

Thu thập dữ liệu việc làm

Người thực hiện: Trần Danh Ngân

Mục tiêu

- Thu thập dữ liệu về tuyển dụng theo từ khóa trên trang web www.vietnamworks.com
- Lưu trữ bằng file excel từ dữ liệu thu thập được

Quy trình thực hiện

1. Tìm hiểu về trang web
2. Tìm hiểu các phương thức thu thập
3. Thu thập dữ liệu
4. Xử lý dữ liệu
5. Xuất filexlsx

Tìm hiểu về trang Web

- Địa chỉ web: <https://www.vietnamworks.com/>
- Sau khi nhập từ khóa tìm kiếm, trang sẽ gửi một request chứa các thông tin về từ khóa, các trường dữ liệu cần thiết
- Lúc này ta sẽ nhận được danh sách các thông tin cần thiết như: Job title, company name, salary, level,...

Tìm hiểu về trang Web

- Hình bên cho thấy phương thức post request và kết quả thu được (kết quả đã được cắt bớt)

Request URL: <https://ms.vietnamworks.com/job-search/v1.0/search>

Request Method: POST

Status Code: 200 OK

Remote Address: 171.244.31.90:443

Referrer Policy: strict-origin-when-cross-origin

```
▼ 0: {jobId: 1538008, jobTitle: "Senior Business Performance Analyst",  
    address: ""  
    alias: "senior-business-performance-analyst-1"  
    approvedOn: "2022-07-04T16:14:28+07:00"  
    ► benefits: [{benefitId: 2, benefitIconName: "fa-user-md", benefitNa  
        canonical: ""  
        companyId: 321984  
        companyLogo: "https://images.vietnamworks.com/pictureofcompany/1a/  
        companyName: "One Mount"  
        companyProfile: ""  
        companySize: ""  
        companySizeId: 0  
        companySizeVI: ""
```


Tìm hiểu về trang Web

- Một số thông tin hữu ích nhưng không được trả về như: các kỹ năng, ngôn ngữ hồ sơ
- Tuy nhiên các thông tin này có thể được tìm thấy tại nội dung của từng công việc, và được liệt kê trong một khung tổng hợp có cấu trúc cố định.



NGÀY ĐĂNG TUYỂN
04/07/2022



CẤP BẬC
Nhân viên



NGÀNH NGHỀ
Tư vấn , Bất động sản ,
Hoạch định/Dự án



KỸ NĂNG
Business Strategic Planning,
Project Management,
Problem Solving Skill, Data
Analysis, Presentation Skill



NGÔN NGỮ TRÌNH BÀY HỒ SƠ
Bất kỳ

Thu thập dữ liệu

Gửi post request

- Gửi post request để lấy các thông tin: Job title, company name, salary, level,...

```
response = requests.post(url=url,  
                           json=query_params)
```

Dùng BeautifulSoup

- Sử dụng cấu trúc của html để tìm ra các thông tin kỹ năng, ngôn ngữ hồ sơ

```
job_page = requests.get(job['jobUrl'])  
soup = BeautifulSoup(job_page.text, 'html.parser')
```

Xử lý dữ liệu

- Sử dụng thư viện pandas để xử lý dữ liệu
- Chuyển đổi một số thông tin để tiện cho việc tra cứu hơn như:
 - Expiration date: chuyển về dạng dd-mm-yyyy
 - Salary: đưa về dạng biểu thị bằng ngôn từ
 - Address: xử lý thành 2 trường city address và address work (chứa địa chỉ đầy đủ)
 - Benefits: Các phúc lợi phân chia bởi dấu chấm phẩy “;”
- Chỉ giữ lại các thông tin cần thiết

Xử lý dữ liệu

Các thông tin đã được thu thập và xử lý

- Job Title
- Company Name
- Salary
- Job Deadline
- Level
- Benefits
- Skills
- Language
- City Address
- Address Work
- URL

Lưu trữ bằng file excel

B	C	D	E	F	G
Keyword	Job Title	Company Name	Salary	Job Deadline	Level
phân tích c	Senior Business Performance Analyst	One Mount	Thương lượng	03-08-2022	Nhân viên
phân tích c	Digital Ads (Mức Lương Từ 8-15 Triệu)	Bệnh Viện Đa Khoa Hồng Ngọc	Thương lượng	24-08-2022	Nhân viên
phân tích c	Chuyên Viên Product Marketing (Lương 10 Triệu)	Công Ty Nhật Linh-LiOA	Từ 700.0\$ đến 1000.0\$	17-08-2022	Nhân viên
phân tích c	Marketing Staff	Công Ty TNHH Công Nghệ CRE Việt Nam (CRE)	Từ 1000.0\$	10-08-2022	Nhân viên
phân tích c	Electrical Technician (Working Location: Hanoi)	Namconson Pipeline Co. (Ncsp)	Thương lượng	24-08-2022	Nhân viên
phân tích c	Business Analyst	Công Ty CP Giải Pháp Kinh Doanh Nền Tảng - I	Từ 642.0\$ đến 1071.0\$	03-08-2022	Nhân viên
phân tích c	Senior Business Analyst Business Intelligence	MISA Jointstock Company	Lên tới 2000.0\$	07-08-2022	Nhân viên
phân tích c	Trưởng Phòng Kinh Doanh	International English Education	Từ 1500.0\$ đến 2000.0\$	17-08-2022	Trưởng phòng
phân tích c	CVCC Quản Trị Dữ Liệu Tài Chính	Techcombank	Thương lượng	12-08-2022	Nhân viên
phân tích c	Chuyên Viên Cao Cấp Đào Tạo (Phụ T	VPBank - https://tuyendung.vpbank.com.vn/	Thương lượng	03-08-2022	Nhân viên
phân tích c	Senior Business Analyst (Martech)	Techcombank	Thương lượng	13-08-2022	Nhân viên
phân tích c	SAP Bw BI Consultant	Công Ty Cổ Phần Tư Vấn Công Nghệ Ngs (Ngs)	Lên tới 2000.0\$	03-08-2022	Nhân viên
phân tích c	Senior IT Business Analyst (Up to \$20	LOTTE Data Communication Company Limite	Từ 1000.0\$ đến 2000.0\$	30-07-2022	Nhân viên
phân tích c	Business Performance Analyst Lead	One Mount	Thương lượng	21-08-2022	Nhân viên
phân tích c	Business Analyst	Niteco (Vietnam) CO., Ltd	Thương lượng	28-07-2022	Nhân viên
phân tích c	[Vinpearl Head Office] Chuyên Viên P	Công Ty Cổ Phần Vinpearl	Thương lượng	30-07-2022	Nhân viên
phân tích c	Chuyên Viên Dữ Liệu_Data Analyst	Công Ty Tài Chính Trách Nhiệm Hữu Hạn MB	Thương lượng	30-07-2022	Nhân viên
phân tích c	Data Analysis (Marketing Research)	Golden Gate Trade & Service JSC	Thương lượng	14-08-2022	Trưởng phòng
phân tích c	Chuyên Viên Phát Triển và Tích Hợp D	Công Ty Cổ Phần Thanh Toán Quốc Gia Việt N	Thương lượng	27-07-2022	Nhân viên
phân tích c	Tmo - Business Intelligent Analyst Ma	Ngân Hàng TMCP Á Châu (ACB)	Thương lượng	17-08-2022	Trưởng phòng

Xin cảm ƠN