

Алгоритм

Теперь для определения профессии нового профиля, для каждого слова из профиля ищем похожие по смыслу слова в графе подобия слов. Мы использовали word2vec из библиотеки gensim

Для быстрого обнаружения слова мы использовали алгоритм Ахо-Корасик, затем если оно было найдено, то мы прибавляли к профессии соответствующий вес с помощью метода tf-idf