

Hauptseminar
Winter Semester 2015/2016

Natural Language Processing with Python: A hands-on
introduction using NLTK

Abstract:

This course provides a hands-on introduction to programming in Python using NLTK. The Natural Language Toolkit NLTK is an open source platform offering transparent access to a broad range of algorithms and resources for computational linguistics.

Instructors:

- Detmar Meurers
 - *Office:* Room 1.28, Blochbau (Wilhelmstr. 19)
 - *Email:* dm@sfs.uni-tuebingen.de
 - *Office hours:* Wednesdays, 10-11 (please arrange a slot by email beforehand)
- Tutor: Aria Omidvar
 - *Email:* omidvar.aria@gmail.com
 - *Office:* Room 1.25, Blochbau (Wilhelmstr. 19)
 - *Office hours:* please arrange by email

Course meets:

- Wednesdays, 8:30–10:00 in 1.13 (SfS, Blochbau, Wilhelmstr. 19)
- Fridays, 8:30-10:00 in 1.13 (SfS, Blochbau, Wilhelmstr. 19)
 - Note: Following the standard rules, missing more than two meetings unexcused, automatically results in failing the class. If you have to miss class for a valid reason, let the instructors know by email **before** class.

Language:

- Course language is English; term papers can also be written in German or French.

Moodle page: <https://moodle02.zdv.uni-tuebingen.de/course/view.php?id=1289>

- Please register for this course in Moodle – all course-related information will be sent through the Moodle course list.

Syllabus (this file):

- html-Version (<http://purl.org/dm/15/ws/hs>)
- pdf-Version (<http://purl.org/dm/15/ws/hs/syllabus.pdf>)

Nature of course and our expectations: This Hauptseminar intends to provide an overview of the concepts and issues involved in research in this domain. Participants are expected to

1. regularly and actively participate in class and read/prepare the material assigned by any of the presenters. (20% of grade)
2. prepare and present a topic (30% of grade)
3. write and submit a term paper in Moodle (50% of grade)
 - approx. 3 weeks of full time work after semester
 - For CL students, term papers have to be written in LaTeX using the CL journal template (<http://cljournal.org/style.html>) with the natbib citation style.

Credits: After successful completion of the course, a Hauptseminar Schein in Core Computational Linguistics is issued, with the following credit points options:

- 6 CP (with presentation), or
- 9 CP (with presentation and term paper)

Academic conduct and misconduct: Research is driven by discussion and free exchange of ideas, motivations, and perspectives. So you are encouraged to work in groups, discuss, and exchange ideas. At the same time, the foundation of the free exchange of ideas is that everyone is open about where they obtained which information. Concretely, this means you are expected to always make explicit when you've worked on something as a team – and keep in mind that being part of a team always means sharing the work.

For text you write, you always have to provide explicit references for any ideas or passages you reuse from somewhere else. Note that this includes text “found” on the web, where you should cite the URL of the web site in case no more official publication is available.

Topics:

We will generally follow the NLTK book (<http://www.nltk.org/book>) with materials added by the presenters wherever useful.

1. Introduction. Language Processing and Python
 - Oct 23 & 28 (Detmar Meurers)
2. Accessing Text Corpora and Lexical Resources
 - Oct 30 (Andreia Rauber, Björn Rudzewitz)
 - Nov 11 (Daniela Stier, Richard Belk)
3. Nov 4, 6: QITL (<http://www2.sfs.uni-tuebingen.de/qitl>)
4. Processing Raw Text (Roshanak Hamidi, Mei-Shin Wu, Julia Koch)
 - Nov 13 & 18
5. Writing Structured Programs (Andreas Daul, Eduard Schaf, Haywood Shannon, Martina Stama-Kirr)
 - Nov 20 & 25 & 27
6. Categorizing and Tagging Words (Natalie Clarius, Kevin Mann, Yevgen Karpenko)
 - Dec 2 & 4
7. Learning to Classify Text (Aria Omidvar, Christian Adam, Niklas Schulze, Melika Azimi)
 - Dec 9 & 16
 - Dec 11 → Workshop in Tübingen “Studying language Learning - From the laboratory to the classroom”
 - Dec 18
8. Extracting Information from Text (Alina Ladygina, Kathrin Adlung, Anastasia Gorbunova, Kanghyun Yu)
 - Jan 8 & Jan 13
9. Analyzing Sentence Structure (Asia Deinekina, Lisa Verena Hiller, Luis Ibargüen, Samuel Solzin)
 - Jan 15 & Jan 20
10. Building Feature Based Grammars (Ben Campbell, Valentin Pickard, Zarah Weiß)
 - Jan 22
11. Analyzing the Meaning of Sentences (Mihael Simonic, Alina Allakhverdieva, Olga Sozinova, Eyal Schejter)

- Jan 27, 29, Feb 3
12. Managing Linguistic Data (Vivian Fresen, Sabrina Galasso, Holger Muth-Hellebrandt, David Bausch)
- Feb 5, 10
13. *Discussion of project ideas for term papers*
- Feb 12

Note: The syllabus is subject to change, as we progress through the semester. So check the online version regularly.

Last update: September 20, 2016