Network Programming in Python I

Justin Ellis MBA

Review

Any Questions?

9.2. A Collection Data Type

So far we have seen built-in types like: int, float, bool, str and we've seen lists. int, float, and bool are considered to be simple or primitive data types because their values are not composed of any smaller parts. They cannot be broken down. On the other hand, strings and lists are different from the others because they are made up of smaller pieces. In the case of strings, they are made up of smaller strings each containing one character.

Types that are comprised of smaller pieces are called **collection data types**. Depending on what we are doing, we may want to treat a collection data type as a single entity (the whole), or we may want to access its parts. This ambiguity is useful.

Strings can be defined as sequential collections of characters. This means that the individual characters that make up the string are assumed to be in a particular order from left to right.

A string that contains no characters, often referred to as the **empty string**, is still considered to be a string. It is simply a sequence of zero characters and is represented by " or "" (two single or two double quotes with nothing in between).

9.3. Operations on Strings

In general, you cannot perform mathematical operations on strings, even if the strings look like numbers. The following are illegal (assuming that message has type string):

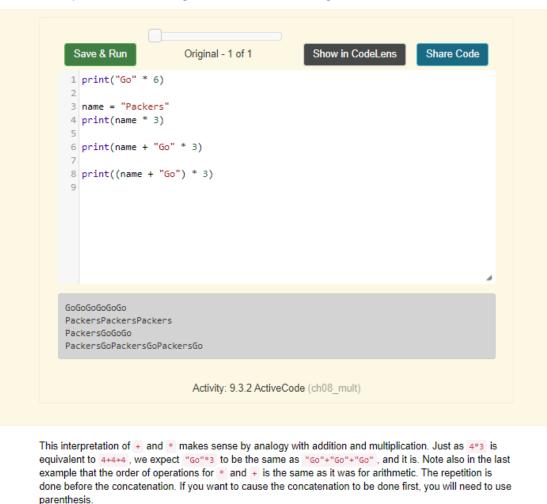
```
message - 1
"Hello" / 123
message * "Hello"
"15" + 2
```

Interestingly, the + operator does work with strings, but for strings, the + operator represents concatenation, not addition. Concatenation means joining the two operands by linking them end-to-end.

Save & Run	Original - 1 of 1	Show in CodeLens	Share Code
1 fruit = "ba 2 bakedGood = 3 print(fruit 4	" nut bread"		
banana nut bread	đ		
	Activity: 9.3.1 ActiveC	ode (ch08_add)	

The output of this program is **banana nut bread**. The space before the word **nut** is part of the string and is necessary to produce the space between the concatenated strings. Take out the space and run it again.

The * operator also works on strings. It performs repetition. For example, 'Fun'*3 is 'FunFunFun'. One of the operands has to be a string and the other has to be an integer.



The indexing operator (Python uses square brackets to enclose the index) selects a single character from a string. The characters are accessed by their position or index value. For example, in the string shown below, the 14 characters are indexed left to right from postion 0 to position 13.

9 10 11 12 13 0 3 4 5 6 78 2 С u t h е r 0 е g e -14 -13 -12 -11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 It is also the case that the positions are named from right to left using negative numbers where -1 is the rightmost index and so on. Note that the character at index 6 (or -8) is the blank character.

5	Save & Run Original - 1 of 1 Show in CodeLens Share Code	
2 3 4 5	<pre>school = "Luther College" m = school[2] print(m) lastchar = school[-1] print(lastchar)</pre>	
t e		
	Activity: 9.4.1 ActiveCode (chp08_index1)	
	ression <pre>school[2] selects the character at index 2 from <pre>school</pre>, and creates a new string <pre>ig just this one character. The variable </pre> refers to the result.</pre>	

Remember that computer scientists often start counting from zero. The letter at index zero of "Luther college" is L. So at position [2] we have the letter t.

If you want the zero-eth letter of a string, you just put 0, or any expression with the value 0, in the brackets. Give it a try.

The expression in brackets is called an **index**. An index specifies a member of an ordered collection. In this case the collection of characters in the string. The index *indicates* which character you want. It can be any integer expression so long as it evaluates to a valid index value.

Note that indexing returns a *string* — Python has no special type for a single character. It is just a string of length 1.

We previously saw that each turtle instance has its own attributes and a number of methods that can be applied to the instance. For example, we wrote tess.right(90) when we wanted the turtle object tess to perform the right method to turn to the right 90 degrees. The "dot notation" is the way we connect the name of an object to the name of a method it can perform.

Strings are also objects. Each string instance has its own attributes and methods. The most important attribute of the string is the collection of characters. There are a wide variety of methods. Try the following program.

Save & Run	Original - 1 of 1	Show in CodeLens	Share Code
1 ss = "Hello, 2 print(ss.upp 3 4 tt = ss.lowe 5 print(tt) 6	er())		
HELLO, WORLD hello, world			
	Activity: 9.5.1 ActiveCo	de (chp08_upper)	

Method Parameters Description

In this example, upper is a method that can be invoked on any string object to create a new string in which all the characters are in uppercase. lower works in a similar fashion changing all characters in the string to lowercase. (The original string ss remains unchanged. A new string tt is created.)

In addition to upper and lower, the following table provides a summary of some other useful string methods. There are a few activecode examples that follow so that you can try them out.

Wethod	Parameters	Description
upper	none	Returns a string in all uppercase
lower	none	Returns a string in all lowercase
capitalize	none	Returns a string with first character capitalized, the rest lower
strip	none	Returns a string with the leading and trailing whitespace removed
Istrip	none	Returns a string with the leading whitespace removed
rstrip	none	Returns a string with the trailing whitespace removed
count	item	Returns the number of occurrences of item
replace	old, new	Replaces all occurrences of old substring with new
center	width	Returns a string centered in a field of width spaces
ljust	width	Returns a string left justified in a field of width spaces
rjust	width	Returns a string right justified in a field of width spaces
find	item	Returns the leftmost index where the substring item is found, or -1 if not found
rfind	item	Returns the rightmost index where the substring item is found, or -1 if not found
index	item	Like find except causes a runtime error if item is not found
rindex	item	Like rfind except causes a runtime error if item is not found
format	substitutions	Involved! See String Format Method, below

You should experiment with these methods so that you understand what they do. Note once again that the methods that return strings do not change the original. You can also consult the Python documentation for strings.

9.5.1. String Format Method

In grade school quizzes a common convention is to use fill-in-the blanks. For instance,



and you can fill in the name of the person greeted, and combine given text with a chosen insertion. *We use this as an analogy:* Python has a similar construction, better called fill-in-the-braces. The string method **format**, makes substitutions into places in a string enclosed in braces. Run this code:

Save & Run	Original - 1 of 1	Show in CodeLens	Share Code
	t('Your name: ') ello {}!'.format(person) g)		

There are several new ideas here!

The string for the format method has a special form, with braces embedded. Such a string is called a *format string*. Places where braces are embedded are replaced by the value of an expression taken from the parameter list for the format method. There are many variations on the syntax between the braces. In this case we use the syntax where the first (and only) location in the string with braces has a substitution made from the first (and only) parameter.

In the code above, this new string is assigned to the identifier greeting, and then the string is printed.

The identifier greeting was introduced to break the operations into a clearer sequence of steps. However, since the value of greeting is only referenced once, it can be eliminated with the more concise version:

Save & Run	Original - 1 of 1	Show in CodeLens	Share Code
ello Justin!			

There are several new ideas here!

The string for the format method has a special form, with braces embedded. Such a string is called a *format string*. Places where braces are embedded are replaced by the value of an expression taken from the parameter list for the format method. There are many variations on the syntax between the braces. In this case we use the syntax where the first (and only) location in the string with braces has a substitution made from the first (and only) parameter.

In the code above, this new string is assigned to the identifier greeting, and then the string is printed.

The identifier greeting was introduced to break the operations into a clearer sequence of steps. However, since the value of greeting is only referenced once, it can be eliminated with the more concise version:

Save & Run	Original - 1 of 1	Show in CodeLens	Share Code
ello Justin!			

There are several new ideas here!

The string for the format method has a special form, with braces embedded. Such a string is called a *format string*. Places where braces are embedded are replaced by the value of an expression taken from the parameter list for the format method. There are many variations on the syntax between the braces. In this case we use the syntax where the first (and only) location in the string with braces has a substitution made from the first (and only) parameter.

In the code above, this new string is assigned to the identifier greeting, and then the string is printed.

The identifier greeting was introduced to break the operations into a clearer sequence of steps. However, since the value of greeting is only referenced once, it can be eliminated with the more concise version:

Save & Run	Original - 1 of 1	Show in CodeLens	Share Code
ello Justin!			

There can be multiple substitutions, with data of any type. Next we use floats. Try original price \$2.50 with a 7% discount:



inside the braces for the monetary values:

There can be multiple substitutions, with data of any type. Next we use floats. Try original price \$2.50 with a 7% discount:



inside the braces for the monetary values:



9.6. Length

The len function, when applied to a string, returns the number of characters in a string.

Save & Run	Original - 1 of 1	Show in CodeLens	Share Code
<pre>1 fruit = "Banana" 2 print(len(fruit)) 3</pre>			
6			
	Activity: 9.6.1 ActiveCo	ode (chp08_len1)	

9.6. Length

The len function, when applied to a string, returns the number of characters in a string.

Save & Run	Original - 1 of 1	Show in CodeLens	Share Code
<pre>1 fruit = "Banana" 2 print(len(fruit)) 3</pre>			
6			
	Activity: 9.6.1 ActiveCo	ode (chp08_len1)	

Alternatively in Python, we can use **negative indices**, which count backward from the end of the string. The expression <u>fruit[-1]</u> yields the last letter, <u>fruit[-2]</u> yields the second to last, and so on. Try it! Most other languages do *not* allow the negative indices, but they are a handy feature of Python!

9.7. The Slice Operator

A substring of a string is called a slice. Selecting a slice is similar to selecting a character:

Save & Run	Original - 1 of 1	Show in CodeLens	Share Code
<pre>1 singers = "Pet 2 print(singers[3 print(singers[4 print(singers[5 </pre>	7:11])		
Peter Paul Mary			
	Activity: 9.7.1 ActiveCo	ode (chp08_slice1)	
cluding the first but exclu not include the characte	ding the last. In other words,	rom the n'th character to the start with the character at inc nay seem counter-intuitive bu	lex n and go up to but

If you omit the first index (before the colon), the slice starts at the beginning of the string. If you omit the second index, the slice goes to the end of the string.

There is no Index Out Of Range exception for a slice. A slice is forgiving and shifts any offending index to something legal.

9.8. String Comparison

The comparison operators also work on strings. To see if two strings are equal you simply write a boolean expression using the equality operator.

<pre>word = "banana" if word == "banana": print("Yes, we have bananas!") else: print("Yes, we have NO bananas!") 6</pre>	
Yes, we have bananas!	

Other comparison operations are useful for putting words in lexicographical order. This is similar to the alphabetical order you would use with a dictionary, except that all the uppercase letters come before all the lowercase letters.

S	Save & Run	Original - 1 of 1	Show in CodeLens	Share Code
2 3 4 5 6 7	<pre>elif word > print(" else:</pre>	banana": (our word, " + word + ", c "banana": (our word, " + word + ", c	omes after banana.")	
9	print((es, we have no bananas!")		4
Υοι	ır word, zebra	a, comes after banana.		

Activity: 9.8.2 ActiveCode (ch08_comp2)

It is probably clear to you that the word *apple* would be less than (come before) the word banana. After all, *a* is before *b* in the alphabet. But what if we consider the words *apple* and *Apple*? Are they the same?

Save & Run	Original - 1 of 1	Show in CodeLens	Share Code
	e" < "banana") e" == "Apple") e" < "Apple")		
True False False			
	Activity: 9.8.3 ActiveCo	de (chp08_ord1)	
	I from our discussion of variable		

It turns out, as you recall from our discussion of variable names, that uppercase and lowercase letters are considered to be different from one another. The way the computer knows they are different is that each character is assigned a unique integer value. "A" is 65, "B" is 66, and "5" is 53. The way you can find out the so-called **ordinal value** for a given character is to use a character function called **ord**.

Save & Run	Original - 1 of 1	Show in CodeLens	Share Code
<pre>1 print(ord("A")) 2 print(ord("B")) 3 print(ord("5")) 4 5 print(ord("a")) 6 print("apple" > 7</pre>	"Apple")		4
65 66 53 97 True			
	Activity: 9.8.4 Active	Code (ch08_ord2)	

When you compare characters or strings to one another, Python converts the characters into their equivalent ordinal values and compares the integers from left to right. As you can see from the example above, "a" is greater than "A" so "apple" is greater than "Apple".

Humans commonly ignore capitalization when comparing two words. However, computers do not. A common way to address this issue is to convert strings to a standard format, such as all lowercase, before performing the comparison.

There is also a similar function called chr that converts integers into their character equivalent.

Save & Run	Original - 1 of 1	Show in CodeLens	Share Code
<pre>1 print(chr(65)) 2 print(chr(66)) 3 4 print(chr(49)) 5 print(chr(53)) 6 7 print("The cha 8 print(ord(" ") 9</pre>	racter for 32 is", chr)	(32), "!!!")	
A B 1 5 The character for 3 32	12 is !!!		
	Activity: 9.8.5 Active		

One thing to note in the last two examples is the fact that the space character has an ordinal value (32). Even though you don't see it, it is an actual character. We sometimes call it a *nonprinting* character.

9.9. Strings are Immutable

One final thing that makes strings different from some other Python collection types is that you are not allowed to modify the individual characters in the collection. It is tempting to use the [] operator on the left side of an assignment, with the intention of changing a character in a string. For example, in the following code, we would like to change the first letter of greeting.

Save & Run	Original - 1 of 1	Show in Code	Lens Share Code
<pre>1 greeting = 1 2 greeting[0] 3 print(greet: 4</pre>		ROR!	
	Activity: 9.9.1 Ac	tiveCode (cg08_imm1)	
Frror			
TypeError: 'str'	does not support item	assignment on line 2	

Instead of producing the output Jello, world!, this code produces the runtime error TypeError: 'str' object does not support item assignment.

Strings are **immutable**, which means you cannot change an existing string. The best you can do is create a new string that is a variation on the original.

<pre>1 greeting = "Hello, world!" 2 newGreeting = 'J' + greeting[1:] 3 print(newGreeting) 4 print(greeting) # same as it was 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5</pre>	ave & Run	Original - 1 of	1	Show in CodeLer	s Share Code	÷
	newGreeting = print(newGree	= 'J' + greeting[1 eting)		las		
	•					
Activity: 9.9.2 ActiveCode (ch08_imm2)		Activity: 9.9.	2 ActiveCode	(ch08_imm2)		

The solution here is to concatenate a new first letter onto a slice of greeting. This operation has no effect on the original string.

Instead of producing the output Jello, world!, this code produces the runtime error TypeError: 'str' object does not support item assignment.

Strings are **immutable**, which means you cannot change an existing string. The best you can do is create a new string that is a variation on the original.

<pre>1 greeting = "Hello, world!" 2 newGreeting = 'J' + greeting[1:] 3 print(newGreeting) 4 print(greeting) # same as it was 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5</pre>	ave & Run	Original - 1 of	1	Show in CodeLer	s Share Code	÷
	newGreeting = print(newGree	= 'J' + greeting[1 eting)		las		
	•					
Activity: 9.9.2 ActiveCode (ch08_imm2)		Activity: 9.9.	2 ActiveCode	(ch08_imm2)		

The solution here is to concatenate a new first letter onto a slice of greeting. This operation has no effect on the original string.

9.10. Traversal and the for Loop: By Item

A lot of computations involve processing a collection one item at a time. For strings this means that we would like to process one character at a time. Often we start at the beginning, select each character in turn, do something to it, and continue until the end. This pattern of processing is called a **traversal**.

We have previously seen that the for statement can iterate over the items of a sequence (a list of names in the case below).



Activity: 9.10.1 ActiveCode (ch08_4)

Recall that the loop variable takes on each value in the sequence of names. The body is performed once for each name. The same was true for the sequence of integers created by the range function.

Save & Run	Original - 1 of 1	Show in CodeLens	Share Code
1 for avalue in 2 print(ava 3	range(10): lue)		
0 1 2 3 4 5 6 7 8 9			
	Activity: 9.10.2 Activ	eCode (ch08_5)	



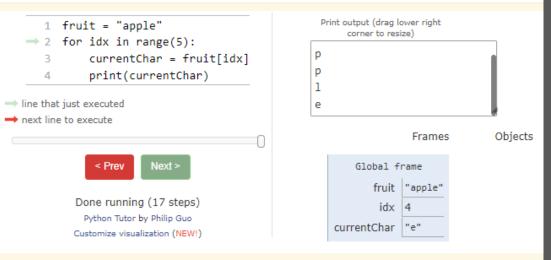
Activity: 9.10.3 ActiveCode (ch08_6)

The loop variable achar is automatically reassigned each character in the string "Go Spot Go". We will refer to this type of sequence iteration as **iteration by item**. Note that it is only possible to process the characters one at a time from left to right.

9.11. Traversal and the for Loop: By Index

It is also possible to use the **range** function to systematically generate the indices of the characters. The **for** loop can then be used to iterate over these positions. These positions can be used together with the indexing operator to access the individual characters in the string.

Consider the following codelens example.



Activity: CodeLens 9.11.1 (ch08_7)

The index positions in "apple" are 0,1,2,3 and 4. This is exactly the same sequence of integers returned by range(5). The first time through the for loop, idx will be 0 and the "a" will be printed. Then, idx will be reassigned to 1 and "p" will be displayed. This will repeat for all the range values up to but not including 5. Since "e" has index 4, this will be exactly right to show all of the characters.

In order to make the iteration more general, we can use the <u>len</u> function to provide the bound for <u>range</u>. This is a very common pattern for traversing any sequence by position. Make sure you understand why the range function behaves correctly when using <u>len</u> of the string as its parameter value.

<pre>1 fruit = "apple" 2 for idx in range(len(fruit)): 3 print(fruit[idx]) 4</pre>	Save & Run	Original - 1 of 1	Show in CodeLens	Share Code
	1 fruit = "ap 2 for idx in 3 print(f	ple" range(len(fruit)):	Show in CodeLens	

Activity: 9 11 2 ActiveCode (cb08 7b)

9.12. Traversal and the while Loop

The while loop can also control the generation of the index values. Remember that the programmer is responsible for setting up the initial condition, making sure that the condition is correct, and making sure that something changes inside the body to guarantee that the condition will eventually fail.

	Save & Run	Original - 1 of 1	Show in CodeLens	Share Code
1 2 3 4 5 6 7	<pre>position = 0 while positio print(fru position</pre>	e" n < len(fruit): it[position]) = position + 1		
a p l e				
		Activity: 9.12.1 Active	Code (ch08_7c)	
condition	n is false, and the	ition < len(fruit), so when body of the loop is not execute ch is the last character in the st	d. The last character access	

9.13. The in and not in operators

The in operator tests if one string is a substring of another:

Save & Run	Original - 1 of 1	Show in CodeLens	Share Code
1 print('p' in 2 print('i' in 3 print('ap' i 4 print('pa' i 5	'apple') n 'apple')		
			4
True False True False			
	Activity: 9.13.1 Active	eCode(chp8 in1)	

The not in operator returns the logical opposite result of in .

Save & Run	Original - 1 of 1	Show in CodeLens	Share Code
1 print('x' not in 2	'apple')		
True			4
	Activity: 9.13.3 ActiveC	code (chp8_in3)	