

This repo contains the data used to develop and test the Tunisian Arabic Automatic Speech Recognition model developed in the following paper :

Our models, allowing to transcribe audio samples in a linguistic mix involving Tunisian Arabic, English and French, and all the data used during training and testing are released for public ...

In this paper, we are concerned with the Tunisian dialect (TD) and propose to survey the availability of corpora for its automatic processing.

In this work, we incorporate Tunisian text data sourced from Tunisiya , a vast corpus of Tunisian Arabic that is openly accessible. We also scrapped code-switched data from various online sources ...

In this work, we incorporate Tunisian text data sourced from Tunisiya , a vast corpus of Tunisian Arabic that is openly accessible. We also scrapped code-switched data from vari-ous online sources ...

It is designed to facilitate the development of Large Language Models (LLMs), chatbots, and translation systems for the Tunisian dialect (Derja/Tounsi), which is a low-resource language ...

In this article, we present a method to create a parallel corpus to build an effective NMT model able to translate into MSA, Tunisian Dialect texts present in social networks. For this, we ...

The tunisian arabic dialect has been chosen as a typical example for an under-resourced Arabic dialect. We propose, in this paper, our first steps to build an automatic speech recognition system for ...

This dataset combines 7,366 positive/negative Tunisian Arabic and Tunisian Romanized alphabet Facebook comments. The same dataset was used to evaluate Tunisian code-switching sentiment ...

In response to the limited availability of paired Text-Speech Tunisian datasets with code-switching, we have built a corpus through meticulous manual annotation.



Tunisian Aggregation Switch DML

Web: <https://www.safireschools.co.za>

