



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

Madame Chloé ARTAUD

Présentera ses travaux intitulés :

« Détection des fraudes : de l'image à la sémantique du contenu »

Spécialité : Informatique et applications

Le 6 février 2019 à 14h00

Lieu :

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

Composition du jury :

**M. BELLOT Patrice
M. CLAVEAU Vincent
Mme DAILLE Béatrice
M. DOUCET Antoine
Mme GOMEZ KRÄMER Petra
M. OGIER Jean-Marc
Mme SERFATY Véronique (*Invitée*)
Mme VINCENT Nicole**

**Professeur, Université Aix Marseille
Chargé de recherche, CNRS, IRISA
Professeure, Université de Nantes
Professeur, la Rochelle Université
Maître de conférences, la Rochelle Université
Professeur, la Rochelle Université
Responsable de domaine scientifique, DGA
Professeure, Université Paris Descartes**

Résumé :

Les entreprises, les administrations, et parfois les particuliers, doivent faire face à de nombreuses fraudes sur les documents qu'ils reçoivent de l'extérieur ou qu'ils traitent en interne. Les factures, les notes de frais, les justificatifs... tout document servant de preuve peut être falsifié dans le but de gagner plus d'argent ou de ne pas en perdre. En France, on estime les pertes dues aux fraudes à plusieurs milliards d'euros par an. Etant donné que le flux de documents échangés, numériques ou papiers, est très important, il serait extrêmement coûteux en temps et en argent de les faire tous vérifier par des experts de la détection des fraudes. C'est pourquoi nous proposons dans notre thèse un système de détection automatique des faux documents.

Si la plupart des travaux en détection automatique des faux documents se concentrent sur des indices graphiques, nous cherchons quant à nous à vérifier les informations textuelles du document afin de détecter des incohérences ou des invraisemblances. Pour cela, nous avons tout d'abord constitué un corpus de tickets de caisse que nous avons numérisés et dont nous avons extrait le texte. Après avoir corrigé les sorties de l'OCR et fait falsifier une partie des documents, nous en avons extrait les informations et nous les avons modélisées dans une ontologie, afin de garder les liens sémantiques entre elles. Les informations ainsi extraites, et augmentées de leurs possibles expansions, peuvent être vérifiées les unes par rapport aux autres au sein du document et à travers la base de connaissances constituée. Les liens sémantiques de l'ontologie permettent également de chercher l'information dans d'autres sources de connaissances, et notamment sur Internet.