Fiche de poste

Intitulé du stage :
Correction et validation de transcriptions HTR de sentences militaires italiennes de la Première Guerre mondiale

Fiche descriptive du stage

Résumé du stage : le ou la stagiaire participera aux activités du projet de recherche *Sur le front de la justice. Une lecture numérique de la justice militaire pendant la Première Guerre mondiale* (FRONT_JUSTICE). Le projet étudie les pratiques de la justice militaire française et italienne pendant la Première Guerre mondiale à partir d'un vaste corpus d'archives judiciaires. Il associe histoire, humanités numériques et traitement automatique des documents afin d'exploiter à grande échelle des sources jusqu'ici difficiles à analyser de manière sérielle.

FRONT_JUSTICE bénéficie depuis 2024 du soutien de la Fondation des Sciences du Patrimoine. Le présent stage est financé par le DIM PAMIR (Patrimoines matériels – innovation, expérimentation et résilience), Région Île-de-France.

Le stage portera principalement sur le corpus italien du projet, constitué de sentences de la justice militaire italienne de la Première Guerre mondiale. Le travail sera réalisé notamment dans la plateforme eScriptorium (https://www.sofer.info), environnement utilisé par le projet pour la reconnaissance automatique des textes et l'entraînement de modèles HTR (Handwritten Text Recognition). L'activité principale consistera **à vérifier et corriger les transcriptions automatiques en confrontant les images des documents historiques des sentences militaires au front avec les textes reconnus.** Les transcriptions corrigées et validées constitueront des données d'entraînement permettant d'améliorer progressivement les modèles de reconnaissance automatique. Le ou la stagiaire pourra également contribuer **à la préparation et à la structuration du corpus en vue de son exploitation historique et numérique.**

Condition d'éligibilité : Le stage est exclusivement réservé aux étudiants et étudiantes inscrits dans un établissement d'enseignement supérieur situé en Île-de-France, conformément aux conditions de financement du DIM PAMIR.

Affectation et environnement de travail

Type de contrat : Stage de Master

Durée : 6 mois temps plein (35 heures la semaine)

Structure d'accueil : Université de Versailles Saint-Quentin-en-Yvelines (UVSQ) – Université Paris-Saclay

Laboratoire : CHCSC – Centre d'histoire culturelle des sociétés contemporaines (EA 2448)

Modalités de travail : Le stage se déroulera principalement à distance, avec des rencontres ponctuelles en présentiel à l'UVSQ.

Responsable scientifique : Giovanni Pietro Vitali

Financement du stage : DIM PAMIR – Région Île-de-France

Missions

Activités principales :

1. Prise en main du corpus italien de FRONT_JUSTICE et de l'environnement eScriptorium.
2. Vérification des transcriptions automatiques à partir des images des sentences militaires italiennes.
3. Correction et validation des transcriptions dans eScriptorium.
4. Constitution de données d'entraînement de qualité pour l'apprentissage automatique et l'amélioration des modèles HTR.
5. Participation, selon l'avancement du stage, à la structuration des données et à la préparation du corpus pour son exploitation historique et numérique.
6. Documentation du travail réalisé afin d'assurer la ré-utilisabilité des données produites dans le cadre du projet.

Formation et accompagnement : Le ou la stagiaire bénéficiera d'une formation initiale et d'un accompagnement tout au long du stage pour l'ensemble des activités mentionnées ci-dessus, notamment la prise en main d'eScriptorium, la vérification et la correction des transcriptions, la constitution de données d'entraînement pour les modèles HTR, ainsi que la structuration et la documentation du corpus. Aucune expérience préalable de ces outils et méthodes n'est donc requise.

Compétences et profil recherché

Profil : étudiant ou étudiante inscrit(e) obligatoirement dans un établissement d'Île-de-France, en Master, de préférence en histoire, humanités numériques, études italiennes, lettres, linguistique, archivistique ou dans une discipline proche.

Compétence linguistique indispensable : le ou la candidat(e) doit être italophone ou justifier d'un niveau d'italien au moins C1. Une excellente compréhension de l'italien écrit est indispensable pour la lecture, la vérification et la correction des sources historiques.

Compétences souhaitées :

1. Intérêt pour l'histoire de la Première Guerre mondiale et/ou pour les sources historiques.
2. Intérêt pour les humanités numériques et le traitement automatique des documents.
3. Rigueur dans la lecture, la transcription et la correction de textes.
4. Une connaissance préalable d'eScriptorium, de la HTR ou des principes de l'apprentissage automatique est appréciée mais n'est pas indispensable.
5. Capacité à travailler de manière autonome tout en s'intégrant à une équipe de recherche interdisciplinaire.

Modalités de candidature

Date de début envisagée : à partir du 1er février 2027 (idéalement 1er février et le 1er avril au plus tard.), sous réserve de confirmation des modalités administratives du DIM PAMIR.

Candidature : CV et courte lettre de motivation à envoyer avant le 16 novembre 2026. Les entretiens se tiendront en visioconférence le 23 novembre 2026.

Date de début du stage : 1er février 2027.

Contact scientifique : Giovanni Pietro Vitali – giovanni.vitali@uvsq.fr

Projet : FRONT_JUSTICE – https://www.pamir.fr/projets-soutenus/front_justice/

Rétribution : Gratification au tarif en vigueur, soit 4,50 € de l'heure.