DYPA

Dynamiques patrimon
et cultur

M. SERGIO TORRES AGUILAR

CHERCHEUR(EUSE)
Docteur
Université de Versailles Saint-Quentin-en-Yvelines > IECI - Institut d'études culturelles et internationales > Laboratoire dynamiques patrimoniales et culturelles (DYPAC) >

Coordonnées

ADRESSE MAIL sergio.torres-aguilar@uvsq.fr

Thèmes de recherche

Sujet de thèse :

Un modèle de reconnaissance automatique des entités nommées et des structures textuelles pour les éditions TEI de textes diplomatiques.

Directeurs de thèse :

- Pierre Chastang (Professeur des universités, Université de Versailles Saint-Quentin-en-Yvelines)
- Xavier Tannier (MCF HDR, Université Paris-Sud)

Description du projet scientifique :

Le projet s'agit de créer un modèle d'interaction afin de générer de manière automatique, pour les éditions numériques de textes anciens, des annotations de contenu et des indexations concernant les structures discursives. En bénéficiant des progrès récents de la NER (Named Entities Recognition), technique de la linguistique informatique, et en utilisant la structure de la TEI (Text Encoding Initiative), modèle d'échange des textes électroniques fondé sur le langage XML, nous souhaitons permettre une reconnaissance de la présence d'entités nommées, produire une annotation sémantique des textes et, en même temps, identifier les structures formelles fixes qui composent leurs parties discursives. Dans ce but, nous utiliserons certains corpus de textes médiévaux : chartes ecclésiastiques et seigneuriales (donations, ventes, etc.), archives de chancellerie, ordonnances royales et documents notariaux. Ces types de textes conviennent particulièrement à cette recherche, car ils combinent des formes discursives et des formules fixes avec une riche information nominale. Nous souhaitons donc explorer les capacités de la NER sur un nouveau type de documents afin de générer un processus d' annotation plus précis, et d'analyser également la structure des textes ; l'objectif étant d'accélérer une tâche très fastidieuse et complexe et de retrouver de manière automatisée un niveau d'information plus avancé et plus structuré que le texte brut.

Activités / CV

Après des premières études au Pérou dans le domaine de la gestion culturelle, Sergio Torres Aguilar a obtenu une licence en histoire complémentée avec des études en langues classiques à l'Université d'Oviedo (2013). Il a ensuite obtenu un double master à Madrid dans le domaine de l'histoire médiévale à l'Université Complutense de Madrid, avec une mémoire sur la cartographie médiévale des XIIe-XVe siècles, et dans les domaines des bibliothèques et des humanités numériques à l'Université Carlos III de Madrid, en travaillant sur la discipline innovante de la paléographie digitale (2014). Il s'a bénéficié, au cours de l'année 2014-2015, d'une bourse d'excellence international Paris-Saclay, ce que lui a permis de poursuivre un master recherche en histoire au sein de l'Université de Versailles Saint-Quentin-en-Yvelines, en travaillant sur le catalogage

numérique et la valorisation du fonds Chappée en partenariat avec la Bibliothèque nationale de France.

Son projet doctorale, dans le cadre de l'IDI (Initiative doctorale interdisciplinaire) de l'Université Paris-Saclay en cotutelle avec DYPAC de l'UVSQ et le laboratoire LIMSI de l'Université Paris-Sud, s'attache a développer un modèle d'interaction entre la TEI (Text Encoding Initiative) et la NER (Named Entities Recognition) visant à l'automatisation du balisage pour les corpus médiévales numérisés à partir de l'identification des entités nommées et à la reconnaissance des structures discursives au sein des éditions diplomatiques numériques.