

02

Working with Files, Texts, and Regular Expressions

Python Programming for Linguists

Ingo Kleiber, 2020

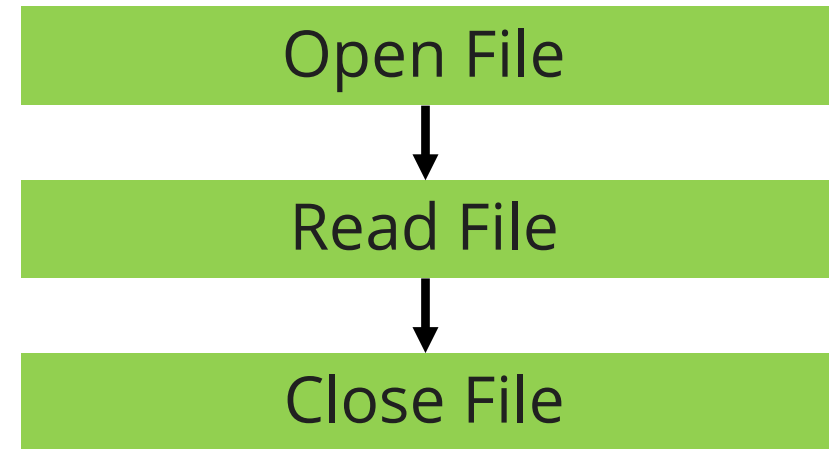
1. Reading and Writing Files

```
with open('file.txt', 'r') as f:  
    data = f.read()
```

```
data = f.read(10)
```

```
data = f.readlines()
```

mode



Character	Meaning
'r'	open for reading (default)
'w'	open for writing, truncating the file first
'x'	open for exclusive creation, failing if the file already exists
'a'	open for writing, appending to the end of the file if it exists
'b'	binary mode
't'	text mode (default)
'+'	open for updating (reading and writing)

<https://docs.python.org/3/library/functions.html#open>

2. Working with Text

Note: Since Python 3.0, str are stored as **Unicode** (UTF-8 by default)

Strings (str objects) are **sequences of characters** (Unicode code points)

`s = 'ABC'`

0	1	2
A	B	C

`print(s[0:2])` → 'AB'

start:stop

2. Working with Text

`s = 'Hello World'`

String Methods

- `s.upper()` & `s.lower()`
- `s.find('World')`
- `s.isdigit()`
- `s.split()`
- `s.replace(find, replace)`
- ...

3. Regular Expressions

```
import re
```

```
re.search(pattern, string)
```

```
re.findall(pattern, string)
```

```
re.split(pattern, string)
```

```
re.sub(pattern, rpattern, string)
```



Match Objects

4. TextDirectory

