Strings
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 will now focus on accessing and processing them.
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.

The `ord()` function can be used to get the Unicode number for any character.

The `chr()` function can be used to get the character for any Unicode number.
### Decimal  Hex  Char | Decimal  Hex  Char | Decimal  Hex  Char | Decimal  Hex  Char
---|---|---|---
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
<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><strong>p</strong></td>
<td></td>
<td></td>
<td><strong>y</strong></td>
<td></td>
<td></td>
<td><strong>t</strong></td>
</tr>
<tr>
<td></td>
<td><strong>h</strong></td>
<td></td>
<td></td>
<td><strong>o</strong></td>
<td></td>
<td><strong>n</strong></td>
</tr>
<tr>
<td><strong>-6</strong></td>
<td><strong>-5</strong></td>
<td><strong>-4</strong></td>
<td><strong>-3</strong></td>
<td><strong>-2</strong></td>
<td><strong>-1</strong></td>
<td></td>
</tr>
</tbody>
</table>
Slicing Strings

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

Strings
Strings and Characters

String Methods

s.method()
Escape Characters

\n
“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.