

0x04. UTF-8 Validation

Algorithm My Playlist Planning/me



Weight: 1

Projects(/projects/current)



Project over - took place from Jul 22, 2024 6:00 AM to Jul 26, 2024 6:00 AM



An auto QA review will automatically make corrections(/to_review)



Evaluation quizzes(/dashboards/my_current_evaluation_quizzes)

In a nutshell...

- **Auto QA review:** 14.0/14 mandatory

- **Altogether:** 100.0%



Curriculums(/dashboards/my_curriculums)

◦ Mandatory: 100.0%

◦ Optional: no optional tasks



Concepts(/concepts)



For the "0x04. UTF-8 Validation" project, you will need to apply your knowledge in bitwise operations, understanding of the UTF-8 encoding scheme, and Python programming skills to validate whether a given dataset represents a valid UTF-8 encoding. Here's a list of concepts and resources that will be



Helpful: Servers(/servers)

>_ Concepts Needed:

Sandboxes(/user_containers/current)

1. Bitwise Operations in Python:



Tools(/dashboards/my_tools)

◦ Understanding how to manipulate bits in Python, including operations like AND (&), OR (|), XOR (^), NOT (~), shifts (<< , >>).

◦ Python Bitwise Operators (/rltoken/BslyYNZIXdyxW3_b0WNOcg)



2. UTF-8 Encoding Scheme

Videos(/dashboards/videos)

◦ Familiarity with the UTF-8 encoding rules, including how characters are encoded into one or more bytes.

◦ Understanding the patterns that represent a valid UTF-8 encoded character.



Peers(/users/peers)

◦ UTF-8 Wikipedia (/rltoken/oqFi6P1hNvp9aSuNv---IQ)

◦ Characters, Symbols, and the Unicode Miracle (/rltoken/d--jVK8sBSlhkosu7pFzdw)



◦ The Absolute Minimum Every Software Developer Absolutely, Positively Must Know About Unicode and Character Sets (/rltoken/9EwaXVds22dSK3lvF5nNCA)

3. Data Representation:

◦ How to represent and work with data at the byte level.

◦ Handling the least significant bits (LSB) of integers to simulate byte data.

4. List Manipulation in Python:

My Profile(/users/my_profile)

◦ Iterating through lists, accessing list elements, and understanding list comprehensions.



Python Lists (/rltoken/TaN91MgmOL80GeOGvmldlw)



5. Boolean Logic:

- (/) Applying logical operations to make decisions within the program.

By studying these concepts and utilizing the resources provided, you will be equipped to tackle the UTF-validation project, effectively applying bitwise operations and logical reasoning to determine the validity of UTF-8 encoded data.



Home(/)



My Planning(/planning/me)

Additional Resource



- Mock Technical Interview (/rltoken/X1IZqipeyegt8pbQ9aXSFQ)
Projects(/projects/current)



Requirements

QA Reviews I can make(/corrections/to_review)

General



- Evaluation quizzes(/dashboards/my_current_evaluation_quizzes)
- Allowed editors: vi, vim, emacs
- All your files will be interpreted/compiled on Ubuntu 20.04 LTS using python3 (version 3.4.3)
- All your files should end with a new line
- The first line of all your files should be exactly `#!/usr/bin/python3`
- Curriculum (/dashboards/my_curriculums)
- A `README.md` file, at the root of the folder of the project, is mandatory
- Your code should use the PEP 8 style (version 1.7.x)
- All your files must be executable
- Concepts(/concepts)



Conference rooms(/dashboards/video_rooms)

Tasks

Servers(/servers)

0. UTF-8 Validation

mandatory



Sandboxes(/user_containers/current)

Score: 100.0% (Checks completed: 100.0%)



Tools(/dashboards/my_tools)

Write a method that determines if a given data set represents a valid UTF-8 encoding.



Video on demand(/dashboards/videos)

- Prototype: `def validUTF8(data)`
- Return: `True` if data is a valid UTF-8 encoding, else return `False`
- A character in UTF-8 can be 1 to 4 bytes long
- The data set can contain multiple characters
- The data will be represented by a list of integers
- Each integer represents 1 byte of data, therefore you only need to handle the 8 least significant bits of each integer



Peers(/users/peers)



Discord(<https://discord.com/app>)



My Profile(/users/my_profile)

```
carrie@ubuntu:~/0x04-utf8_validation$ cat 0-main.py
#!/usr/bin/python3
"""
Main file for testing
"""
Home(/)
validUTF8 = __import__('0-validate_utf8').validUTF8
My Planning(/planning/me)
data = My Planning(/planning/me)
print(validUTF8(data))
Projects(/projects/current)
data = Projects(/projects/current)111, 110, 32, 105, 115, 32, 99, 111, 111, 108, 33]
print(validUTF8(data))
QA Reviews, can make corrections/to_review
data = QA Reviews, can make corrections/to_review
print(validUTF8(data))
? carrie@ubuntu:~/0x04-utf8_validation$ Evaluation quizzes(/dashboards/my_current_evaluation_quizzes)
```

```
carrie@ubuntu:~/0x04-utf8_validation$ ./0-main.py
True
True Curriculums(/dashboards/my_curriculums)
False
carrie@ubuntu:~/0x04-utf8_validation$
Concepts(/concepts)
```

Repo: Conference rooms(/dashboards/video_rooms)

- GitHub repository: alx-interview
- Directory: 0x04-utf8_validation
- Servers(/servers)
- File: 0-validate_utf8.py

> Sandboxes(/user_containers/current)

Check submission View results

🔧 Tools(/dashboards/my_tools)

🎬 Video on demand(/dashboards/videos)

Copyright © 2024 ALX, All rights reserved.

👤 Peers(/users/peers)

🎮 Discord(<https://discord.com/app>)



My Profile(/users/my_profile)