# Аннотация

1. Формулировка идеи (моя работа посвящена, какая теоретическая цель, практическая цель)
2. Постановка задачи (обучить классификаторы анализировать тексты)
3. План действий

Анализ лингвистических характеристик текстов, влияющих на автоматическое определение года его создания на материале ЖЖ.

Год создания текстов – определяем какой язык принадлежит какому году.

В результате анализа характеристик текста, мы можем получить представление, как язык изменялся на протяжении времени. Мы занимаемся дистрибутивной семантикой(дистрибутивное распределние слов в текстах).

Материал – тексты, взятые из ЖЖ с автоматической разметкой текста по годам (2001 – 2020), ВК (2014-2017) для доп. Экспериментов.

Эксперименты:

Экспериментировать с параметрами обучения, с тематическим моделированием. Также предполагается ручной анализ (частотные словари).

Эксперименты:

1. Берем BoW, LogisticReg, в лоб обучаем.
2. TF-IDF
3. n-grams
4. Ансамбли (менять алгоритмы)
5. LDA
6. Применить классификатор, обученный на ЖЖ, к текстам ВК

Гипотезы:

H0: на опредление года будет влиять тематика. Нет эволюции.

H1: не будет влиять тематика, эволюция есть.

Введение

Теоретическая глава:

1. что такое классификация текста
2. обзор по тех.классификациям
3. Какие уже были исследования для определния даты создания текстов

Практическая глава

1. Описание материала (какие тексты, их особенности, сколько их)
2. Описание экспериментов
3. Ручной анализ наиболее влиящий на класс-. слов

Заключение

В результате анализа характеристик текста, мы можем получить представление, как язык изменялся на протяжении времени. Мы занимаемся дистрибутивной семантикой.