**Arabic POS with NetworkX in Articles**

1. **Introduction**

**1.1 Problem Statement**

The goal of this project is Part-of-Speech (POS) tagging for Arabic text and visualizes the relationships between words and their respective POS tags using a graph-based representation.

**1.2 Goals and Objectives**

The main objectives of this project are:

* To process Arabic text and identify each word’s POS tag.
* To use NetworkX to visualize the POS tags and relationships between words.
* To provide an intuitive visual representation for the POS tagging results.

**2. Data Description**

**2.1 Dataset Overview**

This project involves analyzing Arabic text data. The dataset consists of textual samples with different grammatical structures to demonstrate POS tagging.

**2.2 Features and Labels**

* **Features:** Words from Arabic sentences.
* **Labels:** POS tags such as nouns, verbs, adjectives, etc.

**2.3 Training and Testing Split**

| **Data Splits** | **Train** | **Validation** | **Test** |
| --- | --- | --- | --- |
| **Number of examples** | 6075 | 909 | ~~1821~~ |

**3. Baseline Experiments**

**3.1 Objective**

To establish a baseline by evaluating the functionality of the POS tagging model and its ability to correctly tag words in Arabic sentences.

**3.2 Baseline Model(s) Used**

I tried Vanilla RNN and LSTM. In the model architecture I use new layer I didn’t use it before which is **TimeDistributed**

**3.3 Evaluation Metrics**

I used accuracy.

**3.4 Results and Conclusion**

The POS tagging model successfully tags Arabic text and presents the relationships visually. But the best testing accuracy is 70 and I know isn’t a good accuracy and as result, The model misclassified some of POS.

**4. Other Experiments**

**4.1 Steps/Methodology**

* Arabic text is input into the POS tagging model.
* The relationships between words and their POS tags are visualized using NetworkX, where nodes represent words and their tags, and directed edges indicate the association between the two.

**4.2 Results**

The output is a directed graph showing words and their corresponding POS tags. For example:

* Words are represented by one set of nodes.
* Their POS tags are represented by another set of nodes.
* The edges represent the relationships between each word and its POS tag.

**4.3 Conclusion**

The graph-based representation provides a clear and intuitive way to understand POS tagging results, enhancing the analysis of grammatical structures in Arabic sentences.

**5. Overall Conclusion**

**5.1 Summary of Findings**

The project successfully demonstrates the use of POS tagging for Arabic text, coupled with graph-based visualization using NetworkX.

**5.2 Comparison with Benchmarked Results**

The

**6. Tools and Resources**

**6.1 List of Tools**

* **Python**
* **NetworkX** (for graph-based visualizations)
* **Matplotlib** (for generating plots)

**6.2 External Resources Used**

* Arabic POS tagging model (unspecified but potentially derived from a standard NLP library).

**7. Reflection Questions**

**7.1 Biggest Challenge**

* Searching about the best dataset.
* Data preparation wasn’t easy as usual (each sentence has a label) because each sentence has a group of labels. So, data padding wasn’t as I know. So used GPT to help me in this step. I know that the step is easy, but I didn’t face before.

**7.2 Key Learnings**

Graph-based visualization can significantly enhance the understanding of linguistic features in a text. NetworkX proves to be a useful tool for such tasks.