

Reconnaissance et indexation automatique des registres de la chancellerie française (1300-1483)

Christopher Kermorvant

TEKLIA
www.teklia.com

Summary

Les documents manuscrits sont parmi les témoins les plus importants de l'histoire européenne. Ces dernières années, d'importantes collections de manuscrits historiques ont été numérisées et mises à disposition du public et des chercheurs. Cependant, la richesse des informations qu'ils contiennent est encore largement inaccessible car seul les images et quelques méta-données sont disponibles. L'idéal pour les utilisateurs serait de pouvoir faire des recherches textuelles comme pour les livres imprimés modernes (<https://books.google.fr/>). Si les technologies d'analyse de documents historiques et de reconnaissance d'écriture manuscrite sont encore trop peu performantes pour permettre l'utilisation directe de la transcription brute, il est possible de mettre à la disposition des utilisateurs un moteur de recherche textuel basé sur une indexation automatique des images de documents manuscrits. Cette indexation se base sur une transcription automatique mais tire profit de la capacité de la machine à générer des hypothèses reconnaissance multiples et pondérées.

Cette technologie a permis de rendre accessible pour la première fois à la recherche textuelle les registres de la chancellerie royale française (1302 -1483), un des corpus de documents historiques les plus emblématiques pour la France, ouvrant ainsi la voie à de nouvelles méthodes de recherche en histoire : <http://www.himanis.org/>

Biography

Christopher Kermorvant, est ingénieur ENSIIE (1996) et docteur en informatique (2002). Depuis 20 ans, il travaille au développement d'applications utilisant des technologies de Machine Learning. Après une expérience académique en France (doctorant au laboratoire Hubert Curien de l'Université de Saint-Etienne) et à l'étranger (ingénieur de recherche au laboratoire IDIAP de l'Ecole Polytechnique Fédérale de Lausanne, post-doctorant au laboratoire MILA de l'Université de Montréal), il a dirigé pendant 8 ans une équipe de recherche au sein d'A2iA, une PME high-tech, spécialisée dans la reconnaissance d'écriture manuscrite. Avec son équipe, il a développé de nouveaux systèmes de reconnaissance d'écriture basés sur des réseaux de neurones profonds. Ces systèmes se sont classés premier lors d'évaluations internationales de reconnaissance d'écriture en français, anglais, arabe (Rimes, OpenHart, Maurdor) et sont depuis été intégrés dans la gamme des produits de la société A2iA.

Depuis 2015, il travaille en tant qu'expert indépendant en Machine Learning (www.teklia.com) pour aider les entreprises à développer des produits innovants basés sur des technologies de Machine Learning tout en poursuivant des projets de recherche collaborative en Digital Humanities.