

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

МАТЕРИАЛЫ
Международного экономического симпозиума
«Интеграция Сибири в глобальное
социально-экономическое пространство»

Под общей редакцией Е.В. Нехода, М.В. Чикова

Томск
Издательский Дом Томского государственного университета
2020

Г.Ю. Коварж, Е.В. Щекотин

БЛАГОПОЛУЧИЕ ЖИТЕЛЕЙ РЕГИОНОВ СИБИРИ: АНАЛИЗ ДАННЫХ СОЦИАЛЬНЫХ СЕТЕЙ

В статье рассматривается новый источник данных для изучения благополучия населения регионов Сибири – социальные сети. Предложена методология оценка благополучия поданным социальной сети ВК для 10 регионов Сибири по 19 показателям благополучия, рассмотрены преимущества данного метода и его ограничения. Рассчитан агрегированный показатель благополучия – индекс благополучия.

Ключевые слова: благополучие, социальные сети, регионы Сибири, измерение благополучия.

G.Y. Kovarz, E.V. Shchekotin

WELLBEING OF POPULATION OF SIBERIA REGIONS: ANALYSIS OF DATA OF SOCIAL NETWORKS

The article discusses a new source of data for studying the well-being of the population of Siberian regions - social networks. A methodology is proposed for assessing the well-being of a submitted VK social network for 10 regions of Siberia by 19 indicators of well-being, the advantages of this method and its limitations are considered. The aggregate indicator of well-being is calculated - the well-being index.

Keywords: well-being, social networks, regions of Siberia, measurement of well-being.

Научное исследование благополучия неразрывно связано с проблемой его измерения и оценки. В целом, сложившиеся на сегодняшний день теории благополучия подразделяются на две большие группы в зависимости от параметров измерения – объективные подходы и субъективные подходы [1]. Объективные подходы используют различные наборы статистических данных, субъективные подходы измеряют благополучие на основе самоотчетов, суждений людей о своей жизни, которые собираются посредством анкетирования или психологических и социологических экспериментов. В данном исследовании под благополучием мы будем понимать именно субъективную оценку индивидом собственного благополучия.

Развитие информационно-коммуникационных технологий открывает новые возможности для изучения благополучия. Так, например, социальные медиа стали сегодня одним из основных источников данных о настроениях, переживаниях и поведении людей. Социальные медиа как источник информации обладают как достоинствами, так и недостатками в сравнении с опросными методами. По нашему мнению, количественный анализ поведения людей в социальных сетях позволяет получить достаточно полную информацию об их эмоциях и переживаниях и оценить благополучие. В данной статье мы используем метод исследования благополучия, в основе которого лежит использование данных об онлайн активности пользователей в социальных сетях, для сравнительного изучения благополучия населения регионов Сибири.

Методология нашего исследования включала в себя последовательную реализацию следующих этапов:

- 1) разработка модели субъективного благополучия;
- 2) выбор сообществ в социальной сети ВК;
- 3) классификация сообщений и постов в сообществах;
- 4) автоматический анализ контента в отобранных сообществах;
- 5) построение индекса качества жизни для различных регионов России.

Мы выделили 19 индикаторов благополучия, перечень индикаторов представлен в таблице. В исследование были включены 10 субъектов Федерации РФ, входящих в Сибирский федеральный округ. В каждом регионе мы определили 3 самых крупных по численности населения населенных пунктов и для каждого из них выбрали 10 «региональных сообществ» в социальной сети ВК, самой популярной в России, в том числе в Сибири. Таким образом, в исследование были проанализированы 300 городских сообществ. Далее, используя специализированную платформу, разработанную Лабораторией наук о больших данных и проблемах общества Национального исследовательского Томского государственного университета (<https://lk.data.tsu.ru/>), были выгружены все сообщения в найденных сообществах за период с 01 января по 31 декабря 2018 года. Затем сообщения были очищены от «мусора», т.е. сообщений, не представляющих интереса для данного исследования и проведена разметка сообщений по темам и тональности. Посредством алгоритма машинного обучения был проведен анализ всех выгруженных сообщений и рассчитан агрегированный показатель благополучия в регионе – индекс благополучия.

Индекс благополучия для регионов Сибири (по 19 показателям благополучия и итоговый)

	Республика Алтай	Алтайский край	Иркутская область	Кемеровская область	Красноярский край	Новосибирская область	Омская область	Томская область	Республика Тыва	Республика Хакасия
Образование	0,000072	-0,00026	0,000034	-0,00029	-0,0002	-0,00061	-7,9E-05	-0,00011	0,000239	0,000083
ЖКХ	0,000004	-0,00449	-0,00066	-0,00203	-0,00151	-0,00132	-0,00208	-0,00275	-0,00047	-0,00118
Медицина	0,000082	-0,00084	-0,00004	-0,00079	-0,0003	-0,00183	-0,0004	-0,00153	-0,00014	-0,00044
Инфраструктура	0,000418	-0,01279	-0,00295	-0,00349	-0,00296	-0,00169	-0,0043	-0,00937	0,001329	-0,00083
Безопасность (обстановка в городе)	-0,00532	-0,14926	-0,05267	-0,08783	-0,04787	-0,09377	-0,06348	-0,09013	-0,00653	-0,05985
Экология	-3,7E-05	-0,00158	-0,00021	-0,00131	-0,00142	-0,0007	-0,00108	-0,00157	-0,00079	-0,00053
Отношения между людьми	0	-0,00573	-0,00139	-0,00281	-0,00103	-0,00083	-0,00187	-0,00434	-0,00266	-0,00067
Общее эмоциональное состояние	0	-0,00002	0	0	-0,00001	-3E-06	0	0	0	-2,9E-05
Работа	-6,4E-05	-0,00047	-0,00011	-0,00056	-0,00036	-0,00043	-0,00033	-0,00047	-7,9E-05	-0,00045
Магазины	0	-0,00266	-0,00028	-0,00067	-0,00064	-0,00056	-0,0002	-0,00371	-5,9E-05	-0,00026
Налоги	0	-0,00026	-8E-06	-0,00019	-2,7E-05	-0,00019	-0,00019	-5,1E-05	-8E-06	-5,2E-05
Кредитование и предпринимательство	-9E-06	-0,00034	-3,1E-05	-0,00016	-0,00009	-0,00013	0,000001	-0,00011	-1,5E-05	-8,5E-05
Социальная поддержка от государства	0,000107	-0,00053	0,000088	-0,00034	-0,00007	-4,5E-05	-0,0001	-0,00031	0,000283	-0,00025
Свобода СМИ	0	0,000003	0	0	0	0	0	0	0	0
Протестный потенциал	0,000001	-0,00128	-0,00031	-0,00096	-0,00058	-0,00206	-0,00046	-0,00133	0,000002	-0,0003
Свобода выборов	0	-3,6E-05	-0,00001	-0,00013	-0,0002	-0,0002	-6,3E-05	-0,00018	0,00002	-0,00036

	Республика Алтай	Алтайский край	Иркутская область	Кемеровская область	Красноярский край	Новосибирская область	Омская область	Томская область	Республика Тыва	Республика Хакасия
Отношение к власти	0	-0,00009	0,000027	-1,4E-05	-7E-06	-3,7E-05	-6E-06	-6,7E-05	-1,9E-05	-0,00011
Политические решения	0,000063	-0,00095	-0,00034	-0,00096	-0,00069	-0,00084	-0,00056	-0,00056	0,000251	-0,00055
Внутренняя политика	0,000051	-0,00088	-3,7E-05	3,64E-06	-0,00013	-0,0006	-0,00012	-0,00057	0,000118	-0,00054
Итоговый ИБ	-0,00463	-0,18246	-0,0589	-0,10255	-0,05809	-0,10585	-0,07532	-0,11716	-0,00853	-0,0664

Индекс благополучия рассчитывался по формуле:

$$I_{kjt} = A_{kjt}/B_k, \quad (1),$$

где I_{kjt} – это величина Индекса онлайн активности (ИОА) для региона (k) по данному показателю благополучия (j) для определенной тональности сообщения. Индекс онлайн-активности показывает актуальность данной темы сообщения (показателя благополучия) для региона и субъективные оценки пользователями социальной сети состояния данного показателя.

A_{kjt} – показатель онлайн активности в данном регионе по определенному показателю субъективного благополучия, рассчитывается по формуле:

$$A_{kjt} = L+2\times C+5\times R, \quad (2),$$

где L – сумма лайков, которую собрали сообщения, относящиеся к определенному показателю субъективного благополучия, в данном регионе в определенной тональности; C – количество комментариев, которых собрали сообщения, относящиеся к определенному показателю субъективного благополучия, в данном регионе. Мы приравнивали каждый комментарий к 2 лайкам, так как, по нашему мнению, это активное действие пользователя свидетельствует об особой актуальности тематики сообщений для комментатора. В этом случае лайк – это пассивная форма выражения поддержки данного сообщения; R – количество репостов сообщений, относящихся к определенному показателю субъективного благополучия, в данном регионе. Мы приравнивали репост к 5 лайкам, так как по нашему мнению, репост свидетельствует о полной и активной поддержке пользователем данного сообщения. Такое действие означает, что пользователь не просто выражает свое согласие с данным сообщением, но и публично демонстрирует свою солидарность с сообщением. В сравнении с различными формами онлайн активности, репост свидетельствует о наибольшей актуальности темы данного сообщения для пользователя.

B_k – общее количество подписчиков в всех выбранных сообщениях региона. Это показатель позволяет нам оценить относительную величину онлайн-активности для данного региона:

k – количество регионов (в нашем случае 10).

j – тематика сообщений, т.е. индикатор благополучия, который мы включили в модель качества жизни (от 1 до 19) (согласно таблице).

t – тональность сообщения (нейтральная, позитивная или негативная).

Тональность определяется как эмоциональная оценка сообщения – соответственно, позитивная тональность означает, что в сообщении присутствует положительная оценка или выражение одобрения, в отношении той новости или ситуации, о которой идет речь в сообщении, негативная тональность содержит неприятие, возмущение, относительно содержания сообщения, нейтральная тональность означает, что сообщение носит информационный, безоценочный характер.

Мы вычислили значение индекса по каждой тематике и каждой тональности ежемесячно, затем рассчитали среднемесячное значение по каждой тематике и по каждой тональности. Среднемесячные значения рассчитывались путем сложения величины индекса онлайн активности по определенной тональности и деления полученной суммы на 12. Далее, путем простого вычитания среднемесячных значений индекса онлайн активности для позитивной и негативной тональности, мы определили среднюю величину индекса благополучия $ИБ_{kj}$ для данной данного региона по данной тематике/показателю качества жизни:

$$ИБ_{kj\Delta cp} = I_{kj1cp} - I_{kj2cp}, \quad (3)$$

где I_{kj1cp} – индекс онлайн активности в позитивной тональности среднемесячный; I_{kj2cp} – индекс онлайн активности в негативной тональности среднемесячный.

Сообщения нейтральной тональности мы не включали в индекс благополучия. Итоговое общероссийское значение индекса благополучия $ИБ_{итог}$ мы рассчитывали, как сумму индекса по всем тематикам/показателям благополучия для каждого региона:

$$ИБ_{итог} = \sum ИБ_{kj\Delta cp}. \quad (4)$$

Предложенный метод изучения качества жизни хорошо подходит для исследования «цифрового» населения региона, т.е. той части населения региона, которая активно пользуется социальными медиа. К числу преимуществ предлагаемого метода перед традиционными опросными методиками измерения благополучия можно отнести быстроту, относительную дешевизну и масштабность исследования. Дальнейшее развитие метода оценки благополучия на основе данных социальных медиа требует согласования полученных результатов с данными опросных методик, а также выявления объективных факторов, которые оказывают влияние на онлайн оценку качества жизни.

Список использованной литературы

1. Нугаев Р.М., Нугаев М.А. Качество жизни в трудах социологов США // СОЦИС. 2003. № 6. С. 100–105.

Информация об авторах:

Коварж Галина Юрьевна – аналитик Лаборатории наук о больших данных и проблемах общества, Томский государственный университет, Томск; e-mail: kovarzhgalina@gmail.com

Щекотин Евгений Викторович – кандидат философских наук, доцент, доцент кафедры Социологии, Новосибирский государственный университет экономики и управления, Новосибирск; научный сотрудник Лаборатории наук о больших данных и проблемах общества, Томский государственный университет, Томск; e-mail: evgvik1978@mail.ru

Authors:

Kovarzh Galina Yurievna – Analyst, Laboratory of Big Data in Social Sciences, Tomsk State University, Tomsk; e-mail: kovarzhgalina@gmail.com

Shchekotin Evgeniy Viktirovich – candidate of philosophy, docent, Department of Sociology, Novosibirsk State University of Economics, and Management, Novosibirsk; research fellow, Laboratory of Big Data in Social Sciences, Tomsk State University, Tomsk; e-mail: evgvik1978@mail.ru