

## Измерение субъективного благополучия на основе текстов социальных медиа: обзор современных практик

Анна В. Кученкова

*Институт социологии Федерального научно-исследовательского  
социологического центра РАН, Москва, Россия,  
a.v.kuchenkova@gmail.com*

*Аннотация.* Наряду с многочисленными исследованиями субъективного благополучия посредством традиционных социологических методов (в первую очередь, опросов) предпринимаются попытки привлечь в качестве дополнительного источника информации Большие данные, «цифровые следы» (тексты социальных медиа, информацию профилей социальных сетей, статистику поисковых запросов, данные персональных электронных устройств).

На основе обзора зарубежной литературы раскрываются основные практики анализа текстов социальных медиа для измерения субъективного благополучия. В том числе охарактеризован опыт конструирования индекса «Валового национального счастья» для Фейсбука и «Гедонометра» для Твиттера на основе анализа объемов эмотивной лексики и тональности публикаций пользователей сетей. Раскрываются возможности поиска «цифровых следов» удовлетворенности жизнью в текстах социальных медиа.

Выделяются методологические трудности и ограничения в этой области исследований, которые пока не удалось преодолеть: проблема генерализации полученных выводов и валидности конструируемых индексов «счастья» (их соотношения с «реальным» субъективным благополучием, измеряемым посредством опросов). Трудности по измерению субъективного благополучия сопряжены с эффектами самопрезентации в социальных медиа, разной степенью и стратегиями публикационной активности пользователей, несовершенством используемых алгоритмов анализа, пока существенно уступающим «ручному кодированию».

*Ключевые слова:* субъективное благополучие, цифровые следы, цифровые данные, Большие данные

*Для цитирования:* Кученкова А.В. Измерение субъективного благополучия на основе текстов социальных медиа: обзор современных практик // Вестник РГГУ. Серия «Философия. Социология. Искусствоведение». 2020. № 4. С. 92–101. DOI: 10.28995/2073-6401-2020-4-92-101

## Measuring subjective well-being based on social media texts. Overview of modern practices

Anna V. Kuchenkova

*Institute of Sociology of Federal Center*

*of Theoretical and Applied Sociology of the Russian Academy of Science,  
Moscow, Russia, a.v.kuchenkova@gmail.com*

*Abstract.* Along with numerous studies of subjective well-being through sociological methods (first of all, surveys), attempts are being made to use Big Data, “digital footprints” (social media texts, social network profile information, search query statistics, personal electronic device data) as an additional source of information.

Based on a review of foreign literature, the author reveals major practices of the social media texts analysis to measure subjective well-being. Including the experience of constructing the Gross National Happiness index for Facebook and the Hedonometer for Twitter based on the analysis of emotive vocabulary and the tone of publications of network users. Possibilities of searching for “digital traces” of life satisfaction in the social media texts are revealed.

Methodological difficulties and limitations in that area of research, which have not yet been overcome, are highlighted: the issue of the obtained conclusions generalization and the validity of the constructed indices of “happiness” in their correlation with the “real” subjective well-being measured through surveys. Difficulties in measuring subjective well-being are associated with the effects of self-presentation in social media, varying degrees and strategies of users’ publication activity, imperfection of the analysis algorithms that are still inferior to “manual coding”.

*Keywords:* subjective well-being, digital footprints, digital data, big data

*For citation:* Kuchenkova, A.V. (2020), “Measuring subjective well-being based on social media texts. Overview of modern practices”, *RSUH/RGGU Bulletin. “Philosophy. Sociology. Art Studies” Series*, no. 4, pp. 92–101, DOI: 10.28995/2073-6401-2020-4-92-101

Субъективное благополучие – один из ключевых показателей качества жизни, отражающий интегральную оценку человеком собственной жизни, восприятие условий повседневной жизнедеятельности, достаточности/нехватки тех или иных благ, текущего положения в обществе и перспектив. Обращение исследователей именно к субъективным оценкам благополучия было обусловлено не только трудностями по формированию универсального набора его объек-

тивных характеристик, но и их разной значимостью на групповом уровне (в контексте культурных особенностей) и индивидуальном (в зависимости от этапа жизненного цикла, социально-демографического статуса). Традиционным методом изучения субъективного благополучия является анкетный опрос, разработаны разнообразные инструменты измерения (обзор см. [Кученкова 2016]).

Развитие современных информационных технологий, расширяющих методический инструментарий исследователей, являются одним из основных трендов, определяющих вектор развития исследовательской индустрии [Цыбикова 2019]. Примерами таких изменений являются онлайн-исследования, обращение к Большим данным как источнику социальной информации, привлечение «нереактивных» методов сбора данных, поиск «цифровых следов» социальных аспектов жизнедеятельности. Подобные методические новации апробируются на разнообразных предметных полях.

В исследованиях субъективного благополучия (СБ) с привлечением Больших данных можно выделить несколько направлений [Luhmann 2017; Кученкова 2019]: анализ текстов социальных медиа для реконструирования и измерения уровня СБ пользователей; изучение активности и действий в социальных сетях как предикторов СБ; измерение уровня СБ в обществе на основе статистики поисковых запросов в веб-браузерах; поиск детерминант СБ среди практик использования смартфонов; мониторинг уровня СБ индивида на регулярной основе посредством «трекеров счастья» (приложений для смартфонов, позволяющих проводить многократные замеры СБ по стандартным методикам).

Наиболее активно из всего вышеперечисленного развивается анализ постов и статусов в социальных сетях для измерения уровня СБ пользователей сетей безопросными, неактивными способами. Данная работа посвящена обзору основных практик использования подобного рода «цифровых следов» в исследованиях СБ.

### *Тексты социальных медиа как источник информации о субъективном благополучии*

В рамках данного направления исследуются тексты, продуцируемые пользователями социальных сетей (в первую очередь, статусы и посты в Фейсбуке, посты в Твиттере) с точки зрения эмоциональной окраски, тональности используемой лексики, реже – содержания и затрагиваемых тем.

На основе текстовой информации конструируются индексы, отражающие доминирование слов, характеризующих положитель-

ное или отрицательное эмоциональное состояние (эмоциональная, эмотивная лексика). С технической точки зрения подобный анализ реализуется, как правило, двумя способами. Первый связан с использованием составленных лингвистами, «закрытых» словарей эмоциональной лексики. В одном из широко используемых корпусов LIWC (Linguistic Inquiry and Word Count) 407 слов для положительных эмоций (счастье, хороший, легкий, доверие, надежда и др.) и 506 – для отрицательных (ненависть, враг, бояться, плач, горе, несчастье и др.). В постах (и/или статусах) социальных сетей за определенный период времени отыскиваются эти слова и подсчитывается количество их упоминаний. В исследовании на основе данных Фейсбука [Kramer 2010] использовались обновления статусов примерно 100 млн англоязычных пользователей за 2007–2009 гг., для каждого нового статуса вычислялось соотношение слов, выражающих «положительные» и «отрицательные» эмоции. На основе усреднения полученных значений по дням вычислялся индекс «Валового национального счастья» (Gross National Happiness Index). В динамике его значения повышались в праздничные дни (Рождество, День благодарения, Хэллоуин и др.) и выходные, что расценивалось авторами как свидетельство «очевидной» валидности индекса как показателя «всеобщего эмоционального здоровья/благополучия» пользователей социальной сети.

Другой вариант анализа тональности текстов социальных медиа связан с оценкой эмоционального заряда наиболее распространенных и часто встречающихся слов. На этом принципе основан «Гедонометр» Твиттера [Dodds et al. 2011]. Для его конструирования был составлен список из 10 222 наиболее часто используемых в последние десятилетия слов (в англоязычной литературе, песнях, крупнейших газетах). Для каждого из этих слов был оценен его эмоциональный заряд по 9-балльной шкале (усредненная оценка ответов 50 человек на вопрос «Какие эмоции, чувства вызывает у Вас слово?», для этой части исследования на платной основе привлекались пользователи интернет-платформы Amazon's Mechanical Turk). Для постов в Твиттере за определенный период времени (день, час, месяц) вычислялся «средний» уровень эмоционального заряда лексики. На основе анализа твиттов 63 млн пользователей за 2008–2011 гг. (46 млрд слов) были обнаружены циклические закономерности в течение дня и недели (повышение уровня «счастья» в выходные и резкое снижение в понедельник, ежедневное повышение к 20–24 часам и снижение в 5–7 час. утра).

Кроме исследований, основанных на измерении тональности текстов социальных медиа и отражающих в большей степени *аффективную* составляющую субъективного благополучия, предпринимаются

попытки отыскать «цифровые следы» его *когнитивной компоненты* – удовлетворенности жизнью. Например, посредством онлайн-опроса измеряют уровень СБ и анализируют тексты постов и статусов опрашиваемых пользователей за предшествующие полгода. С помощью тематического моделирования (методом Латентного размещения Дирихле – Latent Dirichlet Allocation) в этих текстах ищутся тематические кластеры слов («темы»), употребление которых коррелирует с уровнем удовлетворенности жизнью, измеренном посредством опроса. На индивидуальном уровне [Schwartz et al. 2016] удовлетворенность жизнью чаще сопровождается употреблением слов, характеризующих эмоциональную вовлеченность в повседневную жизнь (потрясающе, супер, завтра и др.), близкие и теплые отношения (семья, друзья, любовь и др.), профессиональное развитие (навыки, обучение, управление и др.), социальные взаимодействия (встреча, конференция, посещать, группа и др.); а неудовлетворенность – ненормативной лексикой, отчуждением (скука, зевать, неинтересный и др.).

В другом исследовании [Yang, Srinivasan 2016] был сформирован перечень выражений, фраз о (не)удовлетворенности жизнью (синонимичных суждениям, входящим в состав Шкалы удовлетворенности жизнью SWLSЭ. Динера<sup>1</sup>), которые отыскивались в текстах 3 млрд твиттов за 2012–2014 гг. Затем были выделены две группы пользователей, в чьих публикациях содержались фразы об удовлетворенности и неудовлетворенности жизнью. Сравнение этих групп показало, что они различаются по активности, тематике публикаций, употребляемой лексике. В содержательном плане последние из упомянутых исследований представляют особый интерес, поскольку реализуют попытку поиска показателей и индикаторов именно когнитивного аспекта СБ в отличие от исследований на основе анализа тональности текстов.

### *Методические ограничения использования текстов социальных медиа для измерения субъективного благополучия*

В онлайн-исследованиях на основе «цифровых следов» помимо угроз валидности статистического вывода (проблема генерализации полученных выводов, отсутствие четких представлений о гене-

---

<sup>1</sup>Одна из самых распространенных англоязычных методик для измерения субъективного благополучия, предложена Э. Динером, представляет собой шкалу суммарных оценок, включает пять суждений об удовлетворенности жизнью.

ральной совокупности, разный уровень проникновения интернета и др. [Девятко 2016]) существуют специфические угрозы валидности и надежности измерения. В том числе эффекты самопрезентации в социальных медиа, разная степень и стратегии публикационной активности участников (большая часть которых выступает в роли «наблюдателей» и продуцирует недостаточные для анализа объемы текстов), несовершенство используемых алгоритмов анализа, пока существенно уступающих «ручному кодированию».

В случае с субъективным благополучием подбор индикаторов («цифровых следов») для когнитивной составляющей (удовлетворенности жизнью) является нетривиальной задачей. Преобладают исследования эмоциональных проявлений в текстах социальных медиа (адаптирующие технологии и методы корпусной лингвистики к конкретному предметному полю). Дискуссионным остается вопрос: как соотносятся данные самоотчетов о субъективном благополучии и «цифровых следов» с точки зрения получаемых результатов и качества измерения, в какой мере «реальное» субъективное благополучие людей отражается в тональности постов и объемах эмотивной лексики?

Для проверки и обоснования валидности «индексов счастья», конструируемых на основе текстов социальных медиа, неоднократно предпринимались попытки оценить корреляцию их значений на индивидуальном уровне с данными, полученными посредством реактивных методов (опроса) и традиционных методик (в частности, Шкалы субъективного благополучия (SWLS Э. Динера)). Процедурно это реализовывалось двумя способами: либо рекрутировались добровольцы, проходившие онлайн-опрос и предоставляющие доступ к своим постам за год-два, предшествовавших опросу; либо использовались данные приложения Фейсбук “my Personality” (функционировавшего с 2007 по 2012 г. и содержащего ответы пользователей на вопросы различных психологических тестов, среди прочего по Шкале субъективного благополучия Э. Динера) для тех, чей доступ к персональной странице и постам был открыт.

Установлено, что объемы эмотивной лексики слабо коррелируют с индивидуальными баллами по Шкале субъективного благополучия Э. Динера. Коэффициент корреляции составляет 0,17 для интенсивности выражения положительных эмоций [Kramer 2010],  $-0,105-0,145$  – для отрицательных [Liu 2015]. Чуть более тесная взаимосвязь обнаруживается в случае, если кроме тональности текстов учитывается их тематика, частота упоминания тех или иных тем (коэффициент корреляции со шкалой SWLS 0,36 [Chen et al. 2017]), либо если данные рассматриваются в динамике на агреги-

рованном, групповом уровне в разрезе недель (корреляция достигает 0,37 [Wang et al. 2014]). Однако при оценке корреляции между объемами эмотивной лексики и баллами по шкале удовлетворенности жизнью с содержательной точки зрения сопоставляются *эмоциональная* и *когнитивная* компоненты субъективного благополучия, что представляется не вполне корректным.

Для предсказания удовлетворенности жизнью на индивидуальном уровне лучшего результата удастся добиться при использовании методов тематического моделирования. Если в модели учитывается и тематика текстов, и отдельная лексика, то корреляция между предсказываемым уровнем СБ и полученным в ходе опроса может достигать 0,566 [Schwartz et al. 2016].

В целом невысокие корреляции между индикаторами СБ в текстах социальных сетей и данными самоотчетов респондентов могут быть связаны с *временным лагом* между ответами на вопросы анкеты и постами / обновлением статуса (который может доходить до одного [Liu 2015] или даже трех лет [Kramer 2010; Chen et al. 2017]). Субъективное благополучие может изменяться в течение этого периода времени. Кроме того, сам факт использования и продолжительность нахождения в социальных сетях могут существенным образом снижать СБ человека.

Проведенный обзор исследований свидетельствует, что выражение эмоций в текстах социальных сетей не позволяет надежно, хорошо измерять и предсказывать «реальное» субъективное благополучие человека (замеряемое посредством анкетных вопросов). Вместе с тем поиск «цифровых следов» его когнитивной составляющей представляется перспективным с точки зрения развития «нереактивных» подходов к изучению субъективного благополучия.

## Литература

- Девятко 2016 – Девятко И.Ф. От «виртуальной лаборатории» до «социального телескопа»: метафоры тематических и методологических инноваций в онлайн-исследованиях // Онлайн-исследования в России: тенденции и перспективы / Под ред. А.В. Шашкина, И.Ф. Девятко, С.Г. Давыдова. М.: Изд-во МИК, 2016. С. 19–34.
- Кученкова 2016 – Кученкова А.В. Социальное самочувствие и субъективное благополучие: соотношение понятий и способов измерения // Вестник РГГУ. Серия «Философия. Социология. Искусствоведение». 2016. № 2 (4). С. 118–127.
- Кученкова 2019 – Кученкова А.В. BigData как источник информации для изучения субъективного благополучия: возможности и ограничения // Будущее социо-

- логического знания и вызовы социальных трансформаций (к 90-летию со дня рождения В.А. Ядова): Сборник материалов. М., 2019. С. 332–336.
- Цыбикова 2019 – *Цыбикова Д.Г.* Глобальные тренды исследовательской индустрии и требования к специалистам // Вестник РГГУ. Серия «Философия. Социология. Искусствоведение». 2019. № 1. С. 130–140. DOI: 10.28995/2073-6401-2019-1-130-140
- Chen et al. 2017 – *Chen L., Gong T., Kosinski M., Stillwell D., Davidson R.L.* Building a profile of subjective well-being for social media users // PLoS ONE. 2017. No 12 (11): e0187278 [Электронный ресурс]. URL: <https://doi.org/10.1371/journal.pone.0187278> (дата обращения 14.08.2020).
- Dodds et al. 2011 – *Dodds P.S., Harris K.D., Kloumann I.M., Bliss C.A., Danforth C.M.* Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter // PLoS ONE. 2011. No. 6 (12): e26752 [Электронный ресурс]. URL: <https://doi.org/10.1371/journal.pone.0026752> (дата обращения 14.08.2020).
- Kramer 2010 – *Kramer A.D.I.* An unobtrusive behavioral model of gross national happiness // Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. April, 2010. Atlanta, Georgia: ACM, 2010. P. 287–290.
- Liu et al. 2015 – *Liu P., Tov W., Kosinski M., Stillwell D.J., Qui L.* Do Facebook status updates reflect subjective well-being? // Cyberpsychological Behavior Social Network – 2015. 2015. Vol. 18. No. 7. P. 373–379.
- Luhmann 2017 – *Luhmann M.* Using big data to study subjective well-being // Current Opinion in Behavioral Sciences. 2017. No. 18. P. 28–33. DOI: 10.1016/j.cobeha.2017.07.006
- Yang, Srinivasan 2016 – *Yang C., Srinivasan P.* Life Satisfaction and the Pursuit of Happiness on Twitter // PLoS ONE. 2016.11(3): e0150881 [Электронный ресурс]. URL: <https://doi.org/10.1371/journal.pone.0150881> (дата обращения 14.08.2020).
- Wang et al. 2014 – *Wang N., Kosinski M., Stillwell D.J., Rust J.* Can well-being be measured using Facebook status updates? Validation of Facebook's gross national happiness index // Social Indicator Research. 2014. No. 115. P. 483–491.
- Schwartz et al. 2016 – *Schwartz H.A., Sap M., Kern M. L., Eichstaedt J.C., Kapelner A., Agrawal M., Blanco E., Dziurzynski L., Park G., Stillwell D., Kosinski M., Seligman M.E., Ungar L.H.* Predicting Individual Well-Being Through the Language of Social Media // Pacific Symposium on Biocomputing / Ed. by R.B. Altman, A.K. Dunker, L. Hunter, M.D. Ritchie, T.A. Murray, T.E. Klein. 2016. P. 516–527 [Электронный ресурс]. URL: [https://doi.org/10.1142/9789814749411\\_0047](https://doi.org/10.1142/9789814749411_0047) (дата обращения 14.08.2020).

## References

---

- Chen, L., Gong, T., Kosinski, M., Stillwell, D. and Davidson, R.L. (2017) “Building a profile of subjective well-being for social media users”, *PLoS ONE*, no. 12 (11): e0187278, available at: <https://doi.org/10.1371/journal.pone.0187278> (Accessed 14 Aug. 2020).

- Devyatko, I.F. (2016), “From “virtual laboratory” to “social telescope”. Metaphors of thematic and methodological innovations in online research”, in Shashkin, A.V., Devyatko, I.F. and Davydov, S.G. (eds.), *Onlain issledovaniya v Rossii: tendentsii i perspektivy* [Online Research in Russia: Trends and Prospects], Izdatel'stvo MIK, Moscow, Russia, pp. 19–34.
- Dodds, P.S., Harris, K.D., Kloumann, I.M., Bliss, C.A. and Danforth, C.M. (2011), “Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter”, *PLoS ONE*, no. 6 (12): e26752, available at: <https://doi.org/10.1371/journal.pone.0026752> (Accessed 14 Aug. 2020).
- Kramer, A.D.I. (2010), “An unobtrusive behavioral model of gross national happiness”, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Atlanta, Georgia: ACM, pp. 287–290.
- Kuchenkova, A.V. (2016), “Social self-perception and subjective well-being. A review of definitions and the measurement models”, *RSUH/RGGU Bulletin. “Philosophy. Sociology. Art Studies” Series*, vol. 4, no. 2, pp. 118–140.
- Kuchenkova, A.V. (2019), “Big Data as a Source of Information for Studying Subjective Well-Being: Opportunities and Limitations”, *“Budushcheye sotsiologicheskogo znaniya I vyzovy sotsial'nykh transformatsiy (k 90-letiyu so dnya rozhdeniya V.A. Yadova)”* [The Future of Sociological Knowledge and the Challenges of Social Transformations (Towards the 90th Anniversary of V.A. Yadov). Collection of papers], Moscow, Russia, pp. 332–336.
- Liu, P., Tov, W., Kosinski, M., Stillwell, D.J. and Qui, L. (2015), “Do Facebook status updates reflect subjective well-being?”, *Cyberpsychological Behavior Social Network – 2015*, vol. 18, no. 7, pp. 373–379.
- Luhmann, M. (2017) “Using big data to study subjective well-being”, *Current Opinion in Behavioral Sciences*, no. 18, pp. 28–33, DOI: 10.1016/j.cobeha.2017.07.006
- Tsybikova, D.G. (2019), “The research industry global trends and the requirements for specialists”, *RSUH/RGGU Bulletin. “Philosophy. Sociology. Art Studies” Series*, no. 1, pp. 130–140, DOI: 10.28995/2073-6401-2019-1-130-140
- Wang, N., Kosinski, M., Stillwell, D.J. and Rust, J. (2014), “Can well-being be measured using Facebook status updates? Validation of Facebook’s gross national happiness index”, *Social Indicator Research*, vol. 115, pp. 483–491.
- Yang, C. and Srinivasan, P. (2016), “Life Satisfaction and the Pursuit of Happiness on Twitter”, *PLoS ONE*, no. 11 (3): e0150881, available at: <https://doi.org/10.1371/journal.pone.0150881> (Accessed 14 Aug. 2020).
- Schwartz, H.A., Sap, M., Kern, M.L., Eichstaedt, J.C., Kapelner, A., Agrawal, M., Blanco, E., Dziurzynski, L., Park, G., Stillwell, D., Kosinski, M., Seligman, M.E. and Ungar, L.H. (2016), “Predicting Individual Well-Being Through the Language of Social Media”, in Altman, R.B., Dunker, A.K., Hunter, L., Ritchie, M.D., Murray, T.A. and Klein, T.E. (eds.), *“Pacific Symposium on Biocomputing”*, pp. 516–527, available at: [https://doi.org/10.1142/9789814749411\\_0047](https://doi.org/10.1142/9789814749411_0047) (Accessed 14.08.2020).

*Информация об авторе*

*Анна В. Кученкова*, кандидат социологических наук, доцент, Федеральный научно-исследовательский социологический центр РАН, Институт социологии, Москва, Россия; 117218, Россия, Москва, ул. Кржижановского, д. 25/35, корп. 5; a.v.kuchenkova@gmail.com

*Information about the author*

*Anna V. Kuchenkova*, Cand. of Sci. (Sociology), associate professor, Federal Center of Theoretical and Applied Sociology of the Russian Academy of Science, Institute of Sociology, Moscow, Russia; bldg. 5, bld. 24/35, Krzhizhanovskii Str., Moscow, Russia, 117218; a.v.kuchenkova@gmail.com