

Assignment for the Session 3 : Language-Specific & Advanced Text Processing Techniques

Objective:

Create a personalized Jupyter Notebook implementing the concepts covered in the session. Use the techniques, tools, and libraries discussed to design an end-to-end text preprocessing pipeline or focus on a specific aspect of text processing. This is your opportunity to express your understanding and creativity.

Assignment Details:

1. Choose Your Focus:

Decide which part of the session you'd like to implement. Examples include:

- Arabic-specific text processing (e.g., diacritization, morphological analysis, dialect handling).
- English-specific text processing (e.g., stemming, lemmatization, handling abbreviations).
- Advanced text handling (e.g., multilingual processing, sentiment analysis preprocessing).

2. Build the Notebook:

- Create a **Jupyter Notebook** using Python.
- Implement the techniques discussed using suitable libraries (e.g., NLTK, SpaCy, PyArabic, Hugging Face).
- Include at least **one language-specific technique** and **one advanced processing technique**.

3. Explain Your Code:

- Add markdown cells to explain what each part of your code does.
- Discuss why you chose certain techniques and how they align with the concepts covered in the session.

4. Evaluate Your Work:

- Test your notebook with sample text data in the language(s) you are working with.
- Reflect on the challenges you faced and how you addressed them.

Evaluation Criteria:

1. Concept Understanding (40%):

Does the notebook demonstrate a good grasp of the concepts covered in the session?

2. Implementation (30%):

How well are the techniques implemented? Are the results accurate and meaningful?

3. Creativity (20%):

Does the notebook reflect personal initiative or innovative approaches?

4. Documentation (10%):

Are markdown explanations clear and insightful?