

PROJET : TAL-BIOInfo (CNU 27)

Informatique :

Traitement automatique des textes biomédicaux, Analyse et intégration de données biologiques

Computer Science:

Biomedical language processing, Analysis and integration of biological data

ARGUMENTAIRES

Enseignement

L'enseignement est l'une des missions fondatrices de l'Université.

À ce titre, la personne recrutée démontrera sa capacité à relier des séquences d'enseignement et des modalités pédagogiques à des objectifs explicites d'apprentissage et d'acquisition de compétences.

La personne recrutée pourra enseigner dans toutes les filières relevant du département d'informatique de la faculté des sciences d'Orsay, aux niveaux licence et master (traditionnel et apprentissage).

Elle enseignera dans ses domaines d'expertise mais pourra également être sollicitée pour renforcer le potentiel d'enseignement du département dans les domaines de base de l'informatique.

Elle pourra être amenée à dispenser une partie de ses enseignements en anglais, notamment dans le cadre de masters mais également en licence.

Au cours de sa carrière, la personne recrutée sera également amenée à participer à la vie du département d'informatique et à prendre des responsabilités (responsabilité de filière, participation aux structures de la Faculté des sciences et de l'Université, contribution au pilotage de projets).

Une partie de l'audition portera sur une proposition de plan de cours niveau licence en lien avec les thématiques de la fiche de poste.

Recherche

Le Laboratoire Interdisciplinaire des Sciences du Numérique (LISN, UMR9015) est un laboratoire fortement pluridisciplinaire situé à Orsay. Il est structuré en cinq départements. La personne recrutée travaillera à l'interface des équipes Information, Langue Ecrite et Signée (ILES : traitement automatique des langues) du département Sciences et Technologie des Langues et de l'équipe Bioinformatique du département Sciences des Données. Elle s'intégrera prioritairement sur les thèmes du traitement des textes biomédicaux et/ou de l'analyse et intégration de données biologiques. Elle y apportera des compétences qui pourront être (liste non exhaustive) en traitement automatique de la langue écrite, en représentation des connaissances, en apprentissage automatique, en algorithmique, en statistiques.

Seront aussi considérés des profils de personnes s'intégrant plus généralement dans l'équipe Information, Langue Ecrite et Signée (ILES) ou dans l'équipe Bioinformatique ainsi que des profils avec une appétence forte pour le traitement automatique des langues et la bioinformatique et s'intégrant dans le département Sciences et Technologie des Langues ou Sciences des Données.

L'objectif général du projet dans lequel s'inscrit ce poste est de construire des ponts entre deux domaines : analyse de textes biomédicaux et intégration de données biologiques. La problématique de l'analyse et de l'intégration de données et de connaissances scientifiques, sous la forme à la fois de données structurées et de textes, est particulièrement prégnante dans le domaine biomédical où la production de données (expériences

biologiques, essais cliniques, dossiers médicaux) et de connaissances (publications) s'est accélérée depuis plusieurs années. Un verrou récurrent pour obtenir une interprétation pertinente du contenu de ces textes spécialisés est de savoir plonger mots, données et connaissances structurées dans un espace commun. Un axe de recherche est l'apprentissage de représentations par des architectures à réseaux de neurones. L'intégration de données biologiques vise quant à elle à exploiter la complémentarité de jeux de données pour permettre la découverte de nouvelles connaissances à partir de multiples jeux de données massifs et hautement hétérogènes. Ce besoin a encore récemment été mis en lumière dans la pandémie de COVID-19.

L'intégration de données nécessite de plus en plus souvent des méthodes de traitement automatique des langues pour en extraire automatiquement des informations qu'ils sont seuls à receler, comme les médicaments testés dans des essais cliniques ou les liens entre gènes et maladies dans les résumés d'articles scientifiques. Traitement automatique des langues et intégration de données biologiques ont aussi un besoin commun de conception de chaînes de traitement FAIR, transparentes et reproductibles. La communauté bio-informatique est fortement structurée autour de bonnes pratiques de ce type, qui pourraient bénéficier au traitement automatique des langues pour mieux reproduire dans le temps une expérience et grâce à la provenance, tracer les données utilisées, le paramétrage des outils, construire des plateformes scientifiques robustes et transparentes. L'enjeu de ce projet est donc de fertiliser mutuellement les domaines d'intégration de données biologiques et d'analyse de textes biomédicaux : les besoins de chaque domaine peuvent trouver des solutions dans l'autre domaine, les solutions de chaque domaine peuvent faire naître de nouveaux problèmes dans l'autre

JOB DESCRIPTION

Teaching

Teaching is one of the University's foundational missions. As such, the recruited person will demonstrate their ability to link teaching sequences and pedagogical modalities to explicit learning and skill acquisition objectives.

The recruited person may teach in all the fields of study of the Computer Science Department of the Orsay Faculty of Sciences, at Bachelor and Master levels (traditional and apprenticeship).

They will teach in their areas of expertise, but may also be called upon to strengthen the teaching potential of the department in basic computer science.

They may be required to provide part of their teaching in English, in particular in the context of Masters courses but also in Bachelor's courses.

During their career, the recruited person will also be required to participate in the life of the Computer Science Department and assume leadership positions (management of "filières", participation in the government structures of the Faculty of Sciences and of the University, contribution to project steering).

Research activities

The LISN laboratory is a multidisciplinary research laboratory in Orsay, France. It is composed of five departments and 17 teams. The recruited person will work at the interface of the Information, Written and Signed Language (ILES: Natural Language Processing) team of the Language Sciences and Technologies department and of the Bioinformatics team of the Data Science department. They will work primarily on the themes of biomedical natural language processing and/or analysis and integration of biological data. They will bring skills that may include (non-exhaustive list) natural language processing, knowledge representation, machine learning, algorithmics, statistics.

We will also consider profiles of people who fit more generally into the Information, Written and Signed Language (ILES) team or into the Bioinformatics team, as well as profiles with a strong interest in natural language processing and bioinformatics and who fit into the Language Sciences and Technologies or Data Science departments.

The general objective of the project of which this position is part is to build bridges between two areas: natural language processing of biomedical texts and integration of biological data. The problem of analyzing and integrating data and scientific knowledge, in the form of both structured data and texts, is particularly relevant in the biomedical field where the production of data (biological experiments, clinical trials, medical records) and knowledge (publications) has accelerated over the last few years. A recurrent challenge to obtain a relevant interpretation of the content of these specialized texts is to know how to embed words, data and

structured knowledge into a common space. One line of research is representation learning with neural network architectures. Biological data integration aims at exploiting the complementarity of datasets to allow the discovery of new knowledge from multiple massive and highly heterogeneous datasets. This need was recently highlighted again in the COVID-19 pandemic.

Data integration increasingly requires natural language processing methods to automatically extract information that is only contained in text, such as drugs tested in clinical trials or gene-disease relationships in scientific abstracts. Natural language processing and biological data integration also share a common need to design FAIR, transparent and reproducible processing pipelines. The bioinformatics community is strongly structured around such good practices, which could benefit natural language processing to better reproduce an experiment over time and thanks to the notion of provenance, trace the data used, the parameterization of the tools, build robust and transparent scientific platforms. The challenge of this project is therefore to mutually fertilize the fields of biological data integration and biomedical text analysis: the needs of each field can find solutions in the other, the solutions of each field can give rise to new problems in the other

Laboratoire(s) d'accueil : (sigle et intitulé détaillé) Laboratoire Interdisciplinaire des Sciences du Numérique (LISN)

Label (UMR, EA, ...)	N°	Nbre de chercheurs	Nbre d'enseignants-chercheurs
UMR	9015	55	108

CONTACTS

Enseignement : presidence-dept-info.sciences@u-psud.fr

Recherche : Pierre Zweigenbaum pz@lisn.fr et Sarah Cohen-Boulakia cohen@lri.fr

Contrat faisant suite à la réussite au concours :

Conformément au décret 2021-1710 du 17 déc. 2021 le candidat retenu sera amené à signer un contrat précisera sa date d'effet et la définition du poste occupé, ainsi que les éléments suivants :

- 1° La dénomination des fonctions exercées, celle de l'unité de recherche ou de la composante d'affectation, ainsi que celle du corps dans lequel l'agent a vocation à être titularisé ;
 - 2° La durée du contrat ;
 - 3° L'intitulé précis du projet de recherche et d'enseignement retenu qui fait l'objet de la convention de recherche et d'enseignement mentionnée à l'article 16 ;
 - 4° Les moyens garantis par l'autorité de recrutement pour la réalisation de ce projet de recherche et d'enseignement ;
 - 5° Le nom et la qualité de la personne désignée en qualité de référent scientifique ;
 - 6° Le montant de la rémunération brute mensuelle ;
 - 7° Les obligations de service d'enseignement et les objectifs à atteindre en matière de recherche ;
 - 8° Le cas échéant, les conditions particulières d'exercice de l'emploi de l'agent, notamment lorsque tout ou partie du projet de recherche et d'enseignement se déroule au sein d'un établissement partenaire.
- Dans un délai de deux mois à compter de la date de signature du contrat, la convention de recherche et d'enseignement prévue à l'article 16 du décret est annexée au contrat.

L'Université Paris-Saclay est l'une des meilleures universités françaises et européennes, à la fois par la qualité de son offre de formation et de son corps enseignant, par la visibilité et la reconnaissance internationale de ses 275 laboratoires de recherche et leurs équipes, ainsi que par l'attention apportée, au quotidien et par tous ses personnels, à l'accueil, l'accompagnement, l'interculturalité et l'épanouissement de ses 65 000 étudiants. L'université Paris-Saclay est constituée de 10 composantes universitaires, de 4 grandes écoles (Agroparistech, CentraleSupélec, Institut d'Optique Graduate School, Ens Paris-Saclay), d'un prestigieux institut de mathématiques (Institut des Hautes Études

Scientifiques) et s'appuie sur 6 des plus puissants organismes de recherche français (CEA, CNRS, Inra, Inria, Inserm et Onera). Elle est associée à deux universités (Université de Versailles Saint-Quentin-en-Yvelines et Université d'Évry Val-d'Essonne) qui fusionneront dans les années à venir et dont les campus jouxtent le territoire du plateau de Saclay et de sa vallée. Ses étudiants, ses enseignants-chercheurs, ses personnels administratifs et techniques et ses partenaires évoluent dans un environnement privilégié, à quelques kilomètres de Paris, où se développent toutes les sciences, les technologies les plus en pointe, l'excellence académique, l'agriculture, le patrimoine historique et un dynamique tissu économique. Ainsi l'Université Paris-Saclay est un établissement de premier plan implanté sur un vaste territoire où il fait bon étudier, vivre et travailler.

Site : <https://www.universite-paris-saclay.fr/fr>

Candidature via l'application GALAXIE :

<https://galaxie.enseignementsup-recherche.gouv.fr/antares/can/astree/index.jsp>