Encodings

David Branner

Hacker School, New York City
March 22, 2013

Terms

- **encoding**

Terms

- **encoding**: process for converting information into patterns or symbols

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec**

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder

Terms

- **encoding**: process for converting information into patterns or symbols — has cognitive parallels
- **character encoding**: process for converting written characters into patterns or symbols
- **codec** = coder-decoder: tool (normally a program or routine) for changing encodings

The key point

Encodings collectively are like a system of types for characters and strings of characters.

They can cause problems if you don't realize that a particular encoding is required or being supplied.

Well known encodings

- ASCII

Well known encodings

- ASCII
- ISO

Well known encodings

- ASCII
- ISO's encodings

Well known encodings

- ASCII
- ISO's encodings: iso-8859-1, etc.
  (ISO = The International Organization for Standardization)

Well known encodings

- ASCII
- ISO's encodings: iso-8859-1, etc.
  (ISO = The International Organization for Standardization)
- Unicode

Well known encodings

- ASCII
- ISO's encodings: iso-8859-1, etc.
  (ISO = The International Organization for Standardization)
- Unicode Consortium's "UTF" encodings

Well known encodings

- ASCII
- ISO's encodings: iso-8859-1, etc.
  (ISO = The International Organization for Standardization)
- Unicode Consortium's "UTF" encodings: utf-8, etc.

What the well-known encodings encode

- ASCII

What the well-known encodings encode

- ASCII: English

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard:

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard:
  - letters `a-z` and `A-Z`

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard:
  - letters `a-z` and `A-Z`
  - numerals `0-9`

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard:
  - letters `a-z` and `A-Z`
  - numerals `0-9`
  - 33 common symbols

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard:
  - letters `a-z` and `A-Z`
  - numerals `0-9`
  - 33 common symbols
- ISO

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard:
  - letters `a-z` and `A-Z`
  - numerals `0-9`
  - 33 common symbols
- ISO:
  - iso-8859-1

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard:
  - letters `a-z` and `A-Z`
  - numerals `0-9`
  - 33 common symbols
- ISO:
  - iso-8859-1: most languages of Western Europe

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard:
  - letters `a`–`z` and `A`–`Z`
  - numerals `0`–`9`
  - 33 common symbols
- ISO:
  - iso-8859-1: most languages of Western Europe, including many Roman letters with diacritics (è, ö, ç, etc.)

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard:
  - letters `a`–`z` and `A`–`Z`
  - numerals `0`–`9`
  - 33 common symbols
- ISO:
  - iso-8859-1: most languages of Western Europe, including many Roman letters with diacritics (è, ö, ç, etc.)
  - iso-8859-2: covers Central Europe: more diacritics (ś, ł, č, etc.)

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard:
  - letters `a-z` and `A-Z`
  - numerals `0-9`
  - 33 common symbols
- ISO:
  - iso-8859-1: most languages of Western Europe, including many Roman letters with diacritics (è, ö, ç, etc.)
  - iso-8859-2: covers Central Europe: more diacritics (ś, ł, č, etc.)
  - iso-8859-5: covers Cyrillic

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard:
  - letters `a`–`z` and `A`–`Z`
  - numerals `0`–`9`
  - 33 common symbols
- ISO:
  - iso-8859-1: most languages of Western Europe, including many Roman letters with diacritics (è, ö, ç, etc.)
  - iso-8859-2: covers Central Europe: more diacritics (ś, ł, č, etc.)
  - iso-8859-5: covers Cyrillic; etc., etc.

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard:
  - letters `a`–`z` and `A`–`Z`
  - numerals `0`–`9`
  - 33 common symbols
- ISO:
  - iso-8859-1: most languages of Western Europe, including many Roman letters with diacritics (è, ö, ç, etc.)
  - iso-8859-2: covers Central Europe: more diacritics (ś, ł, č, etc.)
  - iso-8859-5: covers Cyrillic; etc., etc.
- UTF encodings

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard:
  - letters `a`–`z` and `A`–`Z`
  - numerals `0`–`9`
  - 33 common symbols
- ISO:
  - iso-8859-1: most languages of Western Europe, including many Roman letters with diacritics (è, ö, ç, etc.)
  - iso-8859-2: covers Central Europe: more diacritics (ś, ł, č, etc.)
  - iso-8859-5: covers Cyrillic; etc., etc.
- UTF encodings: "all" characters in all known writing systems

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard:
  - letters `a-z` and `A-Z`
  - numerals `0-9`
  - 33 common symbols
- ISO:
  - iso-8859-1: most languages of Western Europe, including many Roman letters with diacritics (è, ö, ç, etc.)
  - iso-8859-2: covers Central Europe: more diacritics (ś, ł, č, etc.)
  - iso-8859-5: covers Cyrillic; etc., etc.
- UTF encodings: "all" characters in all known writing systems — c. 70K CJK characters

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard:
  - letters `a`–`z` and `A`–`Z`
  - numerals `0`–`9`
  - 33 common symbols
- ISO:
  - iso-8859-1: most languages of Western Europe, including many Roman letters with diacritics (è, ö, ç, etc.)
  - iso-8859-2: covers Central Europe: more diacritics (ś, ł, č, etc.)
  - iso-8859-5: covers Cyrillic; etc., etc.
- UTF encodings: "all" characters in all known writing systems — c. 70K CJK characters, numerous Semitic alphabets and abugidas

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard:
  - letters `a`–`z` and `A`–`Z`
  - numerals `0`–`9`
  - 33 common symbols
- ISO:
  - iso-8859-1: most languages of Western Europe, including many Roman letters with diacritics (è, ö, ç, etc.)
  - iso-8859-2: covers Central Europe: more diacritics (ś, ł, č, etc.)
  - iso-8859-5: covers Cyrillic; etc., etc.
- UTF encodings: "all" characters in all known writing systems — c. 70K CJK characters, numerous Semitic alphabets and abugidas, Linear B

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard:
  - letters `a`–`z` and `A`–`Z`
  - numerals `0`–`9`
  - 33 common symbols
- ISO:
  - iso-8859-1: most languages of Western Europe, including many Roman letters with diacritics (è, ö, ç, etc.)
  - iso-8859-2: covers Central Europe: more diacritics (ś, ł, č, etc.)
  - iso-8859-5: covers Cyrillic; etc., etc.
- UTF encodings: "all" characters in all known writing systems — c. 70K CJK characters, numerous Semitic alphabets and abugidas, Linear B, Yí (Loloish)

What the well-known encodings encode

- ASCII: English — what you find on a standard US keyboard:
  - letters `a`–`z` and `A`–`Z`
  - numerals `0`–`9`
  - 33 common symbols
- ISO:
  - iso-8859-1: most languages of Western Europe, including many Roman letters with diacritics (è, ö, ç, etc.)
  - iso-8859-2: covers Central Europe: more diacritics (ś, ł, č, etc.)
  - iso-8859-5: covers Cyrillic; etc., etc.
- UTF encodings: "all" characters in all known writing systems — c. 70K CJK characters, numerous Semitic alphabets and abugidas, Linear B, Yí (Loloish), the International Phonetic Alphabet, etc., etc., etc.

Character encodings in Python 2

Python 2.7 assumes all strings are ASCII, but this is sometimes concealed from me because the Ipython interpreter tries to convert silently to known encodings (here, utf-8):

```
1 In [1]: a = 'Tiné'
2
3 In [2]: a
4 Out[2]: 'Tin\xc3\xa9'
```

Unicode in Python 2

Python 2.7 also allows strings to be specified as Unicode with a `u` prefix, and here again the Ipython interpreter converts silently, this time to to iso-8859-1 or "latin-1":

```
1 In [1]: a = 'Tiné'
2
3 In [2]: a
4 Out[2]: 'Tin\xc3\xa9'
5
6 In [3]: b = u'Tiné'
7
8 In [4]: b
9 Out[4]: u'Tin\xe9'
```

Explicit encodings in Python 2

Python 2.7 also lets you convert explicitly to known encodings:

```
In [3]: b = u'Tiné'

In [4]: b
Out[4]: u'Tin\xe9'

In [5]: b.encode('utf-8')
Out[5]: 'Tin\xc3\xa9'

In [6]: b.encode('latin-1')
Out[6]: 'Tin\xe9'
```

See `http://docs.python.org/2.7/library/codecs.html#standard-encodings` for a long list of these.

Unicode in Python 3

Python 3 treats all strings as Unicode; the alternative is "bytecode":

```
In [1]: a = 'Tiné'

In [2]: a
Out[2]: 'Tiné'

In [3]: a.encode()
Out[3]: b'Tin\xc3\xa9'

In [4]: type(a.encode())
Out[4]: builtins.bytes
```

Unicode in Python 3

Python 3 treats all strings as Unicode; the alternative is "bytecode":

```
In [1]: a = 'Tiné'

In [2]: a
Out[2]: 'Tiné'

In [3]: a.encode()
Out[3]: b'Tin\xc3\xa9'

In [4]: type(a.encode())
Out[4]: builtins.bytes
```

Explicit encoding as in Python 2 is still possible, but Unicode is now the norm.

Bytecode in Python 3

Many modules in the standard library return bytecode that has to be explicitly converted to the `str` type (i.e., Unicode) before it can be used easily.

Example. . .

Bytecode returned by Python 3 methods

```
In [5]: import urllib.request
        url = 'http://hackerschool.com'
        x = urllib.request.urlopen(url).read()
        type(x)
Out[5]: builtins.bytes
```

To convert to true `str` type:

```
In [6]: y = x.decode()
        type(y)
Out[6]: builtins.str
```

Among the methods that are available to `str` but not to `bytes` are `format()` and `isprintable()`.

Bytecode vs Unicode in Python 3

And if a function (for instance in the `re` module) requires a true string as its argument, you must use `decode()` on bytecode before passing it in:

```
In [7]: import re
        z = re.compile('.')
        z.search(x)
---------------------------------------------------------------------------

TypeError                                 Traceback (most recent call last)
<ipython-input-17-ad5b6e1ce192> in <module>()
----> 1 z.search(x)

TypeError: can't use a string pattern on a bytes-like object
```

Bytecode vs Unicode in Python 3, cont'd

decode() changes it to a true (Unicode) string first — after that, it can be used with all sorts of methods that require a string:

```
1 In [8]: z.search(x.decode())
2 Out[8]: <_sre.SRE_Match at 0x231bbf8>
3
4 In [9]: z.search(x.decode()).group()
5 Out[9]: '<'
```

To recapitulate: the key point

To recapitulate: the key point

Encodings collectively are like a system of types for characters and strings of characters.

They can cause problems if you don't realize that a particular encoding is required or being supplied.

END