

Training: Compendium CE  
 Natural Language Processing in a Nutshell


FORM OF TRAINING	MATERIALS	PRICE	DURATION
Traditional	Hardcopy	400 EUR	2 days
Traditional	Digital materials	400 EUR	2 days
Traditional	CTAB Tablet	550 EUR	2 days
Distance learning	Hardcopy	400 EUR	2 days
Distance learning	Digital materials	400 EUR	2 days
Distance learning	CTAB Tablet	400 EUR	2 days

## LOCATIONS

Krakow - 5 Tatarska Street, II floor, hours: 9:00 am - 4:00 pm

Warsaw - 17 Bielska Street, hours: 9:00 am - 4:00 pm

## TRAINING GOALS:

During NLP in a Nutshell workshop we explain how to deal with most topics related to NLP. In the first part we explain some basic methods to handle strings and documents. Different representations are presented and explained. Spacy Data Model is explained, and we deeply use Spacy for tokenization and intent recognition. Part of the workshop is to build a NLP pipeline. We use some libraries for sentiment analysis and build our own solution. We show how to deal with German and French languages using some commonly known tools.

### Outcomes

Participants will understand...

- main concepts of Natural Language Processing,
- benefits of using Python and Scala for NLP application,
- how to use Natural Language Understanding.

Participants will be able to...

- parse and tokenize documents using commonly used tools in Python,
- use decision trees to do a sentiment analysis in Python,
- build a NLP pipeline, use Scala and Spark for analysis,
- build a spam classifier using NLTK,
- create a custom annotator for a fintech,

- how to use other language than English in NLP

## CONSPECT:

- String and Document Processing<
  - Advanced regular expression used for text extraction and generation in Python.
  - Understanding Spacy workflow
  - Tokenize using Spacy and NLTK
  - Build a tokenizer with Spark and Scala
- Text and Vector Representation
  - Understanding One-hot and Term-frequency Representation
  - Word Vectors
- Text Classification<
  - Intent recognition using Spacy
  - Using scikit for text classification
- Annotators
  - Named Entities Annotators
  - Build custom annotators
- Sentiment Analysis
  - Stanford sentiment analysis tool used for social media texts
  - Build a sentiment analysis method with scikit
- Pipelines
  - Build a NLP pipeline with Kafka, Spark and Spacy
- Languages
  - Grammar models and dependencies. Tagging with Spacy
  - Other language than English.
- Summary and final notes
  - Comparing of available tools and usage

## REQUIREMENTS:

Basic Python and Scala knowledge,

## Difficulty level



## CERTIFICATE:

The participants will obtain certificates signed by Compendium CE.

## TRAINER:

Certified Compendium CE Trainer.