

Project Report: Fake News Detection in Arabic

1. Project Overview

- The project aims to create a robust dataset for detecting fake news in Arabic. The primary goal is to distinguish articles as either fake or real and categorize them by topic.

2. Data Sources and Collection

- Arabic news articles are gathered from reputable fact-checking websites, such as Fatabyyano, to ensure a mix of fake and authentic stories. The articles are collected across various topics, including politics, religion, health, and more, to provide a comprehensive overview of fake news dissemination.

3. Data Annotation

- Each article is classified as fake or real and sorted by topic. This categorization is designed to offer a nuanced perspective on the spread of fake news across different subjects.

4. Scraping Methodology

- The data is scraped from fact-checking websites using BeautifulSoup, a popular Python library for web scraping. This method ensures an efficient and targeted collection of data from the specified sources.

5. Future Work

- While BeautifulSoup is the current scraping method, you have considered using Selenium as an alternative approach. This method could provide additional capabilities for scraping dynamic content and interacting with websites.
- You plan to take a break to update your methods and explore alternative scraping techniques later.

Conclusion

- This project represents a valuable effort to understand and combat the spread of fake news in the Arabic language. By creating a comprehensive dataset, it contributes to the broader goals of information integrity and accuracy. The potential adoption of Selenium and other scraping techniques in the future could further enhance the project's scope and effectiveness.