

Forum:[Hack Forums](#)

Thread: [Python reading file utf-8 uses a lot of memory](#)

Ri?dle - May 4, 2017, 7:11 am

When I add utf-8 encoding it uses about 3000mb of memory (and freezes my PC) the exmaple below works fine,

```
Code pool = ThreadPoolExecutor(max_workers=10)
```

```
with open("data.txt") as f:
```

```
    for line in f:
```

```
        pool.submit(thread, line)
```

```
pool.shutdown(wait=True)but when I add utf-8 encoding it basically crashes my PC
```

```
Code pool = ThreadPoolExecutor(max_workers=10)
```

```
with open("data.txt", encoding='utf8') as f:
```

```
    for line in f:
```

```
        pool.submit(thread, line)
```

```
pool.shutdown(wait=True)
```

HFDone - May 4, 2017, 8:49 am

python 3 is unicode by default maybe its the reason

Ri?dle - May 4, 2017, 9:21 am

(May 4, 2017 08:49 AM)HFDone Wrote: [\[CITE\]](#) python 3 is unicode by default maybe its the reason well I get this error when I dont use encoding='utf8'

Code UnicodeDecodeError: 'charmap' codec can't decode byte 0x8d in position 854: character maps to <undefined>

HFDone - May 4, 2017, 10:18 am

yea as i said its unicode problem try str(line, 'utf-8')

Ri?dle - May 4, 2017, 10:54 am

(May 4, 2017 10:18 AM)HFDone Wrote: [\[CITE\]](#) yea as i said its unicode problem try str(line, 'utf-8') like so?

```
Code with open("data.txt") as f:
```

```
    for line in f:
```

```
        line = str(line, 'utf-8')
```

```
        pool.submit(thread, line)
```

```
pool.shutdown(wait=True)
```

Pillus - May 4, 2017, 11:31 am

Is that file large? Because it will load the file into memory, so you will have to read the file and iterate over the chunks you want to keep the memory usage low.

Look into codecs:

<http://stackoverflow.com/questions/16130904/open-and-codecs-open-in-python-2-7-behave-strangel-y-different>

Ri?dle - May 4, 2017, 2:30 pm

(May 4, 2017 11:31 AM) Pillus Wrote: [\[CITE\]](#) Is that file large? Because it will load the file into memory, so you will have to read the file and iterate over the chunks you want to keep the memory usage low.

Look into codecs:

<http://stackoverflow.com/questions/16130904/open-and-codecs-open-in-python-2-7-behave-strangely-different>

yeah it's a very large file, I'll look at this

Pillus - May 4, 2017, 2:34 pm

(May 4, 2017 02:30 PM) Ri?dle Wrote: [\[CITE\]](#) yeah it's a very large file, I'll look at this

Yeah then that is the problem. A normal file would load the entire thing into memory, so the larger the file, the more memory usage :)

Ri?dle - May 4, 2017, 2:35 pm

(May 4, 2017 02:34 PM) Pillus Wrote: [\[CITE\]](#) Yeah then that is the problem. A normal file would load the entire thing into memory, so the larger the file, the more memory usage :)
thank you <3

Lambasoft - May 4, 2017, 5:59 pm

You're loading the file 10 times in memory, in your case it seems the file size is around 300mb, loading it 10 times (one time in every worker ThreadPoolExecutor(max_workers=10)) requires 3000mb of RAM usage.

I'd suggest either lowering your max_workers, or splitting the file to 10 parts.

Good luck :)