

Collate Module

Source: <http://coding.derkeiler.com/Archive/Python/comp.lang.python/2006-10/msg04303.html>

- *From:* Ron Adam <rrr@xxxxxxxxxxxx>
 - *Date:* Mon, 23 Oct 2006 07:35:30 -0500
-

I've made a few more changes to my little collate module.

There might be better ways to handle the options, or better choices for the options themselves. I tried to keep it as general as possible.

I think it should work with Unicode now too.

Any suggestions in making it faster will be appreciated.

Any minor improvements to wording, spelling, etc.. are also welcome.

Many thanks for all the terrific feed back and suggestions!

Cheers,
Ron

---start---

"""

Collate.py – Sorts lists of strings in various ways depending
options and locale.

Class: Collate(option_string)

Functions:

collate(text, option_string) -> collate in place

result = collated(text, option_string) -> get a collated copy

To use collate with your user locale you need to call setlocale()
before collating:

```
locale.setlocale(locale.LC_ALL, "")
```

This will set several global variables such as string
constants, time and monetary formats, as well as numerical
formatting to your user Locale. Instead of "", you can also

Collate Module

specify a specific locale.

The settings can be restored back to the default 'C' locale with the following:

```
locale.setlocale(locale.LC_ALL, 'C')
```

Example:

```
locale.setlocale(locale.LC_ALL, "")
tlist = [ 'Fruit', 'fruit-apple', 'fruit-bannana',
'nut-pecan', 'Nut' ]
collate(tlist, 'caps_first hyphen_as_space')
```

* For more examples see doctests in function test().

Author: Ron Adam, ron@xxxxxxxxxxxx

```
"""
import re
import locale
import string
__version__ = '0.03 (pre-alpha) 10/23/2006'
```

```
class Collate(object):
    """ A general purpose collator class.
```

```
Use: collate(text, option_string)
result = collated(text, option_string)
```

```
text = a list of strings
option_string = a string of space separated options.
```

Collation can be modified with the following options:

```
CAPS_FIRST -> Aaa, aaa, Bbb, bbb
LOWERCASE_FIRST -> aaa, Aaa, bbb, Bbb
HYPHEN_AS_SPACE -> hyphens separate words
PERIOD_AS_SPACE -> Periods separate numerals
NUMERICAL -> Digit sequences as numerals
IGNORE_COMMAS -> Allows commas in numerals
UNDERSCORE_AS_SPACE -> Underscore as white space
IGNORE_LEADING_WS -> Disregard leading white space
```

```
"""
option
```