

Синтаксический анализ

Для чего нужен

Основа — изучение связей слов в предложении

Можем взвешивать слова подобно TF-IDF/ мешку слов/эмбедингам

Производится на 3 этапе обработки текста, после токенизации и морфологии

Проверка правильности предложений

Структура составляющих

Часть речи

Классы слов с одинаковыми грам. функциями

Группа (*в особом значении) слов

"Расширение" понятия "часть речи"

Пример



Построение графа или дерева, подразумевая что все слова зависят, от глагола/сказуемого

Universal Dependencies

Это соревнование для исследования моделей синтаксического анализа

Модели для построения анализа

Graph-parsing method

Метрики

UAS

LAS

Грамматика зависимостей

Каждое слово зависит от какого-то другого

Грамматика непосредственных составляющих

Слова зависимы друг от друга, но есть группы наиболее зависимых друг от друга слов

Отсюда сразу вырисовывается очевидная проблема, которая сводит на нет использование таких парсеров для языков со свободным порядком слов (русский в том числе)

Инструменты

Stanford Core NLP