

# Random text tool

29 July 2011 · automation, random, selection randomness, testing, tools, unicode

---

I recently blogged about [some of the tools I use](http://www.investigatingsoftware.co.uk/2011_06_01_archive.html) ([http://www.investigatingsoftware.co.uk/2011\\_06\\_01\\_archive.html](http://www.investigatingsoftware.co.uk/2011_06_01_archive.html)), and how some are so useful I keep using them. As I mentioned, randomness is pretty useful, and I have tools to help me generate random text.

A few of my readers requested a copy of my simple random text generating script, so I've decided to open it up for everyone to use and test. It will have bugs, like all software, please send details and I'll try and fix them.

If you are interested in what UTF-8 is and what all that Unicode stuff is about, there is a [great article by Joel Spolsky](http://www.joelonsoftware.com/articles/Unicode.html) (<http://www.joelonsoftware.com/articles/Unicode.html>) that explains all, and the [wikipedia page is ok](http://en.wikipedia.org/wiki/UTF-8) (<http://en.wikipedia.org/wiki/UTF-8>).

To use it...

First [download the script, its on GitHub](https://github.com/phoughton/Fuzz) (<https://github.com/phoughton/Fuzz>). The script is fairly short and is all in one file. You don't have to 'install it', its not a GEM.

Second, make sure you have [Ruby](http://www.ruby-lang.org/en/downloads/) (<http://www.ruby-lang.org/en/downloads/>) version 1.9 or greater. You need version 1.9, because Ruby didn't handle UTF-8 well in older versions.

Thirdly run the script like this:

```
ruby fuzzutf8.rb
```

That will give you some usage examples.

Typically you might use it like this:

```
ruby fuzzutf8.rb -c 100 -utf8_two_byte
δ[]]gϯBн[]Bđµq£Íñ[]ŷ[]ϩ[]Ywψī[]Ϙ[]YтAã:ä
~[]ϣ[]œ[]f[]ϛ[]ϣ[]Y[]ϩ[]È[]ب[]ؤ[]ج[]ξ[]ب[]ز[][]H[]ج[]ü[]ε[]ñ[]ϛ[]ô[]o[]ö[]æ[]ú[]ж[][]'[]R[][]Ù[]h[]-[]´[]ε[]ÿ[][]]H
```

That's a hundred code-points of UTF-8, in the two byte range.

As operating systems sometimes have issues copying and pasting these code-points (e.g. Windows XP) You may prefer to redirect the contents straight to a text file. e.g.:

```
ruby fuzzutf8.rb -c 100 -utf8_two_byte > random_utf8_file.txt
```

You can then treat the output like any other file. You can open it in a text editor, but beware many programs can not handle random UTF8 well, You may wish to save your work first.

The script also allows you to keep a record of the code-point numbers generated in a separate file ( -o option). These can then be replayed at a later date should you need to ( -r option ). The script doesn't do a lot of validation, so you can probably have great fun tripping the script itself up with dodgy inputs and arguments ;-)

I'm confident the script has many bugs, but it is nonetheless quite useful, and has been useful in helping me to locate other bugs. If you are interested in random text generators, There is a windows tool that I've also found useful, called **Babel** ([http://www.testingmentor.com/tools/tool\\_pages/babel.htm](http://www.testingmentor.com/tools/tool_pages/babel.htm)) from TestMentor. Babel allows you to choose the output by character set/codepoint range, which is also interesting.