Customized User-Sensitive Approaches to Inconsistency Management

PhD Position in Bordeaux, France

Keywords

Inconsistency handling, data quality, ontology-based data access, knowledge representation and reasoning

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that real-world data is plagued with quality issues, such as incompleteness (missing information) and errors (false or outdated information).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-based data access (OBDA) 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 OBDA approach is relevant to a wide range of data-intensive applications, and recent industrial projects have witnessed its practical benefits. While OBDA 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 OBDA approach, it is essential to equip OBDA 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 ? While these questions have begun to be explored [6], we are still quite far from having robust and widely applicable techniques for handling data quality in OBDA.

The PhD position 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.

Position & Research Environment

The PhD studentship is a three-year full-time position, with an approximate monthly net salary of 1700€. The position does not include any teaching obligations, but there are opportunities to engage in teaching if desired.

The PhD position is currently available and will remain open until a suitable candidate is found. We welcome applications from students who are currently attending a Master's program, and we can propose a five-month Master's internship on a related topic.

The PhD thesis will be co-supervised by **Meghyn Bienvenu** (LaBRI, Bordeaux) and **Camille Bourgaux** (DI ENS, Paris). The position will be based in Bordeaux in the LaBRI research lab, with regular funded stays in Paris to visit the co-supervisor.

LaBRI (Laboratoire Bordelais de Recherche en Informatique) is a computer science research lab located on the University of Bordeaux Talence campus, which can be easily reached from the city center of Bordeaux by tram. The PhD student will participate in the new research group **RATIO** (Reasoning with data, knowledge and constraints).

Research Topic

The PhD position will focus on the development of a customized user-sensitive approach to data quality in OBDA, in which users can give direction on how to address data quality issues. Inconsistency management policies [7] have been introduced for relational databases to give users control over how errors are resolved, based upon their knowledge, preferences, and intended use of the data. It is appealing to consider such policies for the OBDA setting, but existing definitions and results do not readily transfer.

The first step will be to define a suitable notion of policy and examine its basic properties. Afterwards, the PhD student will develop novel reasoning services and associated reasoning algorithms for managing such policies: How to determine if a policy is well defined, and if it is guaranteed to yield a unique result? How can we aid users in constructing such policies, e.g. by suggesting refinements?

Candidate Profile

At the start of the PhD, the candidate must hold a Master's degree in computer science (or possibly mathematics, if accompanied by relevant computer science experience). As ontologies are expressed using logic-based formalisms, candidates should be familiar and comfortable with first-order logic.

Prior knowledge in one or more of the following areas would be a plus: knowledge representation and reasoning (especially description logics), database theory, Semantic Web (ontologies), theoretical computer science (in particular, computational complexity).

Strong English language skills (reading, writing, & speaking) are expected, but knowledge of French is not required. The working language can be either French or English.

Contact

Potential candidates, either for the PhD position or for a related Master's internship, should contact the two supervisors by email:

- Meghyn Bienvenu (meghyn.bienvenu@labri.fr)
- Camille Bourgaux (camille.Bourgaux@ens.fr)

The email should include a detailed CV and a short description of how the PhD 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. Zakharyaschev. 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] F. Parisi, A. Pugliese, G. Simari, V. S. Subrahmanian and M. V. Martinez, "Policy-based inconsistency management in relational databases," International Journal of Approximate Reasoning, vol. 55, no. 2, pp. 501-528, 2014.