Python for NLP

August 26-30, 2019
LORIA, Nancy

https://synalp.loria.fr/python4nlp

Organisation and Funding

- LIFT (C. Gardent), CNRS GDR
- OLKi Impact Project (C. Cerisara), LUE IDEX

Audience

- Basic Python Required
- Humanity Students (linguists etc.) and researchers
- CS students and researchers
- Industrials

Objective

Learn to

- Retrieve and store textual data from web, api (e.g., Gutenberg books, web pages, social network data)
- Apply linguistic processing (POS tagging, Parsing, NER, etc)
- Compute basic statistics and their visualisation (Nb of sentences, of tokens etc.)
- Apply basic Machine Learning Techniques (Classification, Clustering, Regression)
- Use word embeddings

Program

	9h-10h	10h30-12h	14h-15h	15h30-17h	
Monday, 26 August	Introduction Christophe Cerisara Claire Gardent	Collecting data Yannick Parmentier	Lab Session	Lab Session	Welcome Reception
Tuesday, 27 August		Linguistic Processing Yannick Parmentier	Lab Session	Lab Session	Social event Nancy Art Nouveau Guided Tour and Diner
Wednesday, 28 August	Invited talk Albert Gatt	Analysing textual data Claire Gardent	Lab Session	Lab Session	
Thursday, 29 August	Invited talk Malvina Nissim	Classification and Clustering Claire Gardent	Lab Session	Lab Session	
Friday, 30 August		Using word embeddings Christophe Cerisara	Lab Session	Lab Session	

Collecting Text

- Interaction Web Server / Browser
- What's in a web page
- Processing web pages
- What's an API
- Extracting text from Wikipedia and Social Networks

Processing Text

- Sentence segmentation and tokenization
- Morphological analysis, stemming
- POS tagging
- Named Entity Recognition
- Parsing

Analysing Text

- Descriptive statistics
- Univariate Analysis (distribution, dispersion)
- Bivariate Analysis (Contingency, covariance)
- Vizualisation (scatter plot, box plots, histograms, bar plots)

Classification and Clustering

- What is Machine Learning?
- Extracting Features
- Train/Dev/Test Data
- Supervised and unsupervised learning (Classification, regression, clustering)

Word Embeddings

- What are word embeddings ?
- Downloading and Using word embeddings

Registration

UL Students	0 Euros
Students (<500km)	300 Euros
Students (>500km)	100 euros
Academics	400 euros
Private Sector	800 euros