

LANG, locale, unicode, setup.py and Debian packaging

Source: <http://coding.derkeiler.com/Archive/Python/comp.lang.python/2008-01/msg01379.html>

- *From:* Donn Ingle <donn.ingle@xxxxxxxxx>
 - *Date:* Sat, 12 Jan 2008 10:25:02 +0200
-

Hello,

I hope someone can illuminate this situation for me.

Here's the nutshell:

1. On start I call `locale.setlocale(locale.LC_ALL,"`), the `getlocale`.
2. If this returns "C" or anything without 'utf8' in it, then things start to go downhill:
 - 2a. The app assumes unicode objects internally. i.e. Whenever there is a "string like this" in a var it's supposed to be unicode. Whenever something comes into the app (from a filename, a file's contents, the command-line) it's assumed to be a byte-string that I `decode("utf8")` on before placing it into my objects etc.
 - 2b. Because of 2a and if the locale is not 'utf8 aware' (i.e. "C") I start getting all the old 'ascii' unicode decode errors. This happens at every string operation, at every print command and is almost impossible to fix.
3. I made the decision to check the locale and stop the app if the return from `getlocale` is (None,None).
4. My `setup.py` (`distutils`) also tests locale (because it then loads `gettext` to give localized information to the user during setup).
5. Because it's doing a `raise SystemExit` if the locale is (None,None) which happens if LANG is set to "C", the `setup.py` stops.
6. Someone is helping me to package the app for Debian/Ubuntu. During the bizarre amount of Voodoo they invoke to do that, the `setup.py` is being run and it is breaking out because their LANG is set to "C"
7. I have determined, as best I can, that Python relies on LANG being set to a proper string like `en_ZA.utf8` (`xx_YY.encoding`) and anything else will start Python with the default encoding of 'ascii' thus throwing the entire app into a medieval dustbin as far as i18n goes.
8. Since I can't control the LANG of the user's system, and I am relying on it containing 'utf8' in the locale results.. well I seem to be in a

LANG, locale, unicode, setup.py and Debian packaging

catch-22 here.

Does anyone have some ideas? Is there a universal "proper" locale that we could set a system to *before* the Debian build stuff starts? What would that be – en_US.utf8?

Any words of wisdom would help.

\d

.