

PhD position: Investigating Rich Conflict-Tolerant Semantics and Inconsistency Measures for Imperfect databases

Fully funded PhD position in AI & Databases

Place: Artois University (Lens-France)

Application Deadline: November 30, 2023

Duration: Three years

Funding: French National Research Foundation

Key words: Databases, Artificial Intelligence, Knowledge Representation and Reasoning

The ANR project EXPIDA¹ (EXplainable and parsimonious Preference models to get the most out of Inconsistent DATabases) aims to develop principled and rigorous explainable techniques for dealing with imperfect data. More precisely, EXPIDA aims to design tractable methods for dealing with conflicts in databases by efficiently exploring novel inconsistency-tolerant semantics and quantifying contradictions [6] to answer queries and to draw (high level) explanatory information. While the set of repairs (maximal consistent datasets) is often large for real databases, we aim to explore preference mechanisms (e.g., [4]) in order to retrieve meaningful answers and explanations to identify the reasons of query answers, and to assist end-users to “realize” query outputs. The EXPIDA project aims in addition to be useful for applications intensively relying on multiple heterogeneous data sources. Many such applications are nowadays developed in various domains such as transportation control, health management, social network analysis, data journalism, etc. This research project will advance the state-of-the-art in two major ways: innovations in inconsistency management, preferences and explanation for databases, and developing practical Artificial Intelligence tools for managing inconsistent databases with validations on real data.

This PhD thesis will focus on the following two main aspects. The first one concerns query answering semantics with inconsistent databases. Despite the fact that various methods have been studied, for drawing useful information in presence of conflicts in the propositional and description logic settings [2, 5], we

¹Project website: <https://www.cril.univ-artois.fr/expida/index.html>

are not aware of any existing work studying rich conflict-tolerant relations in the context of databases equipped with flexible preference relations [7], such as the partially ordered relation (e.g., [1, 3]). Then, investigating database query answering under such varying conflict-tolerant methods should be accompanied by studying the computational complexity of the different related problems, which is among the main objectives of this thesis.

The second objective of the thesis is how to handle preferences in conflicting databases. In fact, to answer queries over conflicting databases, it is crucial to express priority among potential repairs in order to select the most optimal candidates. As the number of potential repairs can be (very) large, one may choose to rank repairs according to some preference criteria, and select a small number of the most desirable repairs. Moreover, we note that the preference among the sources of data could be of different nature. Often, conflict-tolerant methods aim to find a stratification inducing a total preorder among all pieces of information. Such a stratification allows to handle more easily inconsistencies in data sources. Nonetheless, this can lead to a comparison of incomparable and independent pieces of information. The main objective here is then to develop a new framework to handle preferences in conflicting databases in order to draw meaningful answers to user queries.

The PhD thesis will be supervised by Said Jabbour (Artois University, CRIL), in collaboration with Badran Raddaoui (Telecom SudParis, SAMOVAR lab) and Yue Ma (Paris Saclay University, LISN).

The application must include:

- Curriculum vitae.
- Transcripts and diplomas for bachelor’s and master’s degrees.
- Cover letter with personal motivation and relevance w.r.t. the requirements of the position.
- Recommendation letters or contact information of at least two referees.

Applications should be submitted via email to Said Jabbour (jabbour@cril.fr), Badran Raddaoui (badran.raddaoui@telecom-sudparis.eu) and Yue Ma (ma@lri.fr) with the subject “Application for EXPIDA PhD 1”.

References

- [1] Sihem Belabbès, Salem Benferhat, and Jan Chomicki. Handling inconsistency in partially preordered ontologies: the elect method. *J. Log. Comput.*, 31(5):1356–1388, 2021.
- [2] Meghyn Bienvenu. Inconsistency handling in ontology-mediated query answering: A progress report. In *DL*, 2019.
- [3] Saïd Jabbour, Yue Ma, and Badran Raddaoui. A framework for reasoning about uncertainty in ontologies. *IEEE Intell. Syst.*, 37(6):27–37, 2022.
- [4] Benny Kimelfeld, Ester Livshits, and Liat Peterfreund. Counting and enumerating preferred database repairs. *Theor. Comput. Sci.*, 837:115–157, 2020.

- [5] Sébastien Konieczny, Pierre Marquis, and Srdjan Vesic. Rational inference relations from maximal consistent subsets selection. In *IJCAI*, pages 1749–1755, 2019.
- [6] Ester Livshits and Benny Kimelfeld. The shapley value of inconsistency measures for functional dependencies. In *ICDT*, pages 15:1–15:19, 2021.
- [7] Slawek Staworko, Jan Chomicki, and Jerzy Marcinkowski. Prioritized repairing and consistent query answering in relational databases. *Ann. Math. Artif. Intell.*, 64(2-3):209–246, 2012.

PhD position: Explainable Query Answering over Inconsistent Databases with Rich Conflict-Tolerant semantics

Fully funded PhD position in AI & Databases

Place: Télécom SudParis - Polytechnic Institute of Paris (Palaiseau-France)

Application Deadline: November 30, 2023

Duration: Three years

Funding: French National Research Foundation

Key words: Databases, Artificial Intelligence, Knowledge Representation and Reasoning, Provenance

The ANR project EXPIDA¹ (EXplainable and parsimonious Preference models to get the most out of Inconsistent DATabases) aims to develop principled and rigorous explainable techniques for dealing with imperfect data. More precisely, EXPIDA aims to design tractable methods for dealing with conflicts in databases by efficiently exploring novel inconsistency-tolerant semantics and quantifying contradictions [6] to answer queries and to draw (high level) explanatory information. While the set of repairs (maximal consistent datasets) is often large for real databases, we aim to explore preference mechanisms (e.g., [5]) in order to retrieve meaningful answers and explanations to identify the reasons of query answers, and to assist end-users to “realize” query outputs. The EXPIDA project aims in addition to be useful for applications intensively relying on multiple heterogeneous data sources. Many such applications are nowadays developed in various domains such as transportation control, health management, social network analysis, data journalism, etc. This research project will advance the state-of-the-art in two major ways: innovations in inconsistency management, preferences and explanation for databases, and developing practical Artificial Intelligence tools for managing inconsistent databases with validations on real data. This PhD thesis will focus on the aspect of explainability in two ways, as presented in what follows.

First, we will consider explanations for query results over inconsistent databases with different conflict-tolerant semantics (e.g., consistent, brave, intersection repair, intersection closed repair, non-objection, nonconsensus based semantics,

¹Project website: <https://www.cril.univ-artois.fr/expida/index.html>

etc.). To this end, we will adapt the notion of lineage (or provenance) [2] in the context of uncertain/inconsistent data and devise mathematical formalisations that will provide the necessary properties for characterising and measuring the ‘quality’ of the explanations. Through the study of causality [7] and argumentation [4, 9, 8, 3] in our setting, we further aim at improving the acceptability and usefulness of the provided explanations by the end-user. Second, we will investigate the complementary problem of explaining missing query results, widely known as Why-Not explanations, which has not been yet addressed in the context of inconsistent databases. In the setting of consistent databases, Why-Not explanations ‘explain’ why certain results are not generated by a query (or a workflow) by means of instance-based (i.e., source tuples), query-based (i.e., query operators) or refinement-based explanations (i.e., corrected query). Close to our problem, [1] has proposed Why-Not provenance polynomials, which may account for probabilistic tuples. It would be interesting to check how such formalisations can be revisited to fit the inconsistent database’s different conflict-tolerant semantics.

The PhD thesis will be supervised by Badran Raddaoui (Telecom SudParis, SAMOVAR), in collaboration with Aikaterini Tzompanaki (CY Cergy University, ETIS lab) and Yue Ma (Paris Saclay University, LISN).

The application must include:

- Curriculum vitae.
- Transcripts and diplomas for bachelor’s and master’s degrees.
- Cover letter with personal motivation and relevance w.r.t. the requirements of the position.
- Recommendation letters or contact information of at least two referees.

Applications should be submitted via email to Badran RADDAOUI badran.raddaoui@telecom-sudparis.eu and Aikaterini TZOMPANAKI aikaterini.tzompanaki@cyu.fr, with the subject “Application for EXPIDA PhD 2”.

References

- [1] Nicole Bidoit, Melanie Herschel, and Katerina Tzompanaki. Immutably answering why-not questions for equivalent conjunctive queries. *Ingénierie des Systèmes d Inf.*, 20(5):27–52, 2015.
- [2] James Cheney, Laura Chiticariu, and Wang Chiew Tan. Provenance in databases: Why, how, and where. *Found. Trends Databases*, 1(4):379–474, 2009.
- [3] Kristijonas Cyras, Antonio Rago, Emanuele Albin, Pietro Baroni, and Francesca Toni. Argumentative XAI: A survey. In *IJCAI*, pages 4392–4399, 2021.
- [4] Phan Minh Dung. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. *Artif. Intell.*, 77(2):321–358, 1995.

- [5] Benny Kimelfeld, Ester Livshits, and Liat Peterfreund. Counting and enumerating preferred database repairs. *Theor. Comput. Sci.*, 837:115–157, 2020.
- [6] Ester Livshits and Benny Kimelfeld. The shapley value of inconsistency measures for functional dependencies. In *ICDT*, pages 15:1–15:19, 2021.
- [7] Alexandra Meliou, Wolfgang Gatterbauer, Katherine F. Moore, and Dan Suciu. The complexity of causality and responsibility for query answers and non-answers. *Proc. VLDB Endow.*, 4(1):34–45, 2010.
- [8] Antonio Rago, Oana Cocarascu, Christos Bechlivanidis, David A. Lagnado, and Francesca Toni. Argumentative explanations for interactive recommendations. *Artif. Intell.*, 296:103506, 2021.
- [9] Antonio Rago, Hengzhi Li, and Francesca Toni. Interactive explanations by conflict resolution via argumentative exchanges. In *KR*, pages 582–592, 2023.