NLP intro

4.11.2020

Фитц С.Ю.

Pipline

- What is?
- Tasks
- Steps for processing
- Applications

NLP tasks

a

- **Content Categorization**: A linguistic document summary that includes content alerts, duplication detection, search, and indexing.
- **Topic Discovery and Modeling**: Captures the themes and meanings of text collections, and applies advanced analytics to the text.
- **Contextual Extraction**: Automatically pulls structured data from text-based sources.
- **Sentiment Analysis:** Identifies the general mood, or subjective opinions, stored in large amounts of text. Useful for opinion mining.

NLP tasks

- **Text-to-Speech and Speech-to-Text Conversion**: Transforms voice commands into text, and vice versa.

- **Document Summarization**: Automatically creates a synopsis, condensing large amounts of text.

- **Machine Translation**: Automatically translates the text or speech of one language into another.

Processing steps

- data obtaining
- tokenization
- lemmatization/stemming
- vectorization(BOW, tf-idf, word2vec, doc2vec)
- modeling

Language models

Probability of word given the previous.

- Probabilistic models
- RNN, LSTM
- word2vec

Sintiments and topics

- Text classification.

- Recommendation Systems.

- Uncovering Themes in Texts.

Chatbots

- Goal-oriented (FAQ bot).
- Convesational (Алиса, Siri).
- NLU + text generation.