

АНАЛИЗ ПРИМЕНЕНИЯ ВЕБ-СКРЕЙПИНГА ДЛЯ ИССЛЕДОВАНИЯ ДЕЛОВЫХ СЕТЕЙ

Колопенко Е. А.

Кафедра экономической информатики,

Белорусский государственный университет информатики и радиоэлектроники

Минск, Республика Беларусь

E-mail: e.kolopenko@bsuir.by

Обозначены тенденции, наблюдаемые в решении задач социально-экономического анализа и управления в статистической науке. Проведен анализ применения веб-скрейпинга для исследования деловых сетей, приведены его методологические, технические и правовые возможности и ограничения.

ВВЕДЕНИЕ

Большие данные, в том числе данные, собранные при помощи технологии веб-скрейпинга, существенно используются в современной социогуманитарной науке. Приведем следующие типы задач социально-экономического анализа и управления, в решении которых одним из методов сбора данных, наряду с традиционными методами, является их извлечение с веб-страниц:

1. Оценка стоимости наблюдаемых товаров или услуг. Примерами могут служить квартиры или компьютерная техника. Становление и развитие цифровой экономики, развитие инфраструктуры информационно-коммуникационных технологий стимулирует использование технологии веб-скрейпинга и в официальной статистике цен. Так, например, Национальный статистический комитет Республики Беларусь использует данные платформ электронной коммерции для регистрации цен на отдельные непродовольственные товары [1].
2. Исследование занятости в цифровой экономике. Веб-скрейпинг может выступать единственным источником данных, например, при выявлении факторов, оказывающих влияние на успешность фрилансера [2].
3. Анализ социальной напряженности и социальных патологий через контент-анализ – текстовый анализ новостей и публикаций.

Перечисленные варианты не исчерпывают всех возможных форм применения веб-скрейпинга в экономике и социологии. В условиях современного информационного общества технология веб-скрейпинга оказывается крайне востребованной при исследовании конкуренции и сотрудничества фирм в рамках более или менее устойчивой сети связей.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Традиционный экономический взгляд на рынок постулирует независимость хозяйственных агентов в принятии решений. Между тем, устойчивые деловые отношения, вызванные близостью в социальном пространстве и независимостью в финансовом и управленческом отношении, транс-

формируются в отношения, которые предполагают длительные обязательства и выработку доверия. Совокупность множественных переплетенных связей в рамках более или менее разветвленной группы независимых хозяйственных агентов называют деловыми сетями. Они основаны на актах материального и символического обмена и служат для распространения деловой информации, формирования репутации, оказания взаимопомощи на формальной или неформальной основе. Деловые сети характеризуются плотностью, степенью централизации и устойчивостью, силой, теснотой и степенью формализации составляющих их деловых связей.

Вовлеченность в деловые сети является частью социального капитала. Для случая низкой степени формализации деловых связей получение экономических выгод от социального капитала в данной его форме и на данном уровне достигается за счет селективных преимуществ на индивидуальной основе [3].

В настоящее время для исследования механизмов взаимного влияния конкуренции и структуры социальных связей распространены конвенциональные источники данных:

- данные официальной статистики (показатели рыночной конъюнктуры, благосостояния населения, качества социальной сферы);
- опросы менеджеров и углубленные интервью с руководителями предприятий;
- экспертно оцениваемые показатели.

В 2015–2022 гг. доля организаций, имевших веб-сайт, выросла на 11,9 п.п. до 71,6% в 2022 г. [4]. Многие организации представлены в таких социальных сетях, как Instagram, LinkedIn. Исходный (априорный) перечень признаков, так или иначе влияющих на характеристики деловых сетей, получение которых возможно с применением технологии веб-скрейпинга, описывает круг деловых партнеров (включенных или не включенных в производство сходных благ – конкурентов и контрагентов), длительность деловых отношений, а также их динамику. Текстовое описание, приводимое в публикациях, может передавать субъективную оценку описываемого объекта. Данные активности в социальных сетях позволяют получить

оценку успешности бизнеса и реакции бизнеса на эффекты среды. Наиболее наглядно это проявляется при изменениях, связанных с действиями других хозяйственных агентов по сохранению или изменению поля сил, или, в еще более явном виде, при изменениях в системе трудовых отношений, системе собственности и политической системе.

Лента новостей и публикаций на веб-сайтах и страницах в социальных сетях позволяют в комбинации с конвенциональными источниками данных провести отвечающий целям исследования анализ на основе подходящих методов эконометрики и машинного обучения. Это позволяет устранить недостатки, присущие обоим методам сбора данных, и обеспечить более полное описание деловых сетей. Объем данных, получаемых с использованием веб-скрейпинга, достаточно велик для того, чтобы можно было получить статистически надежные оценки неизвестных параметров экономико-математических моделей.

Часто веб-скрейпингу противопоставляют Application Programming Interface (API) как более быстрый и, в некоторых случаях, более правомерный способ получения структурированных данных. Наиболее полезными с точки зрения исследования деловых сетей являются API платформ социальных сетей, платформ электронной коммерции, финансовых платформ для получения данных о пользователях, их публикациях и взаимодействиях, данных о конкурентной среде, власти поставщиков, потребителей и государства и данных о финансовом состоянии компаний соответственно. Использованию API присущи определенные недостатки, которые вынуждают исследователей исходить не из теоретико-методологической базы и задач исследования, а из доступности API и конкретных показателей, им предоставляемых.

Применение технологии веб-скрейпинга сопровождается определенными ограничениями.

Становление и развитие онлайн-платформ, с одной стороны, стимулирует развитие альтернативных методов сбора данных и, с другой стороны, способствует совершенствованию методов защиты от веб-скрейпинга. Владельцы платформ могут устанавливать ограничения на объем видимых данных, количество запросов с одного IP-адреса за определенное время, скорость загрузки во время веб-скрейпинга для недопущения несанкционированного доступа, сколько-нибудь значительного ущерба работоспособности сайта и достижения иных целей, обусловленных спецификой конкретного веб-сайта (например, блокирова-

ние автоматизированного увеличения количества подписчиков и комментариев).

Особого внимания требует правовой аспект. В качестве основного закона, регламентирующего использование информации, выступает Закон Республики Беларусь от 10 ноября 2008 г. № 455-З «Об информации, информатизации и защите информации», а в области защиты персональных данных – Закон Республики Беларусь от 7 мая 2021 г. № 99-З «О защите персональных данных». Из установленных законодательством ограничений следует, что веб-скрейпинг правомерен для извлечения любых общедоступных данных, не защищенных авторским правом. Извлечение персональных данных пользователей социальных сетей осуществляется только с их согласия.

ЗАКЛЮЧЕНИЕ

Технология веб-скрейпинга может стать серьезным дополнением к традиционным методам сбора данных для исследования деловых сетей. Несмотря на методологические, технические и правовые ограничения ее применения, у статистиков, экономистов, социологов наблюдается возросший гносеологический интерес к понятию больших данных вообще и научному обоснованию веб-скрейпинга. Учитывая функции и возможности деловых сетей, многочисленность сил, их образующих, и субъектов и явлений, испытывающих их воздействие, крайне важной является задача совершенствования методов исследования, которые бы позволили раскрыть суть взаимосвязи между внедоговорными отношениями и различными факторами, детерминирующими соотношение позиций конкурентов и партнеров.

1. Статистика потребительских цен [Электронный ресурс] // Национальный статистический комитет Республики Беларусь. – Режим доступа: <https://dataportal.belstat.gov.by/osids/rubric-metadata/1063264>. – Дата доступа: 23.10.2024.
2. Вилкова, О. В. К вопросу о научной осмысленности применения веб-скрейпинга как метода сбора данных в социологических исследованиях / О. В. Вилкова // Вестн. Том. гос. ун-та. Философия. Социология. Политология. – 2020. – № 54. – С. 163–175.
3. Радаев, В. В. Новый институциональный подход: построение исследовательской схемы / В. В. Радаев // Журнал социологии и социальной антропологии. – 2001. – Т. 4, № 3. – С. 109–130.
4. Статистический ежегодник Республики Беларусь [Электронный ресурс] // Национальный статистический комитет Республики Беларусь. – Режим доступа: <https://www.belstat.gov.by/upload/iblock/66e/710tc5n78wkj4ulk6er8jxshk1m1k740.pdf>. – Дата доступа: 19.10.2024.