

Offre de stage en Informatique R&D niveau Bac+4/5 ou Master

Titre : Mise en place d'outils de visualisation et de fouille interactive de données pour la transcription et l'indexation de documents anciens.

Mots-clés : Analyse d'images, visualisation et fouille de données, reconnaissance des formes, documents anciens

Résumé du contexte et des objectifs du stage :

Il y a quelques années le LIFAT a obtenu un financement de Google pour la réalisation du Projet Paradiit. Ce financement concernait le développement d'outils innovants d'analyse d'images, afin de permettre une transcription (OCR) et une indexation interactive de documents anciens. Ces documents précieux, datant de la Renaissance, ont été numérisés, notamment par le CESR de Tours afin d'être préservés et mieux valorisés.

Ces travaux ont pour objectif de pallier les difficultés importantes rencontrées par les bibliothèques numériques, qui traitent des documents patrimoniaux antérieurs à 1700. Parmi ces difficultés, on peut citer les performances médiocres des méthodes de reconnaissance de caractères classiques dues à l'importante variation graphique qui précède la stabilisation de l'imprimerie moderne. Pour cela, des techniques d'analyse de redondance de formes et de word spotting peuvent être mises en place.

Nous désirons aujourd'hui poursuivre ces développements dans le cadre d'un stage ayant les objectifs suivants :

Mission :

- Mise en place d'une forge open source et homogénéisation des briques logicielles existantes (Agora – Retro)
- Couplage de Agora/Retro (outils d'extraction et de clustering d'Elements de Contenu développé par le LIFAT) et de l'outil de visualisation Cartolabe <http://cartolabe.fr> (développé à l'INRIA en collaboration entre l'INRIA, le CNRS et Paris-Saclay) et/ou de la librairie UMAP (<https://github.com/lmcinnes/umap>).
- Expérimentations exploitant divers descripteurs de formes et diverses mesures de similarité entre formes (via UMAP)
- Intégration des approches fournissant des résultats pertinents dans une interface permettant leur exploitation par des historiens (cartolabe et/ou Retro).

Compétences requises :

- Connaissances en analyse d'images et analyse des données
- Compétences en C# et C++, aptitudes à modéliser proprement en programmation objet.

Infos sur le projet

<https://sites.google.com/site/paradiitproject/>
<https://bv.hypotheses.org/tag/retro>
<https://bv.hypotheses.org/tag/typography>
<https://hal.archives-ouvertes.fr/hal-01094430/document>

Encadrement

- Jean-Yves RAMEL – LIFAT
- En collaboration avec Jean-Daniel Fekete AVIZ Team - INRIA Saclay

Profil du candidat : Le candidat souhaité est un étudiant en Master ou dernières années d'une formation d'ingénieur en Informatique, avec des bonnes compétences en programmation (langage C# ou/et C++) et des connaissances en analyse d'images et/ou reconnaissance des formes.

Candidature : CV + motivation à envoyer à ramel@univ-tours.fr

Lieu du stage : LIFAT, Polytech-Tours, 64 avenue Jean Portalis, 37200 Tours

Dates / Durée du stage : 4 à 6 mois entre Mars/Avril et Septembre 2019

Indemnités du stage : Indemnités mensuelles légales en vigueur