Результаты встречи от 11.09.2020

1. Было решено использовать разложение векторов по ортогональному базису.
2. Необходимо реализовать определение классов, за счёт кластеризации ключевых слов.
3. Выявлена необходимость в формировании полного списка ключевых слов вручную.
4. Необходимо сделать выборку по поисковому API (в перспективе, не только по яндексу но и по google и bingo).

Было принято решение, протестировать google scholar api. В случае если не получится использовать данное решение, необходимо размечать вручную, на основе текстов всех доступных статей.

1. Необходимо переработать алгоритм определения классов так, чтобы тот при обнаружении новых ключевых слов, помещал его в таблицу в базе данных с более точным значением.

На данный момент при добавлении нового ключевого слова по умолчанию выставляется значение 0.

Необходимо вычислять расстояние до значения соседнего ключевого слова.

1. Было принято решение сделать заглушку, на определение правильности нахождения ключевого слова. На возвращение результата — 1.

Данный этап необходим для дальнейшего расширения, с помощью нейронной сети.

1. Начата реализация обработки фалов pdf. С использование IText – инструмента языка Java.