Faculty of Engineering and Technology

Electrical and Computer Engineering Department

IR With Applications of NLP– ENCS5342

Arabic Question Answering – Project Abstract

Prepared By:

Alaa Saleem – 1200001

Jamila Fawaqa – 1200435

Hala Gholeh – 1201418

Section#: 2

Instructors: Dr. Adnan Yahya

Date of Submission: 4.April.2024

BIRZEIT

April – 2024

Abstract

The goal of this project is to develop an Arabic Question Answering (QA) system using Natural Language Processing (NLP) techniques. The system will use a diverse dataset of Arabic web documents as its knowledge source (specific dataset to be determined from “Masader” website which is an online catalogue for Arabic NLP datasets). The system will be accessible through a user-friendly interface, allowing Arabic-speaking users to input questions via text and receive clear answers, thereby enhancing access to information.

The system will undergo data preprocessing (including tokenization, stemming, lemmatization, and removing stop words) to refine input text quality. Also, it will integrate question analysis to analyze and understand the structure of the question, information retrieval which retrieve a ranked list of documents based on their relevance to the question, and answer extraction from the texts using techniques like named entity recognition (NER) and rule-based approaches.

The evaluation will be done using standard measures including F1-score, recall, and precision. Additionally, the system will identify patterns and areas where it initially find challenges in certain queries (which may involve misinterpretation of user queries and extraction of incorrect information, or overlooking important details) and it will then adjust its processing methods, and refine its algorithms to better handle such queries in the future, using strategies such as contextual understanding, handling negation and affirmation, and user interaction learning, thereby improving the overall performance and accuracy.