

Offre d'emploi | Poste à pourvoir

CDD Ingénieur R&D - *Computer Vision Machine Learning*

Centre d'études supérieures de la Renaissance | Laboratoire d'Informatique de Tours

ANR TypoReF : *Typographie de la Renaissance Française (1470-1640)* ([ANR-22-CE27-0009](#))

Informations générales

- Lieu de travail : LIFAT, Tours, France.
- Date de publication de l'offre : mars 2023
- Type de contrat : CDD Technique/Administratif
- Secteur d'activité : recherche-développement, humanités numériques
- Durée du contrat : 12 mois
- Date d'embauche prévue : 1^{er} septembre 2023
- Quotité de travail : temps complet
- Grade : Ingénieur d'études ou de recherche (Postdoc)
- Rémunération : en vigueur (grilles de la fonction publique), selon expérience
- Niveau d'études souhaité : Bac +5 Informatique ou doctorat
- Date limite de candidature : 22 mai 2023

Contexte de travail

Ce poste s'inscrit dans le cadre du projet TypoReF : *Typographie de la Renaissance Française (1470-1640)* dirigé par Remi Jimenes (CESR, Tours) et financé par l'Agence nationale de Recherche ([ANR-22-CE27-0009](#)). Conçu dans une approche interdisciplinaire, le projet vise à inventorier, décrire et étudier les matériaux typographiques français de la Renaissance en développant une plateforme et des outils d'analyse et d'exploration de corpus de livres numérisés.

La numérisation des livres anciens a débuté au milieu des années 1990. Depuis cette date, des équipes de chercheurs alliant informaticiens et spécialistes du patrimoine écrit ont collaboré pour développer des outils dédiés. Longtemps, l'analyse et la reconnaissance d'images de documents ont été considérées comme des tâches complexes décomposées en 2 sous-tâches indépendantes : l'analyse de la mise en page et la reconnaissance du texte. L'analyse de la mise en page vise à détecter tous les composants de l'image tels que les blocs de texte, les tableaux, les images, les graphiques... La reconnaissance de texte est consacrée à la reconnaissance des caractères à l'intérieur des blocs de texte.

Dans le cadre d'une collaboration établie dès 2003, le [CESR](#) et le [LIFAT](#) ont développé les logiciels [Agora and Retro](#). Les développements ont été prolongés en 2011 et 2012 par deux bourses successives "[Award in Digital Humanities](#)" financées par Google. Ces outils d'analyse et d'indexation de la mise en page adaptés aux images des livres imprimés de la Renaissance ont permis d'extraire et d'indexer automatiquement quelque 13 500 gravures, lettrines, bandeaux et fleurons à partir de 600 numérisations d'ouvrages issues du corpus Fac-similés du programme de recherche [« Bibliothèques Virtuelles Humanistes »](#) (BVH, CESR).

Agora et Retro ont déjà plus de 15 ans et sont basés sur des technologies aujourd'hui dépassées au vu des développements de ces dernières années. Récemment, des architectures profondes ont amélioré les performances obtenues pour les deux sous-tâches. Sur la base de la très forte expérience accumulée au cours des vingt dernières années, le projet TypoReF entend permettre le développement de nouveaux outils d'analyse de la mise en page, d'identification et d'indexation des matériaux

typographiques adaptés aux livres imprimés anciens, en utilisant les technologies les plus puissantes du moment, notamment celles basées sur l'apprentissage profond.

Missions

L'objectif est la mise en place d'une plateforme web de labellisation de contenus. Cette labellisation concerne à la fois la collecte de métadonnées produites par un utilisateur expert et la production de métadonnées résultant d'une analyse des images/pixels. Cette double indexation (humain/machine) doit permettre de rechercher, comparer, regrouper des formes, et ainsi de mettre en évidence des liens difficilement perceptibles.

La plateforme nécessite de développer les fonctionnalités suivantes :

- **Segmentation sémantique des images de documents à l'aide d'architectures d'apprentissage profond** (pour remplacer l'ancienne Agora) : localisation et caractérisation d'éléments de contenu spécifiques à l'intérieur des pages numérisées d'un livre imprimé ancien (blocs de texte, décor gravé...) avec différents niveaux de détail (au sein d'un bloc de texte, on distinguera ainsi des lignes, des caractères) et caractérisation de ces éléments (lettrage, bandeau, marque, caractères, etc.). Chaque élément sera associé à un ensemble de métadonnées permettant de le décrire et de retrouver précisément son origine.
- **Le clustering des éléments de contenu extraits** (pour remplacer l'ancien Rétro) consiste à faire des comparaisons entre les éléments, principalement des ornements gravés (comme les lettrages ou les bandeaux), pour proposer des correspondances entre des formes similaires. La plateforme utilisera différents types d'algorithmes d'apprentissage automatique, notamment l'apprentissage dit « non supervisé ».
- **Importation/exportation des données depuis/vers la base de données** selon des standards tels que le format ALTO ou le protocole IIIF.

Compétences

- Développement logiciel et ingénierie, Python, HTML, CSS, JS
- Apprentissage automatique, vision par ordinateur
- Capacité à travailler en équipe, esprit curieux et rigoureux

Conditions de recrutement

- **Structure de recrutement** : Université de Tours, Laboratoire CESR, UMR 7323 - 59, rue Néricault Destouches BP 12050, 37020 Tours Cedex 1 - <https://cesr.univ-tours.fr/>
- **Lieu de travail** : Laboratoire LIFAT, 64 avenue Jean Portalis 37200 Tours - <http://lifat.univ-tours.fr>
- Ce poste, comme tous les postes de l'université de Tours, est ouvert aux candidats en situation de handicap.

Candidatures

- **Date limite de candidature : 22 mai 2023**
- Les candidatures sont à adresser par email, accompagnées d'un CV (2 pages max.) et d'une lettre de motivation à : Thierry Brouard, Jean-Yves Ramel --> prenom.nom AT univ-tours.fr