TECHNICAL DOCUMENTATION

Title: Final Project – Technical Documentation

Author: Erick Banegas

Course: ITAI2373

Project: NewsBot Intelligence System 2.0

1. Introduction

This project expands the original midterm NewsBot into a complete, production-ready multilingual news analysis system. The goal was to integrate advanced NLP features such as translation, summarization, topic classification, semantic understanding, and a conversational interface into a functional web application.

The system was fully redesigned to follow a modular architecture, improving maintainability, scalability, and readability. Final deployment was completed using Streamlit and made publicly accessible through an Ngrok tunnel.

2. Project Objectives

- The main objectives of the final system were:

- Build a multilingual NLP pipeline capable of processing Spanish and English news articles.

- Implement automatic topic classification using a robust zero-shot model.

- Generate high-quality summaries using transformer-based models.

- Provide natural translation in both directions (ES ↔ EN).

- Integrate all features inside a conversational interface.

- Deploy the system as a web application for real-time use.

3. System Architecture

The system follows a modular architecture stored under the src/ directory:

3.1 Core Modules

multilingual/translator.py

Handles language detection and translation using MarianMT models.

language_models/summarizer.py

Performs summarization using the BART-large CNN model with multilingual support.

analysis/classifier.py

Uses XLM-Roberta-Large-XNLI for multilingual zero-shot classification.

conversation/query_processor.py

Routes user requests to the correct module (translation, summary, classification, etc.).

conversation/chatbot.py

Main interface used by Streamlit to handle user prompts.

4. NLP Models Used

4.1 Translation

Helsinki-NLP/opus-mt-es-en

Helsinki-NLP/opus-mt-en-es

These models provide reliable bilingual translation without needing API keys.

4.2 Summarization

facebook/bart-large-cnn

Produces coherent and concise summaries; supports multilingual input through translation.

4.3 Zero-Shot Topic Classification

joeddav/xlm-roberta-large-xnli

Allows classification of Spanish text into English labels with very high accuracy.

5. Web Application

The application was implemented using Streamlit, allowing interactive modules:

Language detection

Translation (ES → EN, EN → ES)

Topic classification

News summarization

Conversational Q&A

Deployment details:

Hosted inside Google Colab

Exposed externally using Ngrok

Final public URL example:

https://grouseless-dacia-nontaxonomic.ngrok-free.dev

Git Repository

https://github.com/erickxllx/Final-Project-NewsBot-Intelligence-System-2.0

6. How to Run the System

Option A – Local Execution

pip install -r requirements.txt

streamlit run app.py

Option B – Google Colab

Upload project folder

Install dependencies

Run:

streamlit run app.py --server.address=0.0.0.0 --server.port=8501

Create an Ngrok tunnel

7. Limitations

Ngrok links expire after each session.

Large transformer models require GPU for optimal performance.

Summarization relies on translation when text is not in English.

8. Conclusion

This project demonstrates a complete multilingual NLP system with real-time capabilities. The architecture, modular design, and model selection result in a professional, robust solution suitable for real applications and portfolio use.