

Программирование, лекция 15. Бинарные файлы

Кафедра ИУ7 МГТУ им. Н. Э. Баумана,
2022 год

Форматы структурированных текстовых файлов CSV, TSV, PSV...

CSV - comma-separated values:

- value1,value2,,,value6

TSV - tab-separated values

- value1 value2 value6

PSV - pipe-separated values

- value1|value2|| |value6

В Python для поддержки работы с этими форматами предусмотрен модуль **csv**.

hex viewer, hex editor

Программы для просмотра (редактирования) данных (файлов) в “сыром” (raw) виде - как последовательность байтов, в 16-ричной с/с

Классы bytes и bytearray

`bytes` и `bytearray` - классы для представления бинарных данных, “байтовые строки”.

Набор операторов и методов похож на аналогичный у обычных строк.

```
b'bytes'
```

```
'Строка'.encode('utf-8')
```

```
bytes('string', encoding='utf-8')
```

```
bytes([1,2,3,4])
```

```
bytearray(b'string')
```

`bytes` - неизменяемый, `bytearray` - изменяемый

Метод to_bytes()

```
int.to_bytes(length, byteorder, *, signed=False)
```

```
(1024).to_bytes(10, byteorder='big')
```

```
b'\x00\x00\x00\x00\x00\x00\x00\x00\x00\x04\x00'
```

```
(-1024).to_bytes(10, byteorder='big', signed=True)
```

```
b'\xff\xff\xff\xff\xff\xff\xff\xff\xffc\x00'
```

Работа с бинарными файлами

with open('data.bin', 'rb') as f:

 b = f.read(1)

 while b != b'':

 b = f.read(1)

Модуль struct

Формирует упакованные двоичные структуры данных из переменных базовых типов данных и распаковывает их обратно.

Функции:

- `pack(format, v1, v2, ...)`
- `pack_into(format, buffer, offset, v1, v2, ...)`
- `unpack(format, buffer)`
- `unpack_from(format, /, buffer, offset=0)`
- `iter_unpack(format, buffer)`
- `calcsize(format)`

Формат struct

1. Выравнивание:

- @ - нативный, по умолчанию
- = - порядок байт нативный, размер стандартный
- < - порядок байт от младшего к старшему (LE), размер стандартный
- > - порядок байт от старшего к младшему (BE), размер стандартный
- ! - "сетевой" (аналог >)

2. x, c, b, B, ?, h, H, i, l, I, L, q, Q, n, N, e, f, d, s, p, P - различные форматы:

- c - char
- i - int
- f - float
- s - char[]