

PYTHON CHEATSHEET CHƯƠNG 9

KHÁI NIỆM DICTIONARY

Khái niệm Dictionary – Key/Value Mapping

Dictionary là **tập hợp không có thứ tự** của các cặp key-value.

Key: duy nhất, thường là string hoặc số.

Value: có thể thuộc mọi kiểu dữ liệu.

Tạo dictionary:

```
d = {'a': 10, 'b': 20}  
empty = {}
```

Truy xuất:

```
print(d['a']) # 10
```

DICTIONARY NHƯ BỘ ĐẾM (SET OF COUNTERS)

Dictionaries được sử dụng nhiều nhất để **đếm tần suất**.

Mẫu phổ biến (pattern):

```
counts = {}  
for word in words:  
    counts[word] = counts.get(word, 0) + 1
```

.get(key, default) giúp tránh lỗi *KeyError* và viết code gọn hơn.

Ứng dụng:

Đếm từ

Đếm email sender trong file mbox

Đếm ký tự, số lần xuất hiện dữ liệu...

DICTIONARIES VÀ FILES

Kết hợp dictionary + file để xử lý dữ liệu thực tế:

```
fhand = open('mbox-short.txt')  
counts = {}  
for line in fhand:  
    if line.startswith("From "):  
        email = line.split()[1]  
        counts[email] = counts.get(email, 0) + 1
```

Tiếp tục dùng dictionary để:

Đếm số tin nhắn theo domain

Đếm số lần xuất hiện của các giờ gửi mail

Đếm subject, author...

ADVANCED TEXT PARSING

Dictionary thường kết hợp với:

Tách chuỗi (split())

Lọc dòng (startswith(), find())

Phân tích từng thành phần email, domain:

```
domain = email.split('@')[1]
```

Tìm giá trị lớn nhất:

```
bigkey = None  
bigvalue = None  
for k, v in d.items():  
    if bigvalue is None or v > bigvalue:  
        bigkey, bigvalue = k, v
```

DICTIONARY PATTERNS QUAN TRỌNG

Pattern 1: Đếm tần suất

```
d[key] = d.get(key, 0) + 1
```

Pattern 2: Tìm key có giá trị lớn nhất

```
max_key = max(d, key=d.get)
```

Pattern 3: Chuyển dictionary sang list để sort

```
tmp = []  
for k, v in d.items():  
    tmp.append((v, k))  
tmp = sorted(tmp, reverse=True)
```

Pattern 4: Tách dữ liệu dạng dòng

```
line = line.rstrip()  
words = line.split()
```

LOOPING THROUGH DICTIONARIES

Dictionaries không giữ thứ tự → lặp theo key.

Lặp qua keys

```
for key in d:  
    print(key, d[key])
```

Lặp qua items (key + value)

```
for k, v in d.items():  
    print(k, v)
```

Lấy danh sách key/value

```
keys = list(d.keys())  
values = list(d.values())  
items = list(d.items())
```

Sắp xếp dictionary

```
for k in sorted(d):  
    print(k, d[k])
```