

Bài thực hành 5. MỘT SỐ CHỦ ĐỀ MỞ RỘNG

1

A. RÚT TRÍCH DỮ LIỆU TỪ WEB

Bài tập 1

Sử dụng thư viện Requests và thư viện BeautifulSoup để thực hiện các yêu cầu sau.

- a. Crawl **tên của các bài báo khoa học** được công bố trong hội nghị WACV năm 2024 (IEEE/CVF Winter Conference on Applications of Computer Vision) từ đường dẫn:

<https://openaccess.thecvf.com/WACV2024>

- b. Thực hiện một số thao tác xử lý văn bản với tên của các bài báo khoa học để lọc ra các keyword chính:
 - Tách từ dựa trên dấu khoảng trắng.
 - Chuyển sang ký tự thường.
 - Loại bỏ các stopword (từ dừng).
 - Loại bỏ một số từ không quan trọng trong lĩnh vực Công nghệ thông tin.
 - ...
- c. Thống kê tần số xuất hiện của từng keyword.
- d. Chọn loại biểu đồ thích hợp và vẽ biểu đồ biểu diễn tần số xuất hiện của các keyword theo thứ tự giảm dần số lần xuất hiện.
- e. Cho biết top 5 keyword được quan tâm nhiều nhất trong hội nghị này.

Bài tập 2

Sử dụng thư viện Requests và thư viện BeautifulSoup để thực hiện các yêu cầu sau.

- a. Crawl **tên của các tác giả có bài báo khoa học** được công bố trong hội nghị WACV năm 2024.
- b. Thống kê tần số xuất hiện của từng tên tác giả.
- c. Chọn loại biểu đồ thích hợp và vẽ biểu đồ biểu diễn tần số xuất hiện tên của các tác giả theo thứ tự giảm dần số lần xuất hiện.
- d. Cho biết top 3 tác giả có số lượng bài báo khoa học được công bố nhiều nhất trong hội nghị này.

B. XỬ LÝ DỮ LIỆU JSON

Bài tập 3

Sử dụng thư viện Json và file books.json để thực hiện các yêu cầu sau.

- a. Lấy ra tất cả các node con của node book.
- b. Lấy ra tất cả các tên sách (title).
- c. Lấy ra tên sách (title) của quyển sách đầu tiên.
- d. Lấy ra thông tin của hai quyển sách đầu tiên.
- e. Lấy ra tên sách (title) của quyển sách cuối cùng.
- f. Lấy ra tất cả các giá tiền (price).
- g. Lấy ra các giá tiền (price) trên 35.
- h. Lấy ra tên (title) của các quyển sách có giá tiền (price) trên 35.
- i. Lấy ra giá tiền (price) của các quyển sách có giá tiền (price) dưới 30 hoặc trên 40.
- j. Lấy ra tên (title) của các quyển sách có giá tiền (price) nằm trong khoảng từ 40 đến 50.
- k. Lấy ra tên tác giả (author) của các quyển sách.
- l. Lấy ra tên (title) của quyển sách có tên tác giả (author) là James Linn.
- m. Lấy ra tên (title) của quyển sách có tên tác giả (author) là J K. Rowling và giá tiền (price) dưới 50.
- n. Lấy ra giá bán cao nhất trong số các quyển sách.
- o. Lấy ra giá bán trung bình của các quyển sách.
- p. Lấy ra năm xuất bản (year) của quyển sách có tên (title) là Learning XML.

Bài tập 4

Sử dụng thư viện Json và file students.json để thực hiện các yêu cầu sau.

- a. Lấy ra thông tin của các sinh viên trong lớp.
- b. Lấy ra điểm (marks) của các sinh viên trong lớp.
- c. Lấy ra tên (firstname) của các sinh viên có điểm (marks) từ 90 trở lên.
- d. Lấy ra điểm (marks) của sinh viên có họ (lastname) là Kad.
- e. Lấy ra tên (firstname) của sinh viên cuối cùng trong danh sách.
- f. Lấy ra số lượng sinh viên có điểm (marks) từ 90 trở lên.
- g. Lấy ra điểm (marks) trung bình của các sinh viên trong lớp.
- h. Lấy ra họ tên viết liền (firstname, lastname) của sinh viên cuối cùng trong danh sách.

Bài tập 5

Sử dụng thư viện Json và file iris.json để thực hiện các yêu cầu sau.

- a. Lấy ra danh sách chiều dài đài hoa (sepalLength) và tính các đại lượng thống kê mô tả: trung bình, trung vị, tứ phân vị, phương

sai, độ lệch chuẩn, mốt.

- b.** Lấy ra danh sách chiều rộng đài hoa (sepalWidth) và tính các đại lượng thống kê mô tả: trung bình, trung vị, tứ phân vị, phương sai, độ lệch chuẩn, mốt.
- c.** Lấy ra danh sách chiều dài cánh hoa (petalLength) và tính các đại lượng thống kê mô tả: trung bình, trung vị, tứ phân vị, phương sai, độ lệch chuẩn, mốt.
- d.** Lấy ra danh sách chiều rộng cánh hoa (petalWidth) và tính các đại lượng thống kê mô tả: trung bình, trung vị, tứ phân vị, phương sai, độ lệch chuẩn, mốt.
- e.** Lấy ra danh sách chiều dài đài hoa (sepalLength) và vẽ biểu đồ histogram biểu diễn sự phân bố chiều dài đài hoa.
- f.** Lấy ra danh sách chiều rộng đài hoa (sepalWidth) và vẽ biểu đồ histogram biểu diễn sự phân bố chiều rộng đài hoa.
- g.** Lấy ra danh sách chiều dài cánh hoa (petalLength) và vẽ biểu đồ histogram biểu diễn sự phân bố chiều dài cánh hoa.
- h.** Lấy ra danh sách chiều rộng cánh hoa (petalWidth) và vẽ biểu đồ histogram biểu diễn sự phân bố chiều rộng cánh hoa.
- i.** Thống kê số lượng hoa của mỗi loài hoa (species) và vẽ biểu đồ cột biểu diễn sự phân bố các loài hoa.
- j.** Tính hệ số tương quan giữa các cặp thuộc tính dưới đây. Dựa vào giá trị của hệ số tương quan để nêu nhận xét về mối tương quan giữa các cặp đại lượng này.
 - Chiều dài đài hoa (sepalLength) và chiều rộng đài hoa (sepalWidth).
 - Chiều dài đài hoa (sepalLength) và chiều dài cánh hoa (petalLength).
 - Chiều dài đài hoa (sepalLength) và chiều rộng cánh hoa (petalWidth).
 - Chiều rộng đài hoa (sepalWidth) và chiều dài cánh hoa (petalLength).
 - Chiều rộng đài hoa (sepalWidth) và chiều rộng cánh hoa (petalWidth).
 - Chiều dài cánh hoa (petalLength) và chiều rộng cánh hoa (petalWidth).
- k.** Vẽ biểu đồ phân tán giữa các cặp thuộc tính dưới đây. Quan sát sự phân bố của các điểm dữ liệu và ước lượng đường hồi quy tuyến tính. Vẽ các điểm dữ liệu và đường hồi quy trên cùng một

đồ thị.

- Chiều dài đài hoa (sepalLength) và chiều rộng đài hoa (sepalWidth).
- Chiều dài đài hoa (sepalLength) và chiều dài cánh hoa (petalLength).
- Chiều dài đài hoa (sepalLength) và chiều rộng cánh hoa (petalWidth).
- Chiều rộng đài hoa (sepalWidth) và chiều dài cánh hoa (petalLength).
- Chiều rộng đài hoa (sepalWidth) và chiều rộng cánh hoa (petalWidth).
- Chiều dài cánh hoa (petalLength) và chiều rộng cánh hoa (petalWidth).

Bài tập 6

Sử dụng thư viện Json và file departments.json để thực hiện các yêu cầu sau.

- Thống kê số lượng đơn vị của mỗi nhóm công việc (GroupName) và vẽ biểu đồ cột biểu diễn sự phân bố các nhóm công việc.
- Lấy ra tên (Name) của các đơn vị thuộc nhóm công việc (GroupName) Executive General and Administration.
- Lấy ra tên (Name) và nhóm công việc (GroupName) của đơn vị có mã đơn vị (DepartmentID) là 12.

C. VẼ ĐỒ HỌA VỚI THƯ VIỆN TURTLE

Bài tập 7

Di chuyển rùa bằng cách sử dụng các hàm sau:

a) forward(50)

left(90)

forward(100)

right(45)

b) fd(100)

lt(120)

fd(100)

lt(120)

fd(100)

c) fd(100)

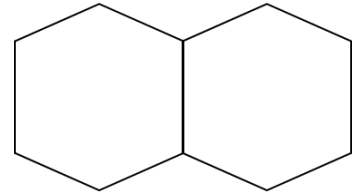
lt(180)

fd(200)

```
lt(180)
fd(100)
lt(90)
fd(50 + 50)
lt(90 * 2)
fd(40 + 80 * 2)
```

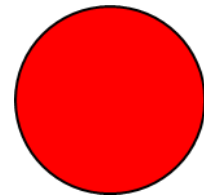
Bài tập 8

Viết chương trình vẽ hình bên cạnh.



Bài tập 9

Viết chương trình vẽ hình bên cạnh.



Bài tập 10

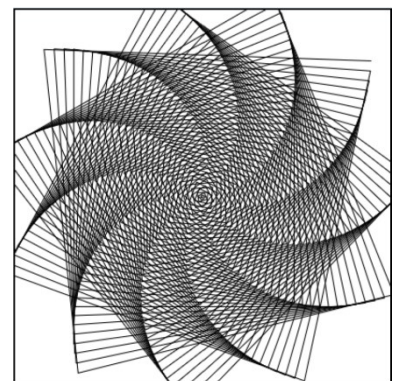
Viết chương trình vẽ các hình sau (mặc định tất cả các cạnh có cùng độ dài và tất cả các góc có số đo bằng nhau):

- a) Hình tam giác đều.
- b) Hình vuông.
- c) Hình lục giác.
- d) Hình bát giác.
- e) Hình ngôi sao.
- f) Hình bình hành.
- g) Hình chữ nhật.



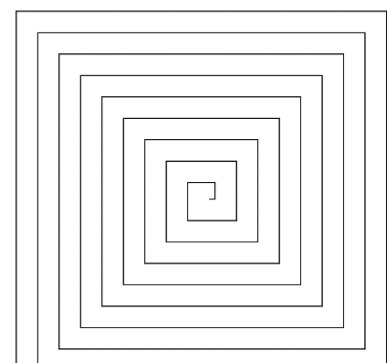
Bài tập 11

Viết chương trình vẽ hình bên cạnh.



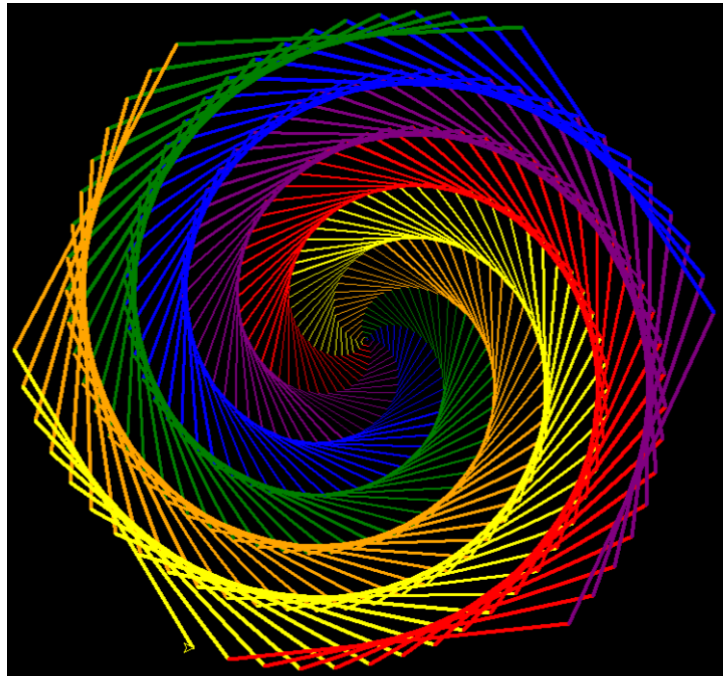
Bài tập 12

Viết chương trình vẽ hình bên cạnh.



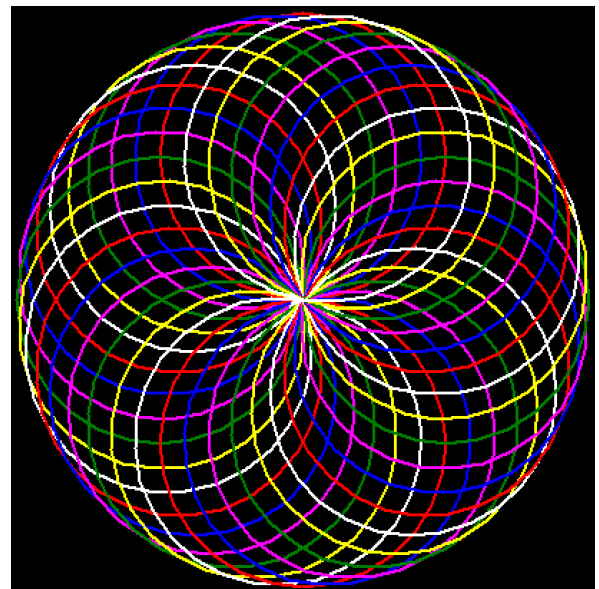
Bài tập 13

Viết chương trình vẽ hình bên cạnh.



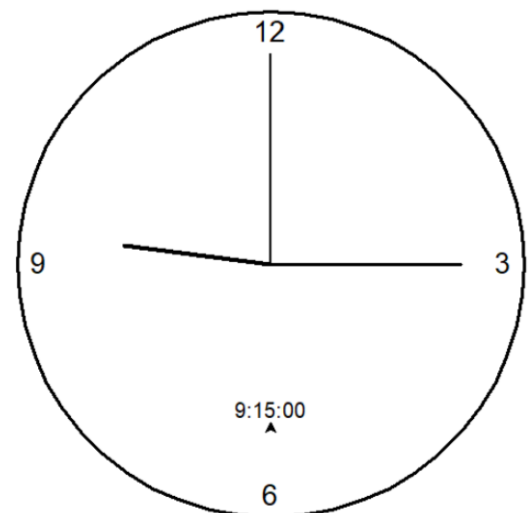
Bài tập 14

Viết chương trình vẽ hình bên cạnh.



Bài tập 15

Viết chương trình vẽ hình bên dưới.



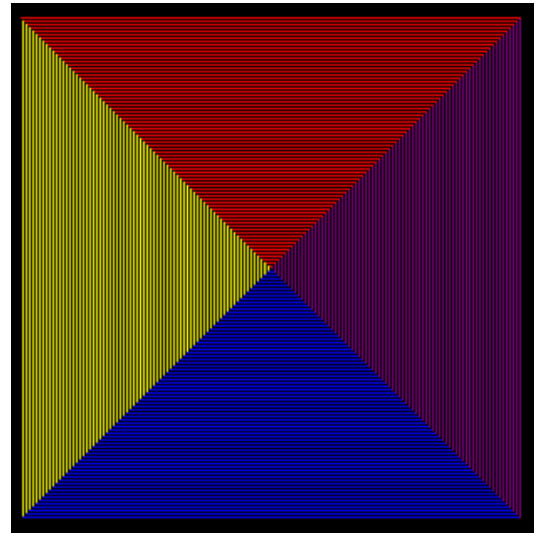
Bài tập 16

Viết chương trình vẽ hình bên dưới.



Bài tập 17

Viết chương trình vẽ hình bên dưới.



Bài tập 18

Viết chương trình vẽ chữ bên dưới.

Cool Python Codes