

## Our NLP API contains

Lots of small "building blocks" for the bigger modules that can be combined to extract tons of useful information:

- Lemmatizer - a module that is capable of extracting lemmas from word forms
- Stemmer - a module for words' stems extraction
- POS Tagger - Part of speech tagger
- CASE Tagger - declension tagger (Slavic languages are usually highly inflectional and there is a lot of essential information inside each word in the text)
- POS Tagger - Part of speech tagger
- NER Tagger - Named Entities tagger, which is capable of finding such entities like Person, Place, Organization, Names, potentially is able to find other entities provided with training data. This tagger is contextual, i.e. not only does it remember the words, but also understands from the word's surroundings where the named entity has to be placed there (takes advantage of lexical and semantical words' features). This characteristics brings about tagger's robustness and makes it universal for further features.
- Spell checker
- Synonyms matching - a module for finding synonyms for the input words
- Intent recognizer - semantic language tool for recognizing user's intent or classifying documents

## Infrastructure for NLP & Chatbots models handling:

- Creating new models with input data
- Recalculation of the existing models with new or updated data
- Model handling: requesting model's state, calculation logs, model deletion and disabling
- Text clusterization - a module for automatic clustering of the raw text to semantically relevant clusters
- NLP Models analysis - automated/semi-automated models' statistics, analysis and evaluation of data inside, e.g. the data cleanliness, categories intersections and etc.