Offre n°2025-08684

Research engineer position on methods and tools for the construction, maintenance and querying of a decentralized knowledge hub in metabolomics

Le descriptif de l'offre ci-dessous est en Anglais

Type de contrat : CDD

Niveau de diplôme exigé : Thèse ou équivalent

Fonction : Ingénieur scientifique contractuel

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 research engineer position takes place within the context of the ANR-SNF MetaboLinkAI project, which aspires to revolutionize the analysis and interpretation of metabolomics data through a multidisciplinary approach that combines a comprehensive knowledge 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 context, this position will focus on the construction and querying of MetaKH, a decentralized, machine-readable knowledge hub federating and linking (1) pre-existing public knowledge and resources relevant for the use cases of the project (e.g. chemical entities description, biochemical pathways, metabolites information, relevant literature), (2) possibly newly created resources or the semantic lifting of existing resources not available in Semantic Web standards, and (3) and mass spectrometry datasets.

Supervisors: Franck Michel, Catherine Faron, Fabien Gandon (University Côte d'Azur, Inria, CNRS)

Mission confiée

The research engineer will be involved in two major contributions of the 2nd work package: "Knowledge representation and management".

First, the research engineer will participate in the creation of a portal and pipeline to support the lifecycle of MetaKH.

Second, the research engineer will take part in the design of a federated query engine capable of querying the distributed knowledge hub, and allowing the service to answer complex, high-level biological questions exploiting decentralized data sources.

In the course of this position, the engineer will collaborate with PhD and postdoc researchers working on the development of AI methods aiming to deal with uncertainty in the data, mine and complement the knowledge hub, and develop an AI research assistant using natural language as an interface to data and knowledge.

Principales activités

Creation of a portal and pipeline to support the lifecycle of MetaKH

The portal must allow users to incrementally integrate, monitor and update reference resources in the knowledge federation (e.g. ChEBI, PubChem, Rhea, SwissLipids, MetaNetX, Pathway Commons, FORUM). This shall involve multiple tasks:

- The development of a domain-specific model to link semantic resources throughout the federation while supporting lack of precision and uncertainty.
- The development and management of a collection of mappings and links between heterogeneous resources. Methods for writing those mappings and links shall range from handcrafting to generative AI models. A git-based life-cycle similar to that of code shall be applied to the produced resources (versioning, issues, publication, continuous integration etc.)
- The continuous monitoring of the integrated resources (typically to integrate new releases).
- The deployment and maintenance of self-hosted mirroring of critical resources.

All of this shall be achieved within the respect of the FAIR principles.

Design of a federated query engine

Designed as a single data access point hiding the federation's complexity from the users, the query engine will leverage the mappings and links across resources (from the first contribution) to dynamically rewrite and expand SPARQL queries so as to query and integrate the multiple knowledge graphs (KG) at runtime.

This shall involve the construction of an index of the federated KGs, possibly reusing and extending the IndeGx framework [Maillot et al, 2023], and the computation of information relevant for writing federated queries such as KG summaries [Aimonier-Davat et al 2024].

Since the goal is to provide an architecture that is scalable, resource efficient, and sustainable in the long-term, an important aspect in this approach will be the level of mapping expressivity to be considered for a trade-off between runtime efficiency and completeness of the results.

[Maillot et al, 2023] IndeGx: A Model and a Framework for Indexing RDF Knowledge Graphs with SPARQL-based Test Suits. Pierre Maillot, Olivier Corby, Catherine Faron, Fabien Gandon, Franck Michel. Journal of Web Semantics, 2023. DOI: ?10.1016/j.websem.2023.100775?. ?hal-03946680?

[Aimonier-Davat et al 2024]. FedUP: Querying Large-Scale Federations of SPARQL Endpoints. Julien Aimonier-Davat, Minh-Hoang Dang, Pascal Molli, Brice Nédelec, Hala Skaf-Molli. *The ACM Web Conference 2024 (WWW '24)*, May 2024, Singapore, Singapore. ?10.1145/3589334.3645704?. ?hal-04538238?

Compétences

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

- Strong experience with Semantic Web standards and technologies
- Experience in distributed data management, querying, crawling, indexing, federating, etc.
- High motivation for scientific research in an open science context
- Good Web development technical skills with knowledge of JavaScript and modern JS frameworks (Node.js, Reactive.js…), REST/RESTful Web services, JSON
- Background knowledge and/or experience in life sciences, biology, metabolomics
- Data science and management expertise
- Language: excellent English oral and writing skills

Other appreciated skills:

- Writing skills and motivation for publication
- Aptitude to work with others and engage in collaborations
- Autonomy and initiative, take on technical decisions within the project and justification of choices
- Remote working capabilities (emails, collaborative tools, trackers, 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

From 2692 € gross monthly (according to degree and experience).

Informations générales

- **Thème/Domaine :** Représentation et traitement des données et des connaissances
  Ingénierie logicielle (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 ans
- **Date limite pour postuler :** 2025-04-26

Contacts

- **Équipe Inria :** WIMMICS
- **Recruteur :**
  Michel Franck / franck.michel@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.

> **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.