**Web Scraping Assignment Report**

Name: Murat Manat

ID: 050128501402

Subject: Анализ и обработка веб данных

# Website Structure

The website used is *http://quotes.toscrape.com/*, which is designed for practicing web scraping. It contains multiple pages with quotes. Each page has several <div class='quote'> blocks, and each block contains:

* Quote text inside <span class='text'>
* Author name inside <small class='author'> - Tags inside <div class='tags'>

# Main Challenges

1. Handling pagination (multiple pages of quotes).
2. Extracting multiple tags for each quote.
3. Avoiding blocking by websites (solved by adding a User-Agent header).

# Solution

* Used the *requests* library to send HTTP requests and download the HTML content.
* Parsed the HTML with *BeautifulSoup*.
* Implemented a loop to scrape the first 3 pages of quotes.
* Extracted the following fields:
* Quote (text of the quote)
* Author (name of the author)
* Tags (keywords associated with the quote)
* Collected the results into a Python list of dictionaries.
* Saved the data into a CSV file (*quotes.csv*) using the csv.DictWriter module.

# Output Example (CSV)

Quote,Author,Tags

"The world as we have created it is a process of our thinking...",Albert Einstein,change,deep-thought

"It is our choices, Harry, that show what we truly are...",J.K. Rowling,abilities,choices

"There are only two ways to live your life...",Albert Einstein,inspirational,life,live

# Final Deliverables

* GitHub repository with the Python script (quotes\_scraper.py)
* CSV file (quotes.csv) with extracted data
* This report in PDF format