

Offre de stage – Annotation de dossiers patient

| | |
|------------------------------|--|
| Objet du stage | Développement d'une approche d'annotation et de classification de compte-rendus d'hospitalisation efficace dans le cas de petits jeux de données annotés |
| Profil recherché | Étudiant.e en école d'ingénieur ou de master (M1 ou M2) BAC+5 en science des données ou informatique |
| Lieu de travail | LIRMM |
| Date de démarrage | À partir de janvier 2020 |
| Durée | 4 à 6 mois |
| Equipe et encadrement | Le stagiaire sera intégré dans une équipe pluridisciplinaire comprenant des enseignants-chercheurs du LIRMM et des personnels du CHU de Montpellier |
| Contexte | <p>Depuis l'informatisation de leur système d'information ces deux dernières décennies, les hôpitaux hébergent de nombreux entrepôts de connaissances, partiellement exploités, notamment des documents textuels faiblement structurés, décrivant les cas passés, aux contenus et formats hétérogènes. Convertir cette masse d'informations sous une forme structurée est un enjeu majeur et constitue le point de départ du développement d'outils d'interrogation automatique adaptés.</p> <p>Dans ce stage, nous nous intéresserons particulièrement aux comptes rendus d'hospitalisation qui contiennent des informations précieuses sur le suivi médical des patients. L'objectif principal consistera à proposer une chaîne de traitements pour classer efficacement les compte-rendus et impliquant le moins possible les professionnels de santé pour constituer des jeux de données annotés.</p> |
| Travail à réaliser | <p>Les approches traditionnelles de classification se concentrent sur le choix du meilleur classifieur et sur la définition des meilleures caractéristiques prises en entrée de ces classifieurs. Dernièrement, pour de nombreuses tâches de classification de textes, les méthodes d'apprentissage profond se sont révélées très efficaces. Cette tendance s'est confirmée avec le succès des <i>word embeddings</i> (Mikolov 2013) pour les données textuelles. Tout récemment, des méthodes comme ULMFiT (Howard 2018), ELMo (Peters 2018) ou BERT (Devlin 2018) généralisent avec succès le mécanisme des <i>word embeddings</i>.</p> <p>Toutefois, une limite de ces approches supervisées est qu'elles nécessitent une forte implication des professionnels de santé, avec un important niveau d'expertise, pour l'étiquetage préalable des données utilisées pour l'apprentissage des modèles. Or, les experts ne sont pas toujours disponibles pour cette tâche rébarbative. Afin d'accompagner au mieux les experts dans cette tâche, nous explorerons des approches semi-supervisée (Li 2018a et Li 2018b).</p> <ul style="list-style-type: none"> - Prise en main des données de l'étude - Rédaction d'un état de l'art des méthodes - Expérimentations des différentes méthodes - Rédaction d'un rapport comparant ces différentes approches |
| Expérience souhaitée | <p>Connaissances : science des données, traitement automatique de la langue, programmation</p> <p>Des qualités rédactionnelles (en anglais et en Français), de communication et de travail en équipe seront indispensables.</p> |
| Gratification | Indemnité légale de stage |
| Candidature | Envoyer un CV, notes des 2 dernières années et une lettre de motivation à : sandra.bringay@lirmm.fr , pascal.poncelet@lirmm.fr et jerome.aze@lirmm.fr |