

Ответы на вопросы рецензента

ВКР: Градуировка тональной шкалы реакций на интернет-новости
(на материале русского, английского и испанского языков)

Исполнитель: Савинцева Дарья Викторовна

Рецензент: Гордеев Станислав Сергеевич

13 июня 2020 г.

1. *При формировании корпусов текстов для русского языка используются комментарии сети ВКонтакте, а для английского и испанского языков — комментарии в YouTube. Не повлияло ли использование разнотипных источников на результаты исследования?*

Мы полагаем, что использование данных из разных социальных сетей несущественно повлияло на результаты исследования, поскольку были выделены как схожие темы для всех корпусов, так и особенные, присущие каждому корпусу в отдельности темы, отсутствующие в других корпусах.

Помимо этого, на момент исследования при анализе русскоязычных новостных каналов YouTube было отмечено, что у некоторых главных новостных каналов закрыты комментарии к видеозаписям (например: канал «Новости на Первом Канале»¹) или для некоторых новостных источников не существует YouTube канала. Поэтому было принято решение собрать русскоязычные данные из сети ВКонтакте, как наиболее популярной в России² для получения корпуса комментариев, достаточного полного для целей исследования.

2. *Еще один вопрос по поводу объема корпусов текстов — их различие не оказало ли существенное влияние на результаты шкалирование?*

На наш взгляд, это различие оказало не критичное влияние на результаты. Перед построением общей шкалы мы выделили основную тенденцию, присущую каждому корпусу в отдельности: темы, вероятность появления которых в комментариях группы меньше, находят больший эмоциональный отклик. В основном, это темы, касающиеся повседневной жизни и насущных проблем пользователей в разных странах, а не глобальных и общих проблем. На основании этого для построения шкалы мы выделили пересекающиеся темы. Также, чтобы сгладить разницу в различии размеров корпусов, за основу тональной шкалы мы взяли среднее, а не абсолютное значение тональности каждой темы в корпусе. Но, безусловно, если бы размеры всех корпусов были примерно одинаковыми, тогда возможно было бы получить более точную оценку тональности для каждой темы в каждом корпусе.

3. *При описании тематической модели текстов корпуса упоминаются разные наборы тем (от 10 до 30) и указывается на параметр их интерпретируемости, в дальнейшем отбираются наборы из 14 тем. Однако каковы эти наборы и результаты интерпретируемости на них?*

Для каждого корпуса были построены модели на 10, 15, 20 и 30 тем, и для дальнейшего исследования мы использовали модели из 15 тем. Нумерация выделенных тем, представленных в таблице 3.1. начинается с 0, а не с 1 (в работе сохранена нумерация, представленная библиотекой gensim).

4. *Тональная интерпретация тем, как следует из рис. 3.3, зависит от источника информации. Ясно, что существуют «демократические» и «консервативные» издания, каков диапазон тональности с учетом политической ангажированности?*

Этот вопрос очень интересный, однако в рамках исследования мы его не касались (на рис. 3.3 представлена лишь вероятность появления той или иной темы в новостном источнике). Мы не акцентировали внимание на политической позиции новостных источников и не анализировали тональность каждой темы в комментариях каждой новостной группы или канала, а брали только общую тональность выделенной темы в корпусе. Возможно, этот вопрос будет актуален для дальнейших, более детальных исследований по теме настоящей работы.

¹https://www.youtube.com/channel/UCKonxxVHzDl55V7a9n_Nlgg/featured, дата обр. 13.06.2020

²<https://br-analytics.ru/blog/social-media-russia-2019/>, <https://gs.seo-auditor.com.ru/socials/>, дата обр. 13.06.2020