



## **AVIS DE PRESENTATION DE THESE EN SOUTENANCE POUR L'OBTENTION DU DIPLOME NATIONAL DE DOCTEUR**

**Monsieur Tien Nam NGUYEN**

Présentera ses travaux intitulés :

**« Segmentation, Reconnaissance et Indexation de caractères dans les documents CHAM »**

Spécialité : Informatique et applications

**Le 12 juillet 2023 à 10h00**

Lieu :

**La Rochelle Université  
Pôle Communication, Multimédia et Réseaux  
Amphithéâtre Michel Crépeau  
44 Av. Albert Einstein  
17000 LA ROCHELLE**

Composition du jury :

**M. BURIE Jean-Christophe  
Mme CLOPPET Florence  
M. LE Thi Lan**

**Mme LEMAITRE Aurélie  
M. RAMEL Jean-Yves  
Mme SCHWEYER Anne-Valérie**

**M. TABBONE Antoine**

**Professeur, La Rochelle Université  
Professeure, Université Paris-Cité  
Associate Professor, Hanoi University of Science and  
Technology  
Maitresse de conférences, Université de Rennes 2  
Professeur, Université de Tours  
Chargée de recherche, Ecole des Hautes Études en  
Sciences Sociales Paris  
Professeur, Université de Lorraine**

### **Résumé :**

Depuis la disparition des royaumes Champa au cours du 19<sup>e</sup> siècle, la langue Cham qui est née et s'est développée à partir de 2<sup>e</sup> siècle, n'est plus vraiment utilisée chez les descendants des Champa. Le manque de transmission des connaissances et des documents de la culture Cham rend, l'étude de cette langue, difficile pour les épigraphistes et les historiens.

Par conséquent, le projet ANR CHAMDOC vise à préserver et proposer des outils pour étudier la langue Cham. Dans cette thèse, nous nous sommes concentrés sur l'analyse de deux types de documents Cham à savoir : des inscriptions, qui ont été gravées sur des stèles en pierre, du 6<sup>e</sup> au 15<sup>e</sup> siècle et des manuscrits datant du 18<sup>e</sup> siècle. Quelques travaux sur la numérisation des inscriptions ont été menés mais aucune étude n'a vraiment été réalisée. La collection de manuscrits, quant à elle, n'a jamais été exploitée. Ces deux collections de documents offrent de nombreux défis pour la communauté scientifique.

Au cours de ces travaux, nous proposons un pipeline complet pour le traitement automatique de ces documents. Celui-ci est basé sur différentes techniques d'analyse d'images de documents.

Les défis rencontrés proviennent des caractéristiques des documents eux-mêmes, mais aussi des spécificités linguistiques du Cham. Une analyse de ces caractéristiques a été menée afin de proposer des solutions adaptées aux inscriptions et aux manuscrits.