

Большие данные

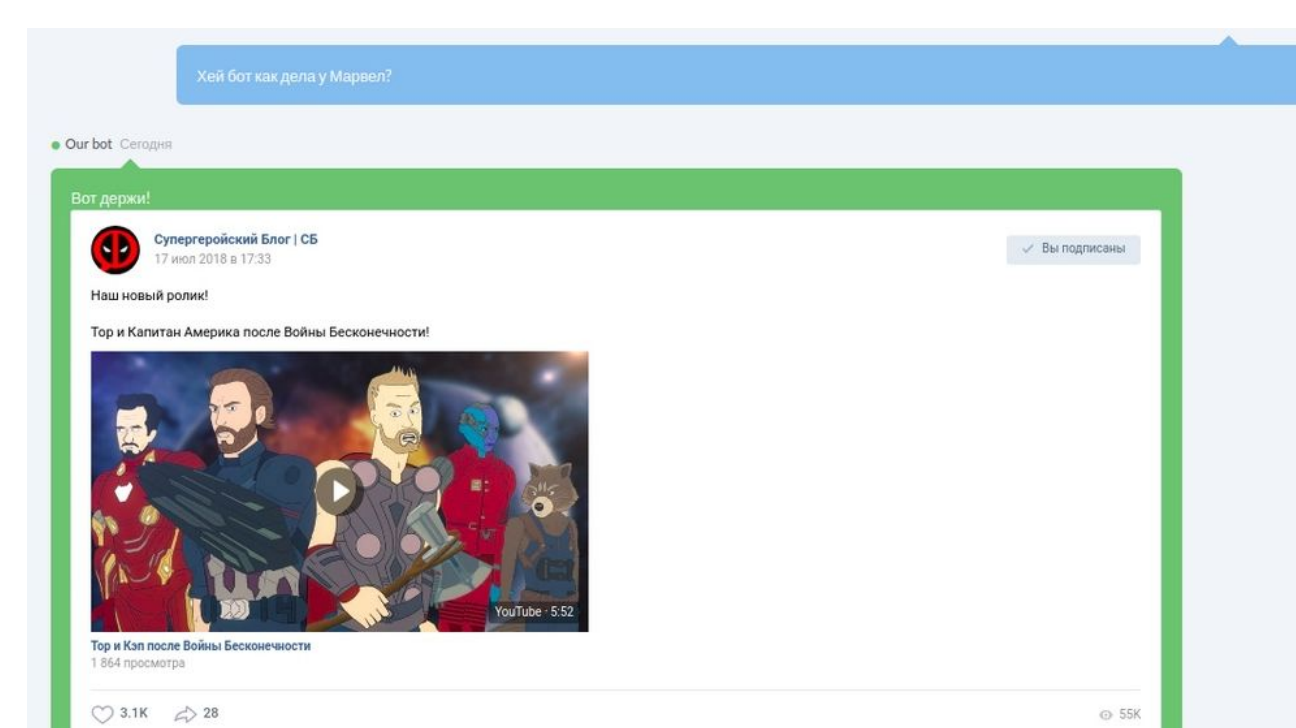
Интеллектуальный помощник колл-центра

ПРОБЛЕМА

- Современный пользователь тратит много времени на поиск нужной и интересной ему информации
- Современные компании имеют значительные финансовые затраты на содержание колл-центров

Мы хотим помочь пользователю, получив по его запросу наиболее релевантный по содержанию пост из социальной сети Вконтакте

При чём тут колл-центр? Имея базу данных конкретной компании можно с легкостью доучить нашу модель как помощника колл-центра

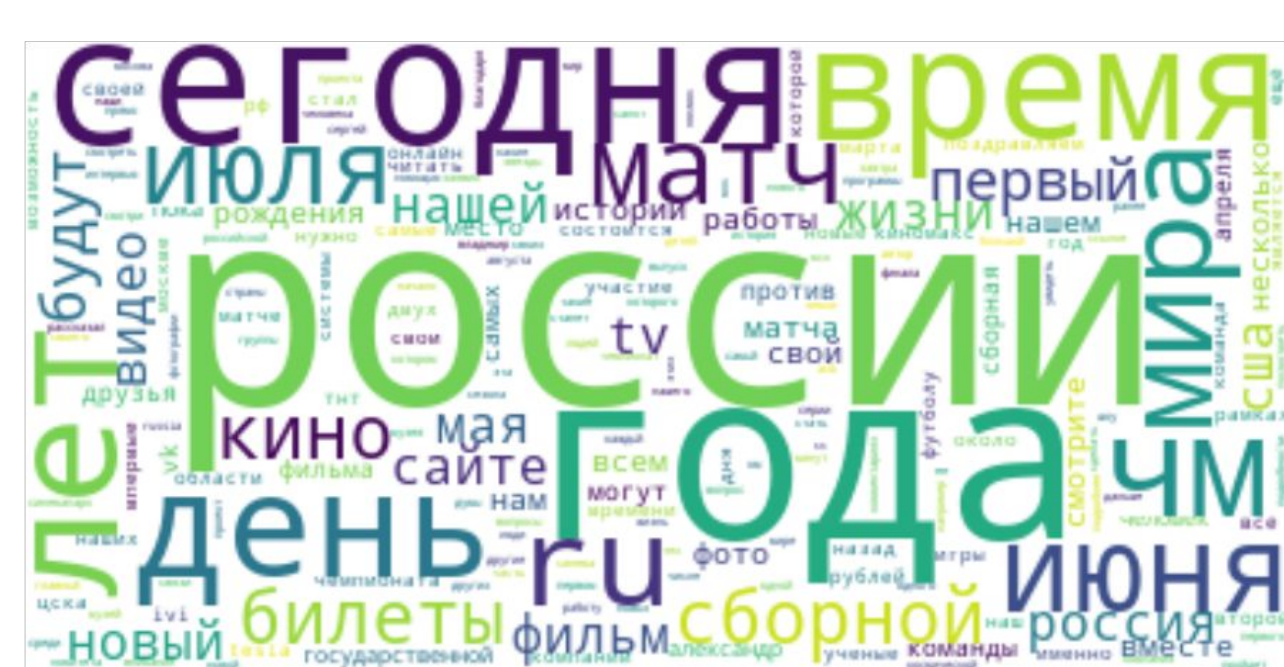


РЕШЕНИЕ

Сбор данных

Не имея данных, необходимых для обучения нашей модели, мы произвели самостоятельный поиск и сбор информации из социальной сети Вконтакте

- Датасет был очищен от мусорных данных
- Была произведена лемматизация слов для увеличения качества выборки
- Были нормализованы объемы данных по каждому классу



Классификация запроса

- Мы использовали трехслойную нелинейную нейронную сеть, классифицирующую текстовый запрос по существующим категориям
- Для выделения численных признаков из естественного текста использовался метод TF-IDF
- Количество выделяемых признаков составило 15000

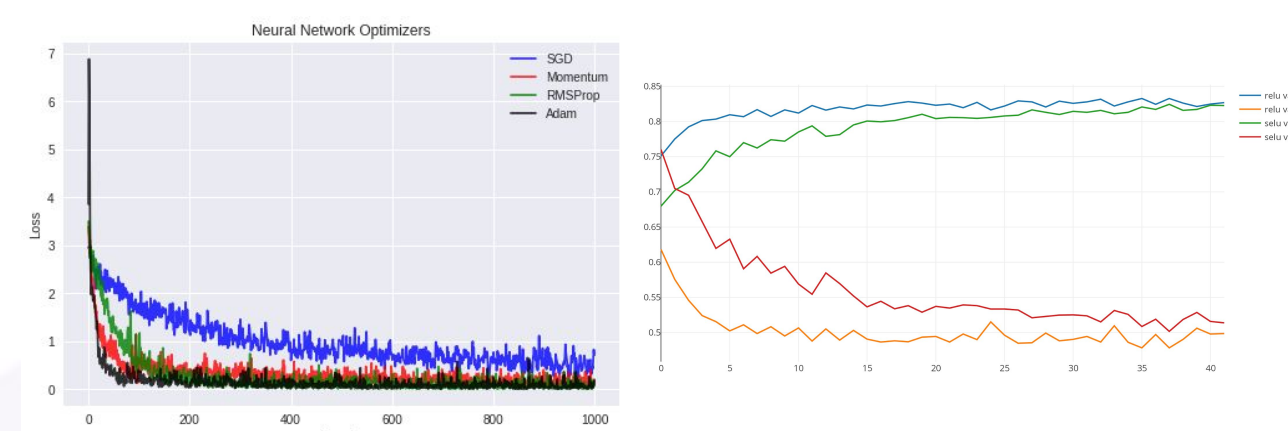
Реализованная модель на тесте имеет **точность 91%**

Модель классифицирует следующие типы запросов:

- Наука и технологии
- Спорт
- Новости
- Развлечения
- Культура и искусство
- Игры и киберспорт

Ранжирование

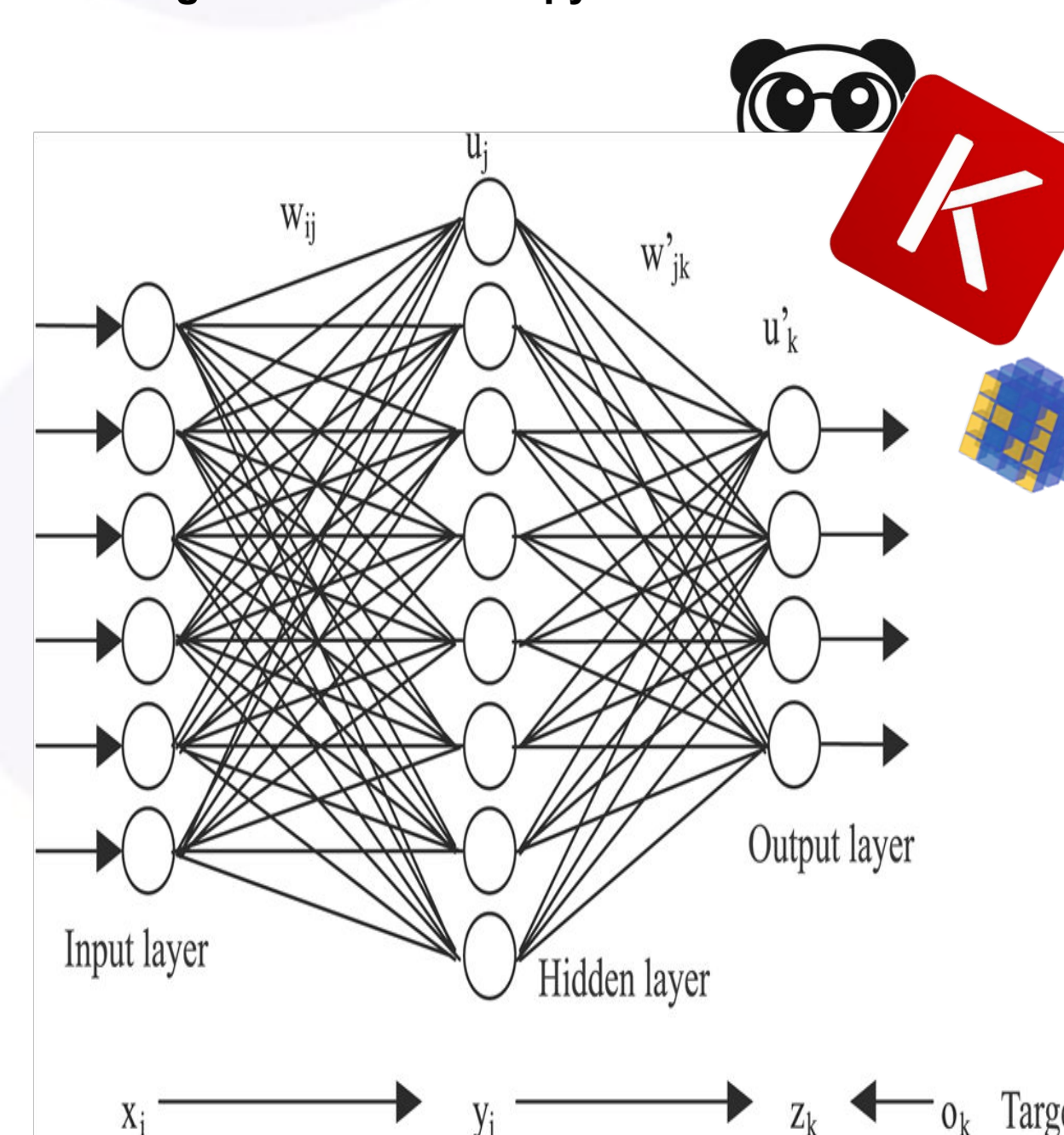
- Мы используем **Word2Vec** для поиска наиболее релевантных записей
- Для составления численных характеристик каждого поста мы используем 40 наиболее часто встречающихся слов в этой записи
- Для каждого поста мы усредняем Word2Vec входящих в него слов, то же проделываем и с запросом
- Поиск осуществляется внутри категории запроса



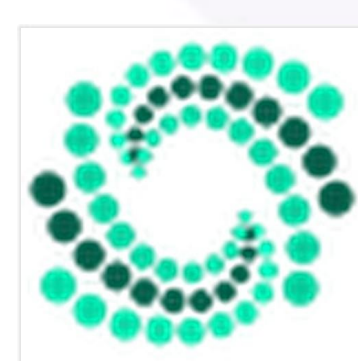
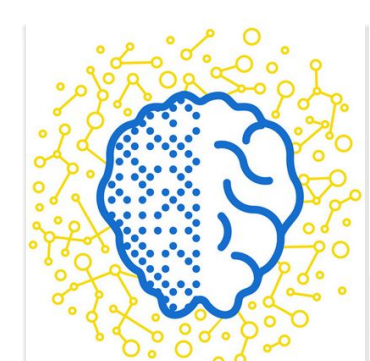
Сеть имеет следующую архитектуру:

- Входной слой
- Полносвязный слой с функцией активации **Tanh**
- Дропаут
- Полносвязный слой с функцией активации **Relu**
- Полносвязный слой с функцией активации **Softmax**

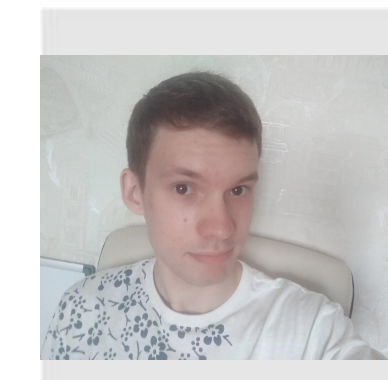
В качестве оптимизатора был выбран **RMSprop**
В качестве функции потерь была выбрана **Categorical cross entropy**



ПАРТНЕРЫ



РУКОВОДИТЕЛИ ПРОЕКТА



Червонцев
Сергей
Сергеевич



Бурцев
Михаил
Сергеевич

ЭКСПЕРТЫ

КОМАНДА

Поднебеснова Яна

Курбанов Малик

Тонкошнуров Никита

Пушкарев Святослав

Шипицин Антон

Новгородов Арсений

Помазенков Илья