Activity_Work with strings in Python

February 21, 2024

1 Activity: Work with strings in Python

1.1 Introduction

Security analysts work with a lot of string data. For example, some security analysts work on creating and updating IDs such as employee IDs and device IDs, which are commonly represented as strings. As another example, certain network activity will be stored as string data. Becoming comfortable working with strings in Python is essential for the work of a security analyst.

In this lab, you'll practice creating Python code and working with strings. You'll work with an employee ID, a device ID, and a URL, all represented as string data.

Tips for completing this lab

As you navigate this lab, keep the following tips in mind:

- `### YOUR CODE HERE ###` indicates where you should write code. Be sure to replace that with your own code before running the code cell.
- Feel free to open the hints for additional guidance as you work on each task.
- To enter your answer to a question, double-click the markdown cell to edit. Be sure to replace the "[Double-click to enter your responses here.]" with your own answer.
- You can save your work manually by clicking File and then Save in the menu bar at the top of the notebook.
- You can download your work locally by clicking File and then Download and then specifying your preferred file format in the menu bar at the top of the notebook.

1.2 Scenario

You're working as a security analyst, and you are responsible for writing programs in Python to automate updating employee IDs, extracting characters from a device ID, and extracting components from a URL.

1.3 Task 1

In your organization, employee IDs are currently either four digits or five digits in length. In this task, you're given a four-digit numeric employee ID stored in a variable called `employee_id`. Convert this to a string format and store the result in the same variable. Later, you'll update this employee ID string so that it complies with a new standardized format.

Complete the following code. Be sure to replace the `### YOUR CODE HERE ###` with your own code before you run the following cell.

[1]:
```
# Assign `employee_id` to a four-digit number as an initial value
employee_id = 4186

# Display the data type of `employee_id`
print("Data type before conversion:", type(employee_id))

# Reassign `employee_id` to the same value but in the form of a string
employee_id = str(employee_id)

# Display the data type of `employee_id` now
print("Data type after conversion:", type(employee_id))
```

```
Data type before conversion: <class 'int'>
Data type after conversion: <class 'str'>
```

Hint 1

Use the `str()` function in Python to convert the initial value of the `employee_id` variable into a string.

Hint 2

Pass `employee_id` into the `str()` function.

Question 1 What do you observe about the data type of `employee_id` the first time it's displayed? What do you observe about the data type of `employee_id` the second time it's displayed (after the variable is reassigned)?

The first time `employee_id` is displayed, its data type is observed to be `int` (integer). After the variable is reassigned by converting it to a string using `str()`, the second time it's displayed, the data type is observed to be `str` (string). The reassignment from an integer to a string is evident from the change in data type in the code.

1.4 Task 2

Imagine that you have just been informed of a new criteria for employee IDs. They must all be five digits long for standardization purposes.

In this task, you will write a conditional statement that displays a message if the length of the employee ID is less than five digits.

Be sure to replace the `### YOUR CODE HERE ###` with your own code before you run the following cell.

[4]:
```
# Assign `employee_id` to a four-digit number as an initial value
employee_id = 4186
```

```
# Reassign `employee_id` to the same value but in the form of a string
employee_id = str(employee_id)

# Conditional statement that displays a message if the length of `employee_id`␣
↪is less than five digits
if len(employee_id) < 5:
    print("This employee ID has less than five digits. It does not meet length␣
↪requirements.")
```

```
This employee ID has less than five digits. It does not meet length
requirements.
```

Hint 1

The `len()` function in Python can be used to get the length of `employee_id`.

Hint 2

Start the conditional statement with the `if` keyword.

Hint 3

To write the condition in the conditional statement, use the `<` comparison operator to check whether the length of `employee_id` is less than `5`. Make sure to place the condition between the `if` and the `:`.

1.5 Task 3

In this task, you'll build upon the previous code. If an employee ID is only four digits, you'll use concatenation to create a five-digit employee ID number.

Concatenation is a process that allows you to merge strings together. The addition operator (`+`) in Python allows you to concatenate two strings.

Write an `if` statement that evaluates whether the length of `employee_id` is less than `5`. When the condition evaluates to `True`, reassign `employee_id` by concatenating `"E"` in front of the four-digit employee ID to create a five character employee ID. Then, display `employee_id` again. Be sure to replace each `### YOUR CODE HERE ###` with your own code before you run the following cell.

[5]:
```
# Assign `employee_id` to a four-digit number as an initial value
employee_id = 4186

# Reassign `employee_id` to the same value but in the form of a string
employee_id = str(employee_id)

# Display the `employee_id` as it currently stands
print("Current employee ID:", employee_id)

# Conditional statement that updates the `employee_id` if its length is less␣
↪than 5 digits
```

```
if len(employee_id) < 5:
    employee_id = "E" + employee_id

# Display the `employee_id` after the update
print("Updated employee ID:", employee_id)
```

```
Current employee ID: 4186
Updated employee ID: E4186
```

Hint 1

To complete the header of the conditional statement, use the `if` keyword to start, the `len()` function to get the length of `employee_id`, and the `<` comparison operator to check whether the length is less than `5`. Make sure to write this before the `:`.

Hint 2

Use the `=` assignment operator to update the value of the `employee_id` variable. Update the value of `employee_id` to the concatenation of `"E"` with the variable's current value. The `"E"` should appear to the left of the current value.

Hint 3

Use the `print()` function to display `employee_id` after the update.

1.6 Task 4

Now you'll move on to the next part of your task. Imagine that the characters in a device ID convey technical information about the device. You'll need to extract characters in specific positions from the device ID. Start off by extracting the fourth character.

The variable `device_id` represents a device ID containing alphanumeric characters; it's already stored as a string.

Be sure to replace the `### YOUR CODE HERE ###` with your own code before you run the following cell.

[6]:
```
# Assign `device_id` to a string that contains alphanumeric characters
device_id = "r262c36"

# Extract the fourth character in `device_id` and display it
print("Fourth character in device_id:", device_id[3])
```

```
Fourth character in device_id: 2
```

Hint 1

Use a pair of square brackets, passing in the appropriate index value, in order to extract the fourth character in `device_id`.

Hint 2

In Python, index values start at `0`.

Hint 3

Given that index values start at `0` in Python, an index value of `3` corresponds to the fourth character in a sequence.

1.7 Task 5

Now you will also need to extract the first through the third characters in the device ID. So take a slice of the device ID. You can achieve this using bracket notation in Python. Then, display the slice to examine the result.

Be sure to replace the `### YOUR CODE HERE ###` with your own code before you run the following cell.

[7]:
```
# Assign `device_id` to a string that contains alphanumeric characters
device_id = "r262c36"

# Extract the first through the third characters in `device_id` and display the␣
 ↪result
print("First through third characters in device_id:", device_id[:3])
```

```
First through third characters in device_id: r26
```

Hint 1

Use a pair of square brackets, passing in the appropriate index values, in order to extract the first through the third characters in `device_id`.

Inside the square brackets, use a `:` to separate the first index value (the starting index value) and second index value (ending index value).

Hint 2

Keep in mind that the second index value passed into bracket notation is exclusive. In other words, the index value passed into the square brackets after the `:` is not included when the string is sliced. The resulting slice will not include the character at that index.

Hint 3

Recall that the second index value passed into bracket notation is exclusive and indexing in Python starts at `0`. The first index value should be `0` and the second index value should be `3`, in order to extract the first through the third characters in `device_id`.

1.8 Task 6

You'll now proceed to the last part of your task. This involves extracting components of a URL.

You'll work with string indices to display various components of a URL that's stored in the URL variable. First, you'll extract and display the protocol of the URL and the `://` characters that follow it using string slicing. Consider that the protocol is in the secure format of `https` when determining the indices for your slice.

Be sure to replace the `### YOUR CODE HERE ###` with your own code before you run the following cell.

[9]:
```
# Assign `url` to a specific URL
url = "https://exampleURL1.com"

# Extract the protocol of `url` along with the syntax following it, and display␣
 ↪the result
protocol_and_syntax = url[:8]

# Display the extracted components
print("Protocol and syntax:", protocol_and_syntax)

#This code uses string slicing to extract the protocol and the syntax following␣
 ↪it (in this case, the :// characters) from the url string and then prints␣
 ↪the extracted components.
```

```
Protocol and syntax: https://
```

Hint 1

Note that `https://` is eight characters long.

Hint 2

Use a pair of square brackets to slice the string stored in `url`, passing in two index values separated by `:`. Keep in mind that the second index value is exclusive and that indexing in Python starts at `0`.

Hint 3

Use the `print()` function to display the slice.

1.9 Task 7

Later in this lab, you'll extract the domain extension. To prepare for this, use the `.index()` method to identify the index where the domain extension `.com` is located in the given URL.

Be sure to replace the `### YOUR CODE HERE ###` with your own code before you run the following cell.

[10]:
```
# Assign `url` to a specific URL
url = "https://exampleURL1.com"

# Display the index where the domain extension ".com" is located in `url`
print("Index of .com in the URL:", url.index(".com"))

#This code uses the .index() method to find the index of the substring .com in␣
 ↪the url and then prints the identified index.
```

```
Index of .com in the URL: 19
```

Hint 1

Apply the `.index()` method to `url` in order to get the appropriate index. The `.index()` method takes in a substring, and if that substring is located in the original string, it returns the index where that substring starts to occur in the original string.

Hint 2

Call `url.index()`, and inside the parentheses, pass in the targeted domain extension as a string.

1.10 Task 8

It's a good idea to save important data in variables when programming. This allows for quick and easy tracking and reuse of information.

Store the output of the `.index()` method in a variable called `ind`, which is short for index. This index represents the position where the domain extension `".com"` starts in the `url`. Be sure to replace the `### YOUR CODE HERE ###` with your own code before you run the following cell. Note that running this cell will not produce an output.

[12]:
```
# Assign `url` to a specific URL
url = "https://exampleURL1.com"

# Assign `ind` to the output of applying `.index()` to `url` in order to␣
↪extract the starting index of ".com" in `url`
ind = url.index(".com")

# This code uses the .index() method to find the index of the substring .com in␣
↪the url and assigns the result to the variable ind.
```

Hint 1

To assign the extracted index to the `ind` variable, use the `.index()` method to the right of the `=` assignment operator.

Hint 2

Call `url.index()`, and inside the parantheses, pass in the targeted domain extension as a string. Be sure to place this to the right of the `=` assignment operator, so that the result is assigned to `ind`.

1.11 Task 9

You can use string slicing to also extract the domain extension of a URL. To do so, you can create a slice. The starting index should be the `ind` variable. This contains the index where the domain extension begins. The ending index should be `ind + 4` (since `".com"` is four characters long). Sometimes, like in this situation, it's easier to express the ending index in relation to the starting index. Examine the following code, run it as is, and observe the output.

[13]:
```
# Assign `url` to a specific URL

```

```
url = "https://exampleURL1.com"

# Assign `ind` to the output of applying `.index()` to `url` in order to␣
 ↪extract the starting index of ".com" in `url`

ind = url.index(".com")

# Extract the domain extension in `url` and display it

print(url[ind:ind+4])
```

```
.com
```

Question 2 What does this code output and why?

The code outputs `com`. This is because the `url[ind:ind+4]` expression uses string slicing method to extract a substring from the `url`. The starting index is determined by the value stored in the variable `ind`, which is the index where the domain extension (".com") begins in the code. The ending index is calculated as `ind + 4`, which includes the characters up to four positions after the starting index. Therefore, the extracted substring corresponds to the domain extension, which is "com" in this case.

1.12 Task 10

Finally, extract the website name from the given URL using string slicing and the `ind` variable that you defined earlier. In the given URL, the website name is `"exampleURL1"`. Be sure to replace the `### YOUR CODE HERE ###` with your own code before you run the following cell.

[14]:
```
# Assign `url` to a specific URL
url = "https://exampleURL1.com"

# Assign `ind` to the output of applying `.index()` to `url` in order to␣
 ↪extract the starting index of ".com" in `url`
ind = url.index(".com")

# Extract the website name in `url` and display it
website_name = url[:ind]

# Display the extracted website name
print("Website name:", website_name)

# This code uses string slicing to extract the substring from the beginning of␣
 ↪the url up to the index stored in the variable ind. The extracted substring␣
 ↪represents the website name, and it is then displayed.
```

```
Website name: https://exampleURL1
```

Hint 1

In order to extract the website name in the given URL, use a pair of square brackets to create a slice of `url`, passing in a start index and an ending index, separating the two with `:`.

Hint 2

The starting index should be set to `8`, since this is the position after the protocol and the `://` syntax ends and where the website name begins. The ending index should be set to the position where the `.com` domain name begins.

1.13 Conclusion

What are your key takeaways from this lab?

Key takeaways from this lab:

1. **String Manipulation:** Understanding how to manipulate strings using string slicing, concatenation, and various string methods like `.index()`.
2. **Variable Assignment:** The importance of assigning meaningful names to variables and using them to store and retrieve information.
3. **String Slicing:** The ability to extract specific portions of a string using string slicing based on indices.
4. **Method Usage:** Utilizing string methods, such as `.index()`, to find the position of a substring within a string.
5. **Conditional Statements:** The use of conditional statements to check and modify data based on certain conditions.
6. **Code Readability:** Writing clear and well-commented code to enhance readability and understanding.

[ ]: