

Assignment N° 2- part 1-

Exercise 1 :

Having the following text:

"AI in 2024 is transforming industries! Isn't it amazing? @TechUpdates
#ArtificialIntelligence 😊"

1. Convert the text to lowercase.
2. Remove punctuation, numbers, and special characters.
3. Tokenize the text.
4. Remove stopwords.
5. Apply lemmatization.
6. Generate the final cleaned list of words.

In the next exercises, you'll need to bring your computers

Exercise 2 :

In these 2 sentences:

- "I can fish."
 - "Can you open the can?"
1. Assign POS tags manually for both sentences.
 2. Use a POS tagging tool to tag the sentences. What do you notice?

Exercise 3 :

"Tesla, led by CEO Elon Musk, reported a 20% increase in quarterly revenue. The company plans to expand its operations in Germany and China."

- 1- Extract and classify entities using spaCy or NLTK.
- 2- List all entities along with their types

Assignment N° 2- part 1-

Exercise 4 :

- 1- Classify different types of coreferences:

"Bill Gates founded Microsoft in 1975. The company became one of the most successful tech firms. Gates stepped down as CEO in 2000."

- 2- Use a coreference resolution tool to confirm your classifications.

Exercise 5:

Generate unigrams, bigrams, and trigrams from the given text.

"NLP focuses on making human language understandable to computers."