



Offer #2023-06824

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

Contract type : Fixed-term contract

Level of qualifications required : PhD or equivalent

Fonction : Post-Doctoral Research Visit

Context

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.

Assignment

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.

Main activities

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.

Skills

- 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.

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 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

General Information

- **Town/city** : PARIS ou ROCQUENCOURT ou SACLAY
- **Inria Center** : [Siège](#)
- **Starting date** : 2024-06-01
- **Duration of contract** : 2 years
- **Deadline to apply** : 2024-06-20

Contacts

- **Inria Team** : MIS-DEFENSE (DIRECTION)
- **Recruiter** :
Maillet Florence / florence.maillet@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.

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.