

Offre de stage/thèse 2021-2024

Titre	Détection d'anomalies dans des bases de données SRH
Niveau du stage	M2 ou dernière année Ingénieur
Date de début/ fin	Env. 4 mois pour le stage - 3 ans pour la thèse
Ville, Pays	Annecy-le-Vieux, France
Laboratoire	LISTIC - Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance - https://www.univ-smb.fr/listic/
Description du sujet de stage	<p>Le stage est préparatoire à une thèse en contrat CIFRE dont le démarrage est prévu au cours de l'automne 2021. Il est donc attendu que les postulants à ce sujet aient une réelle motivation pour poursuivre en thèse de doctorant à l'issue de la période de stage.</p> <p style="text-align: center;">Sujet</p> <p>La révolution numérique et les transformations s'y rattachant mettent les systèmes d'information des entreprises, les progiciels de gestion et plus généralement la donnée au centre de l'activité des entreprises. La fiabilité des données qui y sont enregistrées et qui sont manipulées par ces systèmes est un enjeu majeur car elles conditionnent largement le fonctionnement des entreprises. L'explosion du volume des données rend plus difficile le suivi de qualité et de l'exactitude de celles-ci. Il convient donc de développer des approches automatisées permettant de faciliter la tâche fastidieuse, difficile et coûteuse de la détection de ces anomalies dans les bases de données et les systèmes d'information, qui n'est pas possible de façon manuelle.</p> <p>L'objectif de ce projet et de la thèse est de proposer des méthodologies et des outils logiciels permettant de détecter automatiquement les anomalies ou irrégularités dans une base de données. Ces anomalies peuvent avoir été causées par des erreurs d'introduction, de mauvaise manipulation, des dysfonctionnements matériels ou logiciels ou même par des actions ciblées plus malicieuses. Mais nous ciblons plus particulièrement dans cette thèse les erreurs d'origine humaine (saisies incorrectes, données incomplètes ou manquantes, ...) provenant des opérateurs de l'application concernée. Pour ceci nous souhaitons utiliser des approches de fouilles de données, d'apprentissage machine ainsi que des mécanismes de validation de contraintes. En particulier, les techniques de détection de régularité, ou d'irrégularité dans des flots d'actions ou d'évènements, ainsi que des techniques d'apprentissage, supervisé ou non, sont pertinentes.</p> <p>Le sujet nécessitera une solide connaissance des méthodes statistiques et de science des données, ainsi qu'une connaissance des méthodes d'apprentissage machine et de fouilles de données. La/le candidat-e devra participer à l'intégration des outils qu'il aura développé à des bases de données opérationnelles et valider les méthodes dans un cadre applicatif. Il devra donc interagir avec une équipe de développement ainsi que des experts du métier, rattachés à la base de données. Dans ce cadre, en plus des compétences théoriques, il est nécessaire que la/le</p>

candidat-e est aussi une expérience et un intérêt pour le développement informatique et une connaissance des environnements de programmation Python et Java.

En amont de ce travail de thèse dont le démarrage est prévu au début de l'automne 2021, **un stage est proposé sur le même sujet** permettant de défricher le sujet et de confronter la/le candidat-e aux attentes de ce travail. Ce stage pourra démarrer mi-mai pour une durée de 3-4 mois.

La thèse se fera dans le cadre d'un contrat CIFRE et d'une collaboration entre l'entreprise Cegedim (<https://www.cegedim.fr/>) spécialisée dans les solutions de gestion des entreprises et services et le LISTIC (<https://www.univ-smb.fr/listic>) laboratoire de recherche de l'Université Savoie Mont-Blanc. La/le candidat-e sera principalement localisé-e à Annecy mais doit envisager des déplacements réguliers pour travailler avec les collaborateurs de l'entreprise sur différents sites (Lyon, Paris,...).

Quelques informations sur Cegedim : Fondée en 1969, Cegedim est une entreprise innovante de technologies et de services spécialisée dans la gestion des flux numériques de l'écosystème santé BtoB, ainsi que dans la conception de logiciels métier destinés aux professionnels de santé et de l'assurance.

La thèse se déroulera au sein de l'entité Cegedim SRH qui s'appuie sur sa propre solution TEAMS RH pour offrir à ses clients des prestations d'externalisation adaptées et évolutives de leurs systèmes d'information paie et RH.

Pour faire acte de candidature:

par mel aux personnes de contact en joignant:

- un CV détaillé
- les relevés de notes des 2 dernières années d'études
- une lettre de motivation montrant l'adéquation du profil du/de la candidat-e avec le sujet

Compétences requises	Issu-e d'une formation en informatique. Compétences fortes en analyses de données, science des données, méthodes statistiques, fouilles de données, machine learning, programmation Java et Python. Le/la candidat-e doit maîtriser le français et l'anglais.
Gratification	Selon législation en vigueur
Encadrement / Contacts	Pr. Kavé Salamatian - kave.salamatian@univ-smb.fr Hervé Verjus - herve.verjus@univ-smb.fr