Strings and Characters

\[ s[i] \]
Strings

Strings are one of Python’s primary data types.

Strings can be used to represent anything that can be encoded as text.

Examples: symbols, words, text files, Python programs, and Web pages.

We’ve already been using strings quite a bit and now will focus on accessing and processing them.
Indexing

$s[i]$  Strings are defined as ordered collections of characters

Therefore, we can access these characters by position with “indexing”

Offsets start at 0 and end at one less than the length of the string

Offset can also be specified negatively
## Introduction to Computer Programming

### Strings

#### Strings and Characters

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Python</td>
<td>p</td>
<td>y</td>
<td>t</td>
<td>h</td>
<td>o</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>-6</td>
<td>-5</td>
<td>-4</td>
<td>-3</td>
<td>-2</td>
<td>-1</td>
</tr>
</tbody>
</table>
Characters

ord('a')

All characters have a corresponding character code.

Python uses Unicode to assign character codes.

Unicode is a common standard for encoding thousands of symbols and characters from many languages.
Escape Characters

“Escape characters” are a special notation for whitespace and other unprintable characters.

Used for characters without a standard symbol or symbols Python sets aside.

Escape characters begin with a backslash.

Escape sequences are considered one character.
Slicing Strings

`s[begin:end]`

Slicing is for extracting substrings from a string.

To slice a string, indicate the first character you want and one past the last character.

First index location is always 0.

Last index location is always the length of the string minus one.
Introduction to Computer Programming
CSCI-UA 2

String Functions

s.function()
Regular Expressions

A “regular expression” is a special sequence of characters that helps you match or find other strings

Python has a regular expressions module, “re,” in its standard library

Used to efficiently and compactly describe a set of strings

Regular expressions operators and functions can be combined to perform powerful string processing tasks
Regular Expressions

Operators

xy?
Describes either x or xy (y is optional)

x | y
Describes x or y

x*
Describes an infinite set of strings

x+
Describes an infinite set of one or more strings

Use round brackets to indicate what substring an operator should apply to
Introduction to Computer Programming
CSCI-UA 2

Strings
Strings and Characters