Faculty of Engineering & Technology

Department of Electrical & Computer Engineering

ENCS5342: "Information Retrieval, Web Search and NLP"

Major Assignment (Course Project):

"Arabic Spell Checker with NLP Integration"

Partner names:

Jalila Muadi 1201611

Obada Hattab 1171616

Marian Qasis 1200861

Instructor: Dr. Adnan H. Yahya

Birzeit

March 2024

Abstract

The goal of this project is to create an Arabic spell checker that uses NLP approaches to improve efficiency and accuracy. Taking linguistic trends and context into account, the spell checker for Arabic will recognize misspelled words and offer suggestions for repairs.

To make better correction recommendations, it will make use of NLP techniques like tokenization, stemming, and part-of-speech tagging. For performance study, the system will be assessed using common Arabic language corpora and contrasted with current spell checkers.

Explanation:

The intended purpose of this project is to create a spell checker designed especially for Arabic text. Creating a dataset containing Arabic text from news articles, social media posts, and other public ally accessible sources can be a good place to start.

Next, use fundamental spell-checking algorithms (the Levenshtein algorithm) such phonetic similarity or edit distance to find misspelled words. Subsequently, use natural language processing (NLP) methods such as tokenization to deconstruct text into words, stemming to extract words to their simplest forms, and part-of-speech tagging to comprehend the grammatical context.

Technology:

- Hp, windows11, core i7, Asus windows10 core i7, Hp, windows10, core i7
- Python