Rule Driven Query Expansion

Xinze Lyu and Wei Hu

[1] State Key Laboratory for Novel Software Technology, Nanjing University, China
[2] Department of Computer Science and Technology, Nanjing University, China
`xzlv.nju@gmail.com, whu@nju.edu.cn`

Abstract. Empty answers problem exists when we use SPARQL to access RDF knowledge graph data. Put situations that querying facts inexistent to the real world aside, one reason is that users lack enough knowledge for a particular RDF knowledge graph, that leads to SPARQL queries with wrong formats or inaccurate expressions. However, due to the schema-free nature of RDF data and incompleteness of particular RDF knowledge graphs, even a SPARQL query with correct format can reflect users' intentions accurately, it may fail to get any results. Researches going on in translating the natural language to SPARQL help a lot to address the first problem, but these are inconducive for the second problem which was caused by the faults in structure and content of RDF knowledge graphs. We design a rule-driven framework to alleviate the obstacles caused by the structure and content of RDF knowledge graphs. Specifically, given a SPARQL query, we use knowledge graph oriented rule-learning procedure to take reasoning rules, with the help of these rules, our system return possible results. More importantly, our system shows detailed information with similarity score and rules to explain why our system chooses particular possible answers.

Keywords: SPARQL · Empty Answer · Rule Learning.

1 Problem Statement

1.1 Empty Answer Problem for SPARQL query

Users use SPARQL queries reflecting their intentions to access the data from RDF knowledge graph, reasons for Empty Answer Problem are various, three of which are main ones, 1)the facts that users want to query do not exist in the real world; 2)the formats of SPARQL is wrong, including the namespace, operators and so on; 3)SPARQL queries are accurate and well-formatted, but they do not have exact matches in particular RDF knowledge graphs.

We focus on the third one. There are two factors that contribute to the third problems. One is the SPARQL queries have a high level of constraints, we get an instance from the released resource of [1], there are three constraints for SPARQL query in table 1. This query gets no answers, in fact, the constraints (1) and (2) are redundant, we can infer that "?company" is dbr:Apple_Inc with constraints (3).

Table 1. SPARQL query with high constraints.

SELECT ?company WHERE	
(1)	?company rdfs:type dbo:Company.
(2)	?company dbo:industry dbr:Electronics.
(3)	dbr:IPhone dbp:developer ?company.

1.2   Semantic Parsing

1.3   Paraphrasing

1.4   Similarity Based Method

1.5   Ontology Rule Based Method

1.6   Embedding Based Method

Table 2. Table captions should be placed above the tables.

Heading level	Example	Font size and style
Title (centered)	**Lecture Notes**	14 point, bold
1st-level heading	**1 Introduction**	12 point, bold
2nd-level heading	**2.1 Printing Area**	10 point, bold
3rd-level heading	**Run-in Heading in Bold.** Text follows	10 point, bold
4th-level heading	*Lowest Level Heading.* Text follows	10 point, italic

Displayed equations are centered and set on a separate line.

$$x + y = z \tag{1}$$

Please try to avoid rasterized images for line-art diagrams and schemas. Whenever possible, use vector graphics instead (see Fig. 1).

Theorem 1. *This is a sample theorem. The run-in heading is set in bold, while the following text appears in italics. Definitions, lemmas, propositions, and corollaries are styled the same way.*

Proof. Proofs, examples, and remarks have the initial word in italics, while the following text appears in normal font.

For citations of references, we prefer the use of square brackets and consecutive numbers. Citations using labels or the author/year convention are also acceptable. The following bibliography provides a sample reference list with entries for journal articles [**?**], an LNCS chapter [**?**], a book [**?**], proceedings without editors [**?**], and a homepage [**?**]. Multiple citations are grouped [**?,?,?**], [**?,?,?,?**].
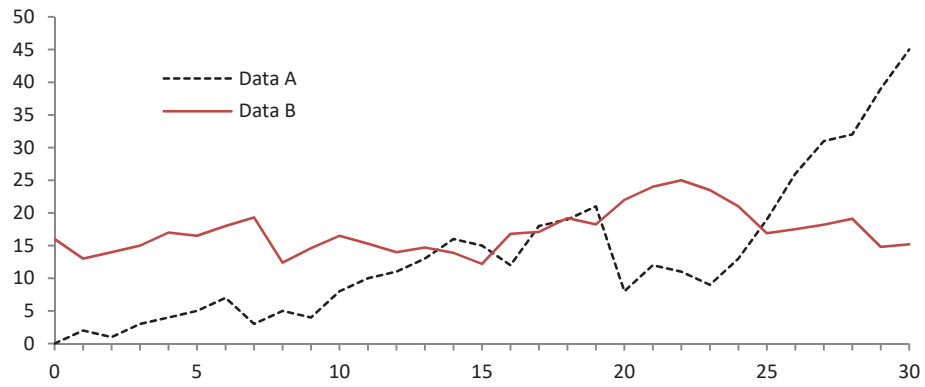
Fig. 1. A figure caption is always placed below the illustration. Please note that short captions are centered, while long ones are justified by the macro package automatically.

References

1. Wang, M., Wang, R., Liu, J., Chen, Y., Zhang, L., Qi, G.: Towards empty answers in sparql: Approximating querying with rdf embedding. In: International Semantic Web Conference. pp. 513–529. Springer (2018)