

# Natural Language Processing 2020/2021

## Test Exam Questions

– Do not turn this page before the official start of the exam! –

First name, Surname: \_\_\_\_\_

Student ID: \_\_\_\_\_

**Program:** Bachelor Data Science and Artificial Intelligence

**Course code:** KEN2570

**Examiner:** Dr. Jan Niehues

**Date/time:** Thursday, 03-06-2021, morning

**Format:** Open book exam (includes slides, notes, assignments etc. printed or handwritten)

**Allowed aides:** Pens, simple (non-programmable) calculator from the DKE-list of allowed calculators.

### Instructions to students:

- The exam consists of 5 questions on 6 pages (including cover page).
- Fill in your name and student ID number on each page, including the cover page.
- Answer every question at the reserved space below the questions. If you run out of space, continue on the back side, and if needed, use the extra blank page.
- Ensure that you properly motivate your answers.
- Do not use red pens, and write in a readable way. Answers that cannot be read easily cannot be graded and may therefore lower your grade.
- You are not allowed to have a communication device within your reach, nor to wear or use a watch.
- You have to return all pages of the exam. You are not allowed to take any sheets, even blank, home.
- If you think a question is ambiguous, or even erroneous, and you cannot ask during the exam to clarify this, explain this in detail in the space reserved for the answer to the question.
- If you have not registered for the exam, your answers will not be graded, and thus handled as invalid.
- **Success!**

The following table will be filled by the examiner:

|                  |    |    |   |    |    |       |
|------------------|----|----|---|----|----|-------|
| Question:        | 1  | 2  | 3 | 4  | 5  | Total |
| Maximum points:  | 10 | 15 | 5 | 20 | 10 | 60    |
| Achieved points: |    |    |   |    |    |       |

**Question 1. (10 Points) Semantics**

## 1. True/False

|                                                                                 | True | False |
|---------------------------------------------------------------------------------|------|-------|
| Laptop is a Hypernym of Computer                                                |      |       |
| Wood ( area with trees vs. material to build a table) is an example of polysemy |      |       |
| Mean (average vs. not nice) is an example of polysemy                           |      |       |

## 2. Skip-gram Model

Given the example sentence “I live in Maastricht” and a skip-gram model with a context of the two previous words and the two next words.

a) What are the training instances for the model? Give the input and output.

| Input | Output |
|-------|--------|
|       |        |
|       |        |
|       |        |
|       |        |
|       |        |
|       |        |
|       |        |
|       |        |
|       |        |
|       |        |
|       |        |
|       |        |
|       |        |

b) If this were all the training data training data, would you benefit from using negative sampling with 10 samples. Explain why or why not?

### 3. Word sense disambiguation

You need to build a word sense disambiguation system for a lexical sample task. Given the sentence “The account at this bank has the highest interest” and the ambiguous words bank(finance institute /river) and interest(interesting topics / financial gain).

- a) How many classifiers will you train? How many output classes does each classifier have?

- b) A friend suggests to use the following classifier: Retrieve the pretrained word embedding for the ambiguous word and then use a logistic regression to predict the logistic regression. What is your opinion to this approach?

- c) How can you improve this model by still using word embeddings and a logistic regression model?

Student name:

Student ID:

Page 4 of 5

NLP

2020/2021

---

**Extra answer sheet.**

Student name:

Student ID:

Page 5 of 5

NLP

2020/2021

---

**Extra answer sheet.**