

pandas 导出 excel 时遇到的非法字符问题

pandas 导出数据到 excel 时，可能会遇到异常字符导致导出失败

```
Traceback (most recent call last):
  File "20150107.py", line 36, in <module>
    encoding='GB18030', index=False)
  File "lib/python3.3/site-packages/pandas/util/decorators.py", line 88, in wrapper
    return func(*args, **kwargs)
  File "lib/python3.3/site-packages/pandas/core/frame.py", line 1260, in to_excel
    startrow=startrow, startcol=startcol)
  File "lib/python3.3/site-packages/pandas/io/excel.py", line 681, in write_cells
    xcell.value = _conv_value(cell.val)
  File "lib/python3.3/site-packages/openpyxl/cell/cell.py", line 353, in value
    self.bind_value(value)
  File "lib/python3.3/site-packages/openpyxl/cell/cell.py", line 267, in bind_value
    self.set_explicit_value(value, self.data_type)
  File "lib/python3.3/site-packages/openpyxl/cell/cell.py", line 238, in set_explicit_value
    value = self.check_string(value)
  File "lib/python3.3/site-packages/openpyxl/cell/cell.py", line 223, in check_string
    raise IllegalCharacterError
openpyxl.exceptions.IllegalCharacterError
```

如果是手动调用 xlwt 这种第三方库除了错可能没法找错误，但是从错误中我们看到错误是由 openpyxl 抛出的，我们试着从 openpyxl 中找解决方案

出错处的代码

```
value = value[:32767]
if next(ILLEGAL_CHARACTERS_RE.finditer(value), None):
    raise IllegalCharacterError
return value
```

意思很明显了，如果找非非法字符则抛出错误，ILLEGAL_CHARACTERS_RE 就定义了非法字符，它是这样的

```
ILLEGAL_CHARACTERS_RE = re.compile(r'[\000-\010][\013-\014][\016-\037]')
```

所以处理 content 时用这个正则去掉非法字符即可

```
content = ILLEGAL_CHARACTERS_RE.sub(r'', content)
```

结尾

openpyxl 大法好，完美处理现在导出 excel 出现的非法字符问题

非法字符大都是一些 ASCII 码，比如上面的那三个范围内的 ASCII 字符在 Excel 都属于非法字符

这些非法字符参见 [ASCII Characters](#)

Published: January 08 2015

Category: Python 13

Tags: Pandas 3 Python 10 Ascii 1

[Tweet](#)

[Follow @gsavl](#)

blog comments powered by Disqus

« [黑天鹅 Black swan : the impact of the highly improbable 电子书笔记](#) | [shaping up with angular js 笔记](#) »
[Back to Top](#)

about

- Cooper Luan - gc.suprs@gmail.com
- github.com/CooperLuan
- twitter.com/gsavl
- [CV\(zh-CN\)](#)
- [CV\(en\)](#)
- [Subscribe to RSS Feed](#)

[加关注](#)

Theme: [the_minimum](#) based on [Jekyll-bootstrap](#).
Powered by [Jekyll](#).