Python Regex Word Boundary

Summary: in this tutorial, you'll learn how to construct regular expressions that match word boundary positions in a string.

Introduction to the Python regex word boundary

A string has the following positions that qualify as word boundaries:

1. Before the first character in the string if the first character is a word character (`\w`).

2. Between two characters in the string if the first character is a word character (`\w`) and the other is not (`\W` – inverse character set of the word character `\w`).

3. After the last character in a string if the last character is the word character (`\w`)

The following picture shows the word boundary positions in the string `"PYTHON 3!"` :

In this example, the `"PYTHON 3!"` string has four word boundary positions:

- Before the letter P (criteria #1)

- After the letter N (criteria #2)

- Before the digit 3 (criteria #2)

- After the digit 3 (criteria #2)

Regular expressions (https://www.pythontutorial.net/python-regex/python-regular-expressions/) use the `\b` to represent a word boundary. For example, you can use the `\b` to match the whole `word` using the following pattern:

```
r'\bword\b'
```

The following example matches the word `Python` in a string:

```
import re

s = 'CPython is the implementation of Python in C'
matches = re.finditer('Python', s)
for match in matches:
    print(match.group())
```

It returns two matches, one in the word `CPython` and another in the word `Python`.

```
Python
Python
```

However, if you use the word boundary `\b`, the program returns one match:

```
import re

s = 'CPython is the implementation of Python in C'
matches = re.finditer(r'\bPython\b', s)
```

```
for match in matches:
    print(match.group())
```

Output:

```
<re.Match object; span=(33, 39), match='Python'>
```

In this example, the `'\bPython\b'` pattern match the whole word `Python` in the string `'CPython is the implementation of Python in C'`.

Summary

- The `\b` represents a word boundary in a string.

- Use the `r'\bword\b'` pattern uses the \b to match the whole `word`