

The ABC of Computational Text Analysis

#10 NLP with Python

Alex Flückiger
06 May 2021

Faculty of Humanities and Social Sciences
University of Lucerne

Recap last Lecture

introduce Python 

- JupyterLab editor
- syntax
- data types

Outline

- organizational
evaluation, mini-project, assignment #3
- Let's getting to serious NLP! ✨
- interactive coding
interrupt, ask, complement

Organizational



Course Evaluation

Tell me...

Please follow the link in the email

- by the University of Lucerne, Faculty of Humanities and Social Sciences
- received on 3 May 2020 (or similar)

Thanks for any constructive feedback,
be it sweet or sour!

Assignment #3

- get/submit via OLAT
starting tomorrow
deadline 14 May 2021, 23:59
- use forum on [OLAT](#)
subscribe to get notifications
- ask friends for support, not solutions

Requirements Mini-Project

present project on 27 May 2021

- executable script
- multiple documents
- compare ...
historically
across actors
- quantitative measures + interpretation

! share your project idea [here](#) by 14 May 2021

Optional Seminar Paper

- writing a seminar paper (6 ECTS)
- get in touch to discuss your idea

A brick wall covered in white tiles with letters, forming words like 'POSSIBLE', 'ELEGANT', and 'KABAREE'. The tiles are arranged in a grid pattern, and the letters are in a simple, sans-serif font. The background is a textured brick wall with some red and brown tones.

A Primer on old school NLP

What is a Word?

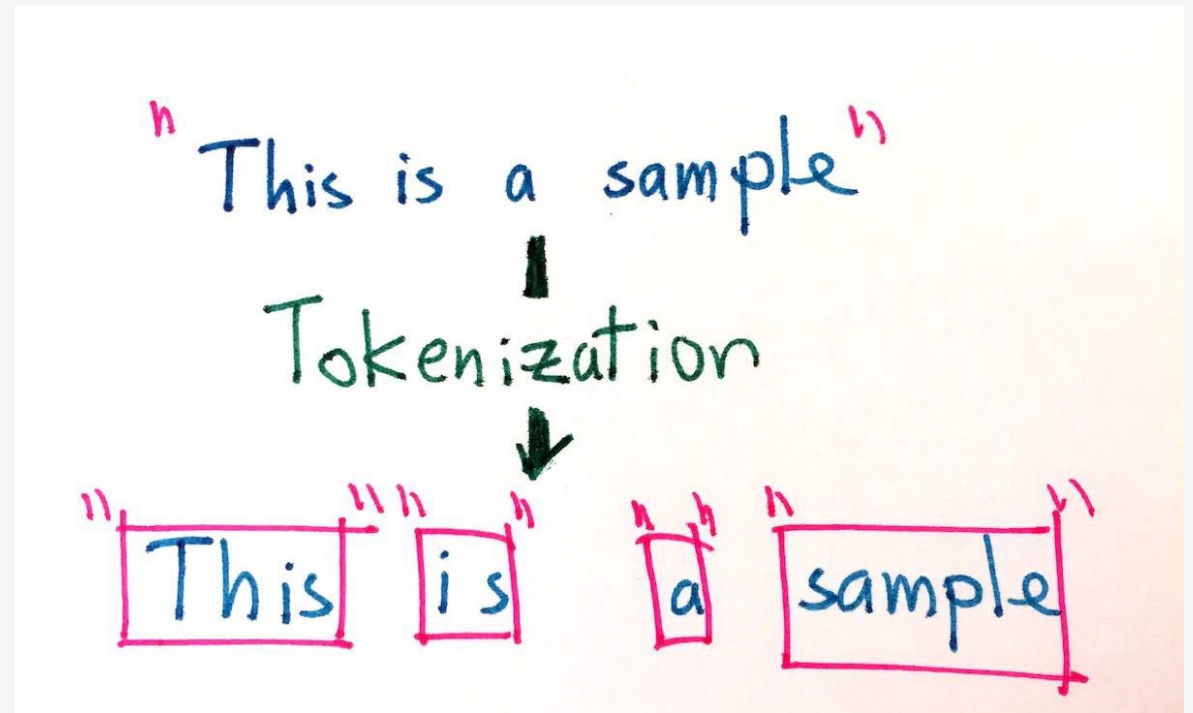
- word ~ segments between whitespace
- yet, there are ...

contractions: *U.S., don't*

collocations: *New York*

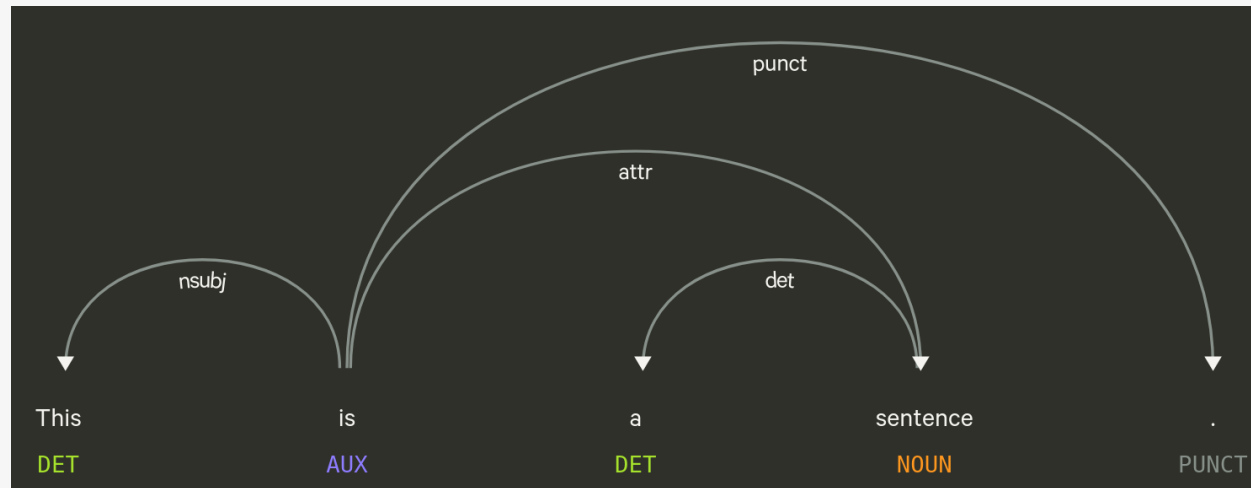
Token

- token ~ computational unit
representation of words
- lemma ~ base form of a word
texts → *text*
goes → *go*
- stop words ~ functional words
lacking deeper meaning
the, a, on, and ...



tokenization (*Medium*)


NLP Processing Steps



Outlook: NLP is on Fire 🔥

- supervised machine learning
- you can do basically anything with modern NLP
train on human-annotated data
- effort, insights and quality may differ
for better or worse

Deep Dive into NLP for Social Science

- check code on GitHub
- run code on Binder 

Resources

tutorials on spaCy

- [official spaCy 101](#)
- [official online course spaCy](#)
- [Hitchhiker's Guide to NLP in spaCy](#)



Questions?

References

Jurafsky, Dan, and James H. Martin. forthcoming. *Speech and Language Processing*. 3rd (Draft of December 30, 2020). London: Prentice Hall. <https://web.stanford.edu/~jurafsky/slp3/>.