

Университет ИТМО, факультет программной инженерии и компьютерной техники  
Двухнедельная отчётная работа по «Информатике»: аннотация к статье

Дата прошёлшей лекции	Номер прошёлшей лекции	Название статьи/главы книги/видеолекции	Дата публикации (не старше 2021 года)	Размер статьи (от 400 слов)	Дата сдачи
11.09.2024	1	Information Theory, Living Systems, Communication Engineering	18.05.2024	~5050	25.09.2024
25.09.2024	2	Research and Development of Data Compression Methods Based on Neural Networks	01.01.2023	~3122	09.10.2024
09.10.2024	3	Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application	03.11.2021	~9800	23.10.2024
	4				
	5				
	6				
	7				

Выполнил(а) Юксель Хамза, № группы P3132, оценка не заполнять  
Фамилия И.О. студента

**Прямая полная ссылка на источник или сокращённая ссылка (bit.ly, tr.im и т.п.)**

<http://ijasca.zuj.edu.jo/PapersUploaded/2021.3.11.pdf>

**Теги, ключевые слова или словосочетания (минимум три слова)**

Web Scraping, Web Crawling, Python Language, Business Intelligence, Data Science, Artificial Intelligence, Big Data, Cloud Computing, Cybersecurity, legal, ethical.

**Перечень фактов, упомянутых в статье (минимум четыре пункта)**

1. Web scraping is a method that allows users to quickly and efficiently collect and compile data from web pages, which is essential for data-driven decision making in businesses
2. Although APIs offer a systematic method for accessing data, they rely on the site owner's permission. When this access is not granted, web scraping serves as a crucial alternative to gather the needed information.
3. The process of web scraping involves three primary interwoven activities: website analysis, website crawling, and data organization, distinguishing it from data mining, which focuses on data analysis.
4. Code reuse and maintenance are critical components of web scraping, as changes in the targeted website's design can disrupt the functionality of scraping programs.
5. Web scraping is often required to delve into more complex aspects of projects, and having a library of reusable code can significantly speed up the scraping process.
6. There are significant ethical and legal controversies surrounding web scraping, including issues of plagiarism and copyright infringement.
7. Web scraping can be applied in various fields, including data science, cybersecurity, and business intelligence.

**Позитивные следствия и/или достоинства описанной в статье технологии (минимум три пункта)**

1. Efficient Data Collection.
2. Cost-Effectiveness.
3. Access to Unstructured Data.
4. Versatile Implementations.
5. Improved Data Quality.

**Негативные следствия и/или недостатки описанной в статье технологии (минимум три пункта)**

1. Ethical and Legal Concerns.
2. Risk of Plagiarism.
3. Dependence on Website Stability.
4. Compliance Challenges.

**Ваши замечания, пожелания преподавателю или анекдот о програмистах<sup>1</sup>**

<sup>1</sup> Наличие этой графы не влияет на оценку

