Проблема низкого количества откликов в опросах не нова для социологии [Назарова, 1998]. Появление онлайн-опросов, использование которых с каждым годом набирает популярность, не позволяет решить эту проблему: онлайн опросы зачастую имеет долю откликов ниже, чем другие виды опросов [Dillman, 2009; Manfreda, 2008]. Становится очевидна необходимость получения социологической информации новыми способами. Одним из таких способов может оказаться извлечение данных из сети интернет или использование информации, собранной крупными компаниями (BigData). Одним из методов извлечения данных из интернета выступает веб-скрапинг. Однако социологи, которые ранее считались новаторами в разработке и совершенствовании методов сбора данных, сейчас используют инновационные методы довольно ограниченно [Farrell, Petersen, 2010]. Возникает закономерный вопрос: почему в социологической науке нереактивные методы извлечения информации из сети интернет не получили широкого распространения? Очевидно, что существуют ограничения таких методов, которые не позволяют социологам широко использовать эти методы в своей практике.

Так, предыдущими исследованиями метода веб-скрапинга в социологии были выделены следующие ограничения: отсутствие подходящего программного обеспечения, неадаптированость методов, а также недостаток у социологов практических навыков по работе с большими объёмами данных [Tinati et al., 2014; Golder, Macy, 2014]. Также важным ограничением признаются вопросы исследовательской этики [Zimmer, 2010; Lazer, Radford 2017]. С точки зрения методологии социологического исследования проблема формирования репрезентативной выборки тоже накладывает ограничения на возможность использования методов извлечения данных из интернета [Grimmer, 2015]. Ещё одним ограничением выступает социальная сконструированность данных, демонстративная природа поведения в сети Интернет, интернет-данные, таким образом, подвержены эффектам социальной желательности [Девятко, 2012].

Однако веб-скрапинг обладает уникальными возможностями, которые могут решить современные проблемы сбора социологической информации. Так, веб-скрапинг позволяет собирать информацию практически о всей генеральной совокупности исследования без участия респондентов в процессе сбора данных. Автоматизированный сбор данных без участия респондента в свою очередь позволяет избежать проблемы низкого отклика. Важным является также то, что веб-скрапинг это один из самых простых методов, позволяющих собирать информацию из сети Интернет. Поэтому в качестве метода извлечения онлайн-данных веб-скрапинг имеет особый интерес для социологов, так как он делает эмпирически возможными извлечение наколенной в интернете информации [Marres, Weltevrede, 2013].

Целью данной работы стало более широкое определение возможностей и ограничений веб-скрапинга как нереактивного метода сбора данных в социологическом исследовании. Возможности и ограничения веб-скрапинга рассматриваются на работе с числовыми и текстовыми данными.

В качестве эмпирического примера по работе с числовыми данными, излеченными из сети Интернет, рассматривается поиск связей между социально обусловленными характеристиками кинофильма как товара и его популярностью у зрителя на данных, собранных с сайта imdb.com (N=36679). Данный эмпирический пример работает в теоретической рамке социологии потребления. В результате проведённого исследования с помощью построения регрессионной модели (R^2=0.749) удалось выявить значимые связи между социально обусловленными характеристиками кинофильма и его популярностью.

Другой эмпирический пример, позволяющий продемонстрировать возможности и ограничения веб-скрапинга, но уже на текстовых данных, предполагает поиск тем, которые возможно выделить в корпусе текстов российских рэп и хип-хоп исполнителей. В качестве данных используются тексты песен, собранные открытого ресурса https://рэп-текст.рф (N=10196). Используемый в работе метод анализа текстовой информации – тематическое моделирование с помощью библиотеки BigARTM, которая даёт большие возможности анализа, используя технологию аддитивной регуляризации тематических моделей. В качестве теоретической рамки здесь принимается рамка гендерной социологии и исследования маскулинности в частности. Так как большинство членов рэп сообщества являются представителями мужского пола, то феномен репрезентации маскулинности представляет здесь интерес. Более того, исследования показывают, что рэп тексты зачастую становятся инструментом демонстрации гегемонной маскулинности [Mohammed-baksh, 2015; Weitzer, 2009; Herd, 2015].

Рассмотрение использования метода веб-скрапинга на эмпирическом примере помогло выделить следующие ранее не обозначенные в литературе ограничения метода при работе с числовыми данными: проблема пропущенных значений и ограничение операционализации теоретических концептов исследования. При работе с текстовыми данными необходимо уделять особое внимание лингвистическим свойствам языка, а также первичной обработке текстов.

В качестве уникальных возможностей веб-скрапинга рассматривается возможность собирать практически всю генеральную совокупность исследования, а также удобство сбора большого количество текстовой информации. Стоит также отметить качество математических моделей, построенных на данных, которые были собраны с помощью веб-скрапинга. Другим преимуществом веб-скрапинга выступает возможность сбора

информации в естественном виде, то есть в том, в котором эта информация была произведена изучаемыми субъектами.

Литература:

1. Девятко И. Ф. Инструментарий онлайн-исследований: попытка каталогизации //Онлайн исследования в России 3.0. – 2012. – С. 17-30.

2. Назарова И. Б. Непроведение опроса и отказ от интервью //Социологический журнал. – 1998. – №. 1-2. – С. С. 161-167.

3. Connell, R. W., & Messerschmidt, J. W. (2005). Hegemonic masculinity: Rethinking the concept. Gender & society, 19(6), 829-859.

4. Dillman D. A. et al. Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet //Social science research. – 2009. – Т. 38. – №. 1. – С. 1-18.

5. Farrell D., Petersen J. C. The growth of internet research methods and the reluctant sociologist //Sociological Inquiry. – 2010. – Т. 80. – №. 1. – С. 114-125.

6. Golder S. A., Macy M. W. Digital footprints: Opportunities and challenges for online social research //Annual Review of Sociology. – 2014. – Т. 40.

7. Grimmer J. We are all social scientists now: how big data, machine learning, and causal inference work together //PS: Political Science & Politics. – 2015. – Т. 48. – №. 1. – С. 80-83.

8. Herd, Denise. "Conflicting paradigms on gender and sexuality in rap music: A systematic review." Sexuality & Culture 19.3 (2015): 577-589.

9. Lazer D., Radford J. Data ex machina: Introduction to big data //Annual Review of Sociology. – 2017. – Т. 43. – С. 19-39.

10. Manfreda K. L. et al. Web surveys versus other survey modes: A meta-analysis comparing response rates //International Journal of Market Research. – 2008. – Т. 50. – №. 1. – С. 79-104.

11. Marres N., Weltevrede E. Scraping the social? Issues in live social research //Journal of Cultural Economy. – 2013. – Т. 6. – №. 3. – С. 313-335.

12. Mohammed-baksh, Sufyan, and Coy Callison. "Hegemonic masculinity in hip-hop music? Difference in brand mention in rap music based on the rapper's gender." Journal of Promotion Management 21.3 (2015): 351-370.

13. Tinati R. et al. Big data: methodological challenges and approaches for sociological analysis //Sociology. – 2014. – Т. 48. – №. 4. – С. 663-681.

14. Weitzer, Ronald, and Charis E. Kubrin. "Misogyny in rap music: A content analysis of prevalence and meanings." Men and Masculinities 12.1 (2009): 3-29.

15. Zimmer M. “But the data is already public”: on the ethics of research in Facebook //Ethics and information technology. – 2010. – Т. 12. – №. 4. – С. 313-325.