

Appel à candidatures

L'UNIVERSITÉ CATHOLIQUE DE LOUVAIN

&

LES ARCHIVES DE L'ÉTAT

recrutent

Un/une spécialiste en **informatique** ou **humanités numériques** m/f/x

1 ETP à durée indéterminée : 50% PostDoc (UCLouvain) et 50% chef de travaux (AGR)

avec un intérêt particulier pour le traitement automatique de documents numériques impliquant à la fois du texte et des images

pour le programme FED-tWIN - Prf-2022-025 - ARKEY

#AI #NLP #ComputerVision #HTR #MetadataExtraction #UX #InfoVis

Contexte

FED-tWIN est un programme de recherche de la Politique scientifique fédérale belge (BELSPO) visant à promouvoir la coopération durable entre les établissements scientifiques fédéraux et les universités belges, par le financement de profils de recherche communs.

Le programme FED-tWIN permet aujourd'hui d'ouvrir une fonction pour le projet « A content-enriched user-centred access key to digital archives » (ARKEY). Il implique d'une part les Archives générales du Royaume et les Archives de l'État dans les Provinces (en bref les « Archives de l'État » ou AGR), institution scientifique fédérale, et d'autre part l'Université catholique de Louvain (UCLouvain).

Objectifs

L'objectif principal du profil de recherche ARKEY est d'optimiser la valorisation numérique des collections d'archives grâce à des outils informatiques pérennes. Il propose (1) la recherche et le développement de moyens d'accès enrichis au contenu numérisé (et nativement numérique), et (2) l'amélioration de l'expérience de navigation au sein des collections d'archives. Il s'appuie sur l'expertise d'une équipe pluridisciplinaire des AGR et de plusieurs groupes de recherche au sein de l'UCLouvain (voir section "Partenaires" ci-dessous). ARKEY vise à apporter une valeur ajoutée pour la

société et le service public en améliorant l'accessibilité et l'intelligibilité des archives : une priorité pour de nombreux chercheurs et chercheuses, et un fondement des États démocratiques.

Actuellement, les AGR et l'UCLouvain possèdent un grand nombre de documents numérisés et nativement numériques provenant d'une grande variété de sources issues d'époques différentes. Cette diversité pose un défi à l'analyse de contenu automatisée, en particulier aux outils de reconnaissance de texte, qui ne sont pas entraînés à de telles variations. Les archives sont également confrontées aux problématiques de stockage, de format, de métadonnées et de navigation des documents numériques : la plupart de ces documents ne sont pas suffisamment mis en lumière. Pour répondre à ces enjeux, ARKEY propose un projet en 3 étapes :

1. Reconnaissance de texte et de structure (*text and layout recognition*) assistée par l'Intelligence Artificielle (IA). ARKEY développera et évaluera des techniques semi-automatiques d'apprentissage machine (*machine learning*) pour l'analyse de contenu adaptées aux documents manuscrits et aux premiers livres imprimés. Ces approches s'appuieront sur les méthodes existantes de reconnaissance optique de caractères (OCR) et de reconnaissance de texte manuscrit (HTR), tout en intégrant les avancées récentes des architectures de type transformer pour le traitement conjoint du texte et de l'image. Une attention particulière sera portée à l'analyse de la structure des documents (*layout analysis*) et à l'extraction d'éléments structurants. ARKEY examinera également dans quelle mesure ces techniques sont applicables ou doivent être adaptées dans un contexte nativement numérique, ce qui permettra de mettre en œuvre les résultats du projet de la manière la plus large possible.

2. Représentation archivistique numérique enrichie par le contenu. Les données extraites de l'analyse de contenu seront exploitées pour enrichir la représentation des documents d'archive. Ce deuxième défi vise dès lors à étudier et à améliorer les méthodes de traitement automatique du langage pour enrichir les fichiers standardisés de description archivistique (EAD) avec des métadonnées générées automatiquement et basées sur la modélisation sémantique, la reconnaissance des entités nommées et l'expansion de requêtes.

3. Navigation orientée utilisateur et adaptée au contexte de recherche. Le troisième défi d'ARKEY est de permettre aux utilisateurs et utilisatrices d'archives d'améliorer leur expérience de navigation au sein des archives en se basant notamment sur les descriptions enrichies décrites ci-dessus. Cela implique la mise en œuvre d'une méthode de conception centrée sur les usages, et visant à élaborer de nouvelles formes d'aides à la recherche et d'outils de visualisation des corpus. En particulier, le projet contribuera à résoudre les deux problèmes suivants : (1) la méconnaissance des représentations archivistiques disponibles et de leurs relations les unes avec les autres, et (2) la difficulté de traduire une question initiale en une recherche spécifique et un scénario de navigation. A cette fin, le projet explorera notamment des approches de recherche sémantique et conversationnelle, par exemple à travers des interfaces interactives combinant recherche vectorielle et génération augmentée par la récupération d'information (*Retrieval-Augmented Generation*).

Partenaires

Les **Archives générales du Royaume et Archives de l'État dans les Provinces** (www.arch.be) sont un établissement scientifique fédéral faisant partie de BELSPO (Belgian Science Policy Office). L'institution comprend les Archives générales du Royaume à Bruxelles, dix-huit dépôts répartis dans tout le pays et le Centre d'Études et Documentation Guerre et Sociétés contemporaines (CegeSoma).

Les Archives de l'État acquièrent et conservent (après tri) les archives d'au moins trente ans des cours et tribunaux, des administrations publiques, de notaires, ainsi que du secteur privé et de particuliers. Elles veillent à ce que les archives publiques soient transférées selon les normes archivistiques. Mettre ces documents d'archives à la disposition du public, tout en respectant la protection de la vie privée, est une des missions primordiales de l'institution. Dans les dix-neuf salles de lecture, les Archives de l'État mettent une infrastructure à la disposition d'un public varié. La valorisation numérique des archives via Internet afin de mieux répondre aux besoins du public est une des priorités de l'institution.

L'**Université catholique de Louvain** (www.uclouvain.be) est une Université d'enseignement et de recherche de la Fédération Wallonie-Bruxelles en Belgique. Elle compte près de 35 000 étudiants et étudiantes, et 3000 chercheurs et chercheuses. Au sein de cette université, le MiiL (Media innovation and intelligibility Lab) est une plateforme interdisciplinaire d'innovation en production et appropriation digitale. Ancrée au sein de l'Institut de Recherche "Langage et Communication" (ILC), cette plateforme rassemble des experts de l'information et de la communication, des linguistes, des ingénieurs informaticiens, des juristes et des économistes, et offre des solutions adaptées à des problématiques portant sur l'intelligibilité des données numériques et leur représentation standardisée, l'expérience-utilisateur (UX), et la communication interactive. Dans le cadre de ce projet, la personne engagée travaillera au sein du MiiL, en étroite collaboration avec plusieurs autres équipes de recherche et services de l'Université, dont les expertises conjointes couvrent les 3 défis décrits ci-dessus. Le CENTAL tout d'abord, centre de traitement automatique du langage, qui rassemble de nombreux chercheurs spécialisés dans le traitement informatique des données textuelles. Le service des archives de l'UCLouvain ensuite, qui conserve les archives définitives de l'université et veille à la bonne gestion de l'ensemble des documents produits ou acquis par l'université. Le GEMCA enfin, centre d'analyse culturelle de la première modernité, qui est un centre de recherche interdisciplinaire rassemblant des chercheurs et chercheuses en littérature, histoire de l'art, histoire et philosophie.

Mission

Le chercheur ou la chercheuse devra contribuer directement à la réalisation des objectifs du projet décrits ci-dessus. Dans le cadre de ce programme FED-tWIN, elle ou il mènera des recherches et développera des méthodes, une expertise et des collaborations en vue de réaliser ces objectifs. Dans ce contexte, le travail de la personne engagée impliquera notamment

- une production scientifique de haut niveau et la publication de résultats des recherches effectuées par les voies appropriées;
- un engagement fort en faveur de la communication scientifique, tant orale qu'écrite, pour un public cible diversifié. Compte tenu du rôle social des AGR, en complément de la production académique sont également attendues des publications de vulgarisation ainsi que l'organisation d'événements de sensibilisation destinés au grand public;
- l'élaboration de dossiers de candidature pour des projets permettant le développement de la recherche scientifique interdisciplinaire et internationale, et la valorisation de la coopération entre la l'UCLouvain et les AGR;
- une activité pédagogique permettant la valorisation des acquis de la recherche dans des activités d'enseignement à l'UCLouvain.

Diplôme

Diplôme de doctorat en informatique, archives numériques, humanités numériques, traitement automatique du langage, ou dans un autre domaine directement pertinent pour la fonction.

Le titre de docteur doit avoir été obtenu au maximum 12 ans avant la date ultime d'introduction des dossiers de candidature¹.

Compétences techniques

Le profil recherché requiert des compétences techniques solides relatives au traitement automatique de documents numériques contenant à la fois du texte et des images.

En particulier, le candidat ou la candidate devra pouvoir démontrer:

- une expertise scientifique reconnue dans le **traitement des documents numérisés**. Cela comprend le traitement d'images (*pattern recognition* et *layout analysis*), la reconnaissance optique de caractères (OCR) et le traitement automatique du langage (NLP),
- une connaissance spécifique des techniques d'**apprentissage automatique** appliquées aux domaines mentionnés ci-dessus (Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), modèles de type *transformers*),
- des compétences de **programmation** de niveau expert en Python ou dans un langage de programmation similaire (une expérience avec PyTorch ou une autre bibliothèque d'apprentissage automatique est requise),
- au moins une expertise de base en **expérience utilisateur (UX)** et en conception centrée sur l'utilisateur, et la flexibilité nécessaire pour approfondir ces compétences.

Compétences relationnelles

La personne recherchée aura :

- des compétences en gestion de projet pour garantir la bonne conduite de la recherche, assurer la coordination entre les modules de travail, et stimuler la coopération avec les collègues des deux organisations si nécessaire,
- une capacité à rechercher des financements supplémentaires et à constituer et gérer une équipe de recherche.
- une attitude orientée vers le travail en équipe
- une forte expérience en *Open Science*, afin de rendre largement accessibles les résultats scientifiques. La personne engagée développera un plan de gestion des données conforme aux principes FAIR (*Findable, Accessible, Interoperable, et Reusable*).

¹ La période visée est prolongée d'un an par période de congé de maternité de trois mois au minimum ou par période ininterrompue de trois mois au minimum de congé parental à temps plein ou congé d'adoption à temps plein, pris par le candidat entre l'obtention du titre de doctorat et la date ultime d'introduction des dossiers de candidature, sans que la prolongation totale puisse excéder un an par enfant. La période visée au premier alinéa est également prolongée de la durée réelle des périodes attestées de maladie de longue durée du candidat ou d'un membre de famille proche du candidat auquel ce dernier a donné des soins médicaux, pour autant qu'il s'agisse de périodes ininterrompues de trois mois au minimum.

- une adhésion au « Code de déontologie de la recherche scientifique en Belgique » qui établit les grands principes d'une pratique scientifique éthique.

Compétences complémentaires

Constituent des atouts :

- un réseau professionnel dans le domaine de l'archivistique ou des humanités numériques et une connaissance des thématiques de recherche dans ces domaines,
- une expertise des normes internationales de description archivistique et de métadonnées (notamment EAD3, RiC, DCAT et DDI), ainsi que de la modélisation des systèmes (en particulier le modèle de référence OAIS) et de leur mise en œuvre dans des systèmes d'information professionnels pour la gestion des archives, l'interopérabilité, l'échange et la valorisation des données.

Connaissance des langues

Une connaissance de base du français est nécessaire dans la perspective d'une bonne intégration au sein des deux institutions. Si, au moment du recrutement, le candidat ou la candidate ne connaît pas ou pas suffisamment le français, l'UCLouvain lui proposera une offre de formation afin d'acquérir les connaissances requises. Le candidat doit également être capable de communiquer en anglais. Compte tenu de l'environnement bilingue des AGR (institution fédérale), la connaissance du néerlandais est également un avantage. L'UCLouvain peut également apporter un soutien à la formation dans ces deux langues.

Informations contractuelles

Le poste proposé est un poste à 100% structuré de la manière suivante :

- Un contrat à durée indéterminée à 50% à prester aux AGR à Bruxelles (rue de Ruysbroeck 2) au grade de chef de travaux (SW2). Échelle barémique SW21 (31.880,00 EUR – 48.350,00 EUR salaire brut non indexé, soit un salaire mensuel brut indexé d'environ 3.100 EUR pour ce mi-temps, sans autre ancienneté que le doctorat exigé). Tous les services prestés dans les services publics comme chercheur / chercheuse peuvent être valorisés.
- Un contrat à durée indéterminée à 50% à prester à l'UCLouvain (50% chercheur ou chercheuse post-doctoral) à Louvain-la-Neuve (ruelle de la Lanterne Magique, 14). Engagement suivant les règles barémiques en cours à l'université et dépendant de l'ancienneté.

Financement garanti sur une durée de 7,5 ans, avec intention de pérennisation du poste.

L'entrée en fonction est prévue dès que possible.

Intéressé(e) ?

Les candidat(e)s doivent transmettre leur dossier de candidature (un seul fichier au format PDF) avant le **10 mai 2026** par courriel à johan.vandereycken@arch.be **et** à antonin.descampe@uclouvain.be (référence : candidature FEDTWIN/ARKEY).

Le dossier de candidature doit se composer d'une lettre de motivation, d'un CV détaillé avec liste des publications, d'une copie du diplôme requis et de deux lettres de recommandation.

Les auteurs des lettres de recommandation sont invités à envoyer directement leur courrier à Antonin Descampe et Johan Van Der Eycken, sans passer par le candidat ou la candidate.

Les candidat(e)s sélectionné(e)s sur base de leur dossier seront invité(e)s pour une audition orale par la Commission conjointe de sélection, prévue dans le courant des mois de mai ou juin.

Pour toute information complémentaire, veuillez prendre contact avec M. Antonin Descampe (antonin.descampe@uclouvain.be) ou M. Johan Van Der Eycken (johan.vandereycken@arch.be).

L'UCLouvain et les AGR veulent créer un environnement de travail dans lequel tous les talents peuvent se développer autant que possible, sans distinction de sexe, d'âge, d'origine culturelle, de nationalité ou de handicap. Pour toute question concernant l'accessibilité ou les possibilités de soutien, veuillez consulter <https://jobs.uclouvain.be/content/ValeursRH/> et/ou contacter le service HR des AGR (pers@arch.be).