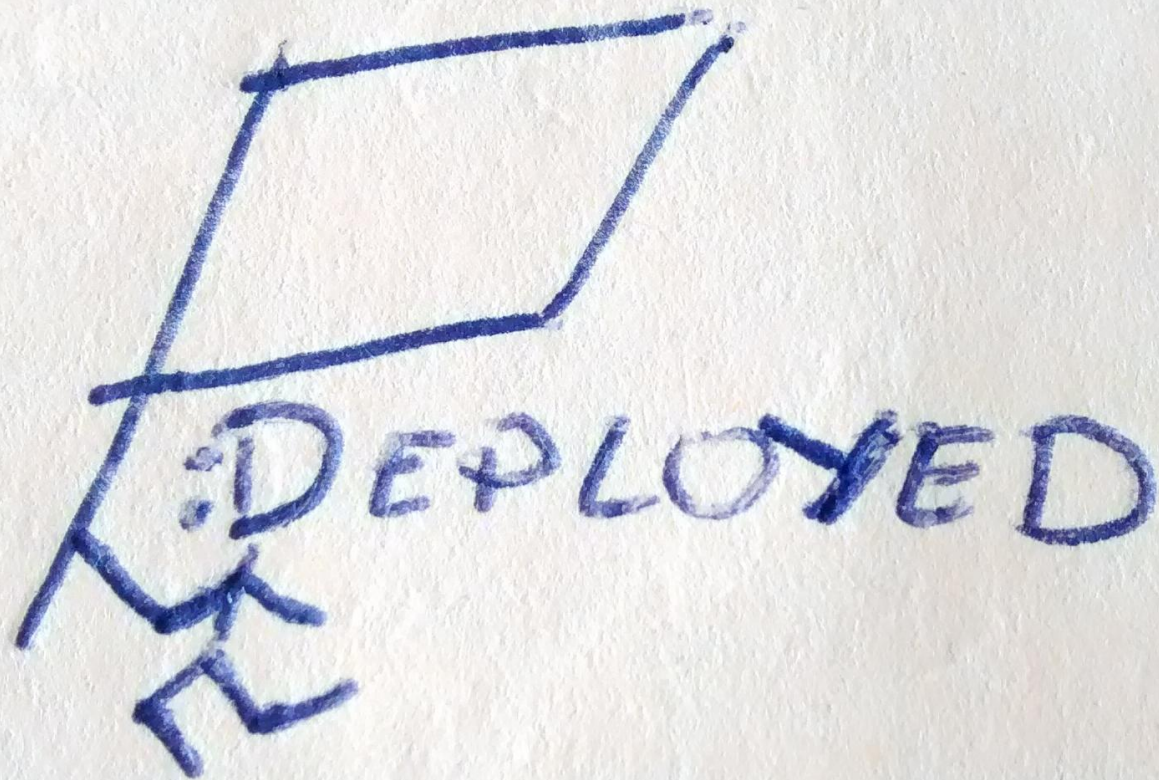


unicode.decode()

“leākīng frōm, your eyes’like liquid pain”

Łukasz Taczuk



DEPLOYED

A hand-drawn diagram in blue ink on a piece of paper. It features a parallelogram with a diagonal line running from the top-left corner to the bottom-right corner. The word "DEPLOYED" is written in capital letters across the middle of the parallelogram, positioned to the right of the diagonal line. Below the word, there are three small, stylized, stacked marks that resemble the letter 'K' or a series of connected 'Z's.

The guessing game:

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

Python 2 / Python 3

Łukasz Taczuk

Python 2

Łukasz Taczuk

Python 2

str = 'tralala'

unicode = u'tralala'

'tralala' is **b'tralala'**

What IS unicode in Python?

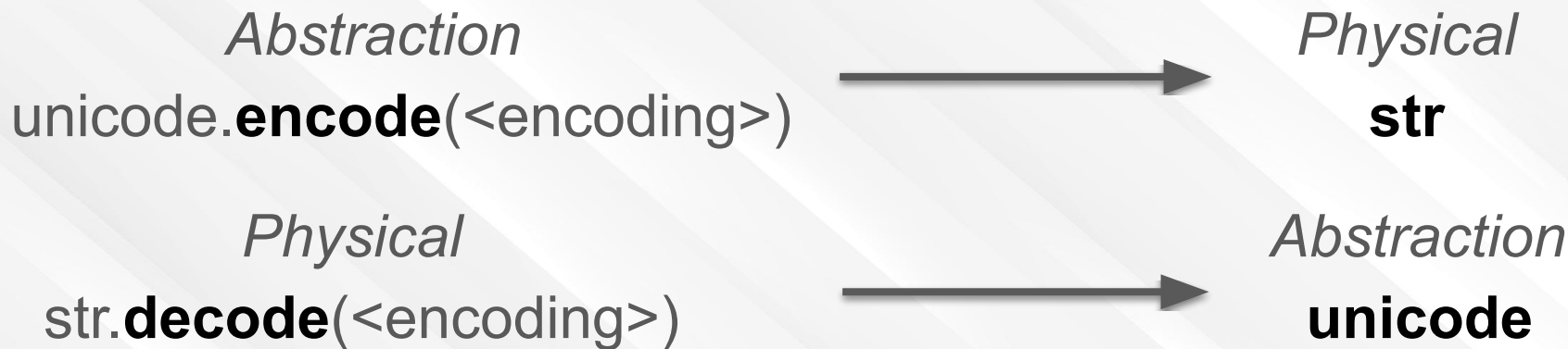
*“I always thought that text in utf-8 was exactly that: Unicode data!”
- Janusz programowania*

Python 2

str = **100
1010
01**

unicode = 

unicode $\Leftrightarrow$ str conversion



unicode $\Leftrightarrow$ str conversion

encode $\rightarrow$ **100**
1010
01

decode $\rightarrow$ 

unicode $\Leftrightarrow$ str conversion

100
1010
01

decode



encode

100
1010
01

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!

Łukasz Taczuk

1.py

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

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

1.py

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

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



2.py

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

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

```
    f.write(u'tralala: {}'.format(u'asdł'))
```

2.py

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

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

```
    f.write(u'tralala: {}'.format(u'asdł'))
```

Traceback (most recent call last):

File "2.py", line 4, in <module>

f.write(u'tralala: {}'.format(u'asdł'))

UnicodeEncodeError: 'ascii' codec can't encode character u'\u0142' in position 12: ordinal not in range(128)

3.py

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

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

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

3.py

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

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

```
    f.write(u'tralala: {}'.format('asd!'))
```

Traceback (most recent call last):

File "3.py", line 4, in <module>

f.write(u'tralala: {}'.format('asd!'))

UnicodeDecodeError: 'ascii' codec can't decode byte 0xc5 in position 3: ordinal not in range(128)

4.py

```
# -*- 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'asd!'))
```

Traceback (most recent call last):

File "4.py", line 4, in <module>

f.write('tralala: {}'.format(u'asd!'))

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"

5.py

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

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

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

5.py

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

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

```
    f.write('tralala: {}'.format('asdł'))
```

OK!

6.py

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

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

```
    f.write(u'tralala: {}'.format(u'asdł'))
```

6.py

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

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

```
    f.write(u'tralala: {}'.format(u'asdł'))
```

Traceback (most recent call last):

File "6.py", line 4, in <module>

f.write(u'tralala: {}'.format(u'asdł'))

UnicodeEncodeError: 'ascii' codec can't encode character u'\u0142' in position 12: ordinal not in range(128)

7.py

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

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

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

7.py

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

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

```
    f.write(u'tralala: {}'.format('asd!'))
```

Traceback (most recent call last):

File "7.py", line 4, in <module>

f.write(u'tralala: {}'.format('asd!'))

UnicodeDecodeError: 'ascii' codec can't decode byte 0xc5 in position 3: ordinal not in range(128)

8.py

```
# -*- 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'asd!'))
```

Traceback (most recent call last):

File "8.py", line 4, in <module>

f.write('tralala: {}'.format(u'asd!'))

UnicodeEncodeError: 'ascii' codec can't encode character u'\u0142' in position 3: ordinal not in range(128)

encode / decode

Łukasz Taczuk

encode.py

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

```
'asdasdł'.encode('utf8')
```

encode.py

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

```
'asdasd!'.encode('utf8')
```

Traceback (most recent call last):

File "encode.py", line 3, in <module>

'asdasd!'.encode('utf8')

UnicodeDecodeError: 'ascii' codec can't decode byte 0xc5 in position 6: ordinal not in range(128)

decode.py

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

```
u'asdasdł'.decode('utf8')
```

decode.py

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

```
u'asdasdł'.decode('utf8')
```

Traceback (most recent call last):

File "decode.py", line 3, in <module>

u'asdasdł'.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

Łukasz Taczuk

Python 3

bytes = b'tralala'

str = 'tralala'

'tralala' is u'tralala'

Python 3

bytes = **100
1010
01**

str = 

Let's do it all one more time! :)

Łukasz Taczuk

1.py

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

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

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

1.py

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

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

```
    f.write(b'tralala: {}'.format(b'asd!'))
```

File "1.py", line 4

```
    f.write(b'tralala: {}'.format(b'asd!'))
```

^

SyntaxError: bytes can only contain ASCII literal characters.

2.py

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

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

2.py

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

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

OK!

3.py

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

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

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


3.py

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

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

```
    f.write('tralala: {}'.format(b'asd!'))
```

File "3.py", line 4

```
    f.write('tralala: {}'.format(b'asd!'))
```

^

SyntaxError: bytes can only contain ASCII literal characters.

4.py

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

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

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

4.py

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

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

```
    f.write(b'tralala: {}'.format('asd!'))
```

Traceback (most recent call last):

File "4.py", line 4, in <module>

```
    f.write(b'tralala: {}'.format('asd!'))
```

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"

5.py

```
# -*- 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'asd!'))
```

File "5.py", line 4

```
    f.write(b'tralala: {}'.format(b'asd!'))
```

^

SyntaxError: bytes can only contain ASCII literal characters.

6.py

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

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

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

6.py

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

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

```
    f.write('tralala: {}'.format('asd!'))
```

Traceback (most recent call last):

File "6.py", line 4, in <module>

```
    f.write('tralala: {}'.format('asd!'))
```

TypeError: a bytes-like object is required, not 'str'

7.py

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

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

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

7.py

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

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

```
    f.write('tralala: {}'.format(b'asd!'))
```

File "7.py", line 4

```
    f.write('tralala: {}'.format(b'asd!'))
```

^

SyntaxError: bytes can only contain ASCII literal characters.

8.py

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

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

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

8.py

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

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

```
    f.write(b'tralala: {}'.format('asd!'))
```

Traceback (most recent call last):

File "8.py", line 4, in <module>

```
    f.write(b'tralala: {}'.format('asd!'))
```

AttributeError: 'bytes' object has no attribute 'format'

encode / decode

Łukasz Taczuk

encode3.py

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

```
b'asdasdł'.encode('utf8')
```

encode3.py

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

```
b'asdasdł'.encode('utf8')
```

File "encode3.py", line 3

```
b'asdasdł'.encode('utf8')
```

^

SyntaxError: bytes can only contain ASCII literal characters.

encode3-bis.py

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

```
b'asdasd'.encode('utf8')
```


encode3-bis.py

```
# -*- 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 -*-
```

```
'asdasd'.decode('utf8')
```

Traceback (most recent call last):

File "decode3.py", line 3, in <module>

'asdasd'.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

Łukasz Taczuk

Outside world
Files
Network

100
1010
01

(bytes)

decode

Your Python 3 Application



(str)

encode

Outside world
Files
Network

100
1010
01

(bytes)

Łukasz Taczuk

The guessing game again:

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


See also:

<https://medium.com/@andreacolangelo/strings-unicode-and-bytes-in-python-3-everything-you-always-wanted-to-know-27dc02ff2686>

<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)

Łukasz Taczuk