

Informatique et applications

Soutiendra publiquement ses travaux de thèse intitulés

Analyse de documents médicaux avec plongements de concepts

dirigés par Monsieur Zied BOURAOUI et Monsieur Jean François CONDOTTA

Soutenance prévue le **vendredi 29 novembre 2024** à 14h00

Lieu : 13 Rue Jean Souvraz, 62300 Lens Faculté des sciences jean perrin Université d'Artois

Salle : des thèses

Composition du jury proposé

M. Miguel COUCEIRO	Universidade de Lisboa	Rapporteur
Mme Afef KACEM	Université de Tunis	Examinatrice
M. Zied BOURAOUI	Université d'Artois	Directeur de thèse
M. Jean-François CONDOTTA	Université d'Artois	Co-directeur de thèse
M. Nasredine SEMMAR	Universite Paris-Saclay	Examineur
M. Frederic KORICHE	Université d'Artois	Examineur
Mme Fatiha SAIS	Université Paris-Saclay	Rapporteure

Résumé :

Pour associer des documents à des étiquettes dans le traitement des documents médicaux, il est essentiel d'apprendre des représentations efficaces qui alignent le contenu des documents avec les étiquettes médicales appropriées. Le défi réside dans la capacité à capturer avec précision les caractéristiques pertinentes du texte médical qui correspondent à ces étiquettes, lesquelles peuvent représenter des concepts médicaux complexes. Cette thèse vise à relever ce défi en se concentrant sur l'apprentissage de représentations pour les documents et les étiquettes. L'objectif est de garantir que les plongements générés pour les textes soient non seulement riches en informations sémantiques, mais également bien alignés avec les étiquettes cibles. Nous employons d'abord un apprentissage contrastif auto-supervisé basé sur des segments pour générer des plongements de documents à partir de textes médicaux. Cependant, étant donné que tous les segments d'un document n'ont pas la même importance pour la classification, nous décomposons davantage la représentation des documents en plongements de concepts, en mettant l'accent sur les entités médicales clés cruciales pour l'étiquetage. De plus, nous tirons parti des connaissances préalables sur les concepts et les entités pour améliorer l'analyse des documents médicaux. Cela est réalisé en distillant des représentations conceptuelles à partir de modèles de langage, en se concentrant sur l'identification des points communs entre les concepts. Cette approche permet un meilleur support pour des tâches telles que le codage automatique des comptes rendus médicaux, la classification de documents, le complétion automatique des ontologies le typage ultra-fine d'entités, améliorant ainsi l'analyse des documents médicaux.