

Subject: [Call For Paper] Scholarly QALD Challenge at ISWC 2023

Question Answering over Scholarly Knowledge Graphs

Web: <https://kgqa.github.io/scholarly-QALD-challenge/2023/>

Task Description

The importance of how typical Web users can access the body of knowledge on the Web grows with the amount of structured data published thereon. Over the past years, there has been an increasing amount of research on interaction paradigms that allow end users to profit from the expressive power of Semantic Web standards while at the same time hiding their complexity behind an intuitive and easy-to-use interface.

However, no ****natural language interface** allows one to access scholarly data** such as papers, authors, institutions, models, or datasets. The key challenge is to translate the users' information needs into a form such that they can be evaluated using standard Semantic Web query processing and inference techniques. Such interfaces would allow users to express arbitrarily complex information needs in an intuitive fashion and, at least in principle, in their own words. The intent of this challenge is thus of great importance and interest to Semantic Web scholars.

By taking part in the challenges sub-tasks, participants will create interfaces to academic and industrial approaches and applications provided by DBLP and ORKG. We will try to bridge the gap between academia and industry to attract junior as well as senior researchers from both worlds leading to a memorable experience. We target to publish system descriptions as proceedings published by CEUR-WS as in past years.

Thus, in its first iteration at ISWC 2023, we have two independent tasks:

- Task 1: SciQA --- Question Answering of Scholarly Knowledge:* This new task introduced this year will use a scholarly data source ORKG (<https://orkg.org/>) dump as a target repository for answering comparative questions.
- Task 2: DBLP-QUAD --- Knowledge Graph Question Answering over DBLP:* For this task, participants will use the DBLP-QUAD dataset, which consists of 10,000 question-SPARQL pairs, and is answerable over the DBLP Knowledge Graph dump released on August 2022.

Important Dates

The tentative timeline for the Scholarly QALD Challenge is as follows. All deadlines are 23:59 AoE (anywhere on Earth).

Papers should be between 5 and 10 pages including references. Papers must be formatted as PDF according to the CEUR-ART format (<https://ceur-ws.org/HOWTOSUBMIT.html>) one-column style. Reviews will be carried out in a single-blind mode. Submission of supplementary source code and data is appreciated. Accepted contributions are planned to

be published in the Open Access CEUR proceedings. At least one author of an accepted paper must register for the workshop to present their work. easychair: <https://easychair.org/my/conference?conf=scholarlyqald23>

Date	Description
2023-06-28	Submission of Systems via HF platform and system description via easychair
2023-07-19	All System Results and Notification of Acceptance
2023-11-06 - 2023-11-10	ISWC Conference, Mandatory Presentation
2023-11-24	Camera-ready submission.
End of 2023	Publication of results.

Datasets

Task 1: SciQA --- Question Answering of Scholarly Knowledge

HF platform link: <https://huggingface.co/datasets/orkg/SciQA>

This new task introduced this year will use a scholarly data source ORKG (<https://orkg.org>) as a target repository for answering comparative questions. KGQA benchmarks and systems were so far mainly geared towards encyclopedic knowledge graphs such as DBpedia and Wikidata.

In this task, we will leverage a novel QA benchmark for scholarly knowledge -- SciQA (<<https://zenodo.org/record/7744048>>), see also <<https://huggingface.co/datasets/orkg/SciQA>>. The benchmark leverages the Open Research Knowledge Graph (ORKG) which includes over 100,000 resources describing complex research contributions. Following a bottom-up methodology, we manually developed a set of 100 questions that can be answered using this knowledge graph. The questions cover a wide range of research fields and question types and are translated into SPARQL queries over the knowledge graph.

The SciQA benchmark represents an extremely challenging task for next-generation QA systems. The 100 hand-crafted questions are significantly more complex to answer than typical common-sense questions.

An example question is: What is the average energy generation for each energy source considered in 5-year intervals in Greenhouse Gas Reduction Scenarios for Germany?

The corresponding SPARQL query includes seven triple patterns, uses eight query components, and is shaped as a tree.

In addition to the 100 hand-crafted questions, we will provide a set of more than 2,000 questions generated from 10 question/query templates to ensure a good balance between question complexity and wider coverage.

Task 2: DBLP-QUAD --- Knowledge Graph Question Answering over DBLP

HF platform link: <https://huggingface.co/datasets/awalesushil/DBLP-QuAD>

For this task, participants will use the DBLP-QUAD dataset (<<https://doi.org/10.5281/zenodo.7554379>>), see also <<https://huggingface.co/datasets/awalesushil/DBLP-QuAD>>, which consists of 10,000 question-SPARQL pairs, and is answerable over the DBLP Knowledge Graph (<https://blog.dblp.org/2022/03/02/dblp-in-rdf/>) and (<https://zenodo.org/record/7638511>).

DBLP is a well-known repository for computer science bibliography and has recently released an RDF dump. This allows users to query it as a knowledge graph. The first subtask is to fetch the right answer from the DBLP KG given the question. The second subtask is entity linking (EL) on the same dataset. The DBLP-QuAD dataset was created using the OVERNIGHT approach, where logical forms are first generated from a KG. Then canonical questions are generated from these logical forms.

Evaluation

For both tasks, we aim to evaluate the participants' approaches using the Hugging Face Evaluate library (<https://huggingface.co/docs/evaluate/index>) through the Hugging Face platform directly. That is, participants upload their models to Hugging Face. The participating systems will be evaluated based on the standard metrics precision, recall, and f-measure.