Parse a text file in Python

Now that you know how to import text files into Python, we're going to take this one step further and learn how to give them a structure.
This will allow us to analyze them more easily.
This process is often referred to as parsing.
Parsing is the process of converting data into a more readable format.
To do this, we're going to put together everything we learned about lists and strings and learn another method for working with strings in Python.
The method we need is the split method.
The split method converts a string into a list.
It does this by separating the string based on a specified character.
Or, if no argument is passed, every time it encounters a whitespace, it separates the string.
So, a split would convert the string "We are learning about parsing!" into this list.
We are using the split method to separate the strings into smaller chunks that we can analyze more easily than one big block of text.
In this video, we'll work with an example of a security log where every line represents a new data point.
To store these points in a list, we want to separate the text based on the new line.
Python considers a new line to be a type of whitespace.
We can use the split method without passing an argument.
We'll start with our code from the previous video.
Remember, we used this code to open a file and then read it into a string.
Now, let's split that string into a list using the split method and then print the output.
After we run it, Python outputs a list of usernames instead of one big string of them.
If we want to save this list, we would need to assign it to another variable.
For example, we can call the variable usernames.
And then we'll run it again.
And now this list can be reused in other code.
Congratulations!
You just learned the basics of parsing a text file in Python.
In the next videos, we're going to be exploring techniques that help us work more in depth with data in Python.
Now that you know how to import text files into Python,.

Revision #1
Created 2023-12-28 16:06:21 UTC by naruzkurai
Updated 2023-12-28 16:06:32 UTC by naruzkurai