

Python Lesson 6: File input and output

vanderbi.it/py

Steve Baskauf



Main goals for today

- Acquire **template code** for input of:
 - plain text file
 - CSV file as list of lists (plus output again)
 - CSV file as list of dictionaries
- Understand the role of **newline** characters (`\n`) as delimiters for lines
- Be aware of the importance of **UTF-8 encoding**
- Understand the connection between CSV text structure and how it's rendered as a table
- Become familiar with csv module `.reader()`, `.writer()`, and `.DictReader()` methods

Some overall points

- I don't expect you to remember the details (just look them up)
- You should understand the big picture so that you can troubleshoot when problems come up
- I recommend either LibreOffice or OpenOffice for working with CSVs. Excel is **NOT** recommended.

UTF-8 character encoding

- **Unicode** is a system for representing 150 character sets by assigning numbers to their characters.
- **UTF-8** is a clever way to encode unicode "code points" (numbers) as bytes in a file.
- **ASCII**, the original encoding scheme for 95 Latin characters, numbers, and symbols is compatible with UTF-8 (all ASCII files are valid UTF-8)
- Escaped characters are used to represent most non-ASCII characters (e.g. `\u00f6`, the unicode codepoint)
- **Always use UTF-8 if it's an option.**

File objects

- A new kind of Python object
- Created by the `open()` function
- Have methods for reading and writing:
 - `.write()`
 - `.read()`
 - `.readlines()`
- Are iterable one line at a time

Writing and reading examples

- The newline character `\n` plays an important role in structuring files
- Newlines are inserted automatically by the `print()` function.
- Newlines are the basis of the iterable pieces of a file object (i.e. lines in the file)
- Newlines are the separator for the `.readlines()` method.

Examining the structure of CSV files

- Tools:
 - code or text editor (VS Code, Atom, or TextEdit/Notepad if you are desperate)
 - spreadsheet editor (LibreOffice, OpenOffice, or Excel if you are desperate)
 - GitHub standard rendering vs. raw
- Use Finder/File Explorer to navigate to the base folder/directory where Python is running (probably your home directory, but wherever the `.py` file is saved)
- On Jupyter notebooks it's the directory where the notebook lives.

Example

- See example on webpage under "CSV files"

```
given_name,family_name,username,student_id  
Jimmy,Zhang,rastaman27,37258  
Ji,Kim,kimji8,44947  
Veronica,Fuentes,shakira<3,19846
```

students.csv - OpenOffice Calc

File Edit View Insert Format Tools Data Window Help

Arial 10 B I U

D4 = 19846

	A	B	C	D	E	F
1	given_name	family_name	username	student_id		
2	Jimmy	Zhang	rastaman27	37258		
3	Ji	Kim	kimji8	44947		
4	Veronica	Fuentes	shakira<3	19846		
5						
6						
7						
8						
9						
10						

Sheet1

Sheet 1 / 1 Default STD Sum=19846 100 %

GitHub rendered page vs. raw

- Go to the "Reading into dictionaries" section of the web page and click on the "this file" link in the second paragraph.
- The default page renders the CSV file as a table
- Click on the Raw button to see the actual contents of the file.

Gotcha notes

- Saving a CSV for the first time is critical for determining the delimiter and character encoding
- Delimiters are sometimes tabs (TSV) or pipes (|)
- The CSV file scheme has special tricks for handling strings that contain the field delimiter (comma) or the text delimiter (double quotes).
- Best not to manage parsing or writing CSVs the hard way – use an editor or Python library functions.
- Excel will always read in CSV text like "1-26" as dates like "January 26". There is no way to turn this off!

The Python csv module

- `.reader()` and `.DictReader()` methods create iterable items by interpreting input text lines as lists (for a table row) or dictionaries (about the subject of the table row).
- To reuse the items row items, append them to a list for later reference.
- If you don't care about the details, just use the `readCsv()` and `readDict()` functions in the examples for input into a list of lists or list of dictionaries respectively.
- See the `writeCsv()` function example to save a list of lists as a CSV (but no error trapping).

Writing and reading examples