07 - Array formatting

1. Array design

So far, we have been limited to a rather cumbersome printout. It's good enough to start with, but it soon becomes unwieldy because some things just can't be done (easily) with it. For a start, you have to negotiate with it not to go to a new line after each printout and not to add spacebetween the things we print out.

If we had a dictionary that said how many kilometres someone had travelled on one of their bicycles,

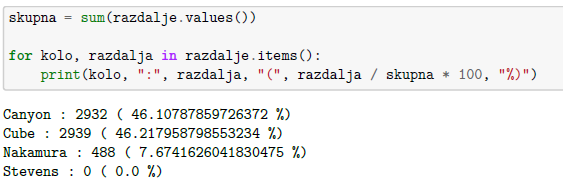


and wanted to print it out, print would add spaces before the colons.

Ein Bild, das Text, Screenshot, Schrift, Reihe enthält.

Automatisch generierte Beschreibung

If you also wanted to write down in brackets what proportion of the journey was made on which bicycle, you would get something like this indigestible:



Too many decimals, those awful spaces ... maybe some alignment would be nice, no?

The bad news: while we can tell print not to use extra spaces, it really can't be bothered with alignment and decimals.

Most of the text printed or otherwise displayed by programs contains fixed text with strings or numbers embedded in it. If the computer displays "Queen.py file not found", the embedded text is "queens.py", which is probably different each time, depending on which file is being (unsuccessfully) searched for. In "Today is 60 Fahrenheit, i.e. 16 degrees Celsius", the numbers 60 and 16 are inserted, depending on what time of year the program is run (and how much we have already managed to

the planet has warmed up), but the rest of the text is always the same.

Most computer languages allow you to construct strings by writing a "constant" string, marking the places where something needs to be inserted. Then we tell it in some way what to insert.

Popular compiled languages have mostly adopted the way C was invented, which is a bit more cumbersome, but newer languages, especially scripting languages, which are closer to the web, usually allow simpler and more powerful ways of assembling strings from patterns.

One of the principles of Python is that there is one and only one obvious way to do each thing. This is good because, as a consequence, all programmers write similar programs and therefore collaborate more easily. But it's in string formatting that this has been subverted: strings can be formatted in three different ways.

This is because certain ideas (not only in Python) have matured over the years, but you can't just discard (older) ways of doing things, because then older programs wouldn't work in newer versions of Python. A serious Python programmer needs to know all three modes, and the last, most modern one will do for us. In fact, I only use this one.

* 1. **F-arrays**

Suppose we have

Ein Bild, das Text, Schrift, Screenshot, Zahl enthält.

Automatisch generierte Beschreibung

Now we would like to make, in fact, a set like this:

Ein Bild, das Text, Schrift, Screenshot, Reihe enthält.

Automatisch generierte Beschreibung

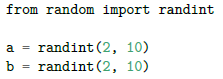
except that we want Python to insert the values of the variables fahr and celz in the places marked {fahr} and {celz}.

This is what f-strings do. An f-string is like a normal string, except that we prefix the first quotation mark with the letter f (as a format). In such strings, the curly brackets have a special meaning: what is in them is (computed and) inserted.

Ein Bild, das Text, Schrift, Screenshot, Reihe enthält.

Automatisch generierte Beschreibung

In a homework assignment - oh, how embarrassing that was! - we wrote





By first outputting and then requesting the answer, we avoided having to laboriously assemble the string when calling the input function:

Ein Bild, das Text, Schrift, Screenshot, Reihe enthält.

Automatisch generierte Beschreibung

Now that we know the f-array method, we know a simpler way:



(Incidentally, let's mention the two older ways of formatting strings that we mentioned at the beginning: in C-style, we would write c = input("What is %sx%s? " % (a, b)), and in the newer style c = input("How many is {}x{}? ".format(a, b)). What they both have in common is that they just mark the places in the string where to insert, and then somewhere else, after the end of the string, we have to list the things we're inserting. With the newer style, we can also add variable names into the wrapped brackets by some weird trick, but it's nowhere near as convenient as the new style.)

Ein Bild, das Text, Schrift, Screenshot, Reihe enthält.

Automatisch generierte Beschreibung

**1.2 Expressions in f-arrays**

It's not just the values of variables that we put into strings: we can write entire expressions in wrapped brackets.

Ein Bild, das Text, Screenshot, Schrift, Reihe enthält.

Automatisch generierte Beschreibung

Now back on the bikes!

Ein Bild, das Text, Screenshot, Schrift, Reihe enthält.

Automatisch generierte Beschreibung

Partial success. We got rid of spaces, but above all the printout has become clearer. Let's compare:



To make it even clearer, we can write

Ein Bild, das Text, Screenshot, Schrift, Zahl enthält.

Automatisch generierte Beschreibung

Now f-string is obviously a pattern into which we insert values.

What remains is the battle with decimals.

**1.3 How we form things**

If we want to tell how a number (or string, or whatever) should be written out, we add a colon to the expression, followed by a description of the format.

This is what we will most often need to format non-integer numbers. It goes like this:

Ein Bild, das Text, Schrift, Screenshot, Reihe enthält.

Automatisch generierte Beschreibung

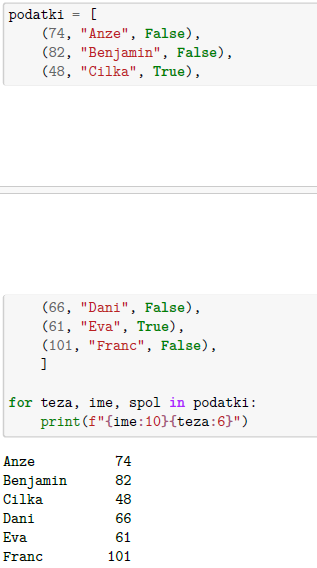
A description of the format is given here 4.1f. Here 4.1 means that we would like to have a four-place format, with one place reserved for the decimal point. For the letter f, imagine that it stands for float. If we leave too little space for a number, the printout will take up more space.

Ein Bild, das Text, Schrift, Screenshot, Reihe enthält.

Automatisch generierte Beschreibung

We asked for one decimal place, and we got it, but the total is wider than four places, because the number is just too long.

Specifying the number of decimal places only makes sense for non-integer numbers. If we are inserting strings or integers, we can only specify the width:



Here, we have not added the letter f after the number of digits, as they are not (non-integers). (There are other letters that could be added to tell Python what it's about, but we can safely forget them here.)

All names are written in ten characters; the missing space is filled with spaces. Numbers are written out in six digits, again with spaces filling the missing space.

The strings are aligned to the left - spaces are added after them. Most of the things we print (strings and numbers are not the only things that can be printed) are aligned like this. Numbers, on the other hand, are aligned to the right.

We can change this by adding <, > or ^ after the colon to tell it to be left, right or centre aligned. The reverse of the above would be achieved by

Ein Bild, das Text, Screenshot, Schrift, Zahl enthält.

Automatisch generierte Beschreibung

This is not very good. Let's add a space.

Ein Bild, das Text, Screenshot, Schrift, Zahl enthält.

Automatisch generierte Beschreibung

This is easier to read, but of course the first printout is still the most beautiful.

The alignment character can be preceded by a symbol to be used for padding - unless, of course, we are happy with a space. Let's align names to the left, numbers to the right, and fill the empty space with dots instead of spaces.

Ein Bild, das Text, Screenshot, Schrift, Zahl enthält.

Automatisch generierte Beschreibung

Any other character can be used instead of dots. When filling space with anything other than spaces like this, we need to add alignment characters (<, > or ^), even when we use them to choose the alignment we would like by default (right for numbers, left for everything else).

We won't go into all the details. There are many more things that can be specified. For numbers, we can ask for the sign to always be displayed, i.e. + when the number is positive. For numbers, we can specify that they should be printed in binary or hexadecimal... Anyone who needs more than that should Google it.

**1.3.1 Designing the wheels**

Back to bicycles. Most of the output is fine, but we would like to output the fractions with fewer decimal places - say one. It is enough to add :.1f.

Ein Bild, das Text, Schrift, Reihe, Screenshot enthält.

Automatisch generierte Beschreibung

The whole story of the bicycles is a bit longer. The information is taken from the file.

Ein Bild, das Text, Schrift, Screenshot enthält.

Automatisch generierte Beschreibung

In the dictionaries of ride, distance and height, we have the number of rides and the total distance and height travelled by each bicycle. What if we could display this in a table?

**Ein Bild, das Text, Screenshot, Schrift, Zahl enthält.

Automatisch generierte Beschreibung**

The new thing is the formatting of the percentages: we don't multiply by 100 and we don't use f (say, above. .1f), but %. Specifically, we wrote {carriage / sk\_carriage:>5.1%}. We right-aligned the quotient carriage / sk\_carriage, outputting it to five places with one decimal place, as a percentage. This means that Python will multiply the value by 100 and add the % sign.

Incidentally, notice that we have given short names to the variables voz, raz and vis, and similar, also short, names to the variables total number of trips, length and distance. Even with such short names, the line is already quite long.

The output already looks quite nominal; it will look even more so if we add column names.

Ein Bild, das Text, Screenshot, Schrift, Zahl enthält.

Automatisch generierte Beschreibung

It's so beautiful, it's a bit kitsch.

07b More on reading files

We have been reading files for a long time, since the second week. What I failed to tell you the other day is that files also have methods. Simply because we haven't come across methods before and there was no need to have them right next to files. Now we can. There are three or four of them, they are uncomplicated and, well, they are not necessarily very useful either. That is why I have kept them from you so far.

We've always read files with a for loop (until we learned about `csv.reader` and `csv.DictReader`, and we'll get to that).

**Ein Bild, das Text, Screenshot, Schrift enthält.

Automatisch generierte Beschreibung**

The file has, as written, methods.

**Ein Bild, das Text, Schrift, Screenshot, weiß enthält.

Automatisch generierte Beschreibung**

One is `read()`; this reads the whole file into a string.

**Ein Bild, das Text, Schrift, Reihe, Screenshot enthält.

Automatisch generierte Beschreibung**

You only have to do it once. The file is now "exhausted".

Ein Bild, das Text, Screenshot, Schrift, weiß enthält.

Automatisch generierte Beschreibung

If we want to do anything else with it, we'll have to reopen it. (There is a way to "roll it back", but we won't do that.)

Ein Bild, das Text, Schrift, Screenshot, weiß enthält.

Automatisch generierte Beschreibung

The second is `readlines()`; this reads all the lines of the file into a list of strings.

Ein Bild, das Text, Screenshot, Schrift, Reihe enthält.

Automatisch generierte Beschreibung

The third is `readline()`. This reads one line.

Ein Bild, das Text, Schrift, Screenshot, Reihe enthält.

Automatisch generierte Beschreibung

One after the other.

Ein Bild, das Text, Screenshot, Schrift, Zahl enthält.

Automatisch generierte Beschreibung

What is read is read and is not returned.

Of these three, you will most often need the last (`readline()`), and sometimes the first (`read()`). If you are interested in using `readlines()`, you will almost certainly be doing yourself a favour by looping over the file.

We'll use `readline()` when the file doesn't have a "homogeneous" format: different lines have different meanings, say, and it's practical for us to read them one by one.

"What is read is read," I wrote with this example in mind:

Ein Bild, das Text, Screenshot, Schrift enthält.

Automatisch generierte Beschreibung

We read the first line in the `column'; maybe it comes in handy, maybe we just wanted to get rid of it. The rest of the lines are read with `for`. The first one won't read to us because it's already read.

The fourth method of note is `close`.

Ein Bild, das Text, Screenshot, Schrift enthält.

Automatisch generierte Beschreibung

Once the file is closed, it can no longer be read :)

Why would anyone close a file? Keep it open! First of all: a program on MS Windows can only have 512 files open at the same time. I don't know about other operating systems. There used to be a limit on the number of files that could be opened at the same time in all programs together ... OK, do we care? Not really.

Python closes files by itself.

Ein Bild, das Text, Screenshot, Schrift, Reihe enthält.

Automatisch generierte Beschreibung

Since `f` now refers to another file and we can't read from the first one anyway, Python closes it.

Even more common is the situation when a file is opened inside a function. When the function is over, the file can no longer be accessed and Python closes it. In short: you won't call `close` very often.

**## Writing files**

After learning how to format strings, we printed out a very nice table.

Ein Bild, das Text, Screenshot, Schrift enthält.

Automatisch generierte Beschreibung

It's so beautiful that you could just write it in a file. The file should be called table-wheels.txt.

Let’s try it.

Ein Bild, das Text, Schrift, Screenshot enthält.

Automatisch generierte Beschreibung

Of course not. There is no file; there is nothing to open. Of course not: we would like to make a file.

We need to give the `open` function one more argument - the way we want to open the file. This can be `"r"` (read), `"w"` (write), or `"a"` (append to an existing file, as \*append\*). (We could add another `t` to say that it is a text file, not a binary file, `b`. We are not interested in that for this object.)

Ein Bild, das Text, Schrift, Screenshot, Reihe enthält.

Automatisch generierte Beschreibung

As `f` is a file, it has the usual file methods.

Ein Bild, das Text, Screenshot, Schrift, Reihe enthält.

Automatisch generierte Beschreibung

Of course not. We write to this file. Nothing that starts with `read` will work. Instead, `write` works for it (which we didn't even show before, when we were still reading, because - you guessed it - it wouldn't work, just like reading doesn't work here).

Ein Bild, das Text, Screenshot, Schrift, Reihe enthält.

Automatisch generierte Beschreibung

Everything is the same as before, but

- Replace `print` with `f.write`,

- add `\n` to the end of each line.

The `print` went to the new line itself, the `write` will not.

There are a couple of differences between `print` and `write`:

- `write` does not go to a new line by itself,

- `print` accepts multiple arguments, and they can be of any type, from strings to numbers to lists to sets to anything. `write` accepts a single argument and it must be a string.

- `print` does not return a result (it returns `None`), `write` returns the number of bytes written in this call. Why? I have no idea. But it's the `64` that's printed at the end: it's 6 + 3 \* 19 = 63 minuses and one \n.

If we now open the `tables-wheels.txt` file, we see that it is ... completely empty. So where did all those `write`s go? Including the last 64?

The computer doesn't write everything to the files all the time, it only writes it as it accumulates. Otherwise it would have to access the disk continuously, line by line; even in the era of SSDs this would be slow. And here ... it just hasn't built up yet. Right, but now we're done. How to tell Python it's done? Simple: close the file. :)



(There is another way: `f.flush()` writes what has accumulated so far, but does not close the file.)

Does this mean that when we write, we will always call `close()` to write the contents of the file? No. I don't usually call it. But that's why we talked about when Python closes a file when we were reading. If you write it inside a function, at the end of the function, when the local variable (say `f`) disappears, the file itself will be closed.

**## Opening files correctly**

I have never given this lecture before, but maybe it is a good time to start. For the last ten years or so, Python files should be opened like this:

Ein Bild, das Text, Screenshot, Schrift, Reihe enthält.

Automatisch generierte Beschreibung

The `with` statement works with things that exist within a block (the official name for this is "context"), and when the execution of the block finishes, they want to be notified because they have some "finishing work" to do. A file is already such a thing: it is opened at the beginning of a block, and when the block is over, it is notified and "closes" itself. For `time f`, we assign the variable `f` to the result of the `open` function.

The same applies to reading: even `csv.reader` should be officially called this way:

```

with open("wheels.txt") as f:

for lines in csv.reader(f):

... and so on

``

Think what you want. Understand if you want. You can open files with the normal `open` call and not close them (as we have been doing since the beginning of the semester), or you can close them with `close`. Ignore what it says in this section. Almost certainly nothing will happen to you. Or you can use `with`.