

Files and Dictionaries

In Python reading files is no problem...

Files

```
>>> f = open( 'spam.txt', 'r' )
      opens the file for reading and calls it f

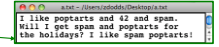
>>> text = f.read()
      reads the whole file and calls it z

alternative...
>>> text = f.read(50)
      reads the next 50 bytes from the file

>>> f.close()
      closes the file (optional)

>>> text
'I like popartars and 42 and spam.\nWill I ...'
      all the text (as one big string)

>>> word_list = text.split()
['I', 'like', 'popartars', ... ]
      text.split() returns a list of each "word"
```



Count how many words are in a given file

```
>>> f = open( 'greeting.txt', 'w' )
      opens the file for writing and calls it z

>>> text = "Hello, world!"

>>> f.write(text)
      Writes string text to z

>>> f.close()
```

```
def wc( filename ):
    """ word-counting program """
    f = open( filename, "r" )
    text = f.read()
    f.close()

    words = text.split()
    print("There are", len(words), "words.")
```

Try file_ops.py

Try wc.py

Dictionaries

1999 Feb 16	2000 Feb 4	Rabbit
1998 Jan 28	1999 Feb 15	Tiger
1997 Feb 7	1998 Jan 27	Ox
1996 Feb 19	1997 Feb 6	Rat

In Python a **dictionary** is a set of **key - value** pairs.

```
>>> d = {}
      creates an empty dictionary, d

>>> d[1996] = 'Rat'
      1996 is the key
      'Rat' is the value

>>> d[1997] = 'Ox'
      1997 is the key
      'Ox' is the value

>>> d
{1996: 'Rat', 1997: 'Ox'}
      ← Curly! And colon!

>>> d[1996]
'Rat'
      Retrieve data like lists...

>>> d[1969]
key error
      It's a sequence where the index can be any immutable key.
```

More on dictionaries

Dictionaries have lots of built-in **methods**, or functions:

```
>>> d = {1996: 'Rat', 1997: 'Ox'}

>>> 1996 in d
True
      in checks if a key is present

>>> 1969 in d
False

>>> len(d)
2
      len returns the # of keys

>>> list(d.keys())
[ 1996, 1997 ]
      d.keys returns a list of all keys

>>> list(d.values())
[ 'Rat', 'Ox' ]
      d.values returns a list of all keys

>>> list(d.items())
[ (1996, 'Rat'), (1997, 'Ox') ]
      d.items returns a list of all key, value pairs
```

```
def wc( filename ):
    """ word-counting program """
    f = open( filename )
    text = f.read()
    f.close()

    words = text.split()
    print("There are", len(words), "words.")

    d = {}
    for w in words:
        if w not in d:
            d[w] = 1
        else:
            d[w] += 1

    print ("There are", len(d), "distinct words.\n")
    return len(d) # return the number of distinct words
```

file handling

word counting

Tracking the number of occurrences of each word with a dictionary, d.

Try word_freq_count.py and dict_wc.py

What if we wanted the number of *different* words in the file? This would be the author's *vocabulary size*, instead of the total word count.

Vocabularists?

Shakespeare used **31,534 different words** and a grand total of 884,647 words counting repetitions (across his works)

<http://www.math.cudenver.edu/~vbruggip/shakespeare.html>

Active vocabulary estimates range from 10,000-60,000.
Passive vocabulary estimates are much higher.

Many Shakespearean contributions:

Adjectives	Shakespeare	
serial	suspicious	baseless
beached	bloodstained	blushing
circumstantial	conspicuous	deafening
disgraceful	domineering	stragg
epileptic	equivocal	eventful
fashionable	foregone	frugal
generous	gloomy	gnarled

One contemporary author in the Oxford Eng. Dictionary... which word?

Vocabularists?

Shakespeare used **31,534 different words** and an estimated vocabulary of 66,534 words (across his works)

<http://kottke.org/10/04/how-many-words-did-shakespeare-know>

Average English speaker knows 10,000 to 20,000 words.

Many Shakespearean contributions:

Adjectives	Shakespeare	
serial	suspicious	baseless
beached	bloodstained	blushing
circumstantial	conspicuous	deafening
disgraceful	domineering	stragg
epileptic	equivocal	eventful
fashionable	foregone	frugal
generous	gloomy	gnarled

"Muggle" goes into Oxford English Dictionary
 JK Rowling's word for non-wizards - "muggle" - has made it into the new edition of the Oxford English Dictionary (OED).
 The draft definition according to the dictionary's website says:
 • **Muggle**: invented by JK Rowling (b. 1965), British author of children's fantasy fiction (see quot. 1997).
 In the fiction of JK Rowling: a person who possesses no magical powers. Hence in allusive and extended uses: a person who lacks a particular skill or ability, or who is regarded as inferior in some way. J. K. Rowling