

How to a read large file, line by line in Python



430



179

I want to iterate over each line of an entire file. One way to do this is by reading the entire file, saving it to a list, then going over the line of interest. This method uses a lot of memory, so I am looking for an alternative.

My code so far:

```
for each_line in fileinput.input(input_file):
    do_something(each_line)

for each_line_again in fileinput.input(input_file):
    do_something(each_line_again)
```

Executing this code gives an error message: `device active`.

Any suggestions?

The purpose is to calculate pair-wise string similarity, meaning for each line in file, I want to calculate the Levenshtein distance with every other line.

python file-read

share improve this question

edited Jun 28 at 11:38



Peter Mortensen

12.8k 19 83 111

asked Nov 4 '11 at 13:26



384X21

2,477 3 11 16

Is that your real code indenting? The second `for` loop runs within the first one? Use four spaces for indentation, please. – eumiro Nov 4 '11 at 13:29

4 Why do you need to read the entire file again for each line? Maybe if you told what you're trying to accomplish someone might be able to suggest a better approach. – JJJ Nov 4 '11 at 13:32

add a comment

10 Answers

active

oldest

votes



1092



The correct, fully Pythonic way to read a file is the following:

```
with open(...) as f:
    for line in f:
        # Do something with 'line'
```

The `with` statement handles opening and closing the file, including if an exception is raised in the inner block. The `for line in f` treats the file object `f` as an iterable, which automatically uses buffered I/O and memory management so you don't have to worry about large files.

There should be one -- and preferably only one -- obvious way to do it.

share improve this answer

edited Jun 28 at 11:45



Peter Mortensen

12.8k 19 83 111

answered Nov 4 '11 at 13:46



Katriel

79.2k 15 98 141

11 yep, this is the best version with python 2.6 and above – Simon Nov 4 '11 at 13:54

3 I personally prefer generators & coroutines for dealing with data pipelines. – jldupont Nov 4 '11 at 14:30

2 what would be the best strategy if a file is a huge text file but with one line and the idea is to process words? – mfcabrera Dec 18 '13 at 14:48

3 Could someone explain how `for line in f:` is working? I mean, how iterating over file object is possible? – haccks Feb 9 '15 at 13:56

7 If you iterate over an object, Python looks up in the list of object methods a special one called `__iter__`, which tells it what to do. File objects define this special method to return an iterator over the lines. (Roughly.) – Katriel Feb 10 '15 at 12:14

show 12 more comments