curated existing literature, both from scientific communities (e.g., SW conferences and journals) and other sources (e.g., awesome SW initiative), to gather metadata about tools. This step aimed to ensure the distributed information about software tools currently scattered among different sources can be collected and structured according to the taxonomy developed previously.

Our initial prototype collects data from three primary sources: (i) recent ISWC/ESWC conferences, (ii) awesome SW initiative, and (iii) OntoCommons report. In total, we have collected almost 150 tools annotated with metadata, including their categorizations. The original annotation information is available as a spreadsheet file[6].

Step 2. Metadata to Wikidata Transformation. We decided to use Wikidata as part of our solution approach due to the flexibility, machine-readability, and nature of crowdsourcing of the Wikidata content development. We believe that our decision will allow for a broader involvement of the community in documenting the available SW tools. Furthermore, it facilitates users with easy access to retrieving and searching for suitable SW tools.

We are currently utilising Open Refine[7] and Wikipedia Quickstatements[8] to transform our spreadsheet data into RDF triples suitable for Wikidata.

Step 3. SW Tools Metadata Visualization Wikidata contains an extensive collection of knowledge on various topics and granularity, which makes it hard for

[6] `https://github.com/semantic-tool-hub/SW-Tool-Hub-data/releases/tag/0.1`
[7] `https://openrefine.org/`
[8] `https://github.com/magnusmanske/quickstatements`

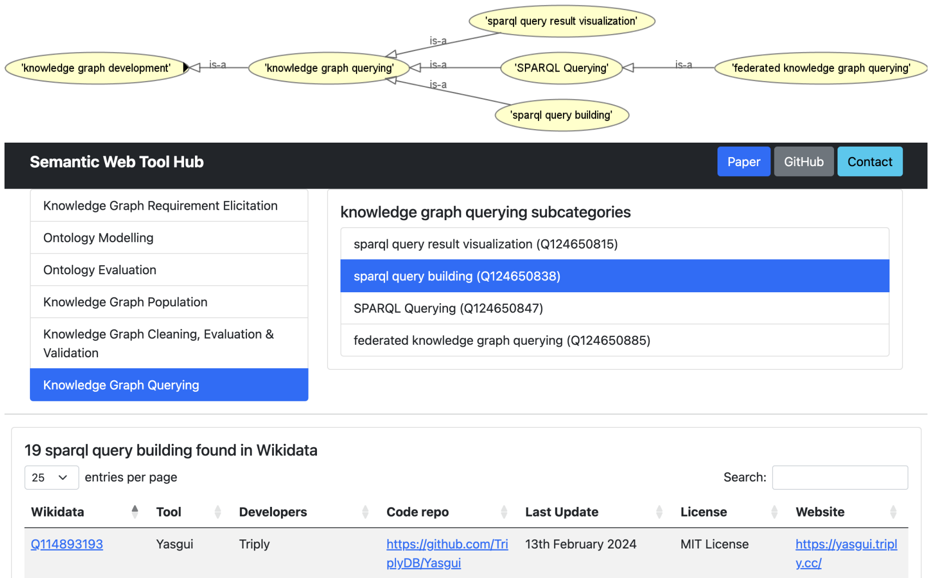

Fig. 2. An excerpt of the Semantic Tool Hub taxonomy (top) & the Semantic Tool Visualization and Search Interface (bottom)

users to browse for relevant information on specific topics quickly. Therefore, a specific interface is needed to help users quickly search for information about specific tools.

In the context of this paper, we have developed a webpage[9] to visualize and search/retrieve the knowledge that we have developed to help users in searching and finding information about SW tools (cf. Fig. 2 bottom).

Step 4. Adding New Data Over Time At the time of publication, the Semantic Tool Hub contains manually curated data collected by the author of this demo. In the future, we hope the semantic web community picks up the idea, and the crowd updates the data collaboratively. Data ingestions and updates can be done directly through the Wikidata interface or using GitHub issues. The GitHub repository contains further information on how to participate[10].

3 Conclusion and Future Work

This paper describes our approach to support a sustainable community-driven documentation and search of SW tools based on Wikidata. We hope the Semantic Tool Hub gets picked up by the semantic web community and establishes itself

[9] `https://semantic-tool-hub.github.io/`
[10] `https://github.com/semantic-tool-hub/SW-Tool-Hub-data`

as an entry point for finding potential software for knowledge graph development efforts.

In the future, we plan to increase the ease of use and means for contributions through a two-way synchronization between Wikidata and the git-based metadata storage. That includes a phase-out of the Excel sheet towards a JSON-based flat-file architecture for managing the various tools, in addition to the existing Wikidata editor UI and GitHub Issues.

References

[1] Martin G. Skjæveland, Laura Ann Slaughter, and Christian Kindermann. *OntoCommons D4.3 - Report on Landscape Analysis of Ontology Engineering Tools*. Apr. 2022. DOI: 10.5281/zenodo.6504670. URL: https://doi.org/10.5281/zenodo.6504670.

[2] Serge Chávez-Feria, Raúl García-Castro, and María Poveda-Villalón. "Chowlk: from UML-based ontology conceptualizations to OWL". In: *European Semantic Web Conference*. Springer. 2022, pp. 338–352.

[3] Daniel Garijo. "WIDOCO: a wizard for documenting ontologies". In: *Proceedings of the 16th International Semantic Web Conference (ISWC2017), Vienna, Austria*. Springer. 2017, pp. 94–102.

[4] Michelle Barker et al. "Introducing the FAIR Principles for research software". en. In: *Scientific Data* 9.1 (Oct. 2022), p. 622. ISSN: 2052-4463. DOI: 10.1038/s41597-022-01710-x. URL: https://www.nature.com/articles/s41597-022-01710-x (visited on 04/11/2024).

[5] Anna Breit et al. "A Lifecycle Framework for Semantic Web Machine Learning Systems". In: *International Conference on Database and Expert Systems Applications*. Springer. 2022, pp. 359–368.