

Centre d'Études Supérieures de la Renaissance (CESR, UMR 6576 du CNRS), Tours

Offre de stage 2011

Pratique et analyse des OCR sur les imprimés anciens

Le stagiaire participera à la réalisation et à l'analyse des tests OCR sur les imprimés anciens du premier volet des projets financés par les bourses Google en 2011, au sein de l'équipe des [Bibliothèques Virtuelles Humanistes](#) (BVH) du Centre d'Études Supérieures de la Renaissance (CESR) et en collaboration avec le [Laboratoire d'Informatique de Tours](#) (LI).

> Voir le site [Google Digital Humanities Awards](#)

> Voir le résumé : [présentation du projet](#)

Objectifs et missions du stage

Deux missions sont à réaliser en parallèle : 1) Participer à l'amélioration des performances de l'OCR « RETRO » (LI Tours) sur les documents numérisés par le CESR (textes imprimés anciens). Le stagiaire devra contrôler le résultat, analyser les erreurs, documenter le processus, effectuer des tests sur d'autres livres, analyser les résultats, etc.; 2) Analyser et comparer les OCR concurrents du marché. Ainsi, il faudra :

- Effectuer les opérations d'étiquetage avec l'OCR RETRO
- Tester et évaluer la reconnaissance automatique de RETRO sur les documents numérisés du CESR
- Réaliser et affiner l'apprentissage de RETRO
- Corriger et améliorer la transcription via l'utilisation de dictionnaires spécifiques du français de la Renaissance
- Tester, comparer et analyser les OCR du marché avec apprentissage (FineReader, Omnipage) ou libres (OCROPUS, Tesseract)
- Affiner l'encodage XML-TEI des transcriptions établies et préparer les versions « quasi-diplomatiques » des textes en vue de leur publication sur le site des BVH
- Effectuer le suivi du projet (tableaux de bord, comptes rendus de réunions, etc.)

Compétences

- Connaissances de la typographie et des graphies anciennes : polices, ligatures et caractères spéciaux (tildes, abréviations)
- Connaissance du langage XML/TEI
- Maîtrise de l'informatique (outils de bureautique) et facilité d'apprentissage de nouveaux outils (interfaces utilisateur d'OCR)
- Connaissance de l'anglais (des notions de latin, grec, italien ou espagnol seraient appréciées)
- Facilités d'intégration et de communication au sein d'une équipe de travail

Contexte

- *Lieu* : Bureau des BVH, CESR, Tours.
- *Période souhaitée* : 2 mois (minimum) à 4 mois, à partir du 15 mars et jusqu'à la fin du mois de juillet.
- *Gratification* : réglementaire, avec indemnité forfaitaire pour les frais de transport.

Diplôme

Niveau master professionnel 1 ou 2 – Cours humanités numériques

Responsable du stage

Marie-Luce Demonet, professeur, directrice-adjointe du CESR, responsable BVH : marie-luce.demonet@univ-tours.fr

Candidature à retourner accompagnée d'un CV et d'une lettre de motivation à :
 Sandrine Breuil, chef de projet BVH : sandrine.breuil@univ-tours.fr / 02 47 36 77 80
 Bibliothèques Virtuelles Humanistes, Centre d'Études Supérieures de la Renaissance - 59 rue
 Néricault-Destouches BP 11328 - 37013 Tours Cedex