

Résumé (250 mots Max) :

L'intégration de l'intelligence artificielle (IA) a révolutionné les interactions numériques, notamment grâce au traitement automatique du langage naturel (TALN). Bien que les avancées soient significatives pour l'anglais, le TALN appliqué à l'arabe demeure complexe en raison de sa morphologie riche, de la diglossie et du manque de ressources. Cette thèse propose BAHI, un cadre de spécification reposant sur quatre piliers : Compréhension Bidirectionnelle, Conception Centrée sur l'Arabe, Haute Performance et Interprétabilité, afin de concevoir des systèmes robustes, adaptés et transparents. Pour valider ce cadre, le système Mukhtassar a été développé, intégrant trois modules : résumé extractif, analyse de sentiment et modélisation de sujets. Basé sur des modèles transformeurs (AraBERT), il inclut des prétraitements spécifiques, des stratégies syntaxiques et des mécanismes d'interprétabilité. Les expérimentations sur des corpus arabes montrent des performances supérieures aux approches traditionnelles. Cette thèse apporte un cadre réutilisable, un système complet et une validation empirique, contribuant à combler le fossé entre IA avancée et besoins linguistiques arabes.

Mots Clés: Traitement Automatique du Langage Naturel en Arabe, Résumé de texte, Analyse de sentiment, Modélisation de sujets, Cadre BAHI (Compréhension bidirectionnelle, Conception centrée sur l'arabe, Haute performance, Interprétabilité), Système Mukhtassar.

Absract (250 words Max) :

Artificial Intelligence (AI) has advanced Natural Language Processing (NLP), enabling systems to interpret and generate human language. However, Arabic NLP remains challenging due to its complex morphology, syntactic variability, diglossia, and limited resources. This thesis introduces BAHI, a specification framework for Arabic NLP built on four pillars: Bidirectional Understanding for contextual semantics, Arabic-Centric Design for tailored preprocessing and modeling, High Performance through empirical validation, and Interpretability for transparent decision-making. To validate BAHI, we developed Mukhtassar, a modular platform comprising extractive summarisation, sentiment analysis, and topic modeling. These modules leverage transformer-based models (AraBERT), Arabic-specific pre-processing, POS-based scoring, and interpretability features such as attention-based sentiment cues and keyword-driven topic explanations. Experiments on Modern Standard Arabic corpora, evaluated using ROUGE, F1-score, accuracy, and coherence metrics, show Mukhtassar outperforming traditional models. Contributions include a reusable framework, a validated system, and empirical evidence bridging AI and Arabic linguistic needs.

Key words : Arabic Natural Language Processing, Text Summarisation, Sentiment Analysis, Topic Modeling, BAHI Framework (Bidirectional Understanding, Arabic-Centric Design, High Performance, Inter-pretability), Mukhtassar System.

CENTRE DES ETUDES DOCTORALES :

SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

FORMATION DOCTORALE : Informatique et Sciences des Données et de l'Information

THÈSE DE DOCTORAT

Présentée Par:

Mlle Ghizlane BOURAHOUAT

Sous le thème

Enhancing ANLP through the BAHI Framework: Specification and realisation through the Mukhtassar System

Soutenue le 14/02/2026 devant le Jury composé de :

Nom et Prénom	Grade	Statut	Etablissement
Pr. Sanaa EL FKIH	PES	Président	ENSIAS, Rabat
Pr. BAINA Amine	PES	Rapporteur	INPT, Rabat
Pr. EL FADDOULI Nour-Eddine	PES	Rapporteur	EMI, Rabat
Pr. HILAL Imane	PH	Rapporteur	ESI, Rabat
Pr. CHERIF Walid	PH	Examineur	ESI, Rabat
Pr. MIKRAM Mounia	PES	Examineur	ESI, Rabat
Pr. Manar ABOUREZQ	PH	Co-Directeur	ESI, Rabat
Pr. Najima DAOUDI	PES	Directeur	ESI, Rabat

LABORATOIRE : (LyRICA) Laboratoire de Recherche en Informatique, Sciences de données et Intelligence Artificielle