20-00-0546-iv Foundations of Language Technology

Handout 1
Getting Started

5. November 2020

The exercises are published every Thursday after the lecture. Try to solve the tasks by yourself. In addition we will upload a video on Thursday explaining solutions for the tasks. Watch this video once your're done, or when you're stuck. On Mondays (except for public holidays) there is a live questions and answers session, where we will discuss the previous homework and answer questions regarding the exercise and the homework exercise.

Each handout also contains homework. Working on these homework tasks and handing them in is **not mandatory**, but you can benefit from that in two ways: You get feedback on your solutions and a better understanding of the subject, and, more important, you have the chance to improve your final grade for the exam. If you hand in your solutions, we will grade them based on their quality. For each homework task, there are points to collect depending on the amount of work you have to put into it. You will find the maximum number of points to gain for a task in its header. There are 10 homework exercise, that can earn you a maximum of 90 points.

In the second half of the semester we offer you to work on a software project as an additional final exercise. Software project details will be announced later. It brings a maximum of 50 points.

Collecting more than 100 of all points will give a 0.3 bonus on the final exam. Note that, due to TU Darmstadt regulations, the exam has to be passed only by counting exam points. Thus, bonus points could be applied exclusively to exams graded at 4.0 or better.

Finally, if you want to hand in your solutions, please use the Moodle system. Please send a Jupyter Notebook for programming parts and plain text or PDF for essay questions. The deadline for the homework is **Thursday, November 12, 18:00 CET**. In case your submission consists of several files, compress these to a `zip`-file. Indicate clearly which submission corresponds to which question. Include comments in your program code to make it easier readable.

Later submissions, submissions in a wrong format or submissions where it is not clear which question is answered will **not be graded**. You can hand in the submission either alone or you can work together with one partner. In case of a group submission, select the other person as a co-author and include both member names in all your documents (PDF as well as Python code).

1.1 Python and NLTK – first contact

Start your Python GUI. If you are using Google Colab open the practice 01 notebook. Then you don't need to install NLTK or download additional data.

Task 1.1 *Enter the code shown in the lecture slides (slide 49 ff.) and consider which results you expect.*

Task 1.2 *As you might have noticed in task 1.1 Python makes a difference between integers and floating point numbers. Explore this a little bit further:*

```python
1//3
1/3
x=1
print("x=", x)
print("x=%f" % x)
print("x = {:f}".format(x))
print("x with precision 2: %.2f" % x)
print("x with precision 2: {:.2f}".format(x))
```

Please keep in mind that there the result depends on the variable type.

Task 1.3 *Strings may be stated in different ways. See the code below.*

```python
hi = 'hi'
greeting = "hi"
```

How could you assign the name "O'Reilly" to a variable? What is the purpose of having different ways to represent a string? Are there more possibilities to assign a string to a variable (search the web)?

1.2   Homework

Today's homework is rather an opportunity for you, to make yourself comfortable with the topics of the lecture. The time is reserved to catch up with different prior knowledge.

Homework 1.1 *(0 Points)*
Find a Python tutorial matching your background, either in the link list provided in the lecture slides or search the web by yourself.

- *If you have experience in another programming language: Work out the differences. Set up a memo, what to keep in mind when implementing in Python.*
- *If you do not have any prior programming experience: Work through the first chapters of any tutorial or book on Python. Work with variables, strings, integers and floats.*

Homework 1.2 *(3 Points)*
Research the web: What are the most challenging tasks in NLP? Check the web pages of global players like IBM, Google, Microsoft, Yahoo and some universities.

Homework 1.3 *(2 Points)*
Make yourself familiar with the Moodle system, in particular how to upload your homework. Upload your solution to Homework 1.2 to Moodle for grading no later than Thursday, November 12, 18:00.