|  |  |
| --- | --- |
| Векторное представление лексических единиц. Семантика в задачах NLP. Дистрибутивные методы. Первые поптыки обучения эмбедингов. Косинусная мера близости. Word2vec. Fasttext. Doc2vec. BPE. Elmo. Кросс-лингвальные подходы к созданию эмбедингов. |  |
| Языковые модели. От n-gram до GPT-2. Классификация текстов: бинарная и многоклассовая. Пространство признаков. Интерпретации. |  |
| Анализ тональности. Латентный семантический анализ. Gensim. Методы снижения размерности. | Copy of Tutorial\_1\_Sentence\_classification\_with\_word\_embeddings.ipynb |
| Извлечение ключевых слов. Методы машинного обучения с учителем и без учителя при решении задачи. Подходы. TextRank. Yake. |  |
| Извлечение именованных сущностей. Автоматический синтаксический анализ. Способы представления синтаксической структуры. Приложения, в которых используется автоматический синтаксический анализ. Синтаксическая омонимия. Автоматический анализ в терминах деревьев зависимостей. Основные подходы. Проект UD | NER with BilSTM with tensorflow |
| Вопросно-ответные системы. Диалоговые агенты. | ChatBot tutorial |
| Нейронный машинный перевод. От статистических методов до нейронного машинного перевода. История развития. Выравнивание и параллельные корпуса. Модели IBM 1-5. Модели seq2seq. Beam-search. Буквенные и словесные модели переводов. Transformer. Метрика blue. | Statistical machine translation  NMT PyTorch |