



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

**Monsieur Edward GIAMPHY**

Présentera ses travaux intitulés :

**« Graphes de connaissances pour la représentation du contexte en traitement  
automatique du langage naturel »**

Spécialité : Informatique et applications

**Le 4 mars 2024 à 10h15**

Lieu :

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

Composition du jury :

**Mme BERTET Karell  
M. BONALD Thomas  
Mme DASHYAN Gohar  
M. DOUCET Antoine  
M. FOURNIER Sébastien  
M. GUILLAUME Jean-Loup  
Mme LARGERON Christine**

**Professeure, La Rochelle Université  
Professeur, Télécom Paris Tech  
Manager de Recherche, Entreprise Preligens  
Professeur, La Rochelle Université  
Professeur, Université Aix Marseille  
Professeur, La Rochelle Université  
Professeure, Université Jean Monnet Saint – Étienne**

### **Résumé :**

La numérisation grandissante de nos sociétés induit une prolifération des textes numérisés. Le traitement automatique du langage naturel (TALN) devient alors indispensable pour gérer le nombre croissant de documents numérisés. Le domaine de la recherche d'information a ainsi émergé pour faciliter l'accessibilité des contenus textuels numérisés recherchés par un utilisateur. Ce dernier usage constitue un comportement qui s'est démocratisé dans le monde socio-économique moderne par l'utilisation massive des moteurs de recherche.

La majorité des approches traitant les textes écrits, reposent sur un découpage au préalable des textes en unités linguistiques, le mot étant l'exemple d'unité linguistique le plus évident. Cette thèse s'interroge sur l'usage du découpage du texte en unités linguistiques et explore des méthodes plus aptes à capturer la proximité sémantique et contextuelle des textes. C'est pourquoi nous étudions l'apport des graphes pour le TALN et notamment leur capacité à représenter de manière structurée les relations complexes entre les concepts contextuels des textes. Nous nous intéressons d'abord aux plongements de graphes, une technique de représentation compressée permettant d'exploiter la structure de graphe. Nous étudions ensuite le potentiel des graphes de connaissances pour la reconnaissance d'entités nommées, une technique qui caractérise contextuellement un texte. Enfin, nous analysons l'impact du bruit dans le texte sur la recherche d'information, un problème peu étudié dans la littérature. Pour conclure, nous proposons une méthode de recherche d'information reposant sur l'utilisation d'un graphe de connaissances et d'une représentation contextuelle des unités linguistiques.