unicode.decode()
The guessing game:

```python
f.write(u'tralala: {}' .format('foo'.encode(u'utf8')))
```
Python 2 / Python 3
Python 2
Python 2

\[ \text{str} = \text{'tralala'} \]

\[ \text{unicode} = \text{u'tralala'} \]

\[ \text{'tralala'} \text{ is } \text{b'tralala'} \]
What IS unicode in Python?

“I always thought that text in utf-8 was exactly that: Unicode data!”
- Janusz programowania

Łukasz Taczuk
Python 2

str = 1010 1010 01

unicode =
unicode ↔ str conversion

Abstraction
unicode.encode(<encoding>)

Physical
str.decode(<encoding>)

Physical
str

Abstraction
unicode

Łukasz Taczuk
unicode ↔ str conversion

encode → 100
       1010
       01

decode → ☁
unicode ⇔ str conversion

decode

encode
unicode ↔ str conversion

unicode.decode(<encoding>)
str.encode(<encoding>)
Automatic type conversion (1)

f.write() - converts to str

yourlibrary.method() - converts to whatever it feels like :)
Automatic type conversion (2)

FOO.format(BAR) - Automatically converts to type(FOO)

FOO % BAR - does the same

Template('$bar').substitute(bar=BAR) - does the same as well
Automatic type conversion (3)

FOO.encode(<encoding>) - Converts to unicode FIRST

FOO.decode(<encoding>) - Converts to str FIRST
Quiz time!
# -*- coding: utf-8 -*-

```python
with open('export.csv', 'w') as f:
    f.write('tralala: {}
```
#coding: utf-8#

with open('export.csv', 'w') as f:
    f.write('tralala: {}'.format('asdl'))

OK!
# -*- coding: utf-8 -*-

with open('export.csv', 'w') as f:
    f.write(u'tralala: {}' .format(u'asdl'))
# -*- coding: utf-8 -*-

with open('export.csv', 'w') as f:
    f.write(u'tralala: {}'.format(u'asdl'))

Traceback (most recent call last):
  File "2.py", line 4, in <module>
    f.write(u'tralala: {}'.format(u'asdl'))
UnicodeEncodeError: 'ascii' codec can't encode character u'\u0142' in position 12: ordinal not in range(128)
# -*- coding: utf-8 -*-

with open('export.csv', 'w') as f:
    f.write(u'tralala: {}'.format('asdl'))
# -*- coding: utf-8 -*-

with open('export.csv', 'w') as f:
    f.write(u'tralala: {}' .format('asdł'))

Traceback (most recent call last):
  File "3.py", line 4, in <module>
    f.write(u'tralala: {}' .format('asdł'))
UnicodeDecodeError: 'ascii' codec can't decode byte 0xc5 in position 3: ordinal not in range(128)
# -*- coding: utf-8 -*-

with open('export.csv', 'w') as f:
    f.write('tralala: {}\'.format(u'asdł'))
4.py

# -*- coding: utf-8 -*-

with open('export.csv', 'w') as f:
    f.write('tralala: {}'.format(u'asdl'))

Traceback (most recent call last):
  File "4.py", line 4, in <module>
    f.write('tralala: {}'.format(u'asdl'))
UnicodeEncodeError: 'ascii' codec can't encode character u'\u0142' in position 3: ordinal not in range(128)
5.py = 1.py "wb"
6.py = 2.py "wb"
7.py = 3.py "wb"
8.py = 4.py "wb"
# -*- coding: utf-8 -*-

with open('export.csv', 'wb') as f:
    f.write('tralala: {}'.format('asdł'))
5.py

# -*- coding: utf-8 -*-

with open('export.csv', 'wb') as f:
    f.write('tralala: {}'.format('asdl'))

OK!
```python
# -*- coding: utf-8 -*-

with open('export.csv', 'wb') as f:
    f.write(u'tralala: {}'.format(u'asdl'))
```
# -*- coding: utf-8 -*-

with open('export.csv', 'wb') as f:
    f.write(u'tralala: {}'.format(u'asdl'))

Traceback (most recent call last):
  File "6.py", line 4, in <module>
    f.write(u'tralala: {}\'.format(u'asdl'))
UnicodeEncodeError: 'ascii' codec can't encode character u'\u0142' in position 12: ordinal not in range(128)
with open('export.csv', 'wb') as f:
    f.write(u'tralala: {}\'.format('asdł'))
# -*- coding: utf-8 -*-

with open('export.csv', 'wb') as f:
    f.write(u'tralala: {}'.format('asdl'))

Traceback (most recent call last):
  File "7.py", line 4, in <module>
    f.write(u'tralala: {}'.format('asdl'))
UnicodeDecodeError: 'ascii' codec can't decode byte 0xc5 in position 3: ordinal not in range(128)
# -*- coding: utf-8 -*-

with open('export.csv', 'wb') as f:
    f.write('tralala: {}'.format(u'asdl'))
8.py

# -*- coding: utf-8 -*-

with open('export.csv', 'wb') as f:
    f.write('tralala: \{}'.format(u'asdl'))

Traceback (most recent call last):
  File "8.py", line 4, in <module>
    f.write('tralala: \{}'.format(u'asdl'))
UnicodeEncodeError: 'ascii' codec can't encode character u'\u0142' in position 3: ordinal not in range(128)
encode / decode
# -*- coding: utf-8 -*-

'asdasdł'.encode('utf8')
encode.py

# -*- coding: utf-8 -*-

'asdasdl'.encode('utf8')

Traceback (most recent call last):
  File "encode.py", line 3, in <module>
    'asdasdl'.encode('utf8')
UnicodeDecodeError: 'ascii' codec can't decode byte 0xc5 in position 6: ordinal not in range(128)
decode.py

# -*- coding: utf-8 -*-

u'asdasd\l'.decode('utf8')
decode.py

# -*- coding: utf-8 -*-

u'asdasdl'.decode('utf8')

Traceback (most recent call last):
  File "decode.py", line 3, in <module>
    u'asdasdl'.decode('utf8')
  File "/usr/lib/python2.7/encodings/utf_8.py", line 16, in decode
    return codecs.utf_8_decode(input, errors, True)
UnicodeEncodeError: 'ascii' codec can't encode character u'\u0142' in position 6: ordinal not in range(128)
An alternative way of writing files to disk

# -*- coding: utf-8 -*-

import codecs

with codecs.open('export.csv', 'w', encoding="utf-8") as f:
    f.write(u'żółw')
Python 3
Python 3

\[\text{bytes} = \text{b'}tralala'\]

\[\text{str} = \text{'tralala'}\]

\['tralala' \text{ is u'}tralala'\]
Python 3

\[
\text{bytes} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 \end{bmatrix}
\]

\[
\text{str} = \text{unknown}
\]
Let’s do it all one more time! :)

Łukasz Taczuk
with open('export.csv', 'w') as f:
    f.write(b'tralala: {}\'.format(b'asdl\'))
```python
# -*- coding: utf-8 -*-

with open('export.csv', 'w') as f:
    f.write(b'tralala: {}' .format(b'asdl'))
```

File "1.py", line 4
    f.write(b'tralala: {}' .format(b'asdl'))
    ^

SyntaxError: bytes can only contain ASCII literal characters.
# -*- coding: utf-8 -*-

with open('export.csv', 'w') as f:
    f.write('tralala: {}
'.format('asdł'))
# -*- coding: utf-8 -*-

with open('export.csv', 'w') as f:
    f.write('tralala: {}
'.format('asdł'))

OK!
3.py

# -*- coding: utf-8 -*-

with open('export.csv', 'w') as f:
    f.write('tralala: {}' .format(b'asdł'))
```python
# -*- coding: utf-8 -*-

with open('export.csv', 'w') as f:
    f.write('tralala: {}' .format(b'asdl'))
```

File "3.py", line 4
```
f.write('tralala: {}' .format(b'asdl'))
^``` 

SyntaxError: bytes can only contain ASCII literal characters.
4.py

# -*- coding: utf-8 -*-

with open('export.csv', 'w') as f:
    f.write(b'tralala: {}'\n           .format('asdł'))
with open('export.csv', 'w') as f:
  f.write(b'tralala: {0}'.format('asdł'))

Traceback (most recent call last):
  File "4.py", line 4, in <module>
    f.write(b'tralala: {0}'.format('asdł'))
AttributeError: 'bytes' object has no attribute 'format'
5.py = 1.py "wb"
6.py = 2.py "wb"
7.py = 3.py "wb"
8.py = 4.py "wb"
# -*- coding: utf-8 -*-

with open('export.csv', 'wb') as f:
    f.write(b'tralala: {}'.format(b'asdł')))
5.py

# -*- coding: utf-8 -*-

with open('export.csv', 'wb') as f:
    f.write(b'tralala: {}'.format(b'asdł'))

File "5.py", line 4
    f.write(b'tralala: {}'\n        format(b'asdł'))

SyntaxError: bytes can only contain ASCII literal characters.
```python
# -*- coding: utf-8 -*-

with open('export.csv', 'wb') as f:
    f.write('tralala: {}'
            .format('asdl'))
```
# -*- coding: utf-8 -*-

with open('export.csv', 'wb') as f:
    f.write('tralala: {}'.format('asdl'))

Traceback (most recent call last):
  File "6.py", line 4, in <module>
    f.write('tralala: {}'.format('asdl'))
TypeError: a bytes-like object is required, not 'str'
# -*- coding: utf-8 -*-

with open('export.csv', 'wb') as f:
    f.write('tralala: {}'.format(b'asdł')))
7.py

# -*- coding: utf-8 -*-

with open('export.csv', 'wb') as f:

    f.write('tralala: {}'.format(b'asdl'))

File "7.py", line 4
    f.write('tralala: {}'.format(b'asdl'))
^ SyntaxError: bytes can only contain ASCII literal characters.
# -*- coding: utf-8 -*-

with open('export.csv', 'wb') as f:
    f.write(b'tralala: {}'.format('asdł'))
with open('export.csv', 'wb') as f:
    f.write(b'tralala: {}'.format('asdł'))
encode / decode
# -*- coding: utf-8 -*-

b'asdasdł'.encode('utf8')
File "encode3.py", line 3
  b'asdasdl'.encode('utf8')
^  
SyntaxError: bytes can only contain ASCII literal characters.
# -*- coding: utf-8 -*-

b'asdasd'.encode('utf8')
Traceback (most recent call last):
  File "encode3-bis.py", line 3, in <module>
    b'asdasd'.encode('utf8')
AttributeError: 'bytes' object has no attribute 'encode'
decode3.py

# -*- coding: utf-8 -*-

'asdasd'.decode('utf8')
decode3.py

# -*- coding: utf-8 -*-

'asdasdl'.decode('utf8')

Traceback (most recent call last):
  File "decode3.py", line 3, in <module>
    'asdasdl'.decode('utf8')
AttributeError: 'str' object has no attribute 'decode'
Python 3 advantages

Does not convert strings to bytes implicitly
Does not convert bytes to strings implicitly
Simply does not contain “dangerous” methods
Catches potential errors at parsing time
Miscellaneous

https://github.com/overfl0/Bulletproof-Arma-Launcher/blob/next/src/utils/unicode_helpers.py

locale.getpreferredencoding()

sys.getfilesystemencoding()

print(...) -> sys.stdout.encoding
Takeaway
Your Python 3 Application

Outside world
Files
Network

100
1010
01
(bytes)

decode

100
1010
01
(bytes)

(str)

encode

Outside world
Files
Network

Łukasz Taczuk
The guessing game again:

```
f.write(u'tralala: {}'.format('foo'.encode(u'utf8'))))
```
See also:


https://nedbatchelder.com/text/unipain.html
Thank you! Any questions?

`unicode('python2') is b'itch'

(the irony is that the above statement evaluates to False)