

Postdoc Position: Principled Approaches to Handling Imperfect Data

Context

Accessing the relevant information 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that real-world data is plagued with quality issues, such as incompleteness (missing information) and errors (false or outdated information).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 & Environment

This is a full-time postdoctoral researcher position. The precise duration of the contract and salary will depend upon the experience of the researcher, but will be either 13 or 14 months with a monthly net salary between 2200€ and 2400€. 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. Located in southwestern France, Bordeaux offers easy access to both the Atlantic coast and the Pyrenees mountains, while being only a two-hour train ride from Paris. Famed for its vineyards, local gastronomy, and well-preserved historic center (a UNESCO World Heritage Site), Bordeaux offers a high quality of life.

The postdoctoral researcher will participate in the 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 research related to the INTENDED Chair's overall theme of developing intelligent, knowledge-based methods for handling imperfect data, which may target either classical or ontology-enriched databases. 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. Examples of ongoing research directions include the integration of qualitative and quantitative preferences into repair-based semantics [7,8], the development of logic-based approaches to entity resolution [9,10], and the definition of holistic frameworks for jointly tackling various data quality issues [11]. We are also interested in implementing our approaches, using SAT solvers [12] and ASP systems [13].

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. In particular, while this position is primarily concerned with foundational aspects of data quality, we also welcome applications from researchers who have solid experience with reasoning and/or database systems and who are interested in developing and implementing efficient algorithms and optimizations for data quality.

Candidate Profile

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

How to Apply

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.

Applications received on or before September 25, 2024 will be given full consideration. The position will remain open until a suitable candidate is found, but a starting date in late 2024 or early 2025 would be preferred.

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 *Proc. 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.
- [7] M. Bienvenu, C. Bourgaux: **Querying and Repairing Inconsistent Prioritized Knowledge Bases: Complexity Analysis and Links with Abstract Argumentation**. In *Proc. of the 17th International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 141-151, 2020.
- [8] M. Bienvenu, C. Bourgaux, R. Jean. **Cost-Based Semantics for Querying Inconsistent Weighted Knowledge Bases**. *Proc. of the 21st International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 2024.
- [9] M. Bienvenu, G. Cima, V. Gutierrez-Basulto. **LACE: A Logical Approach to Collective Entity Resolution**. *Proc. of the 41st ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (PODS)*, 2022.
- [10] M. Bienvenu, G. Cima, V. Gutierrez-Basulto, Y. Ibanez-Garcia. **Combining Global and Local Merges in Logic-based Entity Resolution**. *Proc. of the 20th International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 2023.
- [11] M. Bienvenu, G. Cima, V. Gutierrez-Basulto. **REPLACE: A Logical Framework for Combining Collective Entity Resolution and Repairing**. *Proc. of the 32nd International Joint Conference on Artificial Intelligence (IJCAI)*, 2023.
- [12] M. Bienvenu, C. Bourgaux: **Querying Inconsistent Prioritized Data with ORBITS: Algorithms, Implementation, and Experiments**. *Proc. of the 19th International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 2022.
- [13] Z. Xiang, M. Bienvenu, G. Cima, V. Gutierrez-Basulto, Y. Ibanez-Garcia: **ASPEn: ASP-Based System for Collective Entity Resolution**. *Proc. of the 21st International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 2024.