Inria

Offre n°2025-08760

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

Type de contrat : Fixed-term contract

Niveau de diplôme exigé : Graduate degree or equivalent

Fonction : PhD Position

A propos du centre ou de la direction fonctionnelle

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.

Contexte et atouts du poste

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.

Mission confiée

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.

Principales activités

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.

Compétences

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.)

Avantages

- 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

Rémunération

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

Informations générales

- **Thème/Domaine :** Data and Knowledge Representation and Processing Information system (BAP E)
- **Ville :** Sophia Antipolis
- **Centre Inria :** Centre Inria d'Université Côte d'Azur
- **Date de prise de fonction souhaitée :** 2025-09-01

- **Durée de contrat :** 3 years, 1 month
- **Date limite pour postuler :** 2025-07-31

Contacts

- **Équipe Inria :** WIMMICS
- **Directeur de thèse :**
  Faron Catherine / Catherine.Faron@inria.fr

A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 215 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3900 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 200 start-up. L'institut s'e?orce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

L'essentiel pour réussir

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."

> **Attention**: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

Consignes pour postuler

Sécurité défense :
Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel

scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :
Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.