

Advanced Strings

Many interesting problems involve manipulating sequences of text data. You've learned about strings before, but this activity provides a more in-depth look at what strings can do.

Model 1 Indexing and Slicing

A string is a sequence of characters in single quotes (') or double quotes ("). You can access individual characters using square brackets (e.g., dna[0]).

You can also use *slice notation* (e.g., `dna[4:8]`) to refer to a *range* of characters. All types of sequences (including `list` and `tuple`) support indexing and slicing.

Python code	Shell output
<code>dna = 'CTGACGACTT'</code>	
<code>dna[5]</code>	'G'
<code>dna[10]</code>	IndexError: index out of range
<code>len(dna)</code>	10
<code>dna[:5]</code>	'CTGAC'
<code>dna[5:]</code>	'GACTT'
<code>dna[5:10]</code>	'GACTT'
<code>triplet = dna[2:5]</code>	
<code>print(triplet)</code>	GAC
<code>dna[-5]</code>	'G'
<code>dna[-10]</code>	'C'
<code>dna[: -5]</code>	'CTGAC'
<code>dna[-5:]</code>	'GACTT'
<code>triplet = dna[-4:-1]</code>	
<code>print(triplet)</code>	'ACT'

Questions (20 min)

Start time:

1. What is the *positive index* of each character in the dna string? Check your answers above.

Character:	C	T	G	A	C	G	A	C	T	T
------------	---	---	---	---	---	---	---	---	---	---

Index:

2. What is the *negative index* of each character in the dna string? Check your answers above.

Character:	C	T	G	A	C	G	A	C	T	T
Index:										

3. Based on the previous two questions, what values are `dna[2]` and `dna[-2]`? Explain your answers in general terms.

4. Explain the `IndexError` in Model 1. What is the range of indexes for the dna string?

5. Consider the notation of the operator `[m:n]` for slicing the string.

- a) Is the value at the start of the resulting string the same as the value at index `m`? (i.e., `dna[m]`)
If not, describe what it is.
- b) Is the value at the end of the resulting string the same as the value at index `n`? (i.e., `dna[n]`)
If not, describe what it is.
- c) Explain what the code means when only a single number is referenced when in a slice, such as `[m:]` or `[:n]`.

6. What is the simplest way to get the first three characters of `dna`? What is the simplest way to get the last three characters?

7. Write a Python expression that slices `'GACT'` from `dna` using positive indexes. Then write another expression that slices the same string using negative indexes.

8. Write a Python assignment statement that uses the `len` function to assign the last letter of `dna` to the variable `last`.

9. Write a Python assignment statement that uses a negative index to assign the last letter of dna to the variable last.

Model 2 Common String Methods

Strings have *methods* (built-in functions) that can be called using dot notation. The following table shows three examples: `lower()`, `split()`, and `replace()`.

Python Shell (10 min)

Start time:

Divide into pairs to complete the table. Have one team member run each line in a Python shell. Have another team member record the results. Briefly discuss the results while you work.

Python code	Shell output
<code>dna = 'CTGACGACTT'</code>	
<code>dna.lower()</code>	
<code>print(dna)</code>	
<code>lowercase = dna.lower()</code>	
<code>print(lowercase)</code>	
<code>dnalist = list(dna)</code>	
<code>print(dnalist)</code>	
<code>dnalist.reverse()</code>	
<code>print(dnalist)</code>	
<code>type(dna)</code>	
<code>dna = dna.split('A')</code>	
<code>print(dna)</code>	
<code>type(dna)</code>	
<code>dna.replace('C', 'g')</code>	
<code>print(dna[0])</code>	
<code>type(dna[0])</code>	
<code>dna[0].replace('C', 'g')</code>	
<code>print(dna)</code>	

Questions (15 min)

Start time:

10. Does the `lower` method change the contents of the `dna` string? Justify your answer.
11. Describe the `list` function—what does `list(dna)` return in Model 2?
12. Why is it possible to call the `replace` method on `dna[0]` but not `dna`?
13. Consider the application of a method on a variable:
 - a) Does a string variable change after applying a method? Provide justification.
 - b) Does a list variable change after applying a method? Provide justification.
 - c) Identify the data type that is *immutable* (i.e., the value never changes).
14. Write a single statement to change the final contents of `dna` to `['CTG', 'cc', 'CTT']`. Confirm that your code works in a Python Shell.
15. Why do you think Python has a `replace` method for strings but not for lists?

16. You can view the methods available to an object by calling the `dir()` or `help()` function. Call `dir(dna)` and `help(dna)` in a Python shell, and describe the difference between the two.

17. See <https://docs.python.org/3/library/stdtypes.html#string-methods> for a summary of all string methods. Name several other string methods not shown in Model 2.