

Specifying and reasoning about preferences over inconsistent knowledge bases

Basic Information

Duration: 5 months, ideally between February - August 2021, but there is some flexibility

Supervisors: [Meghyn Bienvenu](#) (LaBRI, Bordeaux), [Camille Bourgaux](#) (DI ENS, Paris)

Host lab: [LaBRI](#) (Laboratoire Bordelais de Recherche en Informatique)

There is a funded PhD position on a related topic (details [here](#))

Context

Accessing the relevant information contained in real-world data to support informed decision making is difficult, time-consuming, and error-prone due to the need to integrate data across multiple heterogeneous sources. Moreover, even if this first hurdle is overcome, a perhaps even more daunting challenge arises: how to obtain reliable insights from imperfect data? It is widely acknowledged that real-world data is plagued with quality issues, such as incompleteness (missing information) and errors (false or outdated information).

The ontology-mediated query answering (OMQA) paradigm [1,2,3] addresses the first challenge by facilitating access to (potentially heterogeneous) data sources through the use of ontologies that specify a convenient user-friendly vocabulary for query formulation (which abstracts from the way the data is stored) and capture domain knowledge that can be exploited at query time, via automated reasoning, to obtain more complete query results. For example, querying for patients with infectious heart disease is non-trivial due to the myriad of ways such a generic condition can manifest, but by leveraging the knowledge formalized in medical ontologies (like SNOMED CT [4]), it is possible to correctly return patients diagnosed with Chagall's disease, toxoplasma myocarditis, etc.

The OMQA approach is relevant to a wide range of data-intensive applications, and recent industrial projects have witnessed its practical benefits. While OMQA systems are growing in maturity [5], they too often fail to address the data quality issue, aside from issuing warnings when inconsistencies are discovered. To widen the applicability of the OMQA approach, it is essential to equip OMQA systems with appropriate mechanisms for handling imperfect data: how to obtain meaningful answers to queries posed over imperfect data, and how best to generate a high-quality version of the data ?

The Master's internship is part of the [INTENDED](#) Chair on Artificial Intelligence, whose aim is to develop intelligent, knowledge-based methods for handling imperfect data. The chair started in September 2020 and has a duration of four years. It is funded by the French National Research Agency (ANR) and the University of Bordeaux.

Research Topic

Several different inconsistency-tolerant semantics have been proposed with the aim of providing meaningful answers to queries posed over inconsistent knowledge bases [6]. Recent work has focused on how to integrate preferences into such semantics in order to exploit information about the relative reliability of facts in the data [7,8].

The aim of this internship is to explore declarative languages for specifying preferences between facts in the OMQA context, an important issue left open in [7]. Specifically, we envision rule-based preference languages, along the lines of "If the data contains Salary(EMP,s1) and Salary(EMP,s2) and $s1 > s2$, then prefer Salary(EMP,s1) to Salary(EMP,s2)" or "If fact1 and fact2 are in contradiction, fact1 was inserted after fact2, and fact2 is not from source A, then prefer fact1 over fact2".

After defining a syntax and semantics for such preference rules, the student will investigate the associated reasoning tasks: Can we decide whether a given set of preference rules always yields an acyclic preference relation? Do the rules always define a total relation (i.e. precisely determine how to correct the data)? How does adopting preference rules rather than assuming an explicit preference relation affect the complexity of query answering under preference-based semantics?

How to apply

Potential candidates should contact the two supervisors by email: Meghyn Bienvenu (meghyn.bienvenu@labri.fr) and Camille Bourgaux (camille.Bourgaux@ens.fr).

The email should include a detailed CV and a short description of how the internship topic relates to their prior experience and research interests.

Bibliography

[1] A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, and R. Rosati. Linking data to ontologies. *Journal of Data Semantics*, 10:133–173, 2008.

[2] G. Xiao, D. Calvanese, R. Kontchakov, D. Lembo, A. Poggi, R. Rosati, and M. Zakharyashev. Ontology-based data access: A survey. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5511–5519, 2018.

[3] M. Bienvenu, M. Ortiz: Ontology-Mediated Query Answering with Data-Tractable Description Logics. *Tutorial Notes of the 11th International Reasoning Web Summer School (LNCS 9203)*, pages 218-307, 2015.

[4] SNOMED CT Website: <https://www.snomed.org/snomed-ct/why-snomed-ct>

[5] D. Calvanese, B. Cogrel, S. Komla-Ebri, R. Kontchakov, D. Lanti, M. Rezk, M. Rodriguez-Muro, G. Xiao, "Ontop: Answering SPARQL queries over relational databases," *Semantic Web Journal*, vol. 8, no. 3, pp. 471-487, 2017.

[6] M. Bienvenu, C. Bourgaux: Inconsistency-Tolerant Querying of Description Logic Knowledge Bases. Tutorial Notes of the 12th International Reasoning Web Summer School (LNCS 9885), pages 156- 202, 2016.

[7] Meghyn Bienvenu, Camille Bourgaux: Querying and Repairing Inconsistent Prioritized Knowledge Bases: Complexity Analysis and Links with Abstract Argumentation. Proc. of KR, 2020.

[8] M. Calautti, S. Greco, C. Molinaro, I. Trubitsyna. Preference-based Inconsistency-Tolerant Query Answering under Existential Rules. Proc. of KR, 2020.