Computational Communication Science 2

Week 1

Lab session

Roeland Dubèl
r.dubel@uva.nl

2025-04-01

Digital Society Minor
University of Amsterdam

Today

1. Introduction
2. Course information
3. Weekly MC-questions
4. Break
5. Weekly exercises

Introduction

Who am I?

- Roeland Dubèl
- PhD candidate
- *Trust in Journalism*
- CS Bachelor and Research Master

Who are you?

Write your name on a piece of paper.

- What is your name?
- What do you study?
- What do you expect to learn in this course?

Course information

Expectations

- Lab sessions are meant for practising with code and asking questions
- You are allowed to miss one tutorial during this course
- Questions about code? Ask them during the tutorial
- $2 \times$ late $= 1$ absence

Graded assignments: Weekly MC-questions (individual)

- Regular multiple choice questions (20%) about the readings and techniques that we discuss
- 4 questions (week 1, 2, 3, 7, and 8)
- Questions about the content of that week
- Questions about readings and codes (but you do not have to code yourself)
- Total of 20 questions, 16 correct answers = full marks

Graded assignments: Report (group)

- In groups of 3 to 4 students write a research report using one of these datasets (see Canvas):
  - News dataset
  - Books dataset
  - Recipe dataset
  - TV Series dataset (large)
- Preprocessing and exploring the dataset + building a recommender system (see Course Manual)
- 30% of your grade
- Deadline = Monday, 6 May, 11:00 AM

Graded assignments: Presentation (group)

- Presenting the research report in class
- Week 4
- 10% of your grade

Graded assignments: Exam (individual)

- Final exam (50%) at the end of the course to show off what you learned
- Individual open book exam in week 9 (no internet + in class!)
  - Tuesday, 27 May
- Your understanding of concepts and computational coding will be tested

Plagiarism

- Attributing code: The following function is copied from https://stackoverflow.com/XXXXX/XXXXX
  - Relevant for the research report
- The use of AI tools such as ChatGPT or Copilot to generate code will be considered fraud
  - Remember the final exam will be in class

Weekly MC-questions

MC-questions Week 1

- Canvas $\rightarrow$ Modules $\rightarrow$ Week 1 $\rightarrow$ MC-questions
- 4 questions, 8 minutes (in silence)
- Afterwards we will discuss the questions

15 minute break

Weekly exercises

Recap preprocessing text data

1. Lowercasing → `.lower()`
2. Removing whitespace → `.strip()`
3. Finding substring → `.find()`
4. Replacing substring → `.replace()`
5. Counting substring → `.count()`
6. Tokenization → `.split()` or `Treebank tokenizer`
7. Stop-word / punctuation removal → `stopwords.words(‘‘english’’)` or `RegexpTokenizer`

Weekly exercises: Week 1

- Form groups of 3
- Go through the weekly exercises (GitHub → week01 → exercise-tutorial → week1-exercises.md)
    - Find the data on Canvas → Modules → Week 1 → data for tutorial exercise
- Ask me for questions!
- The answers are in the same GitHub folder

Next week

- Form groups for the research report (groups of 3 or 4) and choose a dataset
  - Sign up for a dataset (full = full) via Google Docs
- Any questions left?
  - Pose questions for next week via Google Docs
- Google Docs:

  https://tinyurl.com/2994fb3f
- Try the exercises on a dataset
- See you next week!