Ontology-based Data Access: Theory and Practice

Guohui Xiao

KRDB Research Centre

Free University of Bozen-Bolzano

Roman Kontchakov

Department of Computer Science & Inf. Systems

Birkbeck, University of London

http://ontop.inf.unibz.it/ijcai-2018-tutorial

Non-relational Database MongoDB

MongoDB is a popular database storing collections of JSON-like documents:

```
{ "_id": 4,
  "awards": [ {"award": "Rosing Prize", "year": 1999},
              {"award": "Turing Award", "by": "ACM", "year": 2001},
              {"award": "IEEE John von Neumann Medal", "year": 2001, "by": "IEEE"} ],
  "birth": "1926-08-27",
  "contribs": ["OOP", "Simula"],
  "death": "2002-08-10",
  "name": {"first": "Kristen", "last": "Nygaard"}
}
```

MongoDB provides powerful, but unconventional query capabilities. The query below retrieves from the bios collection persons who received two awards in the same year:

```
db.bios.aggregate([
  {$project : {"name": true, "award1": "$awards", "award2": "$awards" }},
  {$unwind: "$award1"}, {$unwind: "$award2"},
  {$project: {"name": true, "award1": true, "award2": true,
              "twoInOneYear": { $and: [ {$eq: ["$award1.year", "$award2.year"]},
                                        {$ne: ["$award1.award", "$award2.award"]} ]}}},
  {$match: {"twoInOneYear": true} },
  {$project : {"firstName": "$name.first",    "lastName": "$name.last" ,
               "awardName1": "$award1.award", "awardName2": "$award2.award",
               "year": "$award1.year" }} ])
```

- How to extend the OBDA framework to support such data?[Botoeva et al., 2016]
- In [Botoeva et al., 2018, ICDT], we show that MongoDB query language is essentially equivalent to **nested relational algebra**

Accessing Spatial Data

- How to access spatial data stored in spatial databases, e.g. PostGIS?
- GeoSPARQL is the GeoSpatial extension of SPARQL, standardized by OGC (Open Geospatial Consortium).

Ontop-spatial extension [Bereta et al., 2016, ISWC]

- Extension of Ontop with GeoSPARQL support
- Enables on-the-fly GeoSPARQL-to-SQL translation
- Was developed as a fork of Ontop v1.16 by University of Athens
- We are planning to reimplement it as part of Ontop based on v3

Accessing Spatial Data

- How to access spatial data stored in spatial databases, e.g. PostGIS?
- GeoSPARQL is the GeoSpatial extension of SPARQL, standardized by OGC (Open Geospatial Consortium).
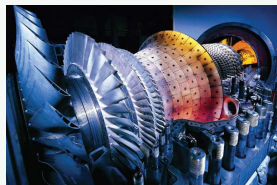
Ontop-spatial extension [Bereta et al., 2016, ISWC]

- Extension of Ontop with GeoSPARQL support
- Enables on-the-fly GeoSPARQL-to-SQL translation
- Was developed as a fork of Ontop v1.16 by University of Athens
- We are planning to reimplement it as part of Ontop based on v3

Example

Siemens remote-diagnostic centres analyse data
by querying aggregated sensor measurements

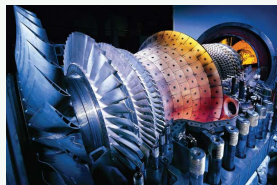| turbine_Id	dateTime	activePower	rotorSpeed	mainFlame	...
	...				
t03	2015-04-04 12:20:48	2	1550	0	
t03	2015-04-04 12:20:49	1.8	1400	null	
t01	2015-04-04 12:20:52	1.7	1350	1	
	...				

Typical query: when an active power trip occurred, i.e., the active
power was above 1.5MW for a period of at least 10 seconds, 3 seconds after which
there was a period of at least one minute where active power was below 0.15MW

- There are different approaches to temporal OBDA
- In [Brandt et al., 2017, AAAI] , we propose a temporal rule language DatalogMTL
  based on the metric temporal language
- Ontop-temporal is under development

Example

Siemens remote-diagnostic centres analyse data
by querying aggregated sensor measurements

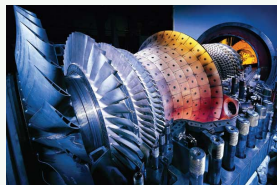| turbine_Id	dateTime	activePower	rotorSpeed	mainFlame	...
	...				
t03	2015-04-04 12:20:48	2	1550	0	
t03	2015-04-04 12:20:49	1.8	1400	null	
t01	2015-04-04 12:20:52	1.7	1350	1	
	...				

Typical query: when an active power trip occurred, i.e., the active
power was above 1.5MW for a period of at least 10 seconds, 3 seconds after which
there was a period of at least one minute where active power was below 0.15MW

- There are different approaches to temporal OBDA
- In [Brandt et al., 2017, AAAI] , we propose a temporal rule language DatalogMTL based on the metric temporal language
- Ontop-temporal is under development

Example

Siemens remote-diagnostic centres analyse data
by querying aggregated sensor measurements

turbine_Id	dateTime	activePower	rotorSpeed	mainFlame	...
	...				
t03	2015-04-04 12:20:48	2	1550	0	
t03	2015-04-04 12:20:49	1.8	1400	null	
t01	2015-04-04 12:20:52	1.7	1350	1	
	...				

Typical query: when an active power trip occurred, i.e., the active
power was above 1.5MW for a period of at least 10 seconds, 3 seconds after which
there was a period of at least one minute where active power was below 0.15MW

- There are different approaches to temporal OBDA
- In [Brandt et al., 2017, AAAI] , we propose a temporal rule language DatalogMTL
  based on the metric temporal language
- Ontop-temporal is under development

Some Perspectives

- Mapping Analysis
- Methodology and tools for ontology and mapping construction
- Further Optimization and Benchmarking
- Data Quality
- Updates
- Streaming data
- Non-Uniform OBDA
- Data Analytics

See our survey [Xiao et al., 2018, IJCAI]

Bereta, K. and M. Koubarakis (2016). "Ontop of Geospatial Databases". In: *Proc. of ISWC*. Vol. 9981. LNCS. Springer, pp. 37–52.

Botoeva, Elena, Diego Calvanese, Benjamin Cogrel, Martin Rezk, and Guohui Xiao (2016). "OBDA Beyond Relational DBs: A Study for MongoDB". In: *Proceedings of the 29th International Workshop on Description Logics, Cape Town, South Africa, April 22-25, 2016*.

Brandt, S. et al. (2017). "Ontology-Based Data Access with a Horn Fragment of Metric Temporal Logic". In: *Proc. of AAAI*, pp. 1070–1076.

Botoeva, Elena, Diego Calvanese, Benjamin Cogrel, and Guohui Xiao (2018). "Expressivity and Complexity of MongoDB Queries". In: *The 21st International Conference on Database Theory (ICDT 2018)*.

Xiao, Guohui, Diego Calvanese, Roman Kontchakov, Domenico Lembo, Antonella Poggi, Riccardo Rosati, and Michael Zakharyaschev (July 2018). "Ontology-Based Data Access: A Survey". In: *IJCAI-18*, pp. 5511–5519.