

Вестник Евразийской науки / The Eurasian Scientific Journal <https://esj.today>

2018, №2, Том 10 / 2018, No 2, Vol 10 <https://esj.today/issue-2-2018.html>

URL статьи: <https://esj.today/PDF/43ITVN218.pdf>

Статья поступила в редакцию 02.04.2018; опубликована 25.05.2018

Ссылка для цитирования этой статьи:

Телевной А.Д., Хлопотов М.В. Исследование тематических профилей и способов расчета вовлеченности аудитории в сообщества социальной сети «ВКонтакте» // Вестник Евразийской науки, 2018 №2, <https://esj.today/PDF/43ITVN218.pdf> (доступ свободный). Загл. с экрана. Яз. рус., англ.

For citation:

Televnoy A.D., Khlopotov M.V. (2018). The study of thematic profiles and ways to calculate audience involvement in community social network «VKontakte». *The Eurasian Scientific Journal*, [online] 2(10). Available at: <https://esj.today/PDF/43ITVN218.pdf> (in Russian)

УДК 004.62

ГРНТИ 20.23.17

Телевной Андрей Дмитриевич

ФГБОУ ВПО «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики», Санкт-Петербург, Россия
Магистрант
E-mail: adtelev@mail.ru

Хлопотов Максим Валерьевич

ФГБОУ ВПО «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики», Санкт-Петербург, Россия
Кандидат технических наук, доцент
E-mail: khlopotov@corp.ifmo.ru

Исследование тематических профилей и способов расчета вовлеченности аудитории в сообществах социальной сети «ВКонтакте»

Аннотация. В статье описано исследование тематических профилей в сообществах социальной сети «ВКонтакте» тематики «Университет» с помощью построения модели латентного размещения Дирихле. Объектами исследования являются онлайн-сообщества социальной сети «ВКонтакте». Для участия в исследовании онлайн-сообщества отбирались по сформулированным авторами критериям. В перечне исследуемых онлайн-сообществ представлены: «Университет ИТМО», «МГИМО Университет», «Казанский федеральный университет», «Сибирский федеральный университет» и «Уральский федеральный университет».

Исследование проводилось с помощью программного пакета для сбора данных из социальной сети «ВКонтакте», разработанного на кафедре Интеллектуальных технологий в гуманитарной сфере Университета Информационных технологий, механики и оптики, а также морфологического анализатора для проведения лемматизации собранных корпусов текстовых данных.

Авторами представлены результаты исследования способов расчета коэффициента вовлеченности аудитории по базе и по охвату на протяжении нескольких временных промежутков (7 дней и 30 дней). Авторами найдена и сформулирована взаимосвязь между

показателями коэффициентов вовлеченности и структурой тематического профиля у одного из анализируемых онлайн-сообществ.

Результаты тематического моделирования позволяют выделять наиболее обсуждаемые темы в онлайн-сообществах и определять публикационную политику их администраторов. Результаты исследования вовлеченности показали, что данный показатель активности аудитории не находится в прямой зависимости от количества подписчиков онлайн-сообществ.

Также по результатам исследования нельзя не отметить взаимосвязь между четкой структурой тематического профиля и максимальными показателями коэффициентов вовлеченности у онлайн-сообщества «Сибирский федеральный университет». Предварительное исследование показателя вовлеченности также можно рекомендовать маркетологам при проведении рекламных кампаний (для выбора сообщества с максимальной вовлеченностью аудитории).

Ключевые слова: анализ социальных сетей; онлайн-сообщество; коэффициент вовлеченности; тематическое моделирование; латентное размещение Дирихле; метрика онлайн-сообществ; графические вероятностные модели; семантические связи

Введение

Социальные сети в наше время являются уникальным источником данных об интересах людей, вследствие чего такая область, как анализ социальных сетей набирает стремительную популярность.

Изучение социальных сетей как научное направление возникло на стыке ряд научных дисциплин – социология, дискретная математика, Computer Science (алгоритмы на графах и сетях). К сожалению, на данный момент четкого математического определения онлайн-сообщества не существует. Обычно про сообщество говорят как про некоторую группу узлов или некоторую группу людей, которые связаны внутри группы сильнее между собой, чем со всем остальным миром [1].

Для анализа социальных данных требуются специализированные программные продукты, многие из которых создаются самими авторами перед проведением очередных исследований в данной области. Например, в научной статье сотрудников Института систем информатики приводится довольно подробное описание разработанного ими программного комплекса для извлечения, обработки и анализа пользовательских данных. Также в данной статье предлагается один из возможных методов определения лидеров мнений, сформулированный на основе адаптации теории социального влияния Латарне [2].

Анализ социальной сети также можно рассматривать под углом исследования связанных структур пользователей и отношений между ними. В статье [3] авторами проведено подобное исследование, объектами которого выступили онлайн-сообщества тематики «чайлдфри» социальной сети «ВКонтакте». Собранные данные были использованы для расчета математических характеристик, после чего визуализированы при построении социальных графов. В качестве результатов заявлена классификация онлайн-сообществ по целям и социальным функциям, а также был составлен «типичный портрет» участника сообщества подобной тематики.

Тематическое моделирование – способ построения модели текстовой коллекции, отражающий переход от совокупности документов, совокупности слов в документах коллекции к набору тем, характеризующих текстовую коллекцию [4].

Построение тематического профиля онлайн-сообщества позволяет определить содержательное наполнение тем, обсуждаемых в рамках данного сообщества. Латентное

размещение Дирихле – графическая вероятностная модель, предложенная в 2003 году Дэвидом Блеем и соавторами [5]. Количество тем в данной модели фиксируется изначально и должно быть задано в явном виде перед запуском алгоритма. На вход тематической модели поступает корпус текстовых данных, каждый из которых является отдельным документом. Результатом работы алгоритма является список тем, выявленных в корпусе и представленным списком первых, наиболее характерных n слов для каждой рассматриваемой темы [6]. Предпочтение данной модели было отдано после знакомства с работой А. Коршунова и А. Гомзина «Тематическое моделирование текстов на естественном языке», в которой авторами был проведен сравнительный обзор различных моделей проведения тематического моделирования, а также описаны способы оценивания их параметров и качества результатов [7].

Вовлеченность (Engagement Rate, ER) – это показатель уровня вовлеченности аудитории в активности компании, который измеряется в процентном соотношении действий к охвату. Данная SMM-метрика изначально появилась для оценки эффективности контентной стратегии и коммуникации в Facebook. Для его подсчета суммарное количество лайков, репостов и комментариев делится на количество подписчиков и постов, которое размещено в анализируемый период [8]. Как правило, подсчет вовлеченности активно используется при анализе необходимости корректирования рекламной активности, исходя из реакции целевой аудитории того или иного онлайн-сообщества [9].

В данной работе были выбраны методы расчета вовлеченности по базе аудитории (общему количеству подписчиков онлайн-сообщества) и охвату аудитории, то есть только тем пользователям, которые видели опубликованные посты (социальная сеть «ВКонтакте» предоставляет возможность оценить охват аудитории поста с помощью так называемого счетчика просмотров) [10].

Формула вычисления вовлеченности аудитории по базе:

$$\text{Daily PER} = \frac{\text{Лайки} + \text{Комментарии} + \text{Репосты за конкретную дату}}{\text{Количество подписчиков страницы в конкретную дату}} \times 100 \%$$

Формула вычисления вовлеченности аудитории по охвату:

$$\text{Daily PER} = \frac{\text{Лайки} + \text{Комментарии} + \text{Репосты за конкретную дату}}{\text{Охват страницы в конкретную дату}} \times 100 \%$$

Теоретико-методологической основой для исследования стала диссертационная работа Ю.Г. Рыкова «Структура социальных связей в виртуальных сообществах: сравнительный анализ онлайн-групп социальной сети «ВКонтакте», в которой автором был предложен сетевой подход к анализу сообществ на основе структурных показателей, а также методикой определения вовлеченности, представленной автором Д. Румянцевой [11, 9].

Цель работы. Исследовать тематические профили и способы расчета вовлеченности аудитории в сообществах социальной сети «ВКонтакте» тематики «Университет» на предмет установления определенных зависимостей.

Базовые положения исследования. Объектом исследования являются онлайн-сообщества социальной сети «ВКонтакте».

Подбор онлайн-сообществ для исследования осуществлялся при помощи поискового фильтра социальной сети. Были сформулированы следующие критерии для участия сообществ в исследовании:

- онлайн-сообщество открытого типа;
- существует возможность комментирования публикаций на стене онлайн-сообщества;

- дата публикация последней записи на стене онлайн-сообщества не позднее 3 дней со дня подбора онлайн-сообществ для исследования.

Таким образом, были подобраны следующие онлайн-сообщества, удовлетворяющие данным критериям (таблица 1).

Таблица 1

Онлайн-сообщества, подобранные для участия в исследовании

| № | Название онлайн-сообщества | Количество подписчиков (по состоянию на 30.03.2018) |
|---|---|---|
| 1 | «Университет ИТМО» (https://vk.com/spbifmo) | 36242 |
| 2 | «МГИМО Университет» (https://vk.com/mgimo) | 27620 |
| 3 | «Казанский федеральный университет» (https://vk.com/kazan_federal_university) | 36121 |
| 4 | «Уральский федеральный университет» (https://vk.com/ural.federal.university) | 32637 |
| 5 | «Сибирский федеральный университет» (https://vk.com/siberianfederal) | 26858 |

Составлено автором

Сбор данных осуществлялся с помощью программного пакета «vkR», разработанного на кафедре ИТГС Университета ИТМО в Санкт-Петербурге, реализованного на осуществлении API-запросов к серверу социальной сети «ВКонтакте».

В выборку данных, получаемых с помощью пакета «vkR», включены текстовые данные со стен онлайн-сообществ и числовые показатели активности аудитории (количество комментариев, лайков, репостов и просмотров).

1 этап исследования. Построение тематических профилей онлайн-сообществ

Построение тематического профиля начинается с этапа сбора текстовых данных со стены онлайн-сообщества. После чего необходимо провести очистку текста от нежелательных элементов (например, html-тегов), лемматизацию (с помощью морфологического анализатора `ru morphology2`) и составление списка стоп-слов (стоп-слова – слова, не несущие смысловой нагрузки). Заданный параметр количества тем равен 7.

Со стен сообществ собирались все существующие документы, начиная с первого дня существования онлайн-сообщества. В таблице 2 представлено количество документов (постов и комментариев к ним), собранных со стен соответствующих онлайн-сообществ.

Таблица 2

Количество документов, собранных со стен онлайн-сообществ

| № | Название онлайн-сообщества | Количество документов |
|---|---|-----------------------|
| 1 | «Университет ИТМО» (https://vk.com/spbifmo) | 6229 |
| 2 | «МГИМО Университет» (https://vk.com/mgimo) | 7277 |
| 3 | «Казанский федеральный университет» (https://vk.com/kazan_federal_university) | 7723 |
| 4 | «Уральский федеральный университет» (https://vk.com/ural.federal.university) | 10824 |
| 5 | «Сибирский федеральный университет» (https://vk.com/siberianfederal) | 7166 |

Составлено автором

После сбора и предварительной обработки корпуса данных проведено построение тематических моделей сообществ с помощью пакета «lda», написанного на языке программирования R.

Пример результата работы данного пакета и последующей визуализации представлен на рисунке.

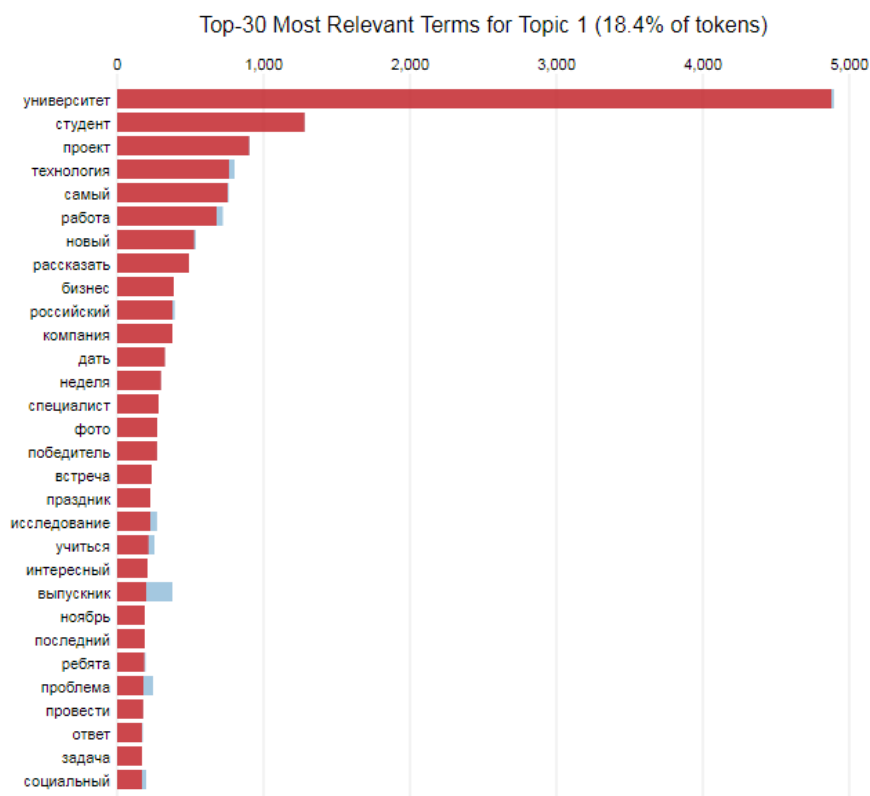


Рисунок. Тематический профиль онлайн-сообщества «Университет ИТМО» (тема 1, разработано автором)

2 этап исследования. Вычисление коэффициента вовлеченности аудитории онлайн-сообществ

На втором этапе исследования производилось вычисление коэффициента вовлеченности аудитории двумя методами: по базе и по охвату.

Для представленных в таблице 1 онлайн-сообществ был проведен сбор числовых показателей активности аудитории за последние 7 дней (24.03.2018-30.03.2018) и 30 дней (01.03.2018-30.03.2018). Стоит отметить, что социальной сетью «ВКонтакте» не предусмотрена возможность просмотра количества подписчиков ранее сегодняшнего дня, поэтому при вычислении вовлеченности по базе учитывалось количество подписчиков за 30 марта 2018 года. Возможная погрешность коэффициента вовлеченности признана незначительной.

Согласно методике определения вовлеченности ключевыми показателями являются численность подписчиков сообщества и количество комментариев, лайков и репостов и просмотров под каждым постом, опубликованным в эту дату.

Пример вычисления коэффициента вовлеченности за последние 7 дней для сообщества «Университет МГИМО» представлен в таблице 3.

Таблица 3
Коэффициент вовлеченности сообщества «Университет МГИМО» за 7 дней

| № | Показатели | 30.03 | 29.03 | 28.03 | 27.03 | 26.03 | 25.03 | 24.03 |
|---|-------------------------|-------|-------|-------|-------|-------|-------|-------|
| 1 | Количество подписчиков | 27620 | | | | | | |
| 3 | Количество комментариев | 3 | 1 | 0 | 1 | 5 | 0 | 5 |
| 4 | Количество лайков | 113 | 95 | 54 | 107 | 34 | 8 | 61 |
| 5 | Количество репостов | 2 | 12 | 4 | 10 | 7 | 2 | 1 |
| 6 | Количество просмотров | 23587 | 11216 | 14167 | 19105 | 10402 | 1821 | 5037 |

| № | Показатели | 30.03 | 29.03 | 28.03 | 27.03 | 26.03 | 25.03 | 24.03 |
|----|----------------------------|-------|-------|-------|-------|-------|-------|-------|
| 7 | Средний охват | 12191 | | | | | | |
| 8 | DPER по базе (за 1 день) | 0.43 | 0.39 | 0.21 | 0.43 | 0.17 | 0.04 | 0.24 |
| 9 | DPER по базе (за 7 дней) | 0.27 | | | | | | |
| 10 | DPER по охвату (за 1 день) | 0.5 | 0.96 | 0.41 | 0.62 | 0.44 | 0.55 | 1.33 |
| 11 | DPER по охвату (за 7 дней) | 0.69 | | | | | | |

Составлено автором

Результаты исследования

Результатом построения тематического профиля является перечень наиболее часто обсуждаемых тем (топиков) в подобранных для исследования онлайн-сообществах (таблица 4). Название той или иной теме давалось по словам, формально занимающим первые позиции в списке.

Таблица 4

Перечень тем, представленных в онлайн-сообществах

| № Темы | «ИТМО» | «МГИМО» | «КФУ» | «УрФУ» | «СФУ» |
|--------|---------------------------|-------------------------|----------------------------|--------------------------|----------------------------|
| 1 | Университет и студент | Международный МГИМО | Университет и студент | Федеральные конкурсы | Представление университета |
| 2 | ВУЗы Петербурга | Университет и студент | Международная школа | Проекты в университете | Университет и студент |
| 3 | Международные мероприятия | Факультеты университета | Работа федер. университета | Абитуриенты и выпускники | Новостные рубрики |
| 4 | Программы мероприятий | Участие в мероприятиях | Расписание курсов | Программы конференций | День в институте |
| 5 | День в истории науки | Встречи на кафедрах | Наука и ученые | Университет и студент | Погода и конкурсы |
| 6 | Вопросы по курсам | Развитие университета | Цели развития университета | Поздравления победителей | Фото студентов |
| 7 | Конкурс среди ученых | Лекции в институте | Проекты и лекции | Заявки на мероприятия | Геометки в Instagram |

Составлено автором

Анализ построенных тематических профилей позволяет сказать о том, что тема «Университет и студент» играет заметную роль в постах и обсуждениях всех онлайн-сообществ. Данный факт можно охарактеризовать как подтверждение существования прямой связи между университетом и студентами хотя бы в рамках социальной сети.

Отдельное место в контенте онлайн-сообществ занимает тема международной деятельности университетов (конкурсы, конференции, олимпиады и другие образовательные мероприятия).

Онлайн-сообщества «Университет ИТМО» и «Казанский федеральный университет» уделяют особое внимание развитию организованных образовательных курсов (Темы «Вопросы по курсам» и «Расписание курсов» соответственно).

Онлайн-сообщества «Уральский федеральный университет» и «Университет ИТМО» регулярно публикуют программы проводимых в рамках данных университетов конференций (Темы «Программы конференций» и «Программы мероприятий» соответственно). Данный подход преимущественен тем, что позволяет ознакомиться с ключевыми положениями образовательных мероприятий, причем не только подписчикам данных онлайн-сообществ, но и сторонним заинтересованным пользователям.

Стоит отметить строгую политику публикации постов у онлайн-сообщества «Сибирского федерального университета», что нашло отражение в появлении у него таких тем, как «Новостные рубрики», «Погода и конкурсы» и «Фото студентов».

Результаты определения коэффициента вовлеченности по базе и по охвату представлены в таблицах 5 и 6.

Таблица 5

Результаты исследования коэффициента вовлеченности за 7 дней

| № | Показатели | «ИТМО» | «МГИМО» | «КФУ» | «УрФУ» | «СФУ» |
|---|----------------|--------|---------|-------|--------|-------|
| 1 | Средний охват | 9960 | 12191 | 6786 | 16414 | 17243 |
| 2 | DPER по базе | 0.17 | 0.27 | 0.15 | 0.23 | 0.69 |
| 3 | DPER по охвату | 0.37 | 0.69 | 0.59 | 0.45 | 1.07 |

Составлено автором

Таблица 6

Результаты исследования коэффициента вовлеченности за 30 дней

| № | Показатели | «ИТМО» | «МГИМО» | «КФУ» | «УрФУ» | «СФУ» |
|---|----------------|--------|---------|-------|--------|-------|
| 1 | Средний охват | 11748 | 8505 | 8023 | 17733 | 19993 |
| 2 | DPER по базе | 0.22 | 0.19 | 0.16 | 0.26 | 0.85 |
| 3 | DPER по охвату | 0.45 | 0.55 | 0.53 | 0.46 | 1.12 |

Составлено автором

Наименьшие показатели среднего охвата аудитории и коэффициента вовлеченности по базе наблюдаются у онлайн-сообщества «Казанский федеральный университет». Значительно выделяются показатели онлайн-сообщества «Сибирский федеральный университет».

Стоит отметить, что у онлайн-сообщества с максимальным количеством подписчиков («Университет ИТМО») одни из минимальных показателей коэффициента вовлеченности и по базе, и по охвату.

Несмотря на относительно высокие показатели среднего охвата аудитории у онлайн-сообщества «Уральский федеральный университет» (второе место среди участвовавших в исследовании), коэффициенты вовлеченностикратно ниже сообщества «Сибирский федеральный университет» (первое место по охвату).

Значение показателя среднего охвата аудитории у онлайн-сообщества «Университет МГИМО» за последний месяц на треть ниже, чем за последние 7 дней. Это можно объяснить тем, что в конце марта 2018 года в данном университете проводился «Столыпинский форум», что повлекло за собой большое количество посвященных мероприятию постов.

Выводы

Результаты построения тематических профилей онлайн-сообществ показали целесообразность применения алгоритма латентного размещения Дирихле. Результаты моделирования позволяют выделять наиболее обсуждаемые темы в онлайн-сообществах и определять публикационную политику их администраторов.

Результаты исследования вовлеченности показали, что показатель вовлеченности аудитории не находится в прямой зависимости от численности и тематики онлайн-сообществ. Все подобранные для исследования онлайн-сообщества имеют примерно равное количество подписчиков (около 30000 человек).

Оба метода расчета коэффициента вовлеченности аудитории имеют право на существование. Целесообразность применения того или иного способа должна определяться в индивидуальном контексте, в зависимости от целей исследования.

Однако, наибольший коэффициент вовлеченности аудитории (равно как и охвата) при минимальном количестве подписчиков наблюдается у онлайн-сообщества «Сибирский федеральный университет», тематический профиль которого имеет строгую структуру и выделяется публикациями развлекательных постов (например, посты с конкурсами, прогнозами погоды на день и фотографиями студентов). Можно порекомендовать администраторам других онлайн-сообществ аналогичным образом пересмотреть политику публикации постов с целью увеличения вовлеченности подписчиков.

По результатам исследования нельзя не отметить взаимосвязь между четкой структурой тематического профиля и максимальными показателями коэффициентов вовлеченности у онлайн-сообщества «Сибирский федеральный университет».

Предварительное исследование показателя вовлеченности также можно рекомендовать маркетологам при проведении рекламных кампаний (для выбора сообщества с максимальной вовлеченностью аудитории).

ЛИТЕРАТУРА

1. Гусарова, Н.Ф. Анализ социальных сетей. Основные понятия и метрики // СПб: Университет ИТМО, 2016 г. 67 с.
2. Т.В. Батура, Ф.А. Мурзин, А.В. Проскуряков. Программный комплекс для анализа данных из социальных сетей // Программные продукты и системы. 2015. В. 4 (112). С. 188-197.
3. Startseva N., Khlopotov M.V., Ivanov S.E. Analysis of the audience of childfree communities in social network "VKontakte" // ACM International Conference Proceeding Series – 2017, Vol. Part F130282, pp. 107-112.
4. О.А. Митрофанова. Моделирование тематики специальных текстов на основе алгоритма LDA [Электронный ресурс] – URL: <http://mfk.spbu.ru/sites/default/files/mitrofanova.pdf>, свободный. – Загл. с экрана. – Яз. рус.
5. D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. Journal of Machine Learning Research, 3:993-1022, January 2003.
6. Седова А.Г., Митрофанова О.А. Тематическое моделирование русскоязычных текстов с опорой на леммы и лексические конструкции // Компьютерная лингвистика и вычислительные онтологии. 2017. В.1. С. 132-144.
7. А. Коршунов, А. Рамзин. Тематическое моделирование текстов на естественном языке // Труды Института системного программирования РАН. 2012. Т. 23. С. 215-242.
8. В. Вититинова. Продвижение в социальных сетях: основные KPI [Электронный ресурс] – URL: https://registratura.ru/blog/prodvizheniya_v_sotsialnykh_setyakh_osnovnye_kpi/, свободный. – Загл. с экрана. – Яз. рус.
9. Д. Румянцева. ER-методы расчета и значение показателей [Электронный ресурс] – URL: <https://www.cossa.ru/155/101492/>, свободный. – Загл. с экрана. – Яз. рус.
10. А. Марфицин. 7 терминов, которые должен знать каждый SMM-менеджер [Электронный ресурс] – URL: <http://www.advertology.ru/article141609.htm>, свободный. – Загл. с экрана. – Яз. рус.
11. Ю.Г. Рыков. Структура социальных связей в виртуальных сообществах: сравнительный анализ онлайн-групп социальной сети «ВКонтакте»: автореферат дис. ... кандидата социологических наук: 22.00.04 / Рыков Юрий Георгиевич; [Место защиты: Нац. исслед. ун-т "Высш. шк. экономики"]. – Москва, 2016. – 27 с.

Televnoy Andrey Dmitrievich

Saint-Petersburg national research university of information technologies, mechanics and optics, Saint-Petersburg, Russia
E-mail: adtelev@mail.ru

Khlopotov Maxim Valerievich

Saint-Petersburg national research university of information technologies, mechanics and optics, Saint-Petersburg, Russia
E-mail: khlopotov@corp.ifmo.ru

The study of thematic profiles and ways to calculate audience involvement in community social network «VKontakte»

Abstract. This article describes the study of thematic profiles in communities social network "VKontakte" flavored "University" by constructing a model of latent Dirichlet placement. The objects of research are online-community "VKontakte" social network. For participation in the study of online communities were selected on criteria for authors. In the list of online communities studied are presented: "University ITMO", "MGIMO University", "Kazan Federal University", "Siberian Federal University" and "Ural Federal University".

The study was conducted using a software package for collecting data from the social network "VKontakte", developed at the Chair of intellectual technologies in Humanities University of information technologies, mechanics and optics, as well as morphological analyzer for spending stemming collected shells of text data.

The authors presented results of a study of ways to calculate the coefficient of audience engagement on the basis and scope for several time periods (days 7 and 30 days). The authors discovered and formulated the relationship between ratios of involvement and the structure of the thematic one of the analysed profile online communities.

Results of a case simulation let you select the most discussed topics in online communities and define their policy publication administrators. The results of the study showed that engagement this indicator activity audience is not in direct proportion to the number of subscribers to online communities.

Also according to the results of the study should be noted the relationship between structured thematic profile and maximum ratios of involvement among the online community "Siberian Federal University". A preliminary study on the measure of involvement also can recommend marketers at carrying out of advertising campaigns (for selecting community with maximum involvement of the audience).

Keywords: analysis of social networks; online community; coefficient of involvement; case simulation; latent Dirichlet placement; metric of online communities; graphical probabilistic models; semantic relationships