University of Zurich[UZH]

Dynamic and Distributed Information Systems

Scholarly QALD at ISWC 2023

NLQxform: A Language Model-based Question to SPARQL Transformer

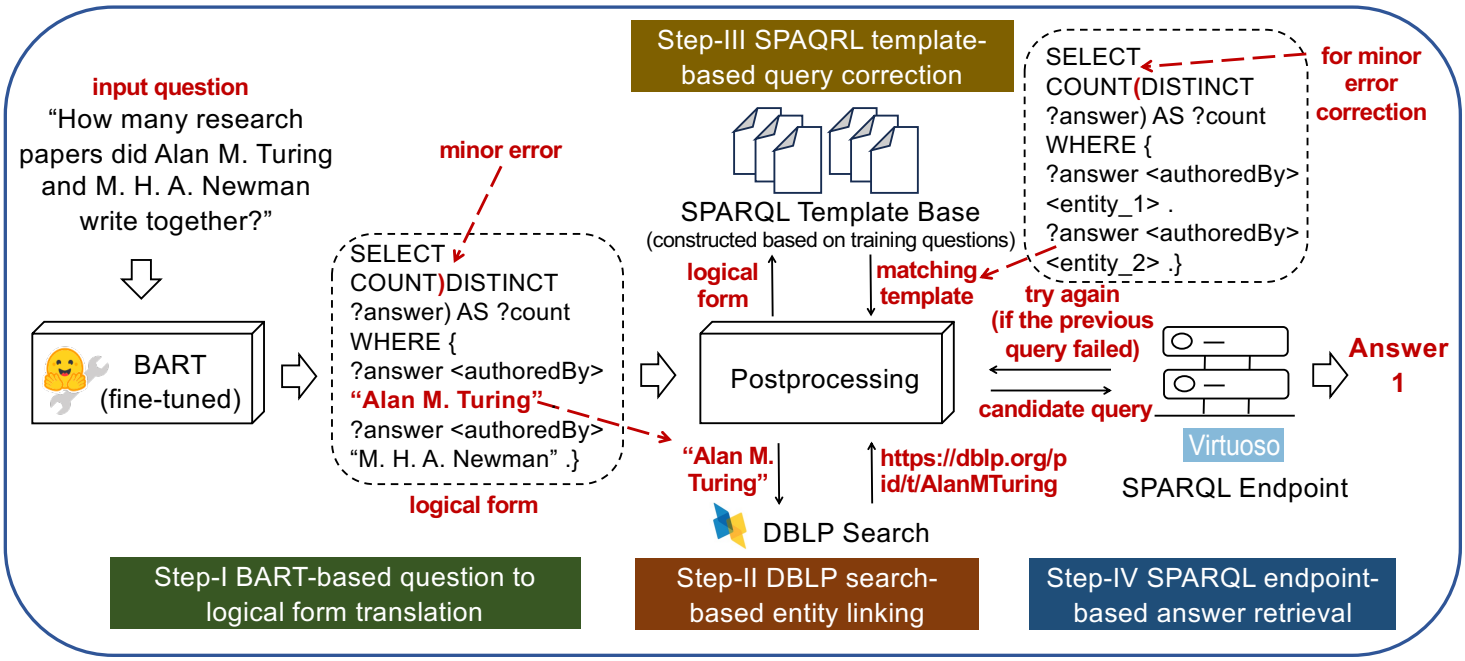
Ruijie Wang, Zhiruo Zhang, Luca Rossetto, Florian Ruosch, Abraham Bernstein

Ruijie Wang
ruijie@ifi.uzh.ch

Introduction

- **Conventional tools for searching scholarly information**, e.g., Google Scholar and DBLP, only support **text search** with limited filtering and sorting. **Scholarly knowledge graphs (KGs)** are more powerful and versatile, as they support **SPARQL queries** with complex constraints and operations. However, scholarly KGs are **significantly less commonly used** than the conventional tools. A main hurdle is the **complexity of the SPARQL language**.
- As part of the **Scholarly QALD Challenge**, this paper presents a **question-answering (QA)** system, called **NLQxform**, which provides an easy-to-use **natural language interface** to facilitate accessing scholarly KGs. NLQxform **fine-tunes** a **transformer-based BART** model and answers given natural language questions by **translating them into executable SPAQRL queries** in four steps, as depicted in the system overview below.

System Overview



Evaluation

Submission	F1 Score (Entity Linking)	F1 Score (Question Answering)
ID-544291	0.8283	0.0000
ID-544863	<u>0.8320</u>	0.0000
ID-557116	**0.8353**	0.0000
ID-545920	0.7100	0.0018
ID-556670	0.6235	0.2175
ID-547129	0.0000	<u>0.6619</u>
ID-557036 (NLQxform)	0.7961	**0.8488**

28.2% ↑

Evaluation results for entity linking and question answering in the Scholarly QALD Challenge **Task 1: DBLP-QuAD — Knowledge Graph Question Answering over DBLP**
(best performance in bold, second best underlined)

- **DBLP-QuAD** as the QA dataset.
- **DBLP Scholarly KG** as the underlying KG.
- NLQxform achieved the **best performance** with a significant improvement over the second-best system on the QA task.

(the website of the challenge)