Example Questions for Exam

Natural Language Processing Technology, Master of AI: L_AAMAALG005

It will be a multiple-choice exam with 25-35 questions.
Only a single answer is correct.

Date: Monday, 30.5., 12.15-14.30
Location: Emergo Blok 1-3 (Amstelveen)
We will follow the standard regulations for exams by the VU:
https://vu.nl/en/student/examinations/exams

You are allowed to use a basic calculator and we will provide scratch paper, if needed.

You will see different types of questions:

- Linguistic analysis questions
- Calculation questions
- Terminology/Functionality questions
- Interpretation/Reflection questions

We provide a few examples.

Linguistic analysis questions:
You will see one or two example sentences, and we will ask about the linguistic characteristics.

Example sentence:
the old train the young .

1) What is the type-token ratio of this sentence?
   A: 1.2
   B: 0.83
   C: 0.8
   D: 0.67

2) How many content words does the sentence contain?
   A: 5
   B: 4
   C: 3
   D: 2

3) What is the correct POS-tag for *train* in the sentence?
   A: ADV
   B: VERB
   C: NOUN
   D: AUX

Calculation Questions:
You see a confusion matrix and will be asked for the correct value for all kinds of metrics (accuracy, precision, recall, F1) for a specific class or for different averages over all classes or for some baselines.

Example: System for Named Entity Recognition

Predicted	B-LOC	B-MISC	B-ORG	B-PER	I-LOC	I-MISC	I-ORG	I-PER	O
Gold									
B-LOC	1348	18	134	70	1	1	7	46	43
B-MISC	28	458	54	47	0	1	8	24	82
B-ORG	184	43	1038	175	1	1	26	57	136
B-PER	68	20	45	1258	0	0	1	141	84
I-LOC	5	0	0	0	155	4	40	30	23
I-MISC	1	8	0	3	3	128	12	23	38
I-ORG	21	8	5	11	45	8	521	111	105
I-PER	4	1	0	13	1	0	15	1093	29
O	29	41	70	101	3	37	36	105	38132

Interpretation Questions:

4) Which statement about the performance of the system is correct?
   A: The system confuses Person and Organization more often than the other named entities (except O).
   B: The system is not better than the majority class.
   C: The system performs worst for the recognition of person names.
   D: For the I-MISC label, the performance is worse than random.

Terminology Questions:

5) Which answer is a prototypical example of variation?
   A: Slang
   B: Semantic scope
   C: Syntactic ambiguity
   D: Stopwords

The Bert tokenizer yields the following tokenization:
['[CLS]', 'i', 'love', 'deadline', '##s', ';', 'i', 'love', 'the', 'who', '##oshi', '##ng', 'noise', 'they', 'make', 'as', 'they', 'go', 'by', '.', '[SEP]']

6) Which symbol indicates a subword split?
   A: [CLS]
   B: [SEP]
   C: ##
   D: ;

Functionality Questions:

6) What are the two main steps in byte-pair encoding?
   A: count and merge
   B: count and flip
   C: split and merge
   D: split and flip

7) Which metric takes the class distribution into account?
   A: Macro-average F1
   B: Weighted- average F1
   C: Accuracy
   D: Inter-annotator agreement

8) Which test type should we use to test if a model can handle slight changes to the input and still give the same prediction?
   A: Directional test
   B: Invariance test
   C: Proportional test
   D: Minimum functionality test

Reflection questions:

9) What is an expected effect of increasing the size of the vocabulary for contextualized language models?

A: A larger vocabulary leads to fewer subword splits in tokenization.
B: A larger vocabulary leads to more subwords splits in tokenization.
C: A larger vocabulary increases the morphological complexity of a language.
D: A larger vocabulary decreases the morphological complexity of a language.

10) Why is the inter-annotator agreement for hate speech detection usually low?

   A: It is difficult to find enough hate speech data.
   B: Hate speech is a subjective concept.
   C: Annotators generate counter-narratives.
   D: The measure for inter-annotator agreement is biased.