



Offre n°2023-06824

Post-Doctoral Research Visit F/M Post-Doctoral Research Visit F/M Information extraction in specialty-language domains CLEE

Type de contrat : Fixed-term contract

Niveau de diplôme exigé : PhD or equivalent

Fonction : Post-Doctoral Research Visit

Contexte et atouts du poste

This post-doc position fits within the roadmap activities of Inria's Defense & Security Department, which is devoted to applications-driven research.

Among the various fields of NLP, information extraction is a crossroad topic that, by focusing on how to turn raw documents into structured data models, echoes the practical needs of many end-users in a broad range of sectors. Information extraction components such as entity recognition or relation extraction are thus key to a number of industrial and general-public applications.

However, whereas information extraction has seen major progresses in the last few years on common language (Wikipedia, news, everyday language), it still lags behind on specialty language, which effectively affects a number of practical applications. Main challenges include unknown words and concepts, unusual phrasings, or differences in the nature of information that is interesting to extract.

The goal of this post-doc is to bridge that gap by developing new methods that enable to model and account for the specificities of a given domain with specialty language, while still benefitting from the models and capabilities developed for the common language.

The first specialty-language domain that has been identified as a test bed for the developed approaches is the scientific literature on chemistry (e.g. ChemRxiv papers). Other domains that are considered for experimentation throughout the period are cybersecurity (e.g. technical documentation) and geopolitics. Inspiration can be drawn from existing work on biomedical NLP, but that domain is not expected to be at the core of the work.

Work can include direct collaboration with other academic or industrial partners of the department.

Mission confiée

The post-doc will focus on developing new algorithmic methods along the following research tracks:

- Automated terminology and concepts extraction
- Identification of new relations that are specific to a domain
- Adapting models (in particular embedding models) to account for extended vocabulary
- Semi-supervised learning to leverage a small amount of in-domain annotations

Special care will be given to the transferability of the methods to other specialty domains, rather than developing approaches that are tailored to one particular domain.

Principales activités

The post-doc will focus on developing new algorithmic methods along the following research tracks:

- Automated terminology and concepts extraction
- Identification of new relations that are specific to a domain
- Adapting models (in particular embedding models) to account for extended vocabulary
- Semi-supervised learning to leverage a small amount of in-domain annotations

Special care will be given to the transferability of the methods to other specialty domains, rather than developing approaches that are tailored to one particular domain.

Compétences

- Holding a PhD (or about to defend) in Natural Language Processing, Computational Linguistics or Computer Science with a specialization in Machine Learning
- Theoretical and practical knowledge of deep learning, as well as traditional machine learning.

- Experience with knowledge-driven or hybrid AI would be appreciated.
- Prior experience on at least one of the following topics: information extraction, semi-supervised learning, domain adaptation, low-resourced NLP
- Strong programming skills (at least Python, git, Linux environment)
- Fluency in English. Knowledge or interest for the French language. Knowledge of a second foreign language would be appreciated.

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 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

Informations générales

- **Ville** : PARIS ou ROCQUENCOURT ou SACLAY
- **Centre Inria** : [Siège](#)
- **Date de prise de fonction souhaitée** : 2024-06-01
- **Durée de contrat** : 2 years
- **Date limite pour postuler** : 2024-06-20

Contacts

- **Équipe Inria** : MIS-DEFENSE (DIRECTION)
- **Recruteur** :
Maillet Florence / florence.maillet@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'efforce 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.