Mentor Help
Ask a mentor on our Q&A platform

Peer Chat 3
Chat with peers and alumni

An excerpt:

> **The Single Most Important Fact About Encodings**

> If you completely forget everything I just explained, please remember one ex fact. It does not make sense to have a string without knowing what encoding longer stick your head in the sand and pretend that “plain” text is ASCII.

> **There Ain’t No Such Thing As Plain Text**

> If you have a string, in memory, in a file, or in an email message, you have to encoding it is in or you cannot interpret it or display it to users correctly.

> Almost every stupid “my website looks like gibberish” or “she can’t read my e accents” problem comes down to one naive programmer who didn’t underst that if you don’t tell me whether a particular string is encoded using UTF-8 or (Latin 1) or Windows 1252 (Western European), you simply cannot display it c figure out where it ends. There are over a hundred encodings and above coc are off.”

What Every Programmer Absolutely, Positively Needs To Know About Character Sets To Work With Text

> An article by Joel Spolsky entitled The Absolute Minimum Every Software Dev Positively Must Know About Unicode and Character Sets (No Excuses!) is a ni the topic and I greatly enjoy reading it every once in a while. I hesitate to refe have trouble understanding encoding problems though since, while entertai on actual technical details. I hope this article can shed some more light on w encoding is and just why all your text screws up when you least need it.

> Any character can be encoded in many different bit sequences and any parti can represent many different characters, depending on which encoding is us them. The reason is simply because different encodings use different numbe characters and different values to represent different characters.”

Unicode and Python

In Python 3, there is:

- one text type: `str`, which holds Unicode data and
- two byte types: `bytes` and `bytearray`

The Stack Overflow answers here explain the different use cases well.

More Information

- If you’re still confused about the difference between character sets and en articles:
  - The difference between UTF-8 and Unicode?
  - More About Unicode in Python 2 and 3