



Intitulé de l'offre de stage

Utilisation d'algorithmes d'Intelligence Artificielle (Traitement Automatique du Langage) pour l'extraction des variables significatives des compte rendus médicaux dans le cadre de la maladie de Horton

Stagiaire en	Intelligence Artificielle (Natural Language Processing)
Affectation	CHU Toulouse
Durée	5 à 6 mois
Rémunération	Selon profil
Date de la publication	01/01/2022
Date d'embauche prévue	mars ou avril 2022 suivant disponibilités
Lieu	CHU, 2 rue Charles Viguerie, 31300 Toulouse, France AKKA Technologies, 7 Blvd Henri Ziegler 31700 BLAGNAC

Le Centre Hospitalier Universitaire de Toulouse

Le Centre Hospitalier Universitaire (CHU) de Toulouse est constitué de plusieurs sites (les principaux étant les sites de Rangueil, Larrey et Purpan, ainsi que celui de l'oncopole en commun avec l'institut Claudius Régaud); il comprend 4 000 médecins et 12 000 personnels hospitaliers. La mission de recherche et d'innovation du CHU fait partie intégrante à la fois de son activité quotidienne et de sa stratégie pour l'avenir. Elle est menée en collaboration étroite avec les facultés et les organismes de recherche que sont notamment l'Institut national de la santé et de la recherche médicale (Inserm) et le Centre national de la recherche scientifique (CNRS).

L'Institut de Recherche en Informatique de Toulouse

L'Institut de Recherche en Informatique de Toulouse (IRIT), une des plus imposantes Unité Mixte de Recherche au niveau national, est l'un des piliers de la recherche en Occitanie avec ses 700 membres, permanents et non-permanents. De par son caractère multi-tutelle (CNRS, Universités toulousaines), son impact scientifique et ses interactions avec les autres domaines, le laboratoire constitue une des forces structurantes du paysage de l'informatique et de ses applications dans le monde du numérique, tant au niveau régional que national.

AKKA Research

AKKA Research est le centre de R&D et d'intégration du groupe AKKA Technologies. Entièrement dédié à l'innovation et à l'anticipation des technologies du futur, AKKA Research travaille autour de trois axes dédiés au domaine du digital (IA, Big Data, cloud computing, IoT,..), aux nouveaux systèmes énergétiques (piles à combustible, batteries, procédé hydrogène) et aux systèmes autonomes (robotique collaborative, véhicules autonomes, vision par ordinateur) en développant des briques techniques qui peuvent être appliquées dans un contexte opérationnel. Pour stimuler l'innovation, nos équipes travaillent en mode agile utilisant des méthodes comme le 'design thinking' pour rapidement créer des prototypes et tester de nouvelles idées.

L'équipe d'accueil

L'informatique, le traitement des données et l'intelligence artificielle sont appelés à prendre une place croissante dans le monde de la recherche médicale. Dans ce cadre, le CHU de Toulouse a créé une équipe de "gestion de la donnée" dont le but est le recueil, l'analyse et le traitement de la donnée médicale, en support aux activités médicales du CHU. Elle collabore activement avec l'IRIT afin d'appliquer les techniques les plus récentes de la science des données et de l'intelligence artificielle dans le domaine de la santé.

L'équipe MELODI de l'IRIT se préoccupe de recherches articulant l'analyse et la formalisation de la langue naturelle d'une part (Traitement Automatique du Langage / Natural Language Processing), la représentation et la modélisation des connaissances d'autre part. L'équipe offre un éventail d'approches de la recherche allant d'études théoriques sur les éléments contribuant à la sémantique et à sa représentation formelle, sur les interactions et les structures de discours, à des études plus expérimentales basées sur des développements logiciels, la constitution de ressources (corpus, corpus annotés, lexiques et ontologies) et des évaluations en corpus. Une partie de ses travaux se situe dans une démarche d'ingénierie des connaissances pour outiller la construction d'applications faisant appel à ces caractérisations sémantiques et à des modèles de connaissances.

Le professeur Laurent Sailer est professeur de Médecine interne, Chef de service au CHU de Toulouse, titulaire d'une thèse de sciences en pharmaco-épidémiologie, vice-président du GEFA et coordinateur de la base de données nationale.

Le stagiaire sera recruté et financé par l'entreprise d'ingénierie AKKA Technologies, et accueilli au sein de son département de recherche et développement ainsi qu'au CHU dans le cadre d'un partenariat de recherche associant le CHU, AKKA et l'IRIT. Il aura l'occasion d'échanger avec la communauté d'experts techniques d'AKKA autour des projets de R&D menés par AKKA et contribuera également à son enrichissement en termes de retour d'expériences par exemple. Enfin, il sera également accompagné par un des référents IA de la société.

Objet du stage

De nombreuses questions se posent sur la prise en charge au long cours de l'artérite à cellules géantes (ACG, anciennement nommée maladie de Horton), justifiant des études dans des cohortes de plusieurs milliers de patients. Le GEFA, groupe français d'étude de l'ACG a créé une base de don-

nées nationale commune à ses différents centres totalisant à ce jour environ 1500 patients atteints de cette maladie. Début 2022, environ 220 patients du CHU de Toulouse auront été inclus dans la base, avec un suivi de 2 à 5 ans. Pour chaque centre, le temps nécessaire à la saisie des données dans la base de données est conséquent. Une aide à la saisie des données dans la base par extraction de données textuelles à partir des comptes rendus (CR) médicaux présents dans le dossier patient informatisé (DPI) du CHU de Toulouse représenterait un gain de productivité important pour les équipes hospitalières.

L'objectif du travail est la mise au point d'algorithmes d'Intelligence Artificielle permettant d'identifier dans les CR les variables d'intérêt permettant de décrire le phénotype clinique, biologique et radiologique initial de la maladie, son traitement et son évolution dans le temps. Ce travail s'appuiera sur une recherche bibliographique et l'étude de faisabilité réalisée en 2020 par un master 2 avec l'équipe du CHU dontil essayera d'étendre les premiers résultats en explorant de nouvelles approches.

Les données à identifier sont pour la plupart présentes dans le Dossier Patient Informatisé (DPI) du CHU de Toulouse. Elles peuvent s'y trouver sous formes structurées (champs spécifique au sein d'un compte-rendu) ou bien sous formes non structurées (au sein d'un texte libre saisi ou dicté par un clinicien). L'enjeu de l'étude est de récupérer ces données via des programmes d'analyse de texte automatisés (NLP). La validation des algorithmes pourra être réalisée à partir du DPI de patients toulousains atteints d'ACG non encore inclus dans la base de données GEFA et à partir de CR anonymisés de patients inclus dans la base GEFA nationale et adressés par les centres partenaires GEFA hors Toulouse au niveau national.

Formation

École d'ingénieur, de préférence avec spécialisation en informatique ou mathématiques (en année de césure ou stage long). Master 2 informatique ou mathématiques appliquées.

Compétences attendues

Des connaissances en NLP, science des données, apprentissage et réseaux de neurones seront appréciées, ainsi que la connaissance d'outils classiques comme Pandas, spaCy ou Scikit-Learn..

Modalité de candidature

- > CV à envoyer à: gefa@stages-medecine-numerique.fr
- > Date limite de candidature : 31/01/2022
- > Encadrants: Professeur Laurent Sailler (chef du service de médecine interne, CHU de Toulouse, UMR INSERM CIC 1436), Philippe Muller (IRIT), Véronique Moriceau (IRIT), Philippe Szczech (chef de projets R&D, AKKA Research), Prosper Burq (CHU), Jean-Marc Alliot

Une première sélection sera effectuée sur la base des CVs reçus. Les candidats seront immédiatement informés du résultat, et ceux qui seront retenus à l'issue de la première sélection seront invités à un ou deux entretiens individuels en téléconférence avant sélection définitive.

L'ensemble des offres de stage est disponible sur <http://www.stages-medecine-numerique.fr>.