Federated Query Optimization Method based on SPARQL Endpoint Features for Efficient Retrieval

Soichiro Watanabe
Master's and Doctoral Programs in Informatics,
Graduate School of Comprehensive Human Sciences
University of Tsukuba, Japan
watanabe.soichiro.qa@alumni.tsukuba.ac.jp

Mitsuharu Nagamori
Faculty of Library,
Information and Media Science
University of Tsukuba, Japan
nagamori@slis.tsukuba.ac.jp

Background : Timeouts in SPARQL query execution

LOD facilitates publishing and sharing metadata in machine-readable formats, enabling data interlinking and integration.

Accessing LOD datasets via the SPARQL query language often leads to timeouts.

Primary causes of query timeouts:

A. Volume of stored data
B. Performance of the SPARQL query engine
C. Composition of the SPARQL query

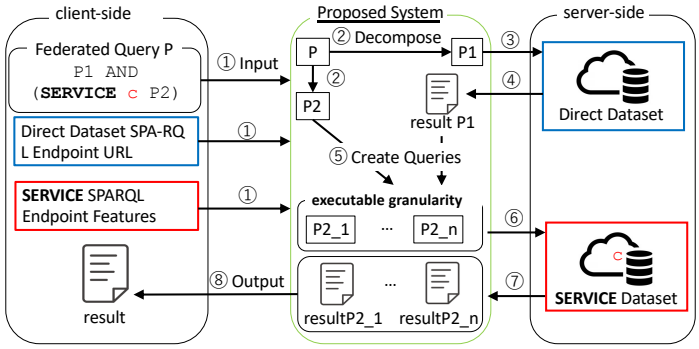
Objective : Retrieving SPARQL results without timeouts

While factors A and B are beyond the user's direct control, this study focuses on understanding and addressing factor C.

Decompose SPARQL queries and rewrite them into queries of executable granularity to mitigate timeouts.

Method: Federated Query Optimization based on SPARQL Endpoint Features for Efficient Retrieval

1. **Iteratively query to gather features for each SPARQL endpoint, specifically:**
   - Supported HTTP request methods (GET or POST)
   - Size constraints for executable SPARQL queries (i.e., HTTP Request Header/Body Size)
2. **Dissect the input Federated Query to align with the previously acquired features:**
   - Utilizing the VALUES clause



Evaluation: Comparison of the three methods using manually and mechanically created queries

- Comparison methods

1. **EXISTING** method
   Execute the federated query as provided.
2. **VALUES** method
   Rewrite the federated query using the VALUES clause.
3. **PROPOSED** method

Accuracy : PROPOSED method is the Best

Run Time : Significantly faster than EXISITING
Almost the same as VALUES

Mechanical Query	number of successes / number of trials, average time(s)		
	1. EXISTING	2. VALUES	3. PROPOSED
assaultlily-rdf	0/10,	0/10,	10/10, 3.60
Cultural Japan	0/10,	0/10,	10/10, 0.21
DBpedia	0/10,	10/10, 2.49	10/10, 3.13
DBpedia Japanese	10/10, 18.01	0/10,	0/10,
Earthquake LOD	10/10, 39.91	0/10,	10/10, 0.92
Japan Search	0/10,	0/10,	10/10, 1.65
jrslod	10/10, 9.80	0/10,	10/10, 0.43
Linked Open Vocabularies	0/10,	0/10,	10/10, 4.23
Media Art Database	0/10,	0/10,	0/10,
Web NDL Authorities	10/10, 9.50	0/10,	10/10, 0.40
Onsen LOD	0/10,	0/10,	10/10, 0.19
OWL de ramen ontology	0/10,	10/10, 5.09	10/10, 5.18
☆pikopiko planet☆space	0/10,	0/10,	0/10,
PrismDB	0/10,	0/10,	10/10, 1.49
Wikidata	0/10,	0/10,	10/10, 44.73

Manual Query	number of successes / number of trials, average time(s)		
	1. EXISTING	2. VALUES	3. PROPOSED
Query1	0/10,	5/10, 63.44	5/10, 72.35
Query2	10/10, 7.36	0/10,	0/10,
Query3	0/10,	0/10,	0/10,
Query4	1/10, 335.41	0/10,	10/10, 61.24
Query5	0/10,	0/10,	10/10, 9.80
Query6	9/10, 437.28	0/10,	10/10, 170.16

Future Work

- Proposed Method needs to be more updated based on the causes of queries that could not be executed
- Evaluation queries will be prepared to assess accuracy further.

MDLAB メタデータ研究室
筑波大学 University of Tsukuba