

Unicode Babel

04 December 2019 · python, unicode

I've written about the joys of Unicode and software development [before](http://www.investigating-software.co.uk/search/label/unicode) (<http://www.investigating-software.co.uk/search/label/unicode>). Using unexpected data in your testing is usually a good way to test for text encoding issues. Finding and fixing these those bugs early could save your team from a host of other related issues and hackery.

Even if you don't expect to have unusual text content, this type of testing can help indicate if all your systems are configured consistently. Failure to do so can result in users seeing the dreaded [Mojibake](https://en.wikipedia.org/wiki/Mojibake) (<https://en.wikipedia.org/wiki/Mojibake>).

I've recently created a python package for generating random Unicode codepoints so they can be incorporated easily into your automated tests and tools. It's called [Unicode Babel](https://pypi.org/project/unicode-babel/#description) (<https://pypi.org/project/unicode-babel/#description>), and can be used to create a simple iterator for supplying 'international' text to your app:

```
from unicode_babel import tools, filters

genny = tools.CodePointGenerator()

for point in genny.random_codepoints(10, filters.filter_out_if_no_name):
    print(point)
```

Will output something like:

```
ق
Q
[]
[]
[]
[]
[]
[]
[]
[]
[]
```

Or you can integrate it with your existing tools like Selenium WebDriver, e.g.:

```
from unicode_babel import tools, filters
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
```

```
browser = webdriver.Chrome()
browser.get("https://www.google.com")

data_genny = tools.CodePointGenerator()
unusual_char = data_genny.get_random_codepoint(filters.filter_out_if_no_name)

search_box = browser.find_element_by_name("q")
search_box.send_keys(unusual_char + Keys.RETURN)
```

I hope it helps you with your testing, send me any [bugs](https://github.com/phoughton/unicode_babel/issues) (https://github.com/phoughton/unicode_babel/issues)!