

Post-doc Deep learning pour la valorisation de journaux anciens

L'équipe Apprentissage (<http://www.litislab.fr/equipe/app/>) du laboratoire LITIS à l'université de Rouen Normandie recrute un(e) post-doc/ingénieur de recherche pour 12 mois dans le cadre du projet ASTURIAS.

Mots clefs : Deep Learning, Réseaux de Neurones Convolutifs (CNN), Lecture automatique d'images de documents, valorisation du patrimoine écrit.

Descriptif du projet

Déployer la puissance des réseaux de neurones convolutionnels profonds afin de plonger les documents anciens dans l'ère numérique.

Le LITIS a conçu et développé une suite logicielle appelée PIVAJ destinée aux archives et bibliothèques qui souhaitent valoriser leurs archives de journaux anciens : comment extraire sections et articles automatiquement de numérisation de journaux ? PIVAJ est actuellement utilisé par plusieurs bibliothèques nationales, notamment celles du Pays de Galles et de Finlande.

Le projet ASTURIAS vise à développer les futures technologies pour PIVAJ, et ciblant de nouvelles fonctionnalités comme l'extraction d'entités nommées, l'extraction de l'ordre de lecture ou le rubricage d'articles (catégorisation par le contenu).



Figure 1 : Classification au niveau pixel de différentes parties d'une page de journal

Missions

Dans le cadre de ce projet, nous proposons un poste de post-doc/ingénieur de recherche de 12 mois en apprentissage machine pour l'analyse d'images de journaux anciens. La personne recrutée sera en charge de deux missions :

Extraction de structure

Dans la continuité des travaux déjà réalisés, une première étape visera l'extraction de la structure des documents. On pourra s'appuyer sur la modélisation de l'ordre de lecture à l'aide de modèles à attention exploitant le modèle optique déjà développé.

Intégration dans PIVAJ

La personne recrutée devra faire évoluer PIVAJ pour intégrer les composants d'extraction développés à l'échelle du projet par l'ensemble des partenaires. Ce qui comprend l'extraction de structures, l'extraction d'entités nommées et la thématisation.

Profil

Idéalement la personne recrutée aura une expérience en apprentissage par réseaux de neurones appliqué à l'analyse d'images, ainsi qu'en développement d'applications. L'expertise en Machine Learning reste cependant prépondérante dans le profil recherché.

Compétences :

1. Machine Learning
 - a. Deep Learning
 - i. CNN/FCN
 - ii. RNN
 - b. TensorFlow/Keras
 - c. Appliqué à des images
 - i. Appliqué à des images de documents
2. Développement d'applications
 - a. D'une certaine taille
 - i. En équipe
 - ii. Avec du code existant
 - b. En C++
3. Lire des articles scientifiques
 - a. Écrire des articles scientifiques

1., 2. et 3. sont nécessaires ; les sous-catégories sont des plus.

Conditions

Le contrat sera de 12 mois et peut débuter dès que possible, avec un salaire dépendant de l'expérience du candidat. La personne recrutée travaillera dans les locaux du LITIS à l'université de Rouen Normandie sur le campus du Madrillet à Saint-Etienne du Rouvray (Seine-Maritime).

Si vous êtes intéressé, contactez asturias-wp1@litislab.fr

