Sawab

Modern Standard and Classical Arabic Track

ثوّب

ثوّب ظاهرة إشكالية في الإملاء العربي: نحو نظام ذكي لتصحيح الأخطاء الإملائية وتعليمها

Abstract :

Arabic spelling errors remain a common challenge for students and language learners, particularly errors related to hamza usage and confusion between ta marbuta and ha. Most existing tools either rely on general-purpose language models without task-specific training or provide corrections without educational explanation.
This project presents Sawab, an intelligent educational system designed to detect and correct common Arabic spelling errors while promoting linguistic understanding. The system employs a fine-tuned Arabic language model based on CAMeLBERT, trained on a labeled dataset to classify input sentences into three categories: hamza errors, ta marbuta/ha errors, or no error.
Once the error type is identified, the system routes the sentence to the appropriate correction module and provides a concise educational explanation of the applied linguistic rule. This approach transforms spelling correction into an interactive learning experience rather than a simple automated fix.
Sawab demonstrates how fine-tuned Arabic language models can be integrated with rule-based educational components to support accurate writing and language learning. The system is designed to be scalable and adaptable for use in educational platforms, writing tools, and Arabic language learning applications.

Idea Fields :

The Sawab project lies at the intersection of Arabic Natural Language Processing, education technology, and intelligent language learning systems. It addresses challenges in Arabic spelling correction by combining machine learning with educational feedback. The project contributes to the development of AI-powered language tools that focus not only on correction accuracy but also on user understanding.
Relevant domains include Arabic NLP research, digital education platforms, writing assistance tools, and language learning applications. Sawab also supports the enhancement of Arabic digital content quality and encourages the adoption of AI-based solutions in linguistically complex environments. By focusing on commonly occurring spelling errors, the project offers a practical and targeted approach to improving Arabic writing skills through intelligent systems.

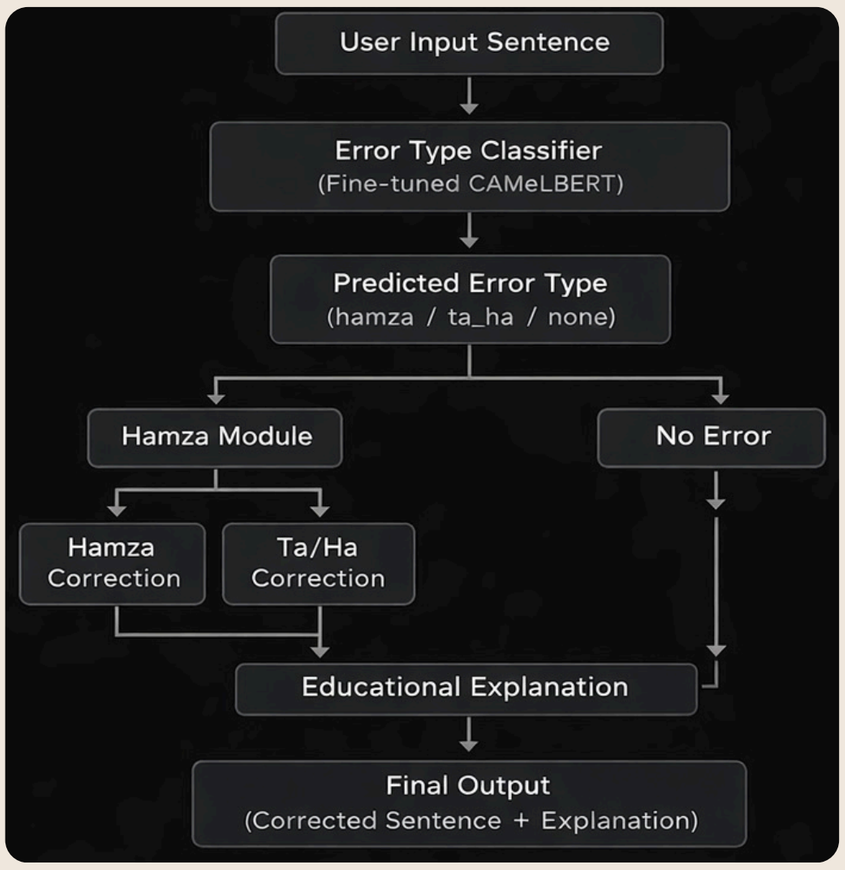
Result :

The fine-tuned CAMeLBERT model demonstrated strong performance in classifying spelling error types across the targeted categories. The system successfully detected and corrected common Arabic spelling errors while providing clear educational explanations. The results confirm the effectiveness of task-specific fine-tuning for Arabic spelling error detection and highlight the system's potential for real-world educational use.

Future Work :

Future work includes expanding the system to support additional Arabic spelling and grammatical error types, such as alif maqsura and diacritics. Increasing the size and diversity of the training dataset is expected to further improve model performance. Additionally, developing a web or mobile-based user interface and introducing personalized feedback based on user proficiency would enhance the system's usability and educational impact.

Methodology :



Team Members :
Leen Ahmed Alsanad