

INTRODUCTION OF DATA SCIENCE

Giảng viên: TS. Bùi Thanh Hùng
Bộ môn Khoa học dữ liệu, Khoa Công nghệ thông tin
Đại học Công nghiệp thành phố Hồ Chí Minh
Email: buithanhhung@iuh.edu.vn
Website: <https://sites.google.com/site/hungthanhbui1980/>

Bài 1:

Hãy tạo một site của bạn bằng google site.

Tham khảo site của thầy: <https://sites.google.com/site/hungthanhbui1980/>

Bài 2:

Đưa ra các đoạn code ví dụ cho các hàm sau trong thư viện re:

`re.findall()`

`re.compile()`

`re.split()`

`re.sub()`

`re.escape()`

`re.search()`

`re.match()`

Bài 3:

3.1 Hãy viết pattern để trích xuất toàn bộ địa chỉ email theo yêu cầu sau:

- Một địa chỉ email bao gồm mọi ký tự, số không phải dấu cách trước dấu “@”
- có dấu “@” phân cách giữa email và loại địa chỉ email.
- mọi ký tự, số không phải dấu cách sau dấu “@” (có thể gồm bất kể các hậu tố khác nhau ví dụ như (.co.uk, .gov.fr, .co.jp, v.v.)

3.2 Áp dụng pattern trên vào 1 đoạn text cụ thể để tìm ra các địa chỉ email trong đoạn text đó.

Bài 4:

Đọc trước cuốn sách Crawling thầy đã gửi