Class Exercise:

RegEx v3

You are given an HTML file containing various elements. Your task is to create a Python class that utilizes regular expressions to perform the following operations:

1. Extract all the URLs from the HTML file.
2. Replace all occurrences of a given tag with a specified replacement string.
3. Split the HTML file into a list of HTML elements.
4. Find and print all the attributes of a specific tag in the HTML file.

Your class should have the following methods:

1. **extract\_urls(html):** This method takes a string html as input and uses regular expressions to find and extract all the URLs present in the HTML. It should return a list of extracted URLs.
2. **replace\_tag(html, tag, replacement):** This method takes a string html, a tag name tag, and a replacement string replacement. It uses the **re.sub** function to replace all occurrences of the specified tag in the HTML with the replacement string. It should return the modified HTML.
3. **split\_elements(html):** This method takes a string html as input and uses regular expressions to split the HTML into a list of HTML elements. Each element should be a separate string in the list. It should return the list of HTML elements.
4. **find\_attributes(html, tag):** This method takes a string html and a tag name tag as input. It uses regular expressions to find all occurrences of the specified tag in the HTML and prints all the attributes of that tag.

import re

class HTMLAnalyzer:

    def extract\_urls(self, html):

# Your code here

    def replace\_tag(self, html, tag, replacement):

# Your code here

    def split\_elements(self, html):

# Your code here

# Example usage:

html\_file = """

<html>

<head>

<title>Sample Page</title>

</head>

<body>

<h1>Welcome</h1>

<p>This is a sample paragraph.</p>

<a href="https://example.com">Visit Example</a>

<img src="image.jpg" alt="Sample Image">

</body>

</html>

"""

analyzer = HTMLAnalyzer()

# Extract URLs

urls = analyzer.extract\_urls(html\_file)

print("Extracted URLs:", urls)

# Replace tag

modified\_html = analyzer.replace\_tag(html\_file, 'h1', '<h2>Modified Heading</h2>')

print("Modified HTML:\n", modified\_html)

# Split elements

elements = analyzer.split\_elements(html\_file)

print("HTML Elements:", elements)

# Find attributes

analyzer.find\_attributes(html\_file, 'a')

Expected output:

Extracted URLs: ['https://example.com']

Modified HTML:

<html>

<head>

<title>Sample Page</title>

</head>

<body>

<h2>Modified Heading</h2>

<p>This is a sample paragraph.</p>

<a href="https://example.com">Visit Example</a>

<img src="image.jpg" alt="Sample Image">

</body>

</html>

HTML Elements: ['<html>', '<head>', '<title>', '</title>', '</head>', '<body>', '<h1>', '</h1>', '<p>', '</p>', '<a href="https://example.com">', '</a>', '<img src="image.jpg" alt="Sample Image">', '</body>', '</html>']

['href']