Complex alignments for better covering ontology heterogeneities

Cassia Trojahn

Institut de Recherche en Informatique de Toulouse
cassia.trojahn@irit.fr

Ontology, Data and Medical Informatics, 27^{th} May 2019 (IRIT)

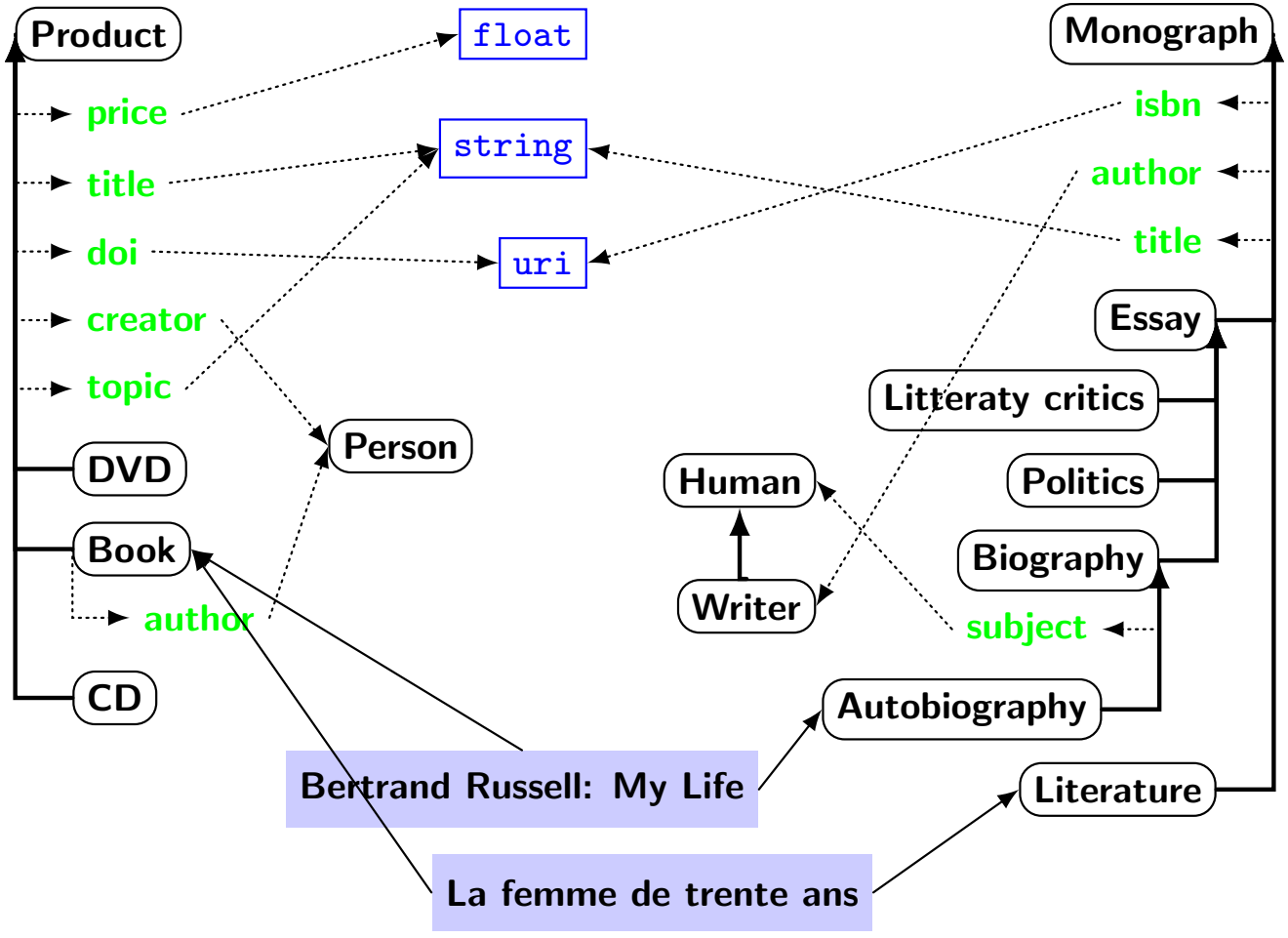
Outline

Outline

Ontology matching problem

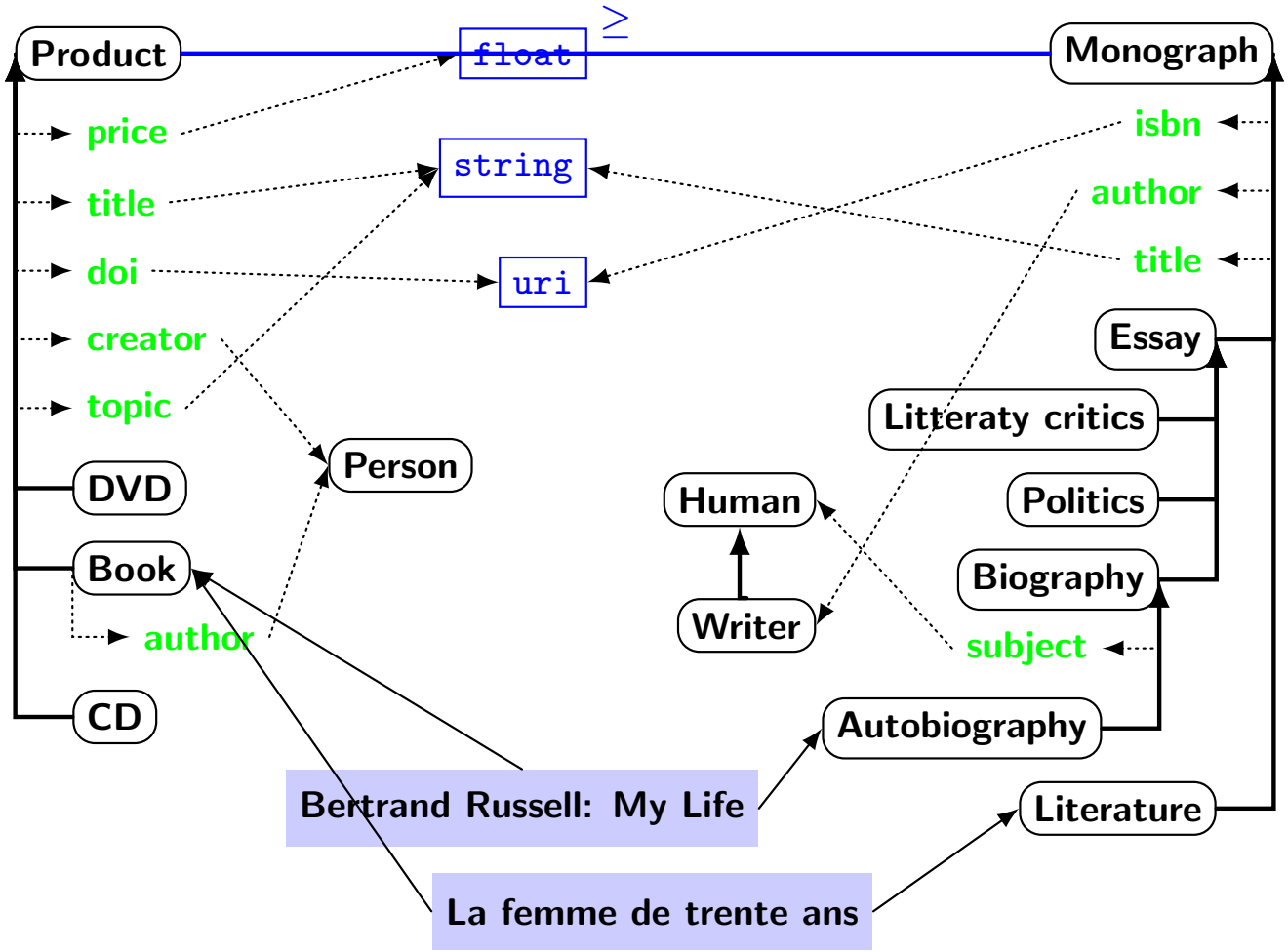
- Different ontology designers, different views on how to model the world
  - different vocabularies for annotating web sites
  - different ontologies for describing related knowledge bases
  - different ontologies describing agents' world
- For many tasks, ontologies have to be aligned

Ontology matching problem



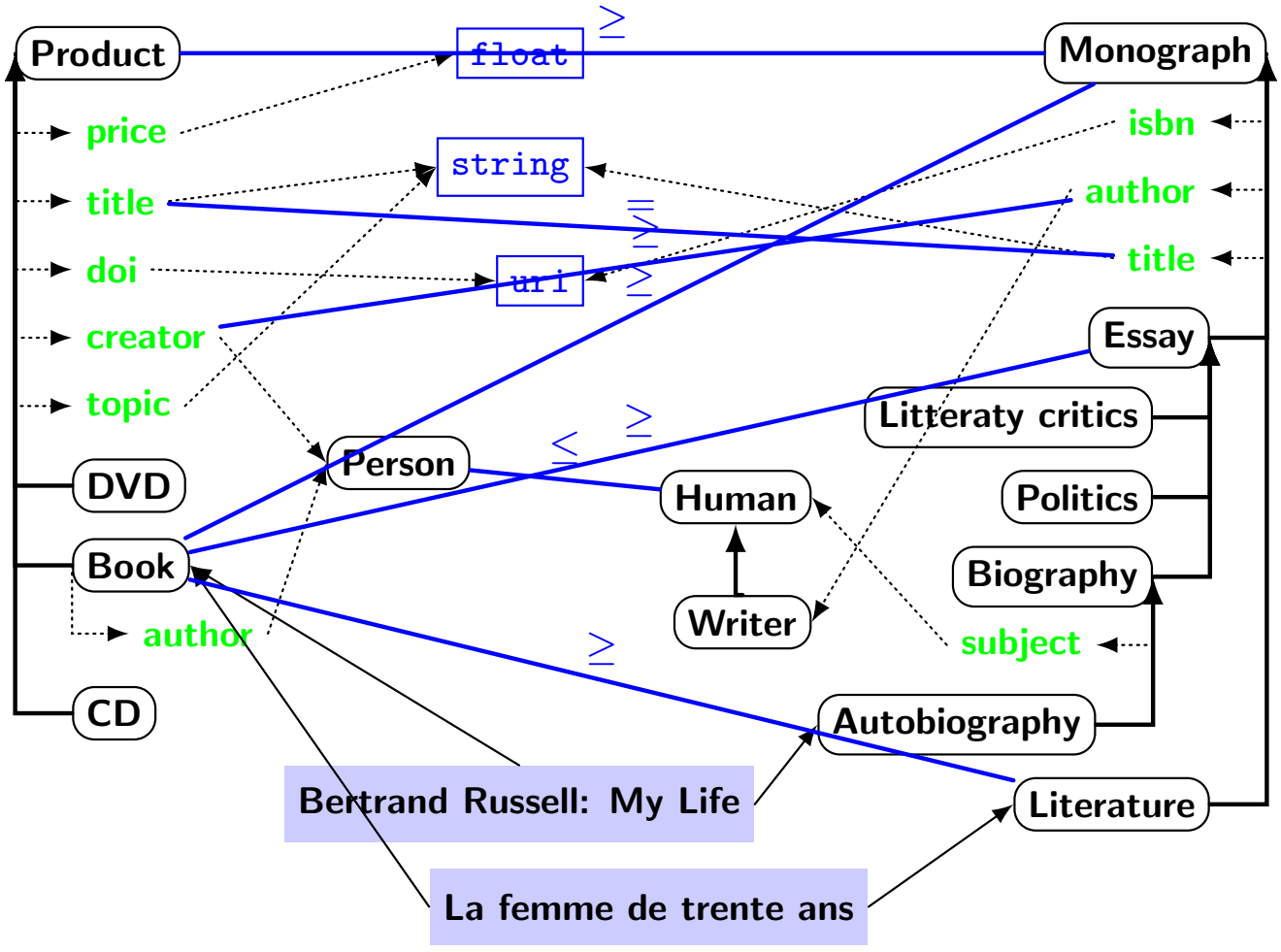
Extracted from [Euzenat and Shvaiko, 2007]

Ontology matching problem



Extracted from [Euzenat and Shvaiko, 2007]

Ontology matching problem



Extracted from [Euzenat and Shvaiko, 2007]

Outline

Main research topics

- Uncertainty and incompleteness in alignments
- Ontology alignment evaluation [Algergawy et al., 2018]
- Domain and top-level ontology matching [Schmidt et al., 2018]
- Multiple alignments visualization [Severo et al., 2017]
- SPARQL query rewriting based on complex alignments [Thiéblin et al., 2018b]
- Complex ontology matching [Thiéblin et al., 2018a]

Outline

Simple vs. complex alignments

- Simple correspondences link single ontology entities
  - $\forall x\ o_s : Product(x) \Leftrightarrow o_t : Monography(x)$
- More expressiveness with complex correspondences
  - $\forall x\ o_s : Autobiography(x) \Leftrightarrow o_t : Book(x) \wedge o_t : author(x, y) \wedge o_t : topic(x, y)$

Complex matching approach (i)

- Discovers complex correspondences between populated ontologies
- Relies on the notion of Competency Questions for Alignment
  - represent the needs of a user with respect to an ontology alignment
  - define (and reduce) the scope of the alignment
  - have to be satisfied over two or more ontologies

Complex matching approach (ii)

1. Takes as input a set of CQAs as SPARQL queries over o_s
2. The answer to each CQA is a set of instances retrieved from a KB described by o_s
3. The instances are matched to those of the KB described by o_t
4. The generation of the correspondence is performed by matching the graph-pattern from the source query to the lexically similar surroundings of the target instances

Proposed approach

SPARQL CQA

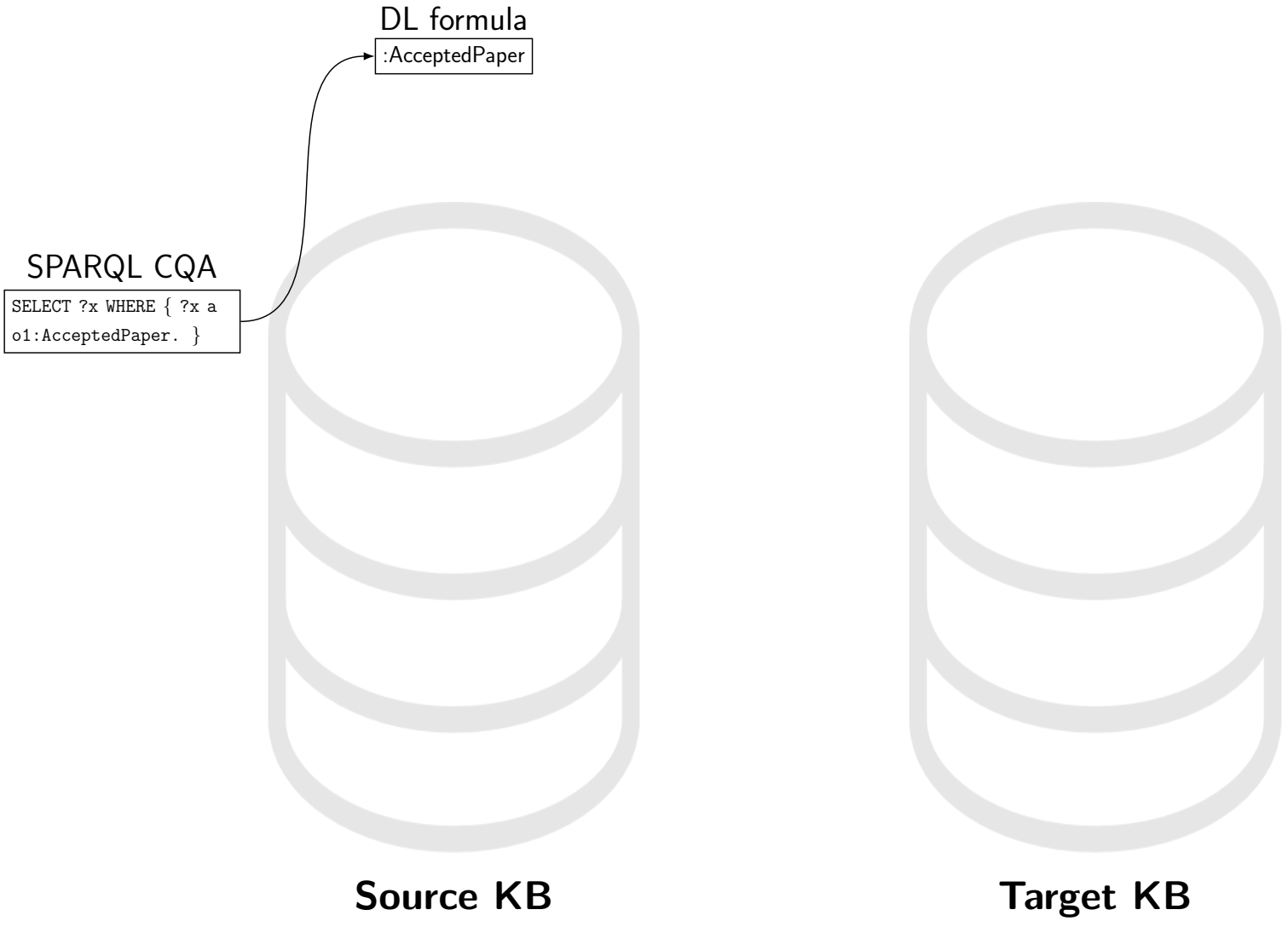
```
SELECT ?x WHERE { ?x a
o1:AcceptedPaper. }
```


Source KB

Target KB

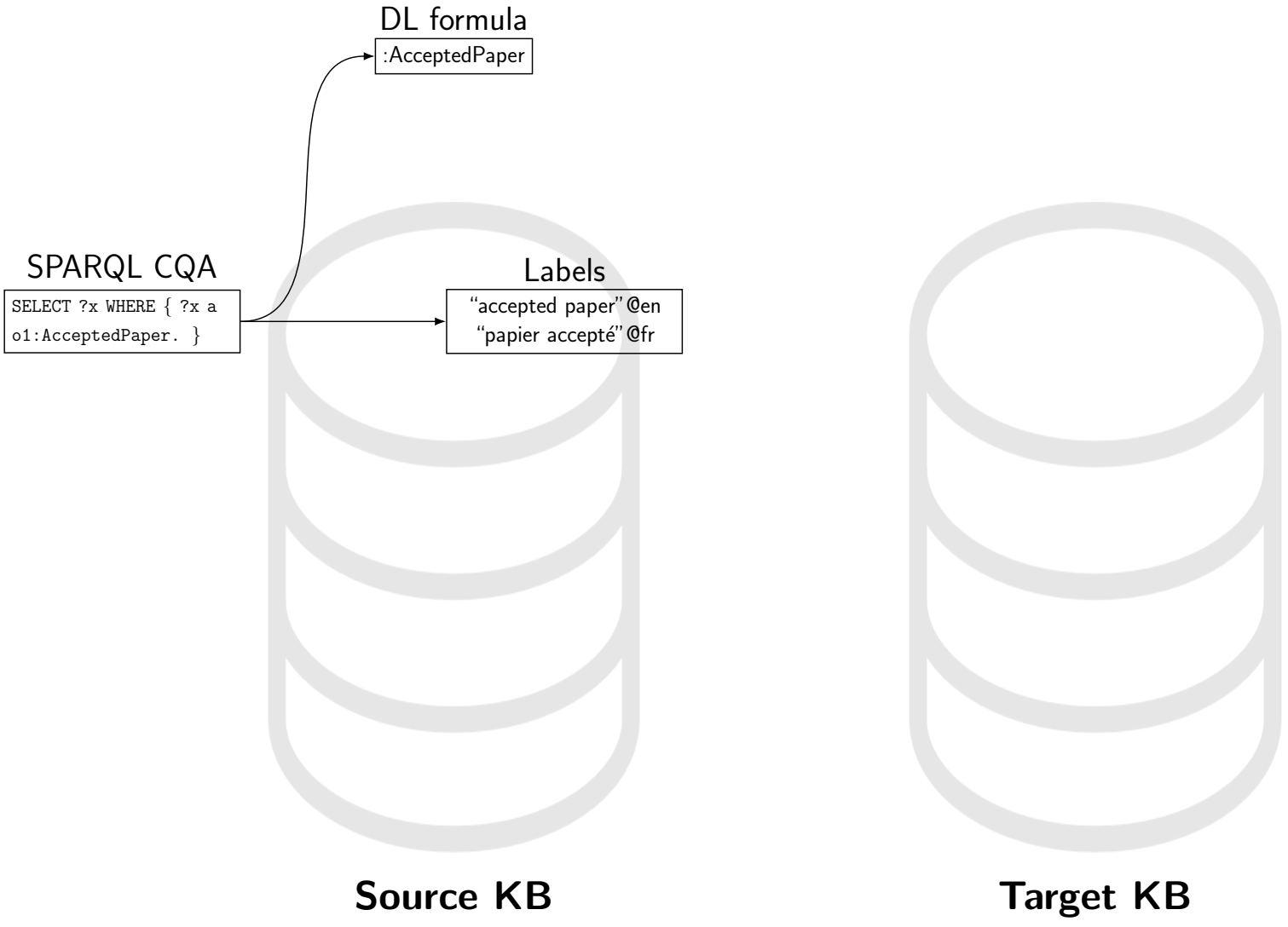
[Thiéblin et al., 2018a]

Proposed approach



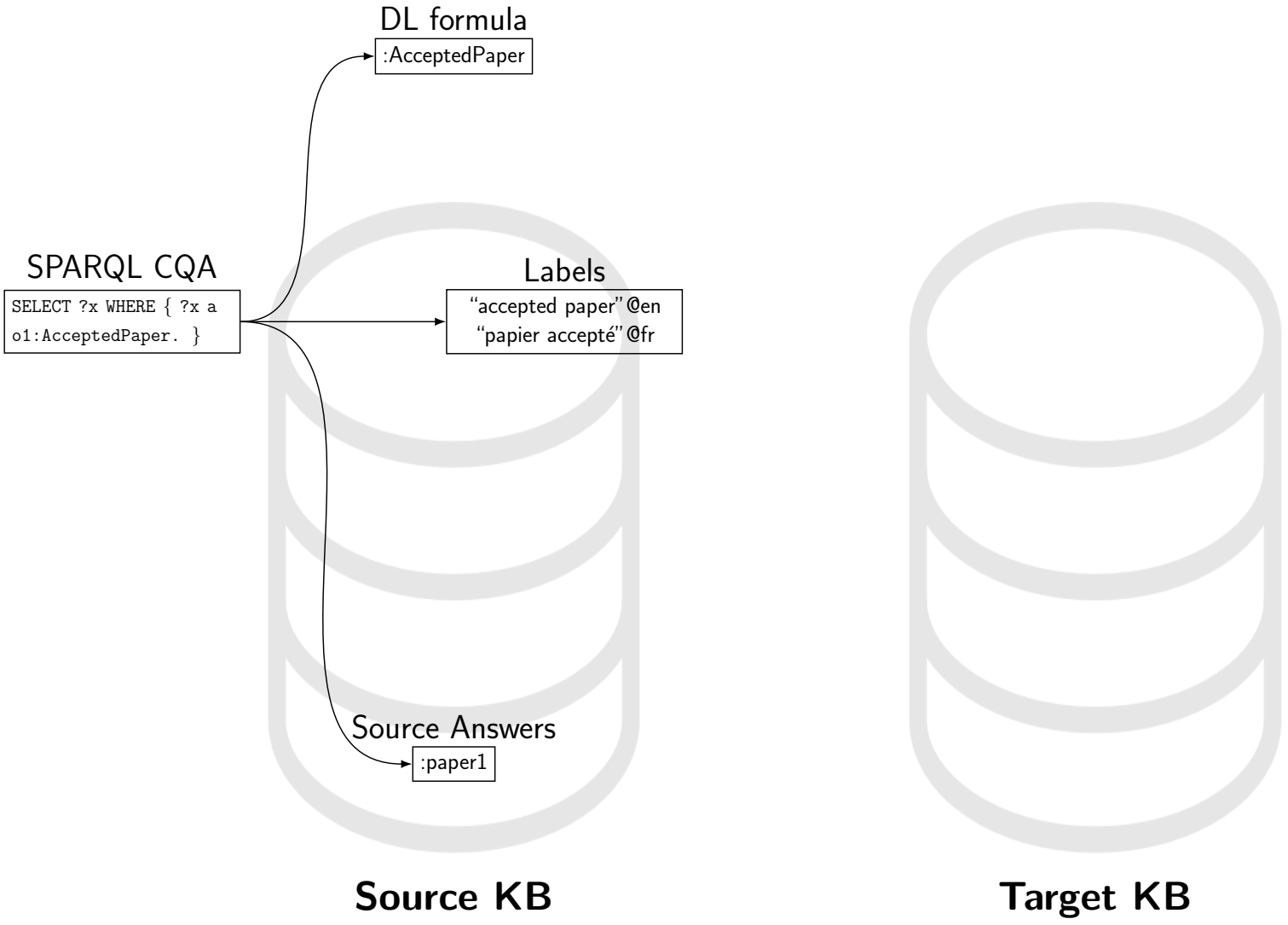
[Thiéblin et al., 2018a]

Proposed approach



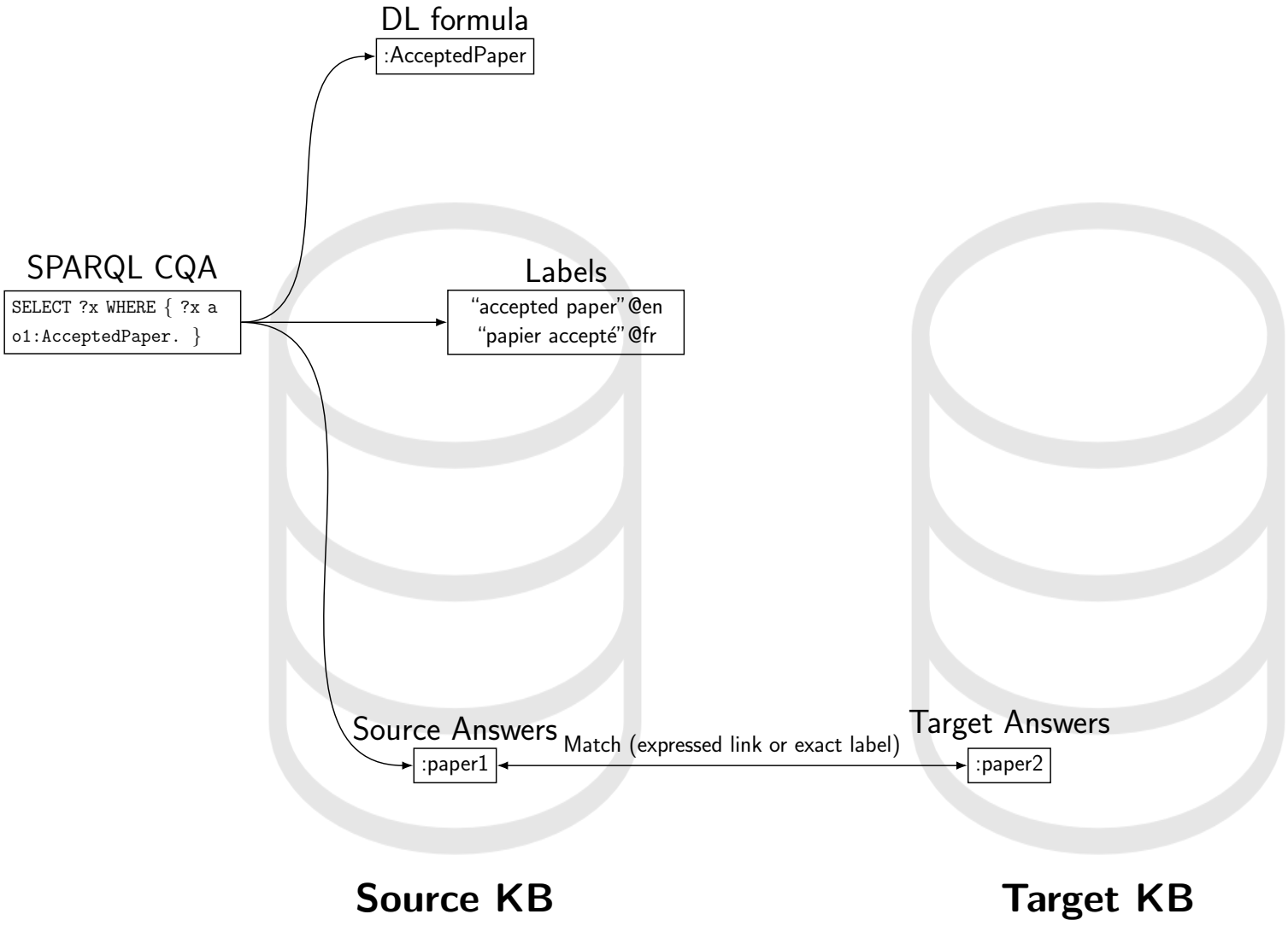
[Thiéblin et al., 2018a]

Proposed approach



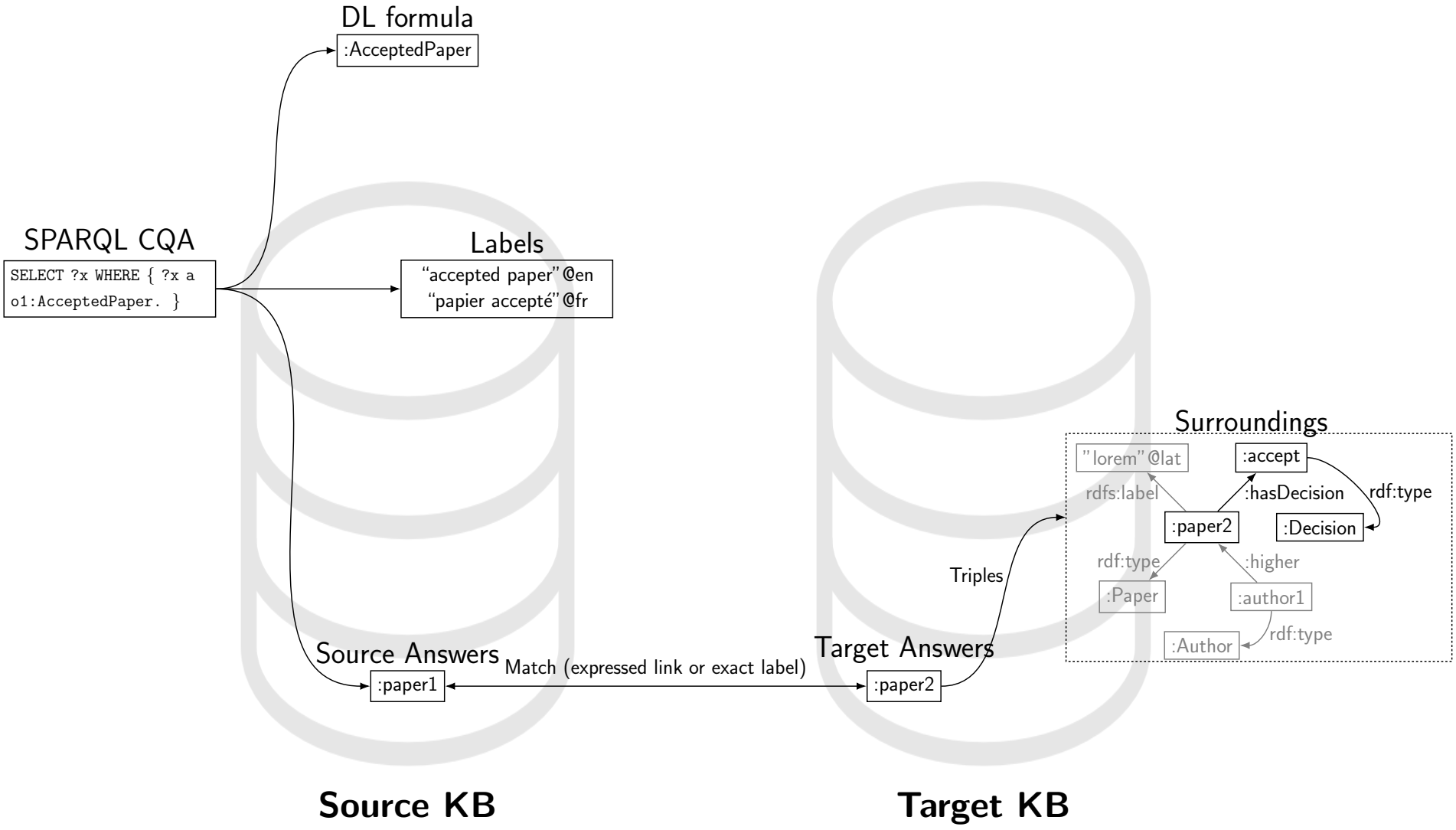
[Thiéblin et al., 2018a]

Proposed approach



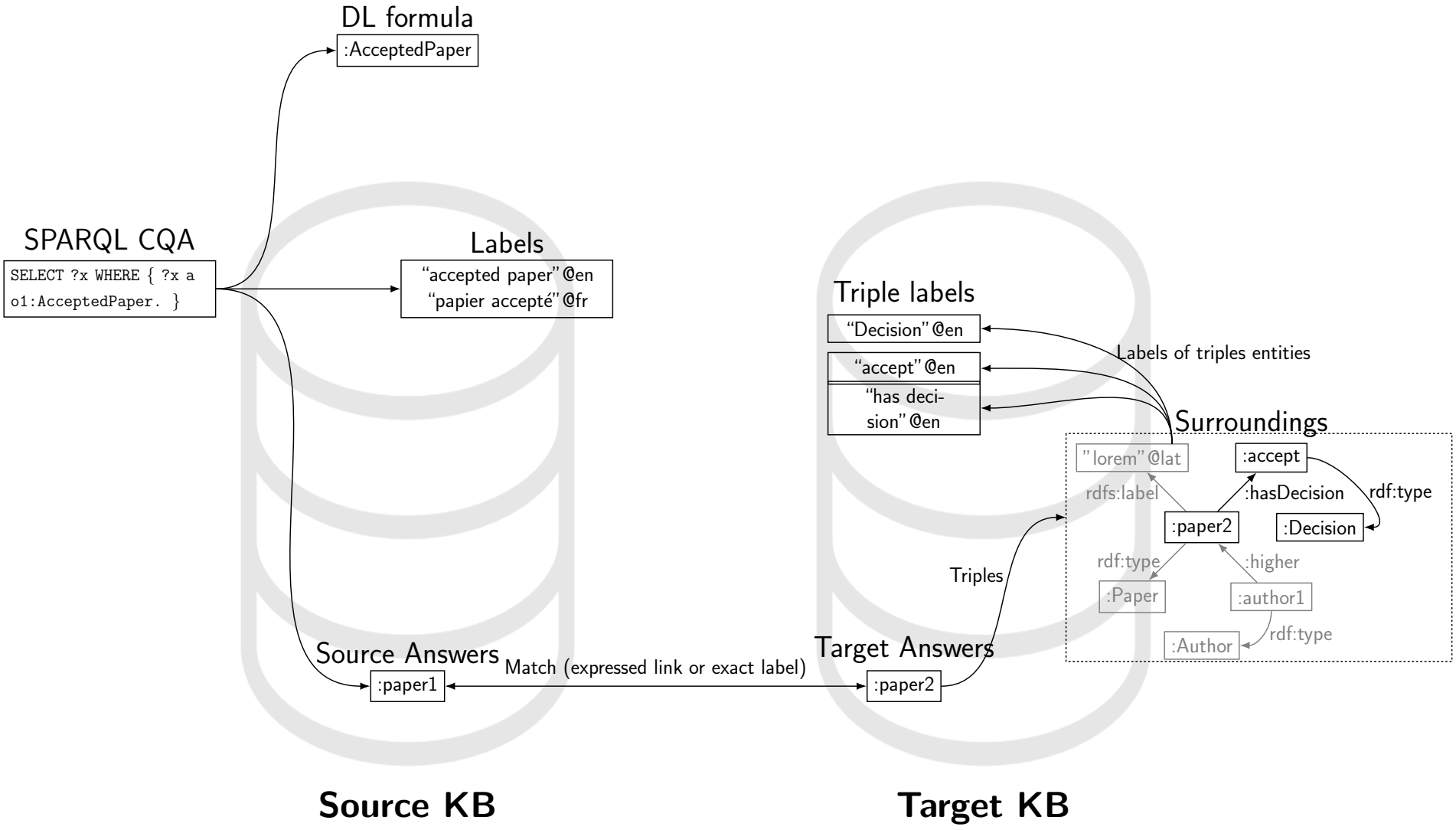
[Thiéblin et al., 2018a]

Proposed approach



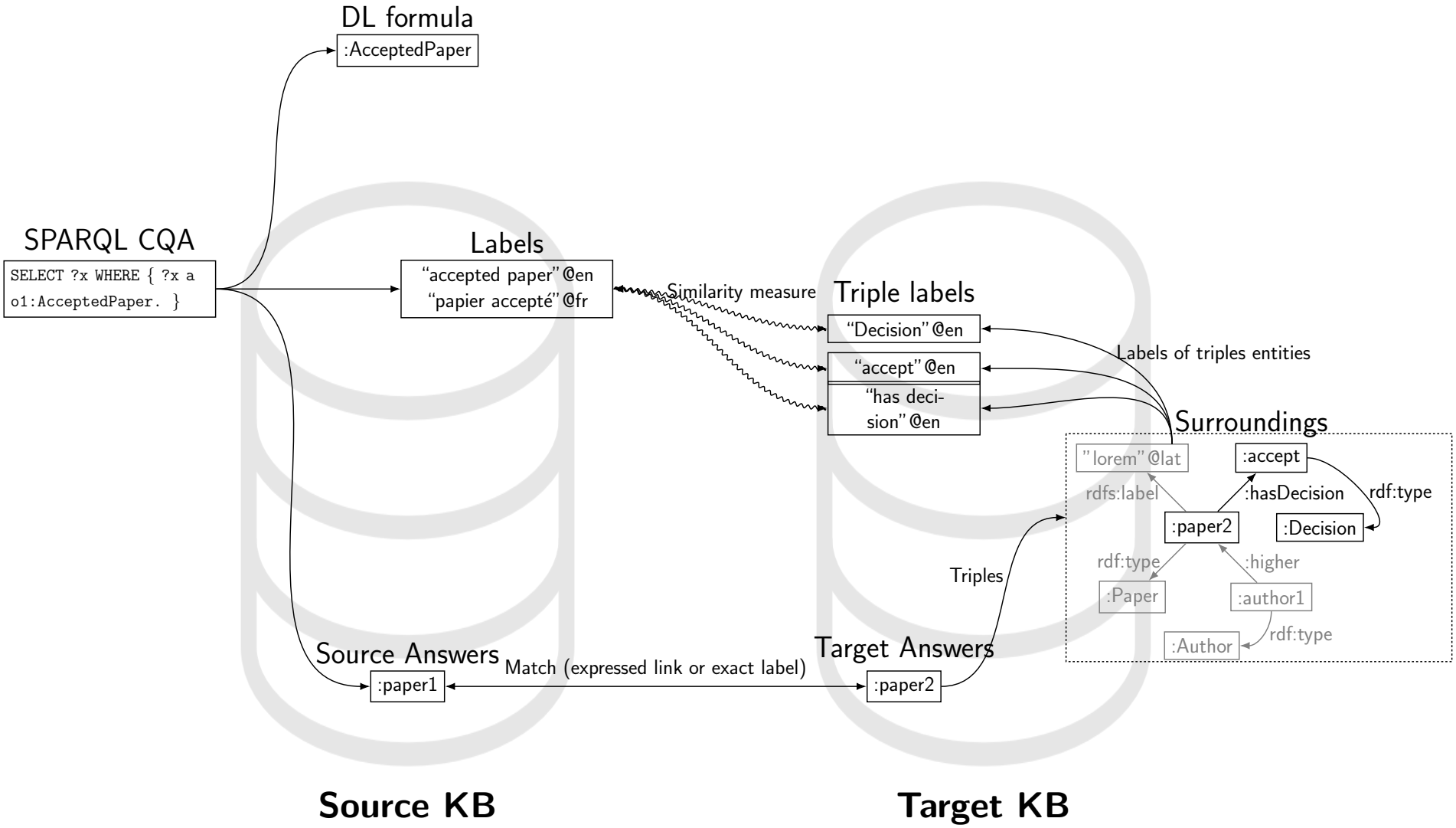
[Thiéblin et al., 2018a]

Proposed approach



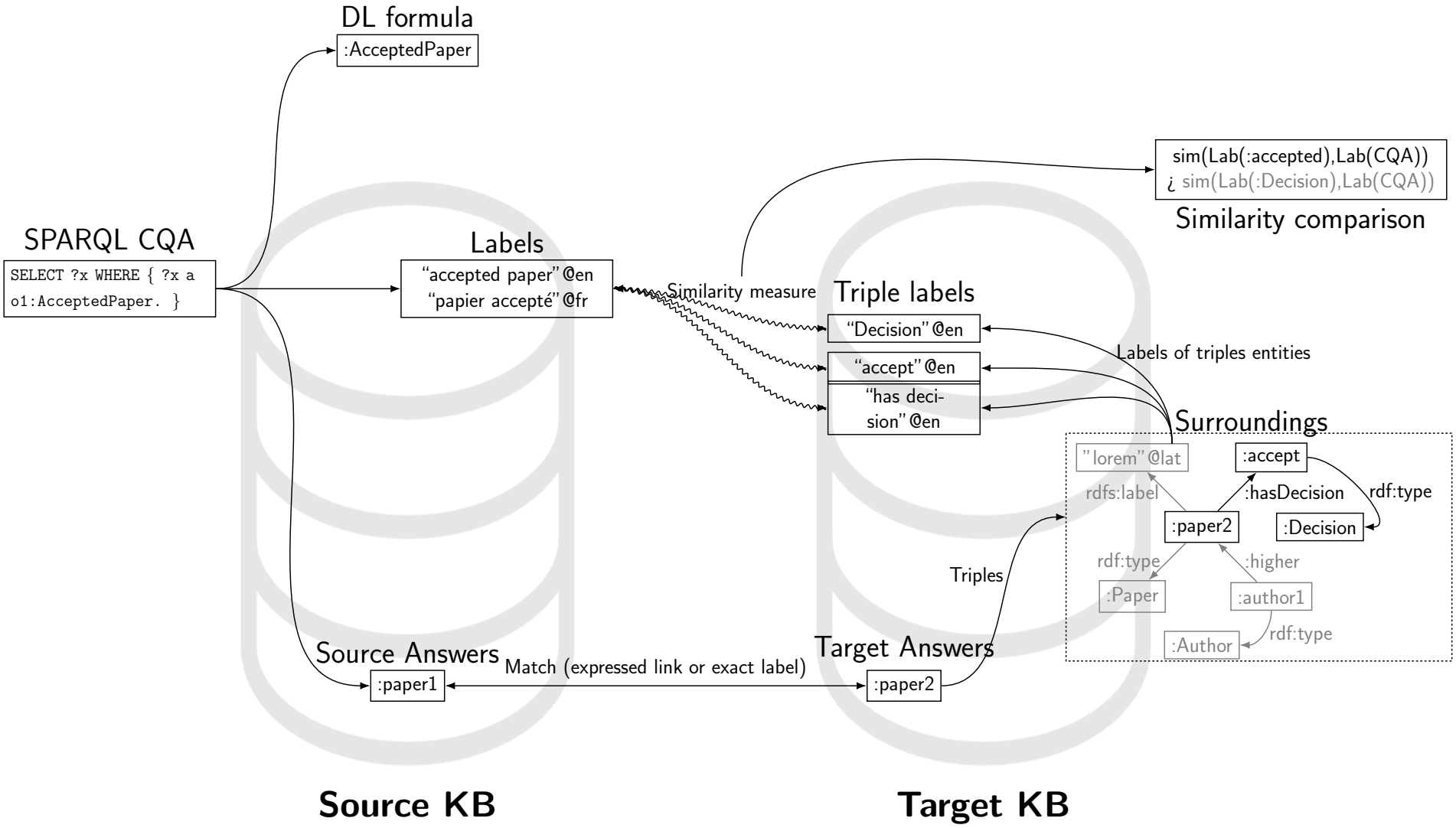
[Thiéblin et al., 2018a]

Proposed approach



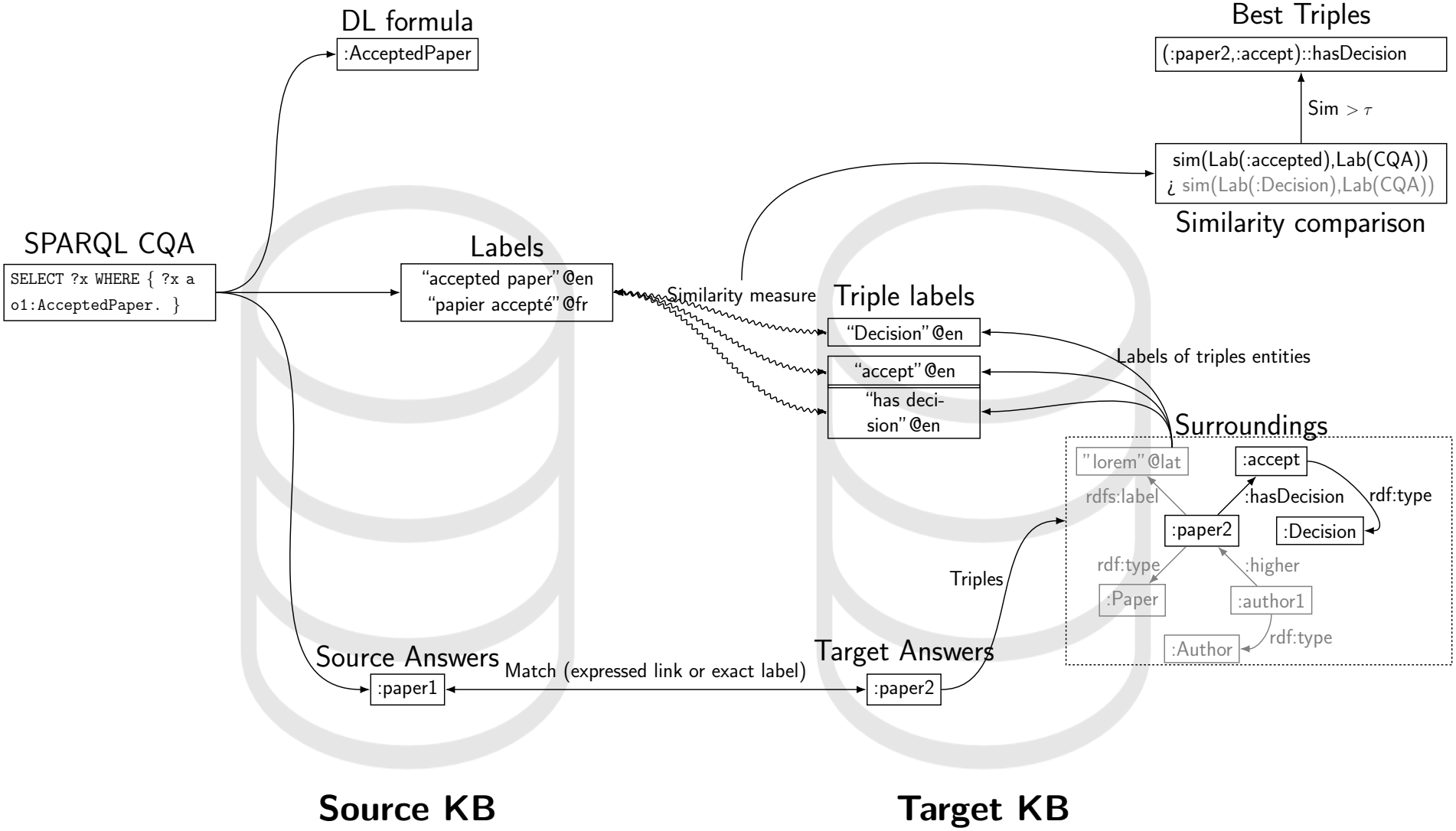
[Thiéblin et al., 2018a]

Proposed approach



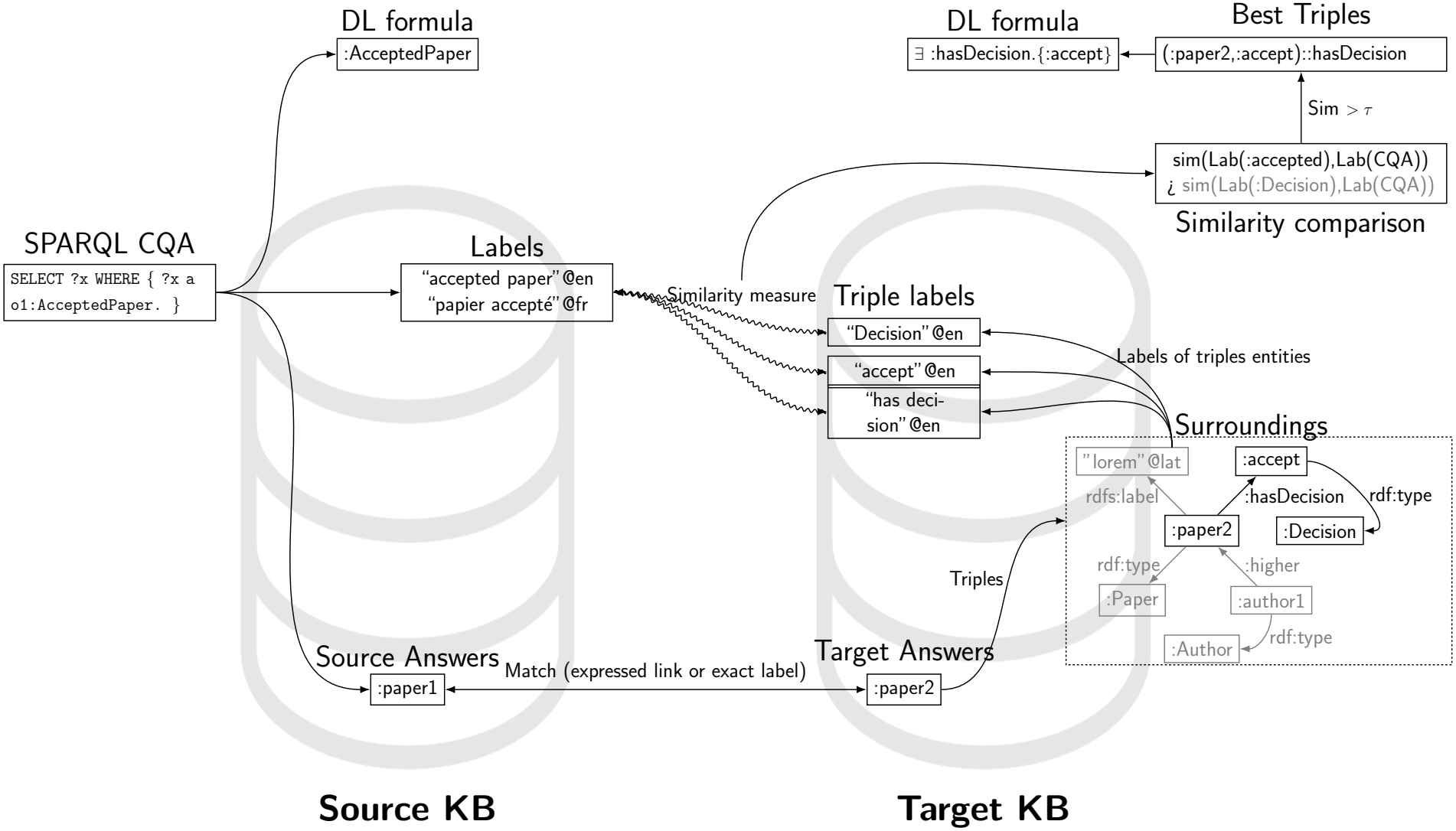
[Thiéblin et al., 2018a]

Proposed approach



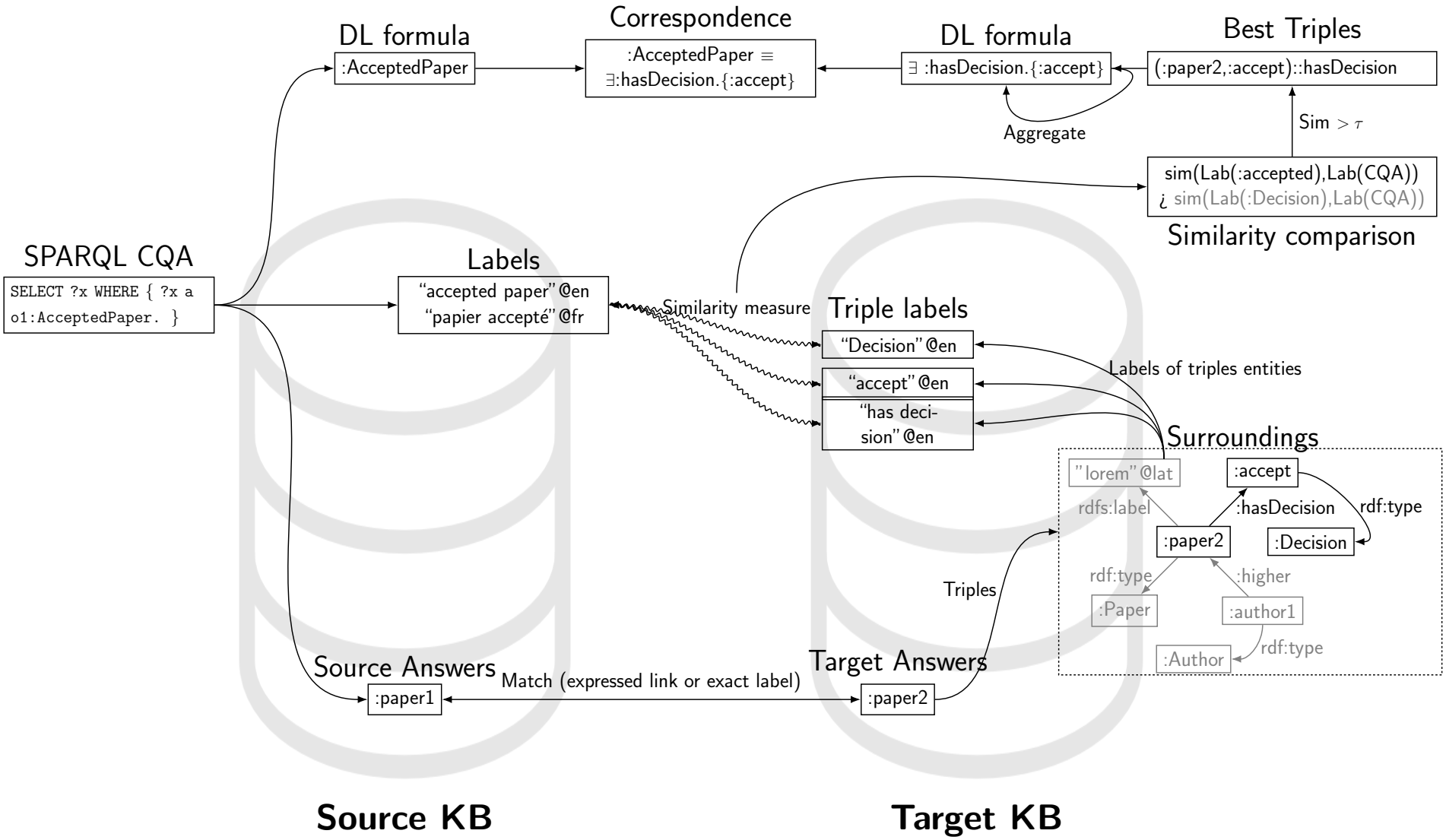
[Thiéblin et al., 2018a]

Proposed approach



[Thiéblin et al., 2018a]

Proposed approach



[Thiéblin et al., 2018a]

First results

OAEI 2018 Complex Taxon Track

Evaluation without CQAs (generated queries)

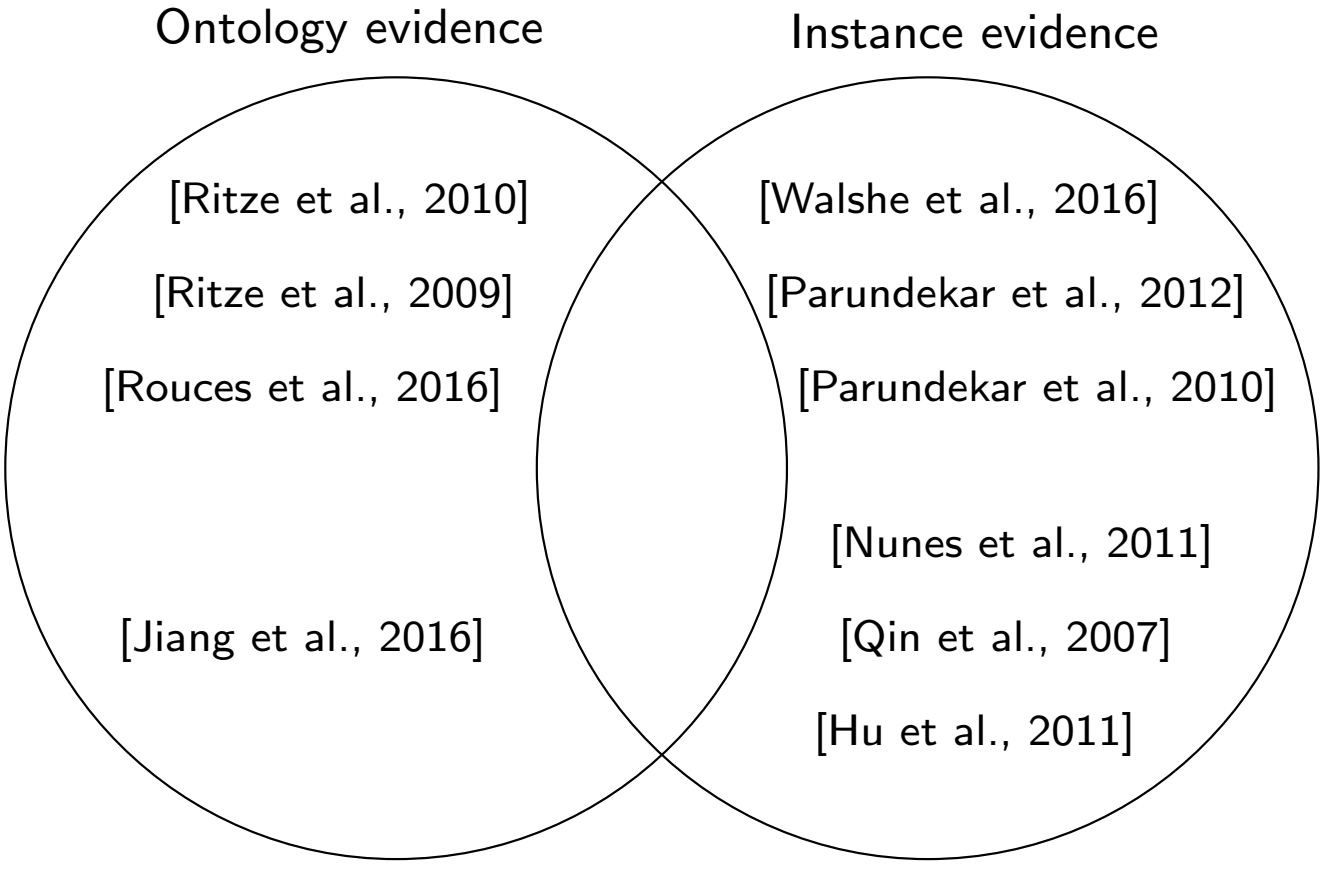
- Precision: 20%
- Queries well rewritten: 13% (best result)

Plant taxonomy evaluation

Evaluation with CQAs (only class expressions)

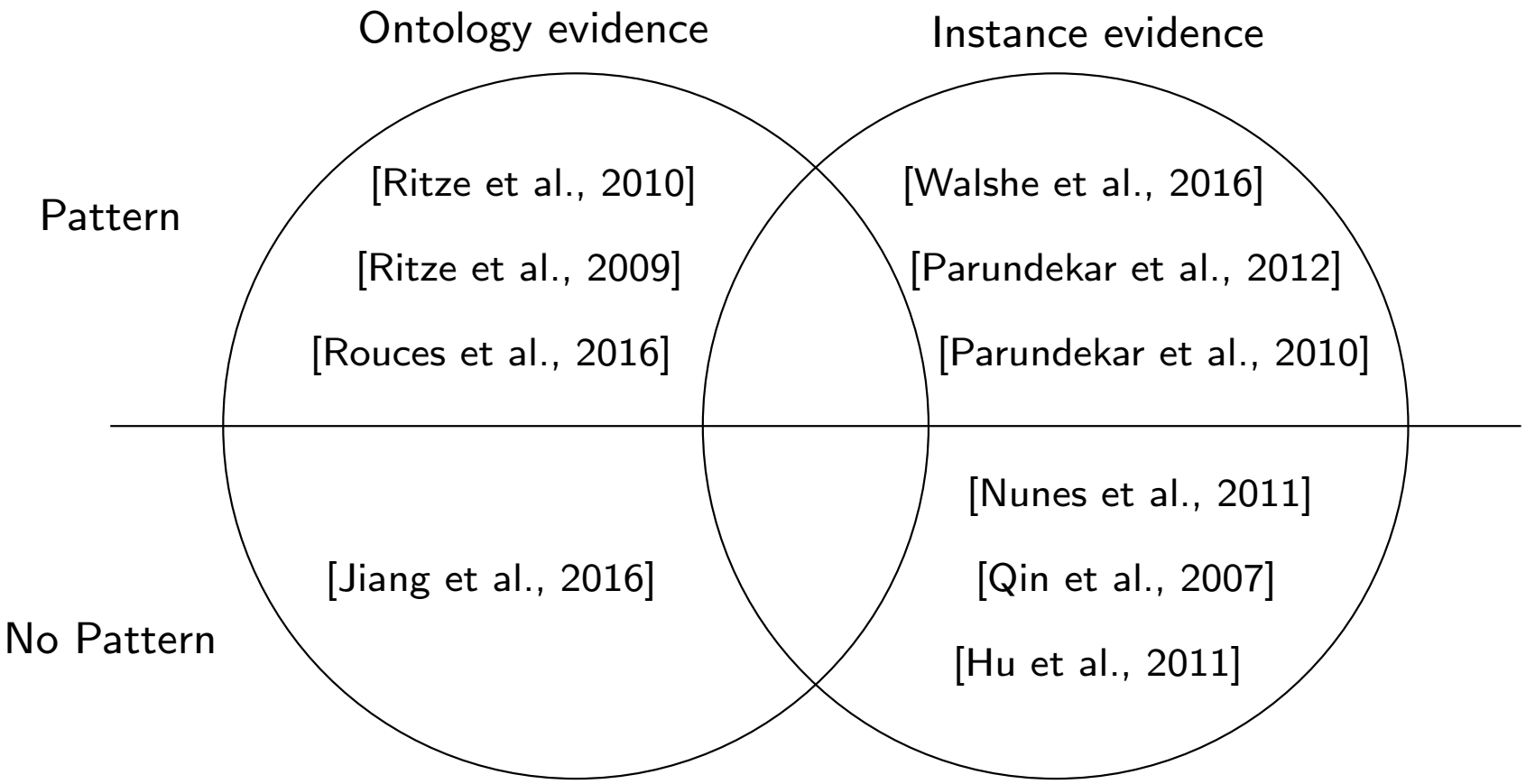
- Precision: 32.8%
- Queries well rewritten: 83.4%

Positioning



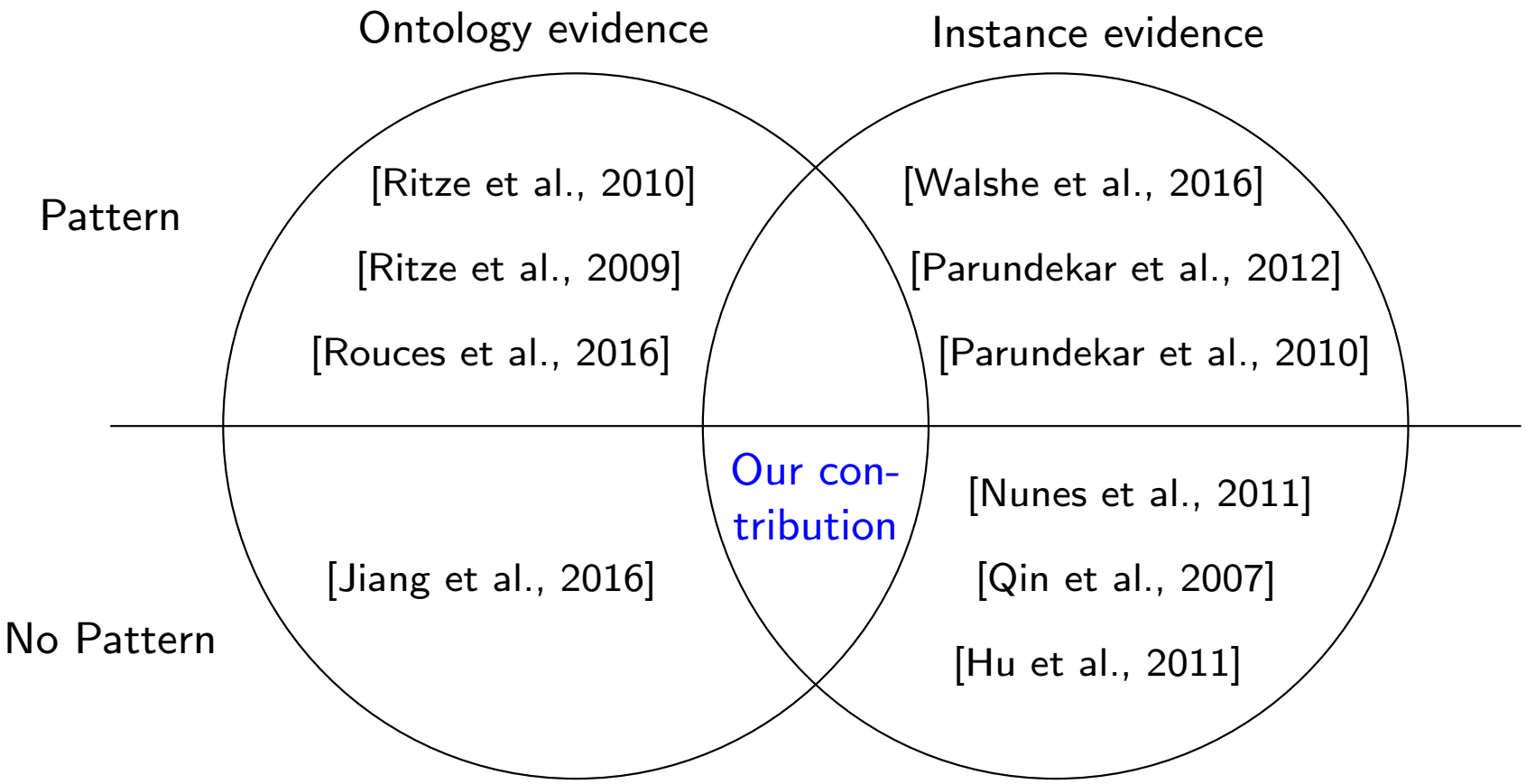
[Thiéblin et al., 2019]

Positioning



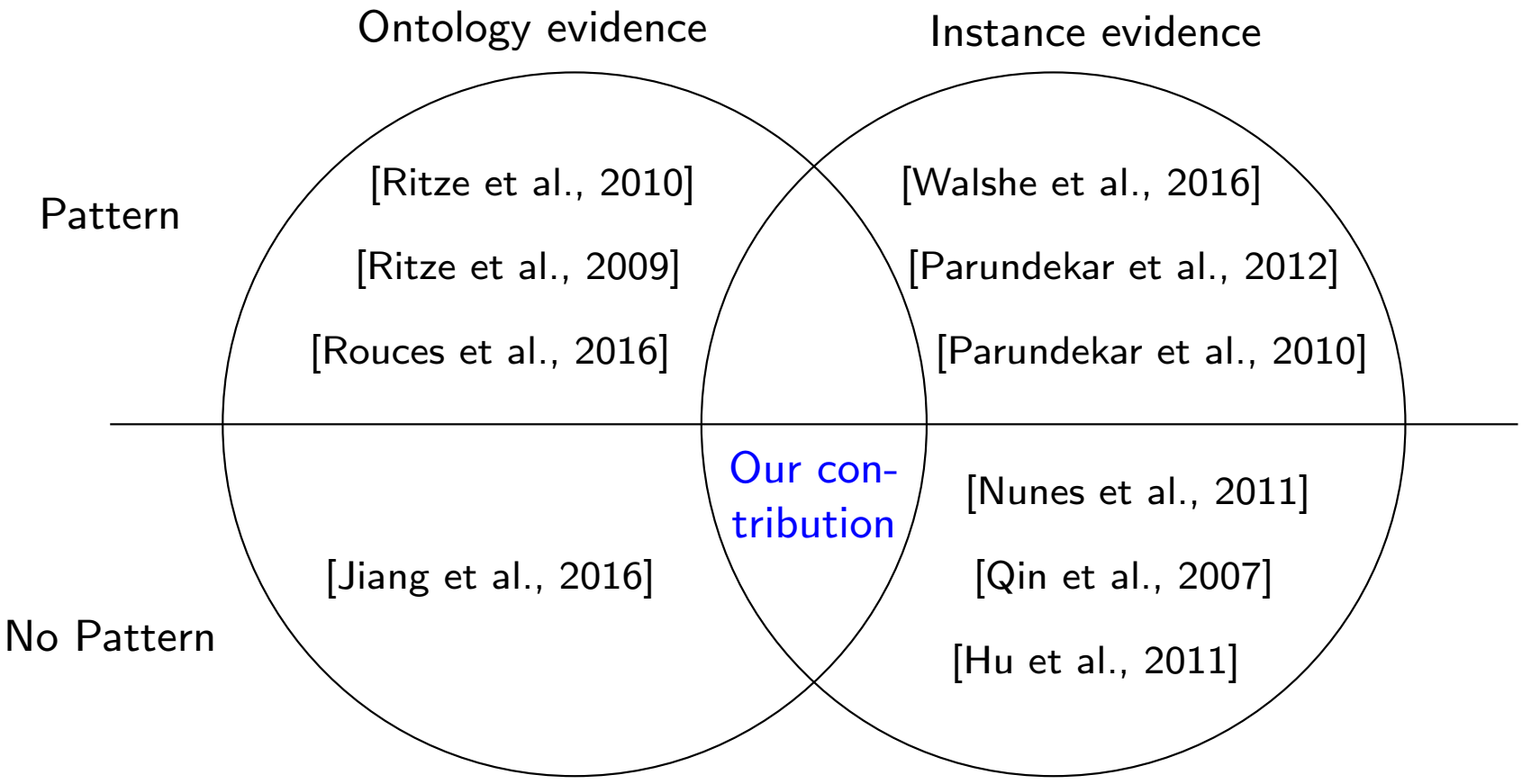
[Thiéblin et al., 2019]

Positioning



[Thiéblin et al., 2019]

Positioning



In terms of user involvement, some propose a user validation *after* the matching process.
[Thiéblin et al., 2019]

Outline

Open questions

- How do compute the confidence of complex correspondences?
- How to involve the user in the loop? How to collaboratively establish complex alignments?
- How to automatically evaluate complex alignments?
- How to deal with complex alignments between multiple ontologies?
- How to reasoning on the uncertainty involving complex alignments and multiple ontologies?

References I

Algergawy, A., Cheatham, M., Faria, D., Ferrara, A., Fundulaki, I., Harrow, I., Hertling, S., Jiménez-Ruiz, E., Karam, N., Khiat, A., Lambrix, P., Li, H., Montanelli, S., Paulheim, H., Pesquita, C., Saveta, T., Schmidt, D., Shvaiko, P., Splendiani, A., Thiéblin, É., Trojahn, C., Vatascinová, J., Zamazal, O., and Zhou, L. (2018).
Results of the ontology alignment evaluation initiative 2018.
In *Proceedings of the 13th International Workshop on Ontology Matching co-located with the 17th International Semantic Web Conference, OM@ISWC 2018, Monterey, CA, USA, October 8, 2018.*, pages 76–116.

Euzenat, J. and Shvaiko, P. (2007).
Ontology Matching.
Springer-Verlag, 1st edition.

Hu, W., Chen, J., Zhang, H., and Qu, Y. (2011).
Learning complex mappings between ontologies.
In *Joint International Semantic Technology Conference*. Springer.

Jiang, S., Lowd, D., Kafle, S., and Dou, D. (2016).
Ontology matching with knowledge rules.
In *Transactions on Large-Scale Data-and Knowledge-Centered Systems*.

Nunes, B. P., Mera, A., Casanova, M. A., Breitman, K. K., and Leme, L. A. P. (2011).
Complex Matching of RDF Datatype Properties.
In *6th OM workshop*.

References II


Parundekar, R., Knoblock, C. A., and Ambite, J. L. (2010).
Linking and building ontologies of linked data.
In *ISWC*, pages 598–614. Springer.


Parundekar, R., Knoblock, C. A., and Ambite, J. L. (2012).
Discovering concept coverings in ontologies of linked data sources.
In *ISWC*, pages 427–443. Springer.


Qin, H., Dou, D., and LePendu, P. (2007).
Discovering executable semantic mappings between ontologies.
In *On the Move to Meaningful Internet Systems*, pages 832–849.


Ritze, D., Meilicke, C., Šváb Zamazal, O., and Stuckenschmidt, H. (2009).
A pattern-based ontology matching approach for detecting complex correspondences.
In *4th ISWC workshop on ontology matching*, pages 25–36.


Ritze, D., Völker, J., Meilicke, C., and Šváb Zamazal, O. (2010).
Linguistic analysis for complex ontology matching.
In *5th workshop on ontology matching*, pages 1–12.


Rouces, J., de Melo, G., and Hose, K. (2016).
Complex Schema Mapping and Linking Data: Beyond Binary Predicates.
In *Proceedings of the WWW 2016 Workshop on Linked Data on the Web (LDOW 2016)*.

References III

Schmidt, D., Basso, R., Trojahn, C., and Vieira, R. (2018).
Matching domain and top-level ontologies exploring word sense disambiguation and word embedding.
In *Emerging Topics in Semantic Technologies - ISWC 2018 Satellite Events [best papers from 13 of the workshops co-located with the ISWC 2018 conference]*, pages 27–38.

Severo, B., Trojahn, C., and Vieira, R. (2017).
VOAR 3.0 : a configurable environment for manipulating multiple ontology alignments.
In *Proceedings of the ISWC 2017 Posters & Demonstrations and Industry Tracks co-located with 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, October 23rd - to - 25th, 2017.*

Thiéblin, É., Haemmerle, O., Hernandez, N., and Trojahn, C. (2019).
Survey on complex ontology matching.
Semantic Web Journal (to appear).

Thiéblin, É., Haemmerlé, O., and Trojahn, C. (2018a).
Complex matching based on competency questions for alignment: a first sketch.
In *Proceedings of the 13th International Workshop on Ontology Matching co-located with the 17th International Semantic Web Conference, OM@ISWC 2018, Monterey, CA, USA, October 8, 2018.*, pages 66–70.

Thiéblin, É., Hernandez, N., Roussey, C., and Trojahn, C. (2018b).
Cross-querying LOD data sets using complex alignments: an experiment using agronomictaxon, agrovoc, dbpedia and TAXREF-LD.
IJMSO, 13(2):104–119.

References IV

Walshe, B., Brennan, R., and O'Sullivan, D. (2016).
Bayes-recce: A bayesian model for detecting restriction class correspondences in linked open data knowledge bases.
International Journal on Semantic Web and Information Systems, 12(2):25–52.