M2 internship (ENS-PSL & Inria Paris) related to logic-based methods for AI and databases

Where: ENS-PSL, Paris

With: Michaël Thomazo (ENS-PSL, Inria) and Meghyn Bienvenu (CNRS, Université de Bordeaux)

When: 5/6 months, starting from February 2026 (flexible)

"Gratification". Follow-up PhD funding (at ENS) secured.

Setting and Motivation A common issue when querying data is that the set of query answers is too large to be examined, let alone understood, by a human, making it more reasonable to present users with a small subset of the answers. It is typically desirable to select a diverse set of query answers, which provide a representative summary of the answer set, rather than choosing a random or arbitrary set, which might contain several highly similar query results. This is especially important when query results are used to support decision making (e.g. one would like to see a variety of different travel itineraries or product models among those satisfying the queried criteria) or data exploration (e.g. once a user has viewed a small, initial set of results, they may request further tuples that are (dis)similar to particular output tuples). Query answer diversity has been explored, both from practical and theoretical angles, by researchers in the database community and remains an active research topic [1,2,3]. Similar notions have also been considered in other areas, such as information retrieval and reasoning systems (SAT, ASP) [4]. Ontology-based data access (OBDA [5], see [6,7] for introductions) is a paradigm that enables users to query heterogeneous data sources through a conceptual model, also known as ontology, without needing to know the details of the underlying data sources. In addition to providing a high-level unified vocabulary for query formulation, the ontology captures domain knowledge, which is exploited by reasoning algorithms to compute the set of all certain query answers (i.e. those that can be inferred by combining information from the data sources and ontology). Although the issue of large result sets arises also in the OBDA setting (especially since the ontology can lead to many additional answers being inferred), the problem of providing diverse answers has not been studied so far in the context of OBDA, and we aim at filling this gap with this internship.

Internship Topic and Goals The first step in the internship will be to gain familiarity with existing approaches for query answer diversity that have been proposed in databases and to adapt them in the simplest possible way to fit the OBDA setting. In a second step, the intern will explore how to define novel notions of diversity specifically targeting the OBDA setting, which will take the ontology axioms and reasoning steps into account. The internship (and the possible follow-up PhD thesis) will aim to make conceptual, algorithmic, and complexity-theoretic contributions: proposing and comparing new formal definitions for query answer diversity in OBDA, devising or adapting algorithms to compute diverse query answers, and analyzing the complexity of the associated computational tasks to identify tractable cases and sources of intractability.

If both the student and supervisors are happy with the collaboration, a PhD funding (at ENS Ulm, Paris) is available to further the study.

Pre-requisites Familiarity and interest in first-order logic, notions of complexity theory. Prior exposure to knowledge representation formalisms, automated reasoning, and/or database theory would be a plus, but is not mandatory.

How to apply   Contact Meghyn Bienvenu (first.last at u-bordeaux.fr) and Michaël Thomazo (first.last at inria.fr), with:

- a CV

- the list of courses followed that have a link with the internship topic;

- a transcript of grades available so far.

References

1 T. Deng, W. Fan. On the complexity of query result diversification. ACM Transactions on Database Systems (TODS), 2014.

2 K. Zheng, H. Wang, Z. Qi, J. Li, and H. Gao. A survey of query result diversification. Knowledge and Information Systems, 2017.

3 T. C. Merkl, R. Pichler, and S. Skritek. Diversity of answers to conjunctive queries. Logical Methods in Computer Science (LMCS), 2025.

4 T. Eiter, E. Erdem, H. Erdogan, M. Fink. Finding similar/diverse solutions in answer set programming. Theory and Practice of Logic Programming (TPLP), 2013.

5 A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, R. Rosati. Linking data to ontologies. Journal of Data Semantics, 2008.

6 M. Bienvenu, M. Ortiz: Ontology-mediated query answering with data-tractable description logics. Reasoning Web, 2015.

7 G. Xiao, D. Calvanese, R. Kontchakov, D. Lembo, A. Poggi, R. Rosati, M. Zakharyaschev. Ontology-based data access: A survey. In Proceedings of IJCAI, 2018.