Mickaël BARON
École Nationale Supérieure de Mécanique et d'Aérotechnique
LIAS – EA 6315
1 avenue Clément Ader BP 40109
86961 Chasseneuil Cedex – France
Tél : +33 (0)5 49 49 80 73
baron@ensma.fr

Extraction semi-automatique d'annotations sémantiques pour la préservation du patrimoine culturel

Intégration et exploitation de thésaurus spécialisés

Localisation : Site du Futuroscope, France
Date de début : mars 2026 (flexible)
Durée : 6 mois
Encadrants : Mickaël BARON, Ali HARIRI et Stéphane JEAN *(l'encadrant dont le nom est souligné est l'encadrant référent : baron@ensma.fr)*
Mots clés : web sémantique, héritage culturel, IA générative

Contexte du stage

L'étude du patrimoine repose sur des données hétérogènes provenant de sources multiples (mesures, croquis, photographies, acquisitions 3D, etc.), souvent difficiles à manipuler en raison de la fragilité ou de l'inaccessibilité des objets. Le projet ANR DIGITALIS [1] vise à développer des outils numériques dédiés à la gestion, à la pérennisation, à la réutilisation et à la visualisation de ces données. Le laboratoire LIAS, partenaire du projet ANR, travaille sur la conception de méthodes et de structures pour gérer des données complexes de manière interopérable.

Dans le cadre du projet ANR DIGITALIS, cette expertise est utilisée pour proposer un modèle de données dédié au patrimoine, permettant aux experts de partager et d'enrichir leurs annotations sur des objets patrimoniaux. Ce besoin s'inscrit dans la continuité de travaux existants, comme le modèle sémantique CIDOC CRM [2], qui vise à assurer la traçabilité et l'annotation des objets numériques dans le domaine historique. Le principal enjeu est d'automatiser l'extraction des données vers ce modèle, une tâche complexe et chronophage, d'autant plus que les historiens maîtrisent peu ce formalisme et que les informaticiens ne sont pas spécialistes du patrimoine.

Le laboratoire LIAS a obtenu des résultats en explorant l'usage de l'IA générative pour transformer des données textuelles issues de fouilles archéologiques en représentations conformes au modèle CIDOC CRM.

Objectifs du stage

Les objectifs du stage sont multiples à vise à compléter les travaux obtenus.

Le premier objectif porte sur le **compromis entre précision, performance et impact en-**

1. https ://digitalis.humanities.science
2. https ://cidoc-crm.org/

Mickaël BARON
École Nationale Supérieure de Mécanique et d'Aérotechnique
LIAS – EA 6315
1 avenue Clément Ader BP 40109
86961 Chasseneuil Cedex – France
Tél : +33 (0)5 49 49 80 73
baron@ensma.fr

vironnemental, ainsi que sur les enjeux de vie privée liés aux déploiements cloud ou locaux. Le stagiaire devra donc reprendre les résultats existants et les évaluer sur différents modèles d'IA générative, qu'ils soient hébergés dans le cloud ou exécutés localement via des modèles à poids ouverts de plus petite taille. Cela permettra au stagiaire de se familiariser avec les concepts de base de l'IA générative et des travaux développés par le laboratoire LIAS dans ce domaine.

Le deuxième objectif porte sur la **capacité à référencer les sources** (mesures, croquis, photographies, acquisitions 3D, etc.) utilisées pour produire les annotations. Cette question, non traitée dans nos travaux initiaux, est pourtant essentielle pour assurer la traçabilité des annotations. Le stagiaire devra ainsi explorer des solutions permettant d'intégrer explicitement la notion de source dans le processus de génération des annotations. Cela permettra au stagiaire de se familiariser avec le modèle CIDOC CRM.

Le troisième objectif, qui constitue la contribution principale de ce stage, porte sur l'**intégration de vocabulaires contrôlés**, tels que des thésaurus, dans le processus de génération des annotations. Cette intégration vise à améliorer la cohérence et la qualité des annotations, notamment en facilitant la gestion des synonymes (par exemple : église <=> lieu de culte).

Ce stage recherche sera encadré par un doctorant travaillant sur des problématiques similaires, ainsi que deux chercheurs du laboratoire LIAS spécialisés dans le domaine de l'IA générative et du patrimoine. Une publication scientifique serait attendue à l'issue du stage, en fonction des résultats obtenus.

À noter enfin que le laboratoire LIAS proposera l'an prochain un financement de thèse dans le domaine de la gestion des données. Le stage constituera une excellente opportunité de découvrir le laboratoire et son environnement de recherche. Le stagiaire intéressé pourra candidater à ce financement, et sa candidature fera l'objet d'une attention particulière.

Profil du candidat

Le candidat doit être en Master 2 en Informatique ou en dernière année de préparation d'un diplôme d'ingénieur spécialité Informatique. Une bonne connaissance du langage de programmation Python et des bibliothèques usuelles d'apprentissage automatique est requise. Le stage se déroulera dans les locaux du LIAS sur le site du Futuroscope.

Documents à fournir

Les deux premiers documents suivants sont indispensables ; sans eux, la candidature sera rejetée.

- Curriculum Vitae ;
- Notes de Master ou équivalent (si possible avec le classement) ;
- Autres documents utiles pour appuyer la candidature (ex. : projets, GitHub).