

## HIMANIS (HISTORICAL MANUSCRIPT INDEXING FOR USER-CONTROLLED SEARCH)

<http://www.irht.cnrs.fr/fr/recherche/les-programmes-de-recherche/himanis>

### Ingénieur de recherche en analyse de sources

**Nature du contrat** : CDD 2 ans

**Date de début de contrat** : 1<sup>er</sup> novembre 2015 – 31 octobre 2017

**Rémunération** : grille de salaire CNRS, niveau IR, approx. 2500 € brut.

#### Missions

Sous la responsabilité du pilote du projet, l'ingénieur(e) aura pour mission d'assurer la qualité et l'utilisation et l'exploitation des données résultant de l'indexation d'images par ordinateur, et d'assurer le suivi du projet.

L'indexation automatique du texte par ordinateur (reconnaissance d'écriture manuscrite) porte sur les images numérisées des registres de la chancellerie royale de France (XIII<sup>e</sup>-XV<sup>e</sup> s.). L'analyse de la qualité portera à la fois sur la qualité textuelle et l'utilisabilité pour les recherches historiques et de diplomatique médiévales.

#### Activités

- Préparer les données pour l'analyse par ordinateur (corpus d'images, dictionnaires linguistiques français et latins, entités nommées, tables de concordance)
- Participer à la définition de besoin pour les interfaces utilisateur
- Tester et évaluer les résultats de l'analyse d'image par ordinateur
- Mener des recherches exploratoires historiques, paléographiques et diplomatiques
- Promouvoir l'usage des données (création d'expositions virtuelles, animation de communauté pour accroître les données validées par utilisateur) en lien avec Europeana
- Assurer le suivi du projet

#### Compétences

- Connaissance exigée du latin, de l'anglais et du français
- Connaissance du moyen français souhaitée
- Connaissance des écritures françaises des XIII<sup>e</sup>-XV<sup>e</sup> siècles (paléographie de lecture)
- Compétences en management de projets
- Connaissance des techniques de codage (TEI-XML)
- Bonne connaissance des outils informatiques notamment des bases de données
- Capacité à travailler en interaction avec les équipes de recherche et les réseaux technologiques du domaine

- Capacité d'autonomie dans la gestion de projets et dans l'animation de communauté

## Contexte

Le poste est à pourvoir à l'Institut de Recherche et d'Histoire des Textes (CNRS UPR 841) dans le cadre du projet HIMANIS (*Historical MANuscript Indexing for user-controlled Search*), piloté par l'IRHT (D. Stutzmann) et associant les universités de Groningen (Pays-Bas, L. Schomaker) et Valencia (Espagne, E. Vidal), ainsi que la société industrielle A2iA (France, R. Messina). Les Archives Nationales, la Bibliothèque nationale de France et Europeana/The European Library sont des partenaires associés.

Le projet vise à :

- indexer automatiquement le texte des registres de la chancellerie royale de France (essentiellement le fonds JJ des Archives nationales et des volumes conservés à la Bibliothèque nationale de France) ;
- créer des services de recherche, de découverte et d'amélioration collaborative à partir de cette indexation automatique ;
- créer des savoirs nouveaux (organisation et écritures de la chancellerie)

### **Extrait du document scientifique du projet:**

Manuscripts are among the most important witnesses to our European shared cultural heritage. In recent years, large quantities of historical handwritten documents are being scanned and made available through web portals. Yet, the wealth of information conveyed by the text captured in these images remains largely inaccessible. General users and researchers more and more expect to query handwritten resources in plain text like printed books, but current handwritten text recognition technology is still far from offering sufficiently accurate transcripts on historical documents. This situation raises the need of searching approaches specifically designed for text images. HIMANIS aims at developing cost-effective solutions for querying large sets of handwritten document images. To this end, innovative keyword spotting, indexing and search methods will be developed, tested, adapted and/or scaled up to meet the real world conditions required. HIMANIS will associate Computer Science (A2iA, UPVLC and RUG), Humanities (IRHT) and Cultural Heritage (The European Library and the Archives Nationales of France) institutions and produce technology to generate new, research-based knowledge from historical manuscripts. As a challenging and particularly interesting case study, the large collection of registers produced by the French royal chancery (14<sup>th</sup>- 15<sup>th</sup> c.) will be used as a test bed. This will allow sustainable use and management of this important cultural resource. Implementing user feedback will also increase our understanding of the meanings that it holds for people and how they perceive, use and interpret it, thereby meeting societal challenges in promoting its use and re-use. The expected outcomes of this project are: a) a new indexing/searching technology for historical manuscripts; b) a new paradigm to study our historical heritage, as conveyed by manuscripts, by using full text search technology. c) a new vision of the raise of nation states in Europe via a new study of the corpus under this paradigm.

## **Sur l'institution d'accueil**

L'Institut de Recherche et d'Histoire des Textes, fondé en 1937, offre un environnement scientifique et humain transdisciplinaire. Il a pour objet la recherche fondamentale sur le manuscrit médiéval et la transmission des textes. C'est un laboratoire interdisciplinaire unique au monde, avec 82 spécialistes (IR, CR, DR, et associés non CNRS) de paléographie, diplomatique, liturgie, littératures latine, française, grecque etc., et des chercheurs de renommée internationale dans chacun de ces domaines.

L'implication dans la recherche transdisciplinaire et numérique y est une tradition solidement ancrée : l'IRHT est le lieu où communiquent des disciplines spécialisées, ailleurs séparées ; l'IRHT a aussi été un précurseur en humanités numériques (publication du *Médiéviste et l'ordinateur* de 1979 à 2005), a fondé le Centre de ressources numériques TELMA et dirige actuellement plusieurs programmes de recherche en humanités numériques.

### Candidature

Candidater par email en envoyant une lettre de motivation et un CV détaillé à [dominique.stutzmann@irht.cnrs.fr](mailto:dominique.stutzmann@irht.cnrs.fr)