

**Сравнение и применение методов извлечения данных веб-страниц**

**Блохина А.Ю. (ИТМО)**

**Научный руководитель – ассистент, Самохин Н.Ю. (ИТМО)**

**Введение.** Программные методы извлечения данных позволяют автоматизировать процесс сбора и структурирования большого массива данных из различных общедоступных источников. Выбор метода извлечения данных – один из ключевых этапов проекта, оказывающий существенное влияние на его результаты. В связи с ростом количества вариантов применения извлекаемых данных, изучение методов извлечения данных веб-страниц и их применение является актуальной и важной задачей [1].

**Основная часть.** В результате развития Интернета и веб-технологий появились различные методы извлечения данных, позволяющих более эффективно осуществлять поиск в сети [2]. Программные методы извлечения данных имеют широкий спектр применения во всех сферах деятельности, требующих обработки и анализа информации из публичных источников.

На данный момент вопрос выбора наилучшего метода извлечения данных веб-страниц остается открытым, поскольку зависит от специфики задачи, требуемой функциональности и доступных ресурсов [3].

Для исследования были выбраны наиболее популярные методы извлечения данных и инструменты их реализации. Каждый из методов был изучен и оценен в контексте его возможностей, ограничений и производительности. В рамках исследования был проведен ряд экспериментов, обосновывающих использование данных инструментов в рамках конкретных требований и условий проекта.

**Выводы.** В ходе исследования и проведения эксперимента выполнен сравнительный анализ, в результате которого выявлены достоинства и недостатки каждого метода извлечения данных. По итогам анализа определен оптимальный метод извлечения данных для получения требуемого результата в рамках поставленной задачи.

**Список использованных источников:**

1. Райан Митчелл. Современный скрапинг веб-сайтов с помощью Python. 2-е межд. издание. – СПб.: Питер, 2021. – С. 9.
2. Morina V., Sejdiu S. Evaluating and comparing web scraping tools and techniques for data collection // 11th UBT Annual International Conference On Computer Science And Engineering. – 2022.
3. Lotfi, C., Srinivasan, S., Ertz, M., Latrous, I. Web Scraping Techniques and Applications: A Literature Review // SCRS Conference Proceedings on Intelligent Systems. – 2021. – С. 381–394.