

Vorschau

Conversion and Annotation

Frage 1 - Question 1 (8 Punkte)

Why is it necessary to convert multilingual text passages?

- To annotate them on distinct linguistic levels.
- To discern all non-English words.
- To prepare their computer-based analysis.
- To format them.

Frage 2 - Question 2 (4 Punkte)

Python is

- an interface that serves to create computational documents.
- an interface that serves to visualise data.
- a programming language
- a template.

Frage 3 - Question 3 (12 Punkte)

Glossary: Match the terms and definitions.

- | | |
|--|-----------------|
| each token is assigned a number | NLP |
| labels the hierarchical ways in which words relate to one another | Pipeline |
| a combination of codes consisting of different taggers which process the text fed to the machine | Tokenization |
| labels indicating what kind of function a word fulfils in a sentence | Token indexing |
| the splitting of a sentence into its parts, called tokens, i.e., words and signs | POS tags |
| a way of researching natural language through computational methods | Dependency tags |

Frage 4 - Question 4 (6 Punkte)

Arrange in the correct order.

- tokenization
- POS
- token indexing
- dependencies

Frage 5 - Question 5 (8 Punkte)

Match the dependency tags which what they stand for.

- | | |
|---------------------------|-------|
| main verb | Dobj |
| subject | Pcomp |
| complement of preposition | ROOT |
| major part of compound | |

Frage 6 - Example Sentence Question 1 (2 Punkte)

What is the third token of the example sentence?

- "ya"
- "Haki"
- "!"
- "Ayaaaana"

Frage 7 - Example Sentence Question 2 (2 Punkte)

What is the POS tag of "Ayaaaana"?

- VERB
- PROP
- ADP
- NOUN

Frage 8 - Example Sentence Question 3 (2 Punkte)

Which token corresponds to index 8?

- "aieee"
- "!"
- "The"
- "Mungu"

Frage 9 - Example Sentence Question 4 (2 Punkte)

What is the dependency tag of "Ayaaaana"?

- ROOT
- prep
- nsubj
- det

Frage 10 - Example Sentence Question 5 (2 Punkte)

What is the root token/head index of "Mungu"?

- 5
- 1
- 14
- 13