Offer #2025-08760

PhD Position F/M Computational approaches for knowledge graph mining and completion dealing with uncertainty

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

About the research centre or Inria department

The Inria centre at Université Côte d'Azur includes 42 research teams and 9 support services. The centre's staff (about 500 people) is made up of scientists of di?erent nationalities, engineers, technicians and administrative staff. The teams are mainly located on the university campuses of Sophia Antipolis and Nice as well as Montpellier, in close collaboration with research and higher education laboratories and establishments (Université Côte d'Azur, CNRS, INRAE, INSERM ...), but also with the regiona economic players.

With a presence in the fields of computational neuroscience and biology, data science and modeling, software engineering and certification, as well as collaborative robotics, the Inria Centre at Université Côte d'Azur is a major player in terms of scientific excellence through its results and collaborations at both European and international levels.

Context

This PhD thesis takes place within the MetaboLinkAI ANR-SNF project, which aspires to revolutionize the analysis and interpretation of metabolomics data through a multidisciplinary approach that combines a comprehensive knowledge graph hub (MetaKH) with cutting-edge artificial intelligence (AI) and machine learning (ML) techniques. The project's main goals are to enhance the querying and ease of use of metabolomics data, improve research efficiency, and stimulate creativity in the field. These objectives are set to surpass current standards by creating an encyclopedic and expandable knowledge base, integrating advanced AI to handle the uncertainties of experimental data, and enabling a broader range of hypothesis testing and evaluation.

Within this project, we will focus on developing innovative methodologies and tools, such as graph mining methods, to enhance data interaction, analysis capabilities, and representation of uncertainty.

One distinctive peculiarity of metabolomics data (and thus MetaKH) is incompleteness, variable confidence and inherent uncertainty. Here, we adopt AI to enhance the completeness and reliability of the KG and to correctly account for uncertainty.

Assignment

Computational approaches for graph mining and completion

Because of the uncertain nature of metabolomics data and associated knowledge, MetaKH will be largely incomplete and partly incorrect. Therefore, it will be crucial to develop a comprehensive computational framework to enhance the quality, completeness and validity to eventually increase the quality of any processing using MetaKH. We propose to adapt heuristic methods and algorithms to discover/induce topological motifs, axioms (OWL), rules (SWRL or SPARQL) or shapes (SHACL) from knowledge graphs (TBox construction/refinement). These will account for the possible uncertainty of knowledge represented in the ABox (as defined in WP3.2). Expert-in-the-loop techniques will also be considered. We will design algorithms and data structures to allow KG queries at different levels of data granularity. The methods will exploit heuristics derived from expert knowledge in combination with semi-succinct and, where needed, approximated data structures. In parallel, we will work on methods for knowledge graph completion, correction and enrichment, to enhance quality and content (ABox refinement). The developed methods will combine deductive reasoning (including analogic), SHACL validation, and link prediction and retraction based on KG embeddings. They will take into account the uncertainty of knowledge as defined in WP3.2. Evaluation will be done by measuring the improvement of KG completeness and validity, and the effectiveness of reasoning by corrupting the KG by adding/removing/perturbing some edges, applying completion/inference/querying, and assessing the impact in comparison with the original KG.

Dealing with (lack of) confidence in KGs

The objective is to develop and integrate a sophisticated framework into semantic web standards for formal representation and reasoning of uncertainty (both ontic and epistemic) in MetaKH, improving data confidence and decision-making processes. Initially, we will review literature to identify adequate models to represent ontic uncertainty (certainly probability theory) and epistemic uncertainty (e.g. possibility theory, Dempster Shafer theory) adequate to represent mass spectrometry observations and metabolomic knowledge. Based on such models, we will propose extensions to Semantic Web standards to express uncertainty, provenance, and temporality metadata, facilitating richer data interpretation and trustworthiness. We will develop algorithms to integrate uncertainty in querying, deduction and embedding in KGs. We will establish criteria for using KGs based on uncertainty and provenance metadata, as well as other types of metadata, enabling users and agents to make informed decisions regarding trust and data application. Algorithms developed in WP3.1 will be extended to integrate uncertainty. Finally, we plan to implement mechanisms for evaluating KG completeness, validity, and reasoning under uncertainty, incorporating expert

feedback and adapting methodologies based on provenance and other metaknowledge types.

Main activities

This thesis will start with a state of the art of the different domains involved, in particular graph-based knowledge representation, KG mining, uncertainty representation and management in KG.

The PhD student is expected to first address computational approaches for MetaKH mining and completion, and then extend these approaches considering the inherent uncertainty of some knowledge in MetaKH, and of the mining approaches and their results.

Expected deliverables are:
[D1] Heuristic methods, data structures and algorithms for KG querying and mining
[D2] Methods and algorithms for KG completion
[D3] Proposal of an extension of SW standards for uncertainty annotation
[D4] Implementation of uncertainty annotation in MetaKH

References

1. Ahmed El Amine Djebri. Uncertainty Management for Linked Data Reliability on the Semantic Web. PhD thesis, Université Côte d'Azur, 2022.
2. Ahmed El Amine Djebri, Andrea G. B. Tettamanzi, and Fabien Gandon. Publishing uncertainty on the semantic web: Blurring the LOD bubbles. In Graph-Based Representation and Reasoning - 24th International Conference on Conceptual Structures, ICCS 2019, Marburg, Germany, July 1-4, 2019, Proceedings, volume 11530 of Lecture Notes in Computer Science, pages 42–56. Springer, 2019.
3. Antonia Ettorre, Anna Bobasheva, Catherine Faron, and Franck Michel. A systematic approach to identify the information captured by knowledge graph embeddings. In WI-IAT '21 : IEEE/WIC/ACM International Conference on Web Intelligence, Melbourne VIC Australia, December 14 - 17, 2021, pages 617–622. ACM, 2021.
4. Rémi Felin. Evolutionary knowledge discovery from RDF data graphs. PhD thesis, Université Côte d'Azur, 2024.
5. Rémi Felin, Catherine Faron, and Andrea G. B. Tettamanzi. A framework to include and exploit probabilistic information in SHACL validation reports. In The Semantic Web - 20th International Conference, ESWC 2023, Hersonissos, Crete, Greece, May 28 - June 1, 2023, Proceedings, volume 13870 of Lecture Notes in Computer Science, pages 91–104. Springer, 2023.
6. Rémi Felin, Pierre Monnin, Catherine Faron, and Andrea G. B. Tettamanzi. An Algorithm Based on Grammatical Evolution for Discovering SHACL Constraints. In EuroGP 2024 - 27th European Conference on Genetic Programming, Genetic Programming – 27th European Conference, EuroGP 2024, Aberystwyth, United Kingdom, April 2024.
7. Thu Huong Nguyen. Mining the semantic Web for OWL axioms. PhD thesis, University of Côte d'Azur, 2021.

8. Andrea G. B. Tettamanzi, Catherine Faron-Zucker, and Fabien Gandon. Possibilistic testing of OWL axioms against RDF data. Int. J. Approx. Reason. , 91 :114–130 , 2017.

Skills

The candidate must hold a Master degree in Informatics / Computer science and must demonstrate
aptitudes or matches with most of the following aspects:

- Competencies and skills in Semantic Web standards and technologies
- Competencies and skills in querying and mining Knowledge Graphs
- High motivation for scientific research in an open science context
- Good development skills
- Writing skills and publication motivation
- Good English oral and writing skills

Soft skills:

- Aptitude to work with others and engage in collaborations
- Autonomy and creativity
- Remote working capabilities (emails, collaborative tools, etc.)

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Remuneration

Gross Salary: 1st year : 2200 € per month, 2nd and 3rd year : 2300 €per month

General Information

- **Theme/Domain :** Data and Knowledge Representation and Processing Information system (BAP E)
- **Town/city :** Sophia Antipolis
- **Inria Center :** Centre Inria d'Université Côte d'Azur
- **Starting date :** 2025-09-01

- **Duration of contract :** 3 years, 1 month
- **Deadline to apply :** 2025-07-31

Contacts

- **Inria Team :** WIMMICS
- **PhD Supervisor :**
  Faron Catherine / Catherine.Faron@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

The keys to success

There you can provide a "broad outline" of the collaborator you are looking for what you consider to be necessary and sufficient, and which may combine :

- tastes and appetencies,
- area of excellence,
- personality or character traits,
- cross-disciplinary knowledge and expertise...

This section enables the more formal list of skills to be completed and 'lightened' (reduced) :

- "Essential qualities in order to fulfil this assignment are feeling at ease in an environment of scientific dynamics and wanting to learn and listen."
- " Passionate about innovation, with expertise in Ruby on Rails development and strong influencing skills. A thesis in the field of **** is a real asset."

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

Instruction to apply

Defence Security :
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST).Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012

relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :
As part of its diversity policy, all Inria positions are accessible to people with disabilities.