

Postdoc Position: Principled Approaches to Handling Imperfect Data

Keywords

Inconsistent, incomplete, & uncertain information, data quality, knowledge representation & reasoning, database theory, ontology-mediated query answering

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that real-world data is plagued with quality issues, such as incompleteness (missing information) and errors (false or outdated information).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. While OMQA systems are growing in maturity [4], 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? While these questions have begun to be explored [5,6], we are still far from having robust and widely applicable techniques for handling data quality in OMQA.

This postdoc 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 is funded by the French National Research Agency (ANR) and the University of Bordeaux.

Position & Research Environment

This is a two-year full-time postdoctoral researcher position, with an approximate monthly net salary of 2100€. The position does not include any mandatory teaching or supervisory activities, but there will be opportunities to engage in teaching and/or student supervision if desired.

The position will be based at LaBRI (Laboratoire Bordelais de Recherche en Informatique), a computer science research lab located on the University of Bordeaux's Talence campus. The lab can be easily reached from the city center of Bordeaux by tram or bike.

The postdoctoral researcher will participate in the new research team RATIO (Reasoning with data, knowledge and constraints), which is part of a much larger Formal Methods

department that brings together dozens of researchers interested in applying logical and automata-theoretic methods to a wide range of problems within computer science.

The postdoctoral researcher will collaborate with Meghyn Bienvenu (leader of the chair) and will have many opportunities to interact with the other INTENDED participants and external collaborators. The chair provides significant resources to finance trips to conferences and to host visiting researchers.

Research Topic

The postdoctoral researcher will engage in foundational research related to the INTENDED Chair's overall theme of developing intelligent, knowledge-based methods for handling imperfect data. Generally speaking, this will involve the definition of formal frameworks for repairing and querying imperfect data, the study of the computational complexity of reasoning, and the development of reasoning algorithms.

The precise research topic will take into account the background and interests of the postdoctoral researcher, so brilliant candidates with an interest for the overall project topic should not hesitate to apply.

Candidate Profile

Applicants should hold (or be close to completing) a PhD degree in computer science and should have a strong research record in knowledge representation and reasoning or database theory (ideally, demonstrated by publications in a top-tier venues).

Familiarity with logic and/or ontology languages is desirable.

Experience and interest in one or more of the following areas would be relevant and welcome: description logics, Semantic Web, automated reasoning, inconsistency handling, data integration, data quality, theoretical computer science (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

Interested candidates should contact Meghyn Bienvenu (meghyn.bienvenu@labri.fr) and provide a detailed CV (with publication list and names of 2-3 references) and a brief description of how the topics relate to their prior experience and research interests.

The position will remain open until a suitable candidate is found, and there is no fixed starting date (preferably early to mid 2022).

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] 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.
- [5] 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.
- [6] M. Bienvenu: A Short Survey on Inconsistency Handling in Ontology-Mediated Query Answering. Künstliche Intelligenz 34(4): 443-451, 2020.