

AVIS RECTIFICATIF DE SOUTENANCE DE THESE

Le Doyen de la Faculté des Sciences Dhar El Mahraz –Fès – annonce que

Mr **FOUADI Hassan**

Soutiendra : **le Mercredi 20/11/2024 à 15h00**

Lieu : **FSDM – Centre Visioconférence**

Une thèse intitulée :

**« Analyse du Texte Arabe Multi-Dialecte Basée sur des Approches
de Deep Learning »**

En vue d'obtenir le **Doctorat**

FD : **Sciences et Technologies de l'Information et de la Communication**
Spécialité : **Informatique**

Devant le jury composé comme suit :

Nom et prénom	Etablissement	Grade	Qualité
Pr TAIRI Hamid	Faculté des Sciences Dhar El Mahraz, USMBA, Fès	PES	Président
Pr BENKIRANE Said	Ecole Supérieure de Technologie, Université Cadi Ayyad, Essaouira	MCH	Rapporteur & Examineur
Pr EL FAZAZY Khalid	Faculté des Sciences Dhar El Mahraz, USMBA, Fès	PES	Rapporteur & Examineur
Pr KHAMJANE Aziz	Ecole Nationale des Sciences Appliquées, Université Abdelmalek Essaâdi, Al Hoceima	MCH	Rapporteur & Examineur
Pr MAHRAZ Mohamed Adnane	Faculté des Sciences Dhar El Mahraz, USMBA, Fès	MCH	Examineur
Pr SABRI My Abdelouahed	Faculté des Sciences Dhar El Mahraz, USMBA, Fès	PES	Examineur
Pr TAZI El Bachir	Faculté Polydisciplinaire, USMBA, Taza	MCH	Examineur
Pr EL MOUBTAHIJ Hicham	Ecole Supérieure de Technologie, Université Ibn Zohr , Agadir	MCH	Co-Directeur de thèse
Pr YAHYAYOUY Ali	Faculté des Sciences Dhar El Mahraz, USMBA, Fès	PES	Directeur de thèse

Résumé :

Ce travail présente des solutions innovantes pour relever les défis de la classification de textes arabes multi-dialectes, une tâche rendue complexe par la diversité linguistique de l'arabe et la rareté des données annotées. Afin de pallier ce manque, plusieurs ensembles de données spécialisés ont été créés, tels que Arabic-Review, Topic_Dial, Identif_Dial et Senti_Dial, couvrant des dialectes comme le Darija marocain, le Dziri algérien, l'Égyptien, le Libanais et le Tunisien. Ces corpus ont servi à adapter et fine-tuner des modèles monodialectaux tels que DarijaBERT et DziriBERT. En comparaison aux modèles génériques comme MARBERT, ces modèles spécialisés ont montré une amélioration significative des performances, soulignant l'importance de personnaliser les modèles selon les spécificités de chaque dialecte pour optimiser la classification des textes arabes. Cette thèse apporte une contribution notable au traitement automatique du langage naturel appliqué à la langue arabe, en combinant la création de ressources linguistiques dédiées et l'utilisation de modèles d'apprentissage profond spécialisés, ouvrant ainsi la voie à des applications plus précises dans divers contextes, comme l'analyse des sentiments, la détection de dialecte et la classification thématique.

Mots clés :

Classification de textes, Analyse de sentiments, Deep Learning, Modèles de langues, BERT, DarijaBERT, DziriBERT, MARBERT

MULTI-DIALECT ARABIC TEXT ANALYSIS BASED ON DEEP LEARNING APPROACHES

Abstract :

This thesis presents innovative solutions to address the challenges of multi-dialect Arabic text classification, a task made complex by the linguistic diversity of Arabic and the scarcity of annotated data. To overcome this limitation, several specialized datasets were created, such as Arabic-Review, Topic_Dial, Identif_Dial, and Senti_Dial, covering dialects including Moroccan Darija, Algerian Dziri, Egyptian, Lebanese, and Tunisian. These corpora were used to adapt and fine-tune monolingual models like DarijaBERT and DziriBERT. Compared to generic models like MARBERT, these specialized models demonstrated a significant performance improvement, highlighting the importance of customizing models to the specific characteristics of each dialect for optimal Arabic text classification. This thesis makes a significant contribution to natural language processing applied to Arabic, combining the creation of dedicated linguistic resources and the use of specialized deep learning models, thus paving the way for more accurate applications in various contexts, such as sentiment analysis, dialect detection, and topic classification.

Key Words :

Text classification, Sentiment analysis, Deep learning, Language models, BERT, DarijaBERT, DziriBERT, MARBERT