

Research Fellow / Ingénieur de recherche en analyse de sources

Institution: Institut de Recherche et d'Histoire des Textes

Location: Paris, France

Application Due: 31 Oct. 2017

Type: Full Time

Duration: 1 year, renewable until 3 years

Starting Date: between January 1, 2018 and April 1, 2018

The **Institut de Recherche et d'Histoire des Textes (CNRS)** invites applications for a Research Fellow (“Ingénieur(e) de recherché en analyse de sources”) for the project “**HORAE - Hours, Analysis, Editions**” (ANR-17-CE38-0008), under the supervision of Mr. Dominique Stutzmann, Principal Investigator.

Position focus

The Research Fellow will ensure the quality, implementation and exploitation of data produced through automated text recognition and automated text identification, and monitor the project progress:

- build the reference data for the project
 - o update and enhance an existing database on books of hours: search for available digitization and IIF manifests; update of core metadata (script, place and time of production...);
 - o prepare a reference corpus in TEI, containing the atomic textual elements of books of hours (verses of psalms, canticles, hymns, etc.; versicles, antiphons, orisons, responsories, prayers in French and Latin, etc.);
 - o order digital reproductions and follow up of digitization orders;
 - o participate in format modelling;
- specify and use ergonomic end user tools:
 - o write functional requirements for ergonomic tools;
 - o validate and correct page classifications, line segmentation, handwritten text recognition;
 - o validate and correct text identifications;
- analyze the textual contents of books of hours
 - o provide formatted scholarly descriptions of manuscript books of hours;
 - o identify and edit hitherto unedited texts;
 - o report on visualization tools (state of the art)
 - o participate in the analysis of textual networks
- participate in the project organization and reporting activities for the French National Research Agency
- communicate and disseminate the ongoing developments and scientific results (research blog, twitter, conferences, articles in peer-reviewed journals)

Required Education, Skills and Experience

- Ph.D. (ABD candidates will be considered) in Medieval History and Manuscript Studies, or a closely related field, and a research concentration in medieval history.



- Attention to detail and strong organizational skills
- Strong interpersonal skills and the ability to work as a team member
- Excellent written and oral communication skills
- Documented scholarship experience
- Languages: English, French, Latin, Middle French
- Knowledge in encoding formats such as TEI-XML is a plus

About the project HORAE

HORAE (Hours - Recognition, Analysis, Editions) is a cross-disciplinary research project studying religious practices and experiences in the late Middle Ages through the books of hours, the absolute medieval best seller. In a public-private partnership, it gathers three partners in the Humanities and Computer Science: the Institut de Recherche et d'Histoire des Textes (IRHT, UPR 841), the private company TEKLIIA and the Laboratoire des Sciences du Numérique de Nantes (LS2N, UMR6004). Books of hours forms, with more than 5000 preserved manuscripts, a vast and crucial ensemble to understand the medieval mindset. Yet, their textual content is very scarcely studied, although the massive production of such a large number of manuscripts is a pivotal cultural and industrial phenomenon to the profound changes in the medieval religious life. Not only a proto-industrial book production arises and the economics of demand shifts to the economics of offer, but also the awakening of the individual and the internalization of faith take places in an increasingly church controlled world. Books of hours are at the same time deluxe objects, implying social staging, and intimate objects, with devotional, memorial and eschatological challenges.

HORAE combines the research and expertise of all three partners in artificial intelligence applied to Computer Vision and image analysis, in Natural Language Processing (NLP), and in book history and religious practices. The research project aims at creating an integrated chain from image treatment to producing new knowledge by placing the end user at the center of the developments, by managing formats, interoperability and long term preservation, but also ergonomics and data visualization. The aims encompass: (1) re-using the many digitized manuscripts, which are online but underused resources; (2) new open source software for Handwritten Text Recognition (HTR); (3) tools for segmentation and plagiarism detection, adapted to the transcription of medieval manuscripts produced by the machine in order to identify the texts in the books of hours; (4) identifying and editing unpublished texts; (5) visualization of manuscript clusters, which share the same textual characteristics in the order of the different parts (Officium Beatae Mariae Virginis, votive offices, suffrages, prayers), but also in the order of textual units to identify the liturgical use; (6) studying the diffusion and circulation of devotional and liturgical texts at the end of the Middle Ages in order to better understand the cultures and faith in the 13th c.-16th c.

With its aims and methods, HORAE profoundly changes the ways of doing research in auxiliary sciences and tackles the challenges of big data. Books of hours have been too scarcely studied until now because they are too numerous, too complex and their text is very repetitive: now, their very number, repetitions and complexities make possible for this project to develop new and efficient technologies and methodologies and gain new knowledge on the Middle Ages.



Host institution

The Institut de Recherche et d'Histoire des Textes is a public research facility created in 1937. It is a cross-disciplinary research environment, with focus on medieval manuscripts and text transmission. It is a unique research lab, with 82 experts in paleography, diplomatics, liturgy, Latin, Greek, French... literatures. IRHT is host to many Digital humanities and cross-disciplinary research projects.

Salary

The successful candidate will be an "Ingénieur de recherche en analyse de sources". Salary : ca. 2500 € gross income / month and ca. 2000 € net income / month (before income tax).

How to Apply:

Submit an application by 31 Oct. 2017 by 15 p.m. (CET) to dominique.stutzmann@irht.cnrs.fr, that includes:

1. Letter of application;
 2. Curriculum vitae
 3. The names and contact information of three professional references
- Candidates with an application file meeting our criteria will be selected for an in-person or videoconference interview. During a video interview, a candidate will meet with 2 researchers for 40 minutes.



Ingénieur de recherche en analyse de sources

Institution: Institut de Recherche et d'Histoire des Textes

Lieu: Paris, France

Date limite de candidature: 31 Oct. 2017

Type: CDD, temps plein

Durée : 1 an, renouvelable jusqu'à 3 ans

Date de prise de poste : entre le 1^{er} janvier et le 1^{er} avril 2018

L'**Institut de Recherche et d'Histoire des Textes (CNRS)** propose un poste d'Ingénieur(e) de recherche en analyse de sources, pour le projet "**HORAE - Hours, Analysis, Editions**" (ANR-17-CE38-0008), sous la direction de M. Dominique Stutzmann, responsable scientifique du projet.

Missions

L'ingénieur(e) de recherche 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 :

- établir les données de référence pour le projet :
 - o mettre à jour et enrichir une base de données, déjà existante, sur les livres d'heures (recherche de numérisation disponibles et de manifests IIF, mise à jour des données descriptives fondamentales, p. ex. lieu et date de production, écriture) ;
 - o préparer un corpus de référence en TEI contenant les textes élémentaires des livres d'heures (versets des psaumes, cantiques, hymnes, antiennes, oraisons, répons, prières en latin et français...) ;
 - o procéder aux commandes de reproductions numériques et suivre les commandes ;
 - o participer au modèle de données ;
- spécifier et utiliser des outils ergonomiques :
 - o écrire les spécifications fonctionnelles pour des outils logiciels ergonomiques.
 - o valider et corriger les classifications de pages, segmentations de lignes, reconnaissance de texte ;
 - o valider et corriger les identifications de textes ;
- analyser le contenu textuel des livres d'heures
 - o fournir des descriptions scientifiques de livres d'heures manuscrits ;
 - o identifier et éditer des textes encore inédits ;
 - o faire un rapport sur les outils de visualisation (état de l'art) ;
 - o participer à l'analyse des réseaux textuels ;
- participer à l'organisation du projet et aux rédactions de rapport pour l'Agence Nationale de la Recherche ;
- communiquer et disséminer les développements en cours et les résultats scientifiques (blog scientifique, twitter, colloques, publications dans revues à comité de lecture) ;



Compétences

- Doctorat (la candidature est possible avec une thèse non encore soutenue ou non encore publiée) en histoire médiévale et étude des manuscrits ;
- Attention au détail, autonomie et capacité d'organisation ;
- Fortes compétences de communication interpersonnelle et capacité au travail d'équipe ;
- Compétences de communication à l'écrit et à l'oral ;
- Expérience académique démontrée ;
- Langues: anglais, français, latin, moyen français
- Connaissance souhaitée des techniques de codage (TEI-XML)

Le projet HORAE

Le projet HORAE (Hours - Recognition, Analysis, Editions) propose d'étudier les pratiques religieuses de la fin du Moyen Âge à travers les livres d'heures, le plus grand best-seller de tout le Moyen Âge et associe trois partenaires en Humanités et Sciences de l'Ingénieur du public et du privé pour une recherche transdisciplinaire : l'Institut de Recherche et d'Histoire des Textes (IRHT, UPR 841), la société TEKLIA et le Laboratoire des Sciences du Numérique de Nantes (LS2N, UMR6004). Les livres d'heures, avec plus de 5 000 témoins conservés, forment en effet un ensemble crucial pour comprendre l'univers mental médiéval. Pourtant leur contenu textuel est très peu étudié, alors que la production d'un si grand nombre de manuscrits est un phénomène culturel et industriel capital qui manifeste les profonds changements du monde religieux du bas Moyen Âge, avec, à la fois, le développement d'une production livresque proto-industrielle et le passage de l'économie de la demande à celle de l'offre, mais aussi l'éveil de l'individu et l'intériorisation de la foi, à une époque où l'encadrement ecclésial devient de plus en plus contraignant. Les livres d'heures sont à la fois des objets de luxe avec des enjeux de mise en scène sociale et des objets intimes, munis de fonctions dévotionnelle, mémorielle, eschatologique.

HORAE combine les compétences et recherches des trois partenaires en intelligence artificielle appliquée à l'analyse d'image par ordinateur, en traitement automatique des langues (TALN), et en histoire du livre et des pratiques religieuses. Le projet a pour ambition de créer une chaîne complète depuis le traitement des images numériques jusqu'à la production des savoirs en plaçant l'utilisateur final au centre des développements, en veillant aux formats, à l'interopérabilité, à la pérennisation des données, mais aussi à l'ergonomie et aux problématiques de visualisation de données massives. Les objectifs comprennent : (1) l'utilisation des nombreux manuscrits numérisés présents en ligne et sous-utilisés ; (2) des nouveaux logiciels open source de reconnaissance d'écriture manuscrite adaptés aux manuscrits médiévaux (HTR, Handwritten Text Recognition) ; (3) des outils de segmentation et de repérage de plagiat adaptés aux manuscrits médiévaux retranscrits par la machine, afin d'identifier les textes transmis par les livres d'heures ; (4) le repérage et l'édition des textes inédits ; (5) la visualisation des 'clusters' de manuscrits présentant les mêmes caractéristiques textuelles, avec l'ordre des différentes parties (petit office de la Vierge, offices votifs, suffrages, prières), mais aussi l'ordre des unités textuelles qui permet de repérer les usages liturgiques ; (6) l'étude de la diffusion et de la circulation des textes dévotionnels et liturgiques transmis par les livres d'heures du Moyen Âge pour mieux comprendre la culture et la foi des XIIIe-XVIe siècles.



Avec ses objectifs et ses méthodes, HORAÉ change radicalement les pratiques des sciences de l'érudition en renversant la problématique de la masse de données. Si les livres d'heures ont été jusqu'ici peu étudiés pour leur texte, c'est parce qu'ils sont trop nombreux, trop répétitifs et trop complexes. C'est précisément le nombre, la répétition des mêmes textes et la complexité qui rendront efficace les technologies développées dans ce projet pour parvenir à de nouveaux outils et obtenir de nouvelles conclusions historiques.

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.

Rémunération

Le candidat recruté sera "Ingénieur de recherche en analyse de sources". La rémunération suit les grilles de salaire CNRS, niveau IR, soit approx. 2500 € brut.

Candidature

Candidater par email à dominique.stutzmann@irht.cnrs.fr avant le 31 octobre 2017 à 15 h (CET), en envoyant :

1. lettre de motivation ;
2. curriculum vitae ;
3. noms et contact de trois références professionnelles.

Les candidats dont le dossier correspond aux critères énoncés seront sélectionnés pour une interview, en personne ou par vidéoconférence, avec deux chercheurs pendant 40 minutes.

Date souhaitée de prise de poste: entre le 1er janvier et le 1^{er} mars 2018

