

## **ЦИФРОВЫЕ СЛЕДЫ ИЛИ BIG DATA В ОПТИКЕ СОЦИОЛОГИИ**

Появление нового исследовательского направления – вычислительной социологии, позволяет использовать цифровые следы или Big Data для решения социологических задач. В настоящей публикации рассмотрим классификацию направлений анализа цифровых данных, а также обсудим несколько важных методологических и организационных аспектов работы с большими данными.

Ключевые слова. цифровые следы, цифровые данные, большие данные, вычислительная социология.

**I.E. Finkelshtein, Samara University**

## **DIGITAL FOOTPRINTS IN SOCIOLOGY OPTICS**

The emergence of a new research direction – computational sociology – allows the use of digital footprints or Big Data to solve sociological problems. This publication will be considered the classification of directions for the analysis of digital data and discussed several important methodological and organizational aspects of the work on big data.

Keywords. digitalfootprints, digital data, bigdata, computationsociology.

Цифровизация современной жизни открывает для социолога новые горизонты изучения человеческого поведения, которые можно рассматривать в трех аспектах: (1) цифровая реальность как новое исследовательское поле; (2) цифровые следы как новый источник данных для социологического анализа; (3) новые методы сбора, обработки и анализа данных. Эти три аспекта сформировали новое направление исследования – цифровая социология, или вычислительная социальная наука, в рамках которого анализ цифровых данных осуществляется посредством компьютерных технологий: различные технологии работы с большими данными (Big Data), методы машинного обучения, нейронные сети [1; 2]. Новое предметное поле, новые данные, новые методы ставят перед профессиональным сообществом много вопросов: как изучать цифровую реальность – традиционными для социологии методами сбора и анализа или методами вычислительной социологии; каковы перспективы использования новых для социологии методов анализа; какие

исследовательские вопросы можно решить, используя цифровые технологии анализа больших данных; может ли социолог сам овладеть методами вычислительной социологии.

В настоящей работе остановимся на проблеме использования новых технологий работы с большими данными для решения социологических задач.

Каждый современный человек «производит» огромное количество цифровых следов, общаясь в социальных сетях, используя поисковые системы для интернет-серфинга или различные цифровые гаджеты, осуществляя покупку в интернет магазине и т.д. Все эти разнообразные, с большой скоростью накапливающиеся и меняющиеся данные и называют большими (BigData), подразумевая их объем, измеряющийся в терабайтах. Еще одной важной характеристикой больших данных является их неструктурированность. Крайне редко исследователь работает с уже упорядоченными базами данных (dataset). Поэтому первой задачей при работе с цифровыми следами – это их сбор и упорядочивание, создание «опрятного вида» [3]. Для поиска и отбора релевантных данных используются программы-парсеры (например, Java Script, PHP). Для решения задач по обработке, анализу и визуализации данных в социологии чаще всего используют язык программирования Python и R.

В поисках ответа на вопрос, может ли социолог освоить языки программирования, сложилось две позиции. Первая позиция – желательно, чтобы специалисты-социологи самостоятельно разрабатывали все алгоритмы поиска, структурирования, обработки и анализа данных, поскольку они понимают методологические, содержательные аспекты исследования. Вторая позиция – такие исследования возможны в коллаборации со специалистами в области работы с большими данными. В такой ситуации социолог должен понимать, какие исследовательские задачи могут быть реализованы при анализе больших данных, хотя бы в общих чертах иметь представление о технологии работы с большими данными, чтобы разработать словарь слов и словосочетаний для поиска релевантных документов и максимально четко обозначить круг задач для специалиста по программированию. Примеры успешных исследовательских проектов в рамках вычислительной социологии есть и в коллективах, состоящих из социологов, и в коллаборациях социологов и программистов.

Методы вычислительной социологии позволяют осуществлять различные направления анализа при работе с цифровыми данными:

(1) анализ геолокации (может использоваться в городских исследованиях, при изучении мобильности);

(2) анализ интернет активности, запросов, просмотров, лайков (может использоваться в исследованиях интернет поведения, протестной активности, предпочтений);

(3) анализ взаимодействия между группами, сообществами, отдельными представителями групп (например, изучение виртуальных диаспор [4], онлайн сообществ пациентов);

(4) анализ высказываний в социальных сетях (может использоваться в исследованиях протестной активности, проявлений онлайн экстремизма);

(5) анализ СМИ.

К тому же технологии работы с большими данными позволяют осуществлять анализ по половозрастной структуре пользователей, анализ динамики изучаемых процессов, составлять рейтинг по территориальному признаку [5].

В перспективе мы будем наблюдать рост интереса социологов к большим данным, развитие методологических подходов к изучению социального поведения на основе цифровых данных. Связано это с ограниченностью традиционных социологических методов для решения некоторых исследовательских задач. Как раз с такой ситуацией столкнулась Елизавета Сивак и ее коллеги при изучении мобильности детей. Для решения задач исследования сначала был проведен традиционный анкетный опрос, затем опрос с использованием карт для обозначения локаций, следующим этапом – онлайн опрос с использованием цифровых карт. Но все эти традиционные опросы, даже с применением современных технологий оказались недостаточно эффективными. Тогда социологи обратились к западному опыту исследований с использованием языков программирования – через фиксацию контактов в социальных сетях, демографических данных и местоположения. Ценность такого подхода в том, что можно получать ретроспективные данные без риска ошибки. При использовании традиционных опросных методах исследователи постоянно сталкивались с проблемой фиксации ретроспективных данных – подростки не могли точно вспомнить все свои перемещения [6]. Проблема изучения сенситивных тем или социально неодобряемых позиций тоже может решаться отказом от использования опросных методов в пользу анализа больших данных или их одновременным использованием для последующего сопоставления результатов.

Безусловно, любой метод сбора и анализа имеет свои возможности и ограничения. Поэтому в некоторых случаях одновременное или последовательное использование разных методологических подходов и

методов позволит скорректировать интерпретацию, повысить валидность полученных результатов.

**Список литературы:**

1. Edelman A., Wolff T., Montagne D., Bail C.A. Computational social science and sociology // *Annual Review of Sociology*. 2020. Т.46. Р. 61-81.

2. Мещерякова Н.Н. Методология познания цифрового общества // *Цифровая социология*. 2020. Т.3. №2. С.17-26.

3. Смирнов В.А. Новые компетенции социолога в эпоху больших данных // *Мониторинг общественного мнения*. 2015. №2. С. 44-54.

4. Стариков В.С., Иванова А.А., Ни М.Л. Транснационализм в режиме Online: Как исследовать миграционные процессы на больших массивах данных? // *Мониторинг общественного мнения: Экономические и социальные перемены*. 2018. №5. С. 213-232.

5. Бродовская Е.В., Домбровская А.Ю., Карзубов Д.Н., Синяков А.В. Развитие методологии и методики интеллектуального поиска цифровых маркеров политических процессов в социальных медиа // *Мониторинг*. 2017. №5 (141). С. 79-104.

6. Ledford H. Computing humanity: How data from Facebook, Twitter and other sources are revolutionizing social science // *Nature*. 2020. Vol. 582. Р. 328-330.