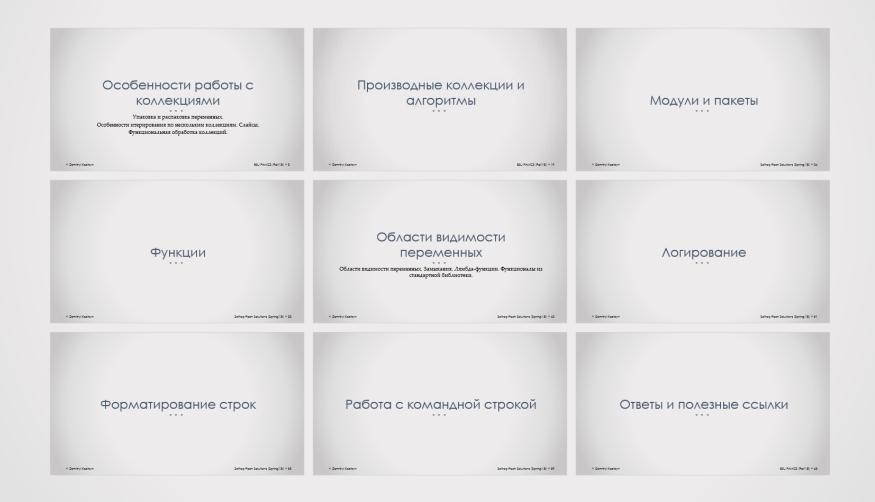
Python

Лекция 2 Преподаватель: Дмитрий Косицин BSU FAMCS (Fall'18)



Dzmitryi Kasitsyn

Особенности работы с коллекциями

Упаковка и распаковка переменных. Особенности итерирования по нескольким коллекциям. Слайсы. Функциональная обработка коллекций.

Упаковка переменных

Упаковка – создание кортежа / списка / т.п.

```
>>> x = [1, 2]

>>> x0 = x[0]

>>> x1 = x[1]

>>> print(x[0], x[1])

1, 2
```

Пример распаковки кортежа (пары значений):

```
>>> for x, y in zip(range(5), range(5, 10)): >>> print(x, y)
```

Dzmitryi Kasitsyn

Распаковка переменных

Кортеж можно распаковать автоматически:

```
>>> x0, x1 = x
>>> print(x[0], x[1])
1, 2
```

Для обмена переменных местами – упаковать и распаковать в другом порядке:

```
>>> x[0], x[1] = (x[1], x[0]) # скобки не обязательны >>> print(x[0], x[1])
```

Dzmitryi Kasitsyn

Распаковка переменных

Важно! Присваивание выполняется справа налево, но подвыражения в присваивании – в порядке следования:

```
>>> x = [1, 2]
>>> i = 0
>>> i, x[i] = 1, 1
>>> print(x)
[1, 1]
```

Распаковывать можно list, tuple, set, а также итераторы.

■ Dzmitryi Kasitsyn
 BSU FAMCS (Fall'18)
 ■ 6

Распаковка переменных

Допускается вложенность:

```
>>> ((x, _), z) = [[1, 2], 3]
>>> print x, z
1, 3
```

Важно! Символ ',' (запятая) является элементом синтаксиса не только кортежей – например, при бросании исключений распаковывать кортеж неявно нельзя.

В Python 3 допустима частичная распаковка (<u>PEP-3132</u>):

```
>>> head, *middle, tail = range(5)
>>> print head, middle, tail
0, [1, 2, 3], 4
```

Итерирование по нескольким коллекциям

Итерирование по двум коллекциям – функция **zip**:

```
>>> for x, y in zip([1, 2, 3], [-3, -2, -1]): >>> print(x % y == 0, end=' ')
```

False True True

Важно! zip возвращает новый список кортежей в Python 2.x и генератор в Python 3.x.

Вопрос: что будет, если коллекции разной длины?

enumerate (x) – сокращение zip (range(len(x), x))

Слайсы

Можно обращаться сразу к нескольким элементам коллекции:

```
>>> x = list(range(10))
>>> print(x[0:10:2])
0 2 4 6 8
```

Замечание. Вспомните range(0, 10, 2)

Слайсы за пределами коллекции или некорректные:

```
>>> x[100:110]
[] # тип соответствует типу переменной х
>>> x[0:10:-2]
[]
```

Способы задания слайсов

Следующие записи эквивалентны:

- x [9:0:-2]
- $\times [-1:0:-2]$
- x[:None:-2]

Слайс – это объект slice

```
>>> x[slice(9, 0, -2)] == x[9:0:-2]
```

True

Замечания по работе со слайсами

Слайсу можно дать имя и связать с переменной

```
>>> even_indices_slice = slice(None, None, 2)
>>> print(x[even_indices_slice])
```

Слайсам можно присваивать, в том числе iterable с иным количеством элементов:

```
>>> x = [1, 2, 3]
>>> x[0:2] = [7, 6, 5]
>>> print(x)
[7, 6, 5, 3]
```

Замечания по работе со словарями

Особенности создания словарей с ключами, имеющими одинаковый хэш:

```
>>> x = {1: 'a', True: 'b', 1.0: 'c'}
>>> assert x == {1.0: 'c'}
```

Методы для итерирования в Python 2.x:

- **keys**(), **values**(), **items**() возвращают списки ключей, значений и пар ключзначение
- iterkeys(), itervalues(), iteritems() возвращают, соответственно, итераторы
- viewkeys(), viewvalues(), viewitems() возвращают view-объекты

B Python 3.х методы **keys**(), **values**() и **items**(), возвращают на самом деле *view*-объекты. Такие объекты отражают изменения в исходной коллекции.

Замечания по работе со списками

Три способа создать список, содержащий три списка:

```
>>> x = [[], [], []]

>>> y = [[]] * 3

>>> z = []

>>> for _ in range(3):

>>> z.append([])

([[], [], []], [], []], []], [[], []])

# использовать ";" (semicolon) нельзя!

>>> x[0].append(1); y[0].append(2); z[0].append(3)
```

Dzmitryi Kasitsyn

Замечания по работе со списками

```
>>> print(x, y, z)
([[1], [], []], [[2], [2], [2]], [[3], [], []])
```

Wow!.. Лучше использовать другой способ!

```
>>> y_pretty = [[] for _ in range(3)]
>>> y_pretty[0].append(2)
>>> print(y_pretty)
[[2], [], []]
```

Такая конструкция называется list-comprehension.

Функциональный подход

Основные функции для работы с последовательностями:

- map применить функцию к каждому элементу последовательности;
- **filter** оставить только те элементы, для которых переданная функция возвращает **True** (предикатом может быть **None**);
- all возвращает True, если все элементы преобразуются к True;
- any возвращает True, если хотя бы один элемент True.

Важно! В Python 2.х функции **map** и **filter** возвращают списки, в то время как в Python 3.х – генераторы.

Comprehensions

Comprehensions допускают вложенные **for** и одно **if** выражение:

```
>>> def is_odd(x):
>>> return bool(x % 2)

>>> s1 = filter(is_odd, range(10))
>>> s2 = [x for x in range(10) if is_odd(x)]
>>> assert s1 == s2
```

Tuple-comprehension нету. Выражение с круглыми скобками – генератор

```
>>> s3 = list(x for x in range(10) if is_odd(x))
>>> assert s1 == s2 == s3
```

Comprehensions

Есть comprehension-выражения для создания словарей и множеств:

```
>>> d = {x: y**2
...     for x in range(2)
...     for y in range(x + 1)}
>>> assert d == {0: 0, 1: 1}

>>> s = {1 for _ in range(10)}
>>> assert s == set([1])
```

Замечание. Перед **for** может стоять любое допустимое выражение, в том числе содержащее некоторое преобразование (см. у**2) или условие.

Замечание. Comprehension-выражения допускают произвольное количество **for** и **if**, причем *не* обязательно поочередно: после **for** допустимо несколько **if** ов.

Statement del

Statement **del** имеет несколько смысловых нагрузок:

- «разрывает связь» между переменной и объектом (здесь также задействуется счетчик ссылок объекта)
- удаляет элемент, атрибут или слайс (за удаление отвечает сам объект)

```
>>> class X(object):
>>> a = None
>>>
>>> del X.a
>>>
>>> y, z = [1, 2, 3], {'x': 0}
>>> del y[0], z['x']
```

Bonpoc: как с помощью **del** очистить список, не удалив при этом объект?

Производные коллекции и алгоритмы

Встроенные коллекции

В Python реализованы следующие коллекции:

- deque дек, двухсторонняя очередь
- **defaultdict** словарь, который возвращает значение по умолчанию в случае отсутствия ключа
- Counter реализация defaultdict, когда для всех ключей значение по умолчанию ноль
- OrderedDict словарь, сохраняющий порядок вставки элементов
- namedtuple именованный кортеж
- Queue потокобезопасная очередь
- array массив, хранящий данные определенного C-совместимого типа

Подробнее o defaultdict

Подробнее o defaultdict:

```
>>> import collections
>>> def f(): return 0
>>> x = collections.defaultdict(f)
>>> print(x[2])
0
```

Важно! Конструктор **defaultdict** требует не число, а объект, при вызове которого будет возвращаться объект.

Dzmitryi Kasitsyn

Подробнее o namedtuple

Подробнее o namedtuple (именованный кортеж):

```
>>> Point = collections.namedtuple('Point', ['x', 'y'])
>>> p = Point(1, 2)
>>> print(p.x == p[0] == 1 and p.y == p[1] == 2)
True
```

Методы namedtuple:

- _fields вернет имена полей ('x', 'y')
- _asdict() вернет OrderedDict с соответствующими ключами и значениями
- _replace(x=new_value, ...) вернет namedtuple с замененными значениями

Алгоритмы стандартной библиотеки

В стандартной библиотеке реализованы алгоритмы по работе с кучей и упорядоченным списком:

- **heapq** модуль, содержащий функции по созданию кучи (heap), добавлению элементов, взятию k-максимальных
- **bisect** модуль, содержащий функцию бинарного поиска элемента по списку, а также вставки элемента в упорядоченную последовательность

Dzmitryi Kasitsyn

Модули и пакеты

Схема импорта модулей

При вызове «import x» происходит следующее:

- 1. find module
- 2. load module # an object loaded_module is created
 - create an object type of ModuleType
 - o read source
 - o compile source
 - o execute source
- 3. sys.modules['x'] = loaded_module
- 4. <this_module>.x = loaded_module

Загрузка и перезагрузка модуля

reload(loaded_module) – перезагружает модуль, не создавая новый объект.

Сравните:

```
>>> from module import x
>>> # then use x directly
>>> import module
>>> # then use module.x
```

Важно! Это не то же самое, что удалить модуль и заново загрузить. В случае **reload** (importlib.reload в Python 3) все объекты в модуле *пересоздаются*! Более того, reload имеет множество подводных камней.

Загрузка модулей

При загрузке создаются и вызываются default-значения функций (вопрос: что произойдет?):

```
>>> def f(x=g()):
>>> pass

>>> def g():
>>> return 0
```

Для изменения поведения при исполнении модуля от поведения при импорте используется следующее:

```
>>> if __name__ == '__main__':
>>>  # code to be executed if module is an entry point
```

Импорт модулей

Модуль можно загружать по имени:

```
>>> import importlib
>>> module_instance = importlib.import_module('module_name')
```

Есть возможность загружать source, compiled (.pyc) и dynamic (.pyd, .so) модули по полному пути (см. *imp* в Python 2 и *importlib.util* в Python 3)

Очередность загрузки:

- package
- module
- namespace (<u>PEP 420</u>, Python 3.3+)

Пакеты и пространства имен

Package – папка с модулями, где присутствует файл __init__.py. **Namespace** – файл __init__.py отсутствует.

Разница будет для вложенных папок: package вложенный обнаружится, а для namespace нужно явно прописать путь в sys.path.

Напоминание. sys.path.append(x) равносилен sys.path += x, но отличается от sys.path = sys.path + x

Замечание. Узнать имя файла из модуля можно обратившись к переменной __file__, имя модуля – к __name__.

Дополнительные возможности

Модули можно загружать прямо из *zip-*архива.

Можно использовать относительные импорты (РЕР-328).

Для moduleX.py верны следующие относительные импорты:

```
package/
                                     from .moduleY import spam
  __init__.py
                                     from .moduleY import spam as ham
 subpackage1/
                                     from . import moduleY
   __init__.py
                                     from ..subpackage1 import moduleY
   moduleX.py
   moduleY.py
                                     from ..subpackage2.moduleZ import eggs
 subpackage2/
                                     from ..moduleA import foo
   __init__.py
                                     from ...package import bar
   moduleZ.py
                                     from ...sys import path
 moduleA.py
```

Дополнительные возможности. Замечания

Относительные импорты не столь распространены ввиду худшей переносимости.

У модуля также может присутствовать *docstring* – его следует располагать вверху файла в тройных кавычках.

Модуль __future__ является директивой компилятору создать .pyc файл, используя другие инструкции.

Существуют дополнительные механизмы: path hooks, metapath, module finders and loaders, etc.



Синтаксис функций

Определение функции:

В Python нету перегрузки функций – используются значения по умолчанию и динамическая типизация.

Значение по умолчанию вычисляется единожды при определении функции, а потому *должно* быть неизменяемым.

Вызов функций

Примеры вызовов:

```
>>> def f(x, y=0, *args, **kwargs):
>>> return x, y, args, kwargs

>>> f() # TypeError
>>> f(1) # x: 1, y: 0, args: tuple(), kwargs: {}
>>> f(1, 2) # x: 1, y: 2, args: tuple(), kwargs: {}
>>> f(1, 2, 3) # x: 1, y: 2, args: tuple(3), kwargs: {}
```

Допустима распаковка аргументов при вызове функции:

```
>>> f(*(1, 2, 3, 4))
# x: 1, y: 2, args: tuple(3, 4), kwargs: {}
```

Передача аргументов по ключевым словам

В Python 3.5+ (<u>PEP-448</u>) допустима передача нескольких аргументов для распаковки:

```
>>> f(*(1, 2, 3), *(5, 6))
# x: 1, y: 2, args: tuple(3, 5, 6), kwargs: {}
```

Аргументы можно передавать по ключевым словам (порядок произвольный):

```
>>> f(y=1, x=2) # x: 2, y: 1, args: tuple(), kwargs: {}
```

Минусы:

- возможно более медленное выполнение (<u>Issue 27574</u>)
- проблемы с переименованием

Передача аргументов по ключевым словам

Переданные по ключевым словам аргументы, для которых нет имен, помещаются в kwargs:

```
>>> f(x=1, z=2) # x: 1, y: 0, args: tuple(), kwargs: {'z': 2}
```

Замечание. Порядок kwargs гарантируется с Python 3.6 (<u>PEP-468</u>).

Допустима распаковка аргументов по ключевым словам:

```
>>> f(**{'x': 1, 'z': 2})
# x: 1, y: 0, args: tuple(), kwargs: {'z': 2}
```

Исключительно ключевые аргументы

В Python 3 функции могут принимать аргументы исключительно по ключевому слову (<u>PEP-3102</u>):

```
>>> def f(*skipped, some_value=0):
>>> pass
>>> f(1, 2, 3)  # skipped: (1, 2, 3), some_value: 0
>>> f(some_value=100)  # skipped: tuple(), some_value: 100
```

Важно! Если значение по умолчанию не указано, функция обязана вызываться с данным ключевым аргументом, иначе возникнет **TypeError**.

Замечание. Имя variadic аргумента может быть опущено.

Алгоритм маппинга аргументов

Значения назначаются аргументам функции по порядку:

- На позиционные слоты
- Ha variadic аргумент (в кортеж неименованных элементов)
- Переданные по ключевым словам либо на позиционные, либо в словарь

Важно! Если аргумент позиционный аргумент не был передан или ключевой аргумент был передан более 1 раза, возникнет **TypeError**.

Bonpoc: что будет, если имя variadic аргумента опущено (вместо *args оставлен просто символ "*"), но в функцию переданы variadic аргументы?

Полный алгоритм маппинга аргументов также описан в <u>PEP-3102</u>.

Документация функций

Описание функции и их аргументов производится в docstring (<u>PEP-257</u>): def complex (real=0.0, imag=0.0):
"""Form a complex number.

```
Keyword arguments:
real -- the real part (default 0.0)
imag -- the imaginary part (default 0.0)
"""
# some code here ...
```

Документация может быть получена вызовом функции **help**(f) или взятием аргумента f.__doc__

Области видимости переменных

Области видимости переменных. Замыкания. Лямбда-функции. Функционалы из стандартной библиотеки.

Области видимости переменных

Области видимости переменных определяются функциями:

- built-in встроенные общедоступные имена (доступны через модуль builtins или __buitlins__, например, sum, abs и т.д.)
- global переменные, определенные глобально для модуля
- enclosing переменные, определенные в родительской функции
- *local* локальные для функции переменные

Локальные переменные в функциях могут в них свободно изменяться, enclosing, global и built-in – только читаться (<u>PEP-227</u>).

Области видимости переменных

Пример:

```
>>> abs(2) # built-in
>>> abs = dir # global, overrides
>>> def f():
>>> abs = sum # enclosing
>>> def g():
>>> abs = max # local
```

Для справки. (Нужно крайне редко). Для переопределения **abs** из функции **g** в функции **f** используется ключевое слово *nonlocal*, для переопределения глобальной переменной **abs** – ключевое слово *global* (<u>PEP-3104</u>).

```
>>> global abs
>>> abs = max # переопределит abs, глобальный для модуля
```

Переменные в циклах

Циклы *не имеют* своей области видимости: как только переменная была создана, она становится доступной и после цикла.

Замечание. В Python 3.4 и ниже так же «утекали» переменные из comprehension-выражений. Баг был исправлен в Python 3.5.

```
>>> x = [i for i in range(10)]
>>> print(i)
NameError: name 'i' is not defined # Python 3.6
```

Локальные и глобальные переменные

В Python можно получить доступ ко всем локальным и глобальным переменным:

- locals() словарь видимых локальных переменных
- globals() аналогичный словарь глобальных переменных

Словари автоматически обновляются интерпретатором.

Вопрос: можно ли модифицируя эти словари добавлять, изменять или удалять соответствующие переменные?

Замыкания

В функции доступны переменные, определенные уровнями выше – они замыкаются.

```
>>> def make_adder(x):
>>> def adder(y):
>>> return x + y
>>> return adder
>>> add_five = make_adder(5)
>>> add_five(10) # 15
```

Важно! Значение замкнутой переменной получается *каждый раз* при вычислении выражения.

Пример замыкания

```
>>> x = 2
>>> def make adder():
      def adder(y):
>>>
          return x + y
>>>
>>> return adder
>>> add x = make adder()
>>> add x(-2) # 0
>>> del x # delete 'x' - unbind variable name with object
>>> add x(-2) # NameError: name 'x' is not defined
```

Lambda-функции

Lambda-функции в Python допускают в себе одно лишь выражение: lambda arguments: expression

Эквивалентно:

def <lambda_name>(arguments):

return expression

Bonpoc: как будет выглядеть **lambda**, которая ничего не принимает и не возвращает?

Важно! С точки зрения bytecode **lambda** аналогична функции с тем же кодом, но при использовании **def** объект-функция еще получает имя.

Пример lambda-функций

Функция, возвращающая сумму аргументов:

```
>>> lambda x, y: x + y
```

Пример списка lambda-функций:

```
>>> collection_of_lambdas = [lambda: i*i for i in range(6)]
>>>
>>> for f in collection_of_lambdas:
>>> print(f())
```

Вопрос: что будет выведено в результате выполнения?

Пример lambda-функций

Поскольку вычисление происходит run-time, то для всех созданных функций значение **i** будет равно **5**. Переменная **i** была «захвачена» в comprehension-выражении, хоть и вне этого выражения она недоступна (в Python 3.6).

Для «захвата» значения можно создать локальную для lambda копию:

```
>>> lambdas = [lambda i=i: i*i for i in range(6)]
```

В модуле **operator** (<u>Py2</u>, <u>Py3</u>) есть множество функционалов, которыми можно пользоваться наряду с **lambda-**функциями:

```
>>> import operator
>>> # аналог: lambda x, y: x + y
>>> operator.add # operator. add
```

Замечания по операторам

Помимо операторов арифметических операций и операций сравнения, есть функционалы для работы с атрибутами и элементами коллекций.

```
f = operator.attrgetter('name.first', 'name.last')
# the call f(b) returns (b.name.first, b.name.last)

g = operator.itemgetter(2, 5, 3)
# the call g(r) returns (r[2], r[5], r[3])

h = operator.methodcaller('name', 'foo', bar=1)
# the call h(b) returns b.name('foo', bar=1)
```

Функционалы и lambda-функции наряду с обычными функциями используются как аргументы, выполняющие некоторое действие, например, в **map**, **filter**.

Логирование

Логирование

```
import sys
import logging
import datetime

logger = logging.getLogger(__name__)
logger.setLevel(logging.INFO)
logger.addHandler(logging.handlers.StreamHandler(sys.stdout))

logger.info("current year's %d", datetime.datetime.today().year)
```

Замечание. В Python 3 Handlers и Filters расположены во вложенных модулях, а содержимое модуля datetime перемещено.

Логирование: подробности

Объекты Logger объявлены в модуле logging.

Создать новый logger можно вызовом getLogger, передав ему некотрое имя.

Замечание. Если передать имя существующего логгера в метод **getLogger**, то будет возвращен уже существующий логгер с таким именем.

Логировать сообщение можно с помощью методов debug, info, warning, error, critical или общего log.

Установить уровень чувствительности (verbosity) можно методов setLevel.

Логирование: подробности

Добавить обработчик можно методом addHandler, фильтр – addFilter.

Реализованные обработчики: StreamHandler, FileHandler, RotatingFileHandler, SocketHandler, etc.

Замечание. Bce handler'ы и filterer'ы имеют базовые классы - logging.Handler и logging.Filterer.

Конфигурация логгера может быть сохранена в файле и загружена с помощью logging.config.dictConfig.

Подерживается printf-style форматирование:

```
>>> 'number of %.3f values in %s is %d' % (0.1234, 'some object', 3) number of 0.123 values in some object is 3
```

Есть возможность использовать именованные аргументы:

```
>>> 'number of %(name)s is %(count)d' % {'name': 'names', 'count': 2} number of names is 2
```

Поддерживается новый стиль форматирования строк:

А еще можно:

- аналогично использовать именованные аргументы: "{name}"
- индексировать аргументы: "{items[0]}"
- обращаться к атрибутам: "{point.x}"
- опускать индексы: " { } { } "
- повторять и менять местами индексы: " { 1 } { 0 } { 1 } "

Вопрос: что будет, если не совпадает количество аргументов для подстановки? А если нету такого именованного аргумента?

Замечание. Если вам нужно подставить множество локальных переменных, можно использовать словари **locals**() и **globals**(), определенные интерпретатором:

```
>>> x = 2
>>> "{x}".format(**locals())
2
```

Вопрос: верно ли, что так хитро можно менять значения локальных переменных?

Замечание. В Python 3 добавлен метод **format_map**, чтобы передавать словарь не распаковывая.

Работа с командной строкой

Парсинг аргументов командной строки

```
import argparse
parser = argparse.ArgumentParser(description='Process some
integers.')
parser.add argument('integers', type=int, nargs='+',
                    help='an integer for the accumulator')
parser.add argument('--sum', dest='accumulate',
                    action='store const',
                    const=sum, default=max,
                    help='sum the integers (default: find the
max) ')
args = parser.parse args()
print(args.accumulate(args.integers))
```

Для парсинга есть **ArgumentParser** в модуле **argparse**.

Методы parse_args и parse_known_args принимают некоторый список аргументов (по умолчанию sys.argv), парсят его и возвращают объект Namespace.

Произвольную строку запуска можно разбить на список с помощью модуля **shlex**.

Информации об аргументе добавляется с помощью add_argument:

- имена переменных через '-' (dash), '--' (double dash) или без них
- dest имя переменной, в которой хранится значение
- *type* преобразование типа
- action действие при получении аргумента (store, store_true, append, etc.)
- *nargs* количество аргументов (1, 1 и более, 0 и более)
- default значение по умолчанию (если не передан)
- required обязательный аргумент
- *choices* список возможных значений
- *help* описание аргумента

Ответы и полезные ссылки

• • •

Dzmitryi Kasitsyn

В случае передачи в метод **zip** коллекций (итераторов по коллекциям) разной длины, итерирование закончится при достижении конца наименьшей коллекции.

В случае итераторов важно, какой итератор будет исчерпан первым.

Очистить список, не удалив его самого, можно так:

```
>>> x = []
>>> del x[:]
>>> x[:] = [] # эквивалентная запись
```

Строки модуля интерпретируются последовательно. При попытке создания объекта **f** – функции аргументы по умолчанию будут также интерпретированы и сохранены в данном объекте. Поскольку функция **g** объявлена ниже, произойдет исключение **NameError**, что такого имени нету. Обратите внимание, что значения аргументов по умолчанию создаются *только один раз* при загрузке модуля.

Dzmitryi Kasitsyn

Аргументы функций

Если функция имеет неименованный variadic параметр (Python 3), но variadic аргументы переданы, то произойдет исключение **TypeError**.

Модифицировать локальные переменные через словарь **locals**() крайне не рекомендуется, хоть исключения и не будет. Следует только получать значения.

Lambda-функция

Пустая lambda-функция имеет следующий вид:

>>> lambda: None

Форматирование строк

Если количество аргументов для подстановки не совпадает с количеством в шаблоне или один из требуемых именованных аргументов не передан, произойдет исключение **TypeError**.

Однако для подстановки можно передать словарь, в котором значений больше, чем требуется. Ошибки в таком случае не будет.

Полезные ссылки

Множество shortcuts можно найти в книге Pilgrim, M. Dive Into Python 3.

Подробнее механизм импортов описан здесь (Python 3):

- https://docs.python.org/3.7/library/modules.html
- https://docs.python.org/3.7/tutorial/modules.html
- https://docs.python.org/3.7/reference/import.html

Для Python 2 документация расположена по следующим ссылкам:

- https://docs.python.org/2.7/library/modules.html
- https://docs.python.org/2.7/tutorial/modules.html