Appel à candidature – Bourse de thèse

« Des lacs de données à l'analyse de documents textuels »

Sous la responsabilité du Professeur Jérôme Darmont, le/la doctorant·e travaillera dans le cadre du projet AURA-PMI, mené en collaboration avec le laboratoire COACTIS et financé par la Région Auvergne-Rhône-Alpes.

Contexte : Pour continuer à se développer, les PME industrielles doivent constamment améliorer la qualité et l'étendue de leur offre, leur proposition de valeur. Les possibilités de plateformes numériques, d'économie collaborative et de réseaux étendus se multiplient ; la proposition d'une offre industrielle digitale devient cruciale pour une industrie française en retard sur d'autres pays. Des travaux récents ont, en outre, souligné l'importance pour les PMI de saisir l'opportunité de la servicisation. Digitalisation (utilisation en interne et en externe des technologies numériques) et servicisation (transformation de « l'offre produit » en offre combinée de « produits et services », voire une offre d'« usage ») sont, dès lors, deux nouveaux leviers de création de valeur à leur disposition, qui impliquent de connaître parfaitement les besoins du client en ayant recours à l'échange de données et à la fouille de données (*data mining*).

L'objectif du projet AURA-PMI est d'analyser les enjeux et les risques, en termes de croissance et de rentabilité, des virages stratégiques des PME industrielles de la région AURA, induits par la digitalisation et la servicisation. Ses thématiques de recherche incluent les conditions d'évolution des systèmes sociotechniques, notamment celui du système d'informatique décisionnelle.

Sujet de thèse : Les lacs de données (*data lakes*), baptisés ainsi par Dixon (2010), sont une manière, née avec les mégadonnées (*big data*), de stocker des données variées et diversement structurées dans leur format natif en vue de les analyser (*reporting*, visualisation, fouille de données…). C'est un concept qui s'oppose à celui d'entrepôt de données, qui est une base de données décisionnelle intégrée, très structurée et orientée sur un sujet précis, mais qui a l'inconvénient de diviser les données en silos étanches (Stein & Morrison, 2014).

Toutefois, tout le monde s'accorde pour dire qu'un lac de données doit être bien conçu sous peine de devenir un marécage (*data swamp*) inexploitable (Inmon, 2016). En revanche, les solutions pour y parvenir sont peu ou prou inexistantes dans la littérature et relèvent à l'heure actuelle de pratiques industrielles peu divulguées et souvent orientées vers le stockage des données, mais peu vers leur exploitation.

Contact secrétariat : 5 avenue Pierre Mendès-France - 69676 Bron Cedex
Téléphone : +33 (0)4 78 77 31 54 - Télécopie : +33 (0)4 78 77 23 75
Courriel : secretariat@eric.univ-lyon2.fr
http://eric.univ-lyon2.fr

L'objectif de cette thèse est d'aborder les problèmes de recherche liés à la mise en œuvre d'un lac de données essentiellement textuelles (Madera & Laurent, 2016) : qualité des données en entrée, stockage et indexation des données, analyse des données (OLAP, fouille). Il s'agit de plus de mener un travail pluridisciplinaire dès les premières étapes de conception, afin d'intégrer les besoins, contraintes et méthodologies des chercheurs en gestion (notamment d'un·e doctorant·e du laboratoire COACTIS travaillant sur le même projet) et de garantir leur appropriation du lac de données et de son exploitation.

Lieu d'exercice : Laboratoire ERIC, Campus Porte des Alpes

École doctorale : InfoMaths ED 512

Démarrage du financement : Septembre 2018

Rémunération : Montant d'un contrat doctoral classique

Contact : Envoyer CV et lettre de motivation à jerome.darmont@univ-lyon2.fr.

Références :

- Dixon J. (2010). Pentaho, Hadoop, and Data Lakes. James Dixon's Blog. https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes/
- Inmon B. (2016). Data Lake Architecture: Designing the Data Lake and Avoiding the Garbage Dump. Technics Publications.
- Madera C., Laurent A. (2016). The next information architecture evolution: the data lake wave. 8th International Conference on Management of Digital Ecosystems (MEDES 2016), Biarritz, France: 174-180.
- Stein B., Morrison, A. (2014). The enterprise data lake: Better integration and deeper analytics. Technology Forecast, 1. http://www.pwc.com/us/en/technology-forecast/2014/cloud-computing/assets/pdf/pwc-technology-forecast-data-lakes.pdf