

**Проблемы выявления зависимости между позитивными и негативными  
русскоязычными новостями и выбором страны для эмиграции**

**Научный руководитель – Линченко Андрей Александрович**

***Алексеева Елизавета Александровна***

*Студент (бакалавр)*

Финансовый университет при Правительстве РФ, Липецкий филиал, Кафедра  
математики и информатики, Липецк, Россия

*E-mail: lippro48@gmail.com*

Согласно данным Росстата [n2], отток граждан из России начал происходить с 2011 года. В 2017 году количество выехавших из страны составило 377155 человек. Самыми популярными странами дальнего зарубежья для миграции являются Китай, КНДР и Германия.

Одним из основных источников формирования мнения о стране являются новостные порталы на родном языке потенциального эмигранта. Именно из-за их доминирующего присутствия создается первоначальный образ о стране. Не последнюю роль здесь играют и поисковые системы, выдающие по запросам те или иные новостные ресурсы. Согласно данным Statcounter [n5], лидирующими поисковыми системами в России являются Yandex и Google. Если проанализировать количество запросов на русском языке с ключевым словом «название страны» и выбрать раздел «новости», то статистика в «Google Trends» [n3] выдаст заинтересованность русскоязычных пользователей новостями о КНР и США в большей степени. Германия и КНДР находятся в среднем диапазоне. Однако данная статистика не говорит, какое мнение о стране формируется у пользователей. Сравнение запросов «Китай» и «США» дает практически одинаковый диапазон, однако граждане РФ в большей степени эмигрируют в Китай, а не в США: в 2017 году 7600 человек против 1452. Поэтому поднимается вопрос о необходимости анализа информационных настроений в русскоязычной сети и их влияния на миграцию населения.

Для проведения синтаксического и лингвистического анализа необходим сбор данных за 8 лет. В данном случае необходим набор заголовков, относящихся к странам-лидерам по эмиграции граждан из РФ и менее приоритетным странам для миграции. Для этого есть фильтры как в Google поисковике, так и в Yandex. Однако автоматизировать данный процесс целиком не получится, поскольку компания Google, как и Facebook «стала придерживаться правил, прописанных в «General Data Protection Regulation» [n1]. Поисковая система Yandex построена таким образом, что открытые API в ней недоступны, поэтому автоматизировать сбор данных не получится.

Несмотря на это можно извлечь необходимые заголовки неэффективным способом - в «ручном режиме». Однако проблемы не заканчиваются, ведь необходимо провести саму лингвистическую оценку заголовков. Для англоязычных источников существуют разные библиотеки, которые способны провести подобный анализ. Так, например, использование пакета «scikit-learn» для Python. С помощью данного пакета можно создать необходимую нейросеть с учителем, которая выстраивает дерево взаимосвязей и формирует оценочную характеристику, возвращая результаты в процентах [n4]. Тем не менее, работать с русифицированными данными он не может.

Таким образом, на сегодняшний день основными проблемами при выявлении взаимосвязи влияния новостей на потенциальных эмигрантов в автоматизированном режиме

закljučаются в отсутствие открытого API у поисковых систем и неумение работать оценочных библиотек с русифицированными источниками. Однако гипотетическое влияние информационных настроек значительно сказывается на потенциальной миграции.

### Источники и литература

- 1) Алексеева Е.А. Проблемы формирования big data на основе открытых пользовательских данных в социальных сетях // Центральный научный вестник ТЗ. No. 22s(63s). ISSN: 2499-9989
- 2) Федеральная служба государственной статистики. [http://www.gks.ru/wps/wcm/connect/rosstat\\_main/rosstat/ru/statistics/population/demography/#](http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/population/demography/#)
- 3) Google Trends. <https://trends.google.ru/trends/>
- 4) Habr. com <https://habr.com/ru/post/263171/>
- 5) Statcounter. <http://gs.statcounter.com/search-engine-market-share/all/russian-federation#monthly-201805-201805-map>