

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH  
TRƯỜNG ĐẠI HỌC BÁCH KHOA  
KHOA KHOA HỌC - KỸ THUẬT MÁY TÍNH



# ĐỒ ÁN KỸ THUẬT LẬP TRÌNH CO1031

---

## Project 2

---

Lớp: L01  
GVHD: Lưu Quan Huân  
SV thực hiện: Nguyễn Thế Viễn 1814764  
Nguyễn Ngọc Thuần 1814217  
Phạm Đức Duy Anh 1810814  
Huỳnh Ngọc Tấn 1813952  
Bùi Hữu Đang 1811828

Tp. Hồ Chí Minh, Tháng 1/2021



## Mục lục

<b>1</b>	<b>MỤC ĐÍCH, YÊU CẦU</b>	<b>2</b>
<b>2</b>	<b>GIỚI THIỆU ĐỀ TÀI</b>	<b>2</b>
<b>3</b>	<b>TÌM HIỂU VÀ HIỆN THỰC CÁC CÔNG CỤ CRAWL DỮ LIỆU</b>	<b>3</b>
3.1	Thư viện sử dụng cho việc crawl dữ liệu . . . . .	3
3.1.1	Scrapy . . . . .	3
3.1.2	Beautiful Soup . . . . .	3
3.1.3	Selenium . . . . .	4
3.2	Công cụ Crawl dữ liệu . . . . .	5
3.2.1	Format data . . . . .	5
3.2.2	JSON . . . . .	7
3.2.3	Ưu điểm của JSON . . . . .	7
3.2.4	Mô hình hóa data . . . . .	8
3.2.5	Cách kết nối Mysql sử dụng python . . . . .	10
<b>4</b>	<b>Phân tích dữ liệu thu thập được từ các trang</b>	<b>10</b>
4.1	Sơ đồ phân bố các tin tuyển dụng theo giá trị tiền lương . . . . .	10
4.2	Sơ đồ phân bố các tin tuyển dụng theo vị trí địa lý . . . . .	11
4.3	Sơ đồ phân bố giá trị tiền lương trung bình theo vị trí địa lý . . . . .	12
4.4	Các công cụ hỗ trợ phân tích . . . . .	12
<b>5</b>	<b>NHỮNG KHÓ KHĂN GẶP PHẢI</b>	<b>13</b>
5.1	Cài đặt thư viện scrapy của python . . . . .	13
5.2	File HTML bị lỗi dẫn đến việc tìm các tag của file khó khăn . . . . .	14
5.3	Trang web yêu cầu đăng nhập để hiển thị thêm thông tin . . . . .	14
<b>6</b>	<b>Đề xuất ứng dụng</b>	<b>15</b>
<b>7</b>	<b>LINK PROJECT</b>	<b>15</b>
	<b>Tài liệu</b>	<b>16</b>

# 1 MỤC ĐÍCH, YÊU CẦU

- Deep learning, Machine Learning hay AI đã và đang là một chủ đề đang được quan tâm với nhiều ứng dụng trong thực tiễn và các nghiên cứu mang tính học thuật cao. Tuy nhiên, việc triển khai các ứng dụng, các mô hình cũng như việc tiến hành nghiên cứu gặp nhiều khó khăn. Một mặt, do các bộ dữ liệu huấn luyện hạn chế, về cả số lượng và không đảm bảo về chất lượng. Các bộ dữ liệu hiện có cũng hạn chế, đa phần bằng các ngôn ngữ không phải tiếng Việt.
- Xuất phát từ nhu cầu thực tế của việc thực hiện tác đề tài, đồ án môn học sau này hoặc khi làm LVTN liên quan đến dữ liệu, sinh viên gặp nhiều khó khăn trong việc thu thập dữ liệu mẫu (Dataset) để triển khai các mô hình Machine Learning, Deep Learning hoặc cơ bản hơn là thu thập dữ liệu làm dữ liệu mẫu cho một ứng dụng, hệ thống đã xây dựng. Trong bài tập lớn này, SV sẽ có cơ hội hiện thực việc thu thập các dữ liệu này, làm quen với các thao tác tiền xử lý, phân loại, lưu trữ và biểu diễn một tập dữ liệu (Data set).

# 2 GIỚI THIỆU ĐỀ TÀI

- Xây dựng công cụ thu thập thông tin và phân loại các tin tuyển dụng từ:
  - [www.vietnamworks.com/](http://www.vietnamworks.com/)
  - [www.vietsingworks.com](http://www.vietsingworks.com)
  - <http://careerbuilder.vn/>
  - <http://mywork.com.vn/>
  - <http://1001vieclam.com/itviec>

## 3 TÌM HIỂU VÀ HIỆN THỰC CÁC CÔNG CỤ CRAWL DỮ LIỆU

### 3.1 Thư viện sử dụng cho việc crawl dữ liệu

#### 3.1.1 Scrappy

- Scrappy là một khung ứng dụng để thu thập dữ liệu các trang web và trích xuất dữ liệu có cấu trúc có thể được sử dụng cho nhiều ứng dụng hữu ích, như khai thác dữ liệu để giám sát, kiểm tra tự động, xử lý thông tin hoặc lưu trữ lịch sử.
- Scrappy cung cấp rất nhiều tính năng mạnh mẽ để việc crawl trở nên dễ dàng và hiệu quả, chẳng hạn như:
  - Hỗ trợ tích hợp để chọn và trích xuất dữ liệu từ các nguồn HTML / XML bằng cách sử dụng bộ chọn CSS mở rộng và biểu thức XPath, với các phương thức trợ giúp để trích xuất bằng cách sử dụng biểu thức chính quy.
  - Một bảng điều khiển trình bao tương tác (IPython nhận thức) để thử các biểu thức CSS và XPath để quét dữ liệu, rất hữu ích khi viết hoặc gỡ lỗi trình thu thập thông tin.
- Hướng dẫn cài đặt Scrappy:  
Cài đặt các gói của python, sau đó gõ lệnh trên terminal:

```
1 pip install Scrappy
```

#### 3.1.2 BeautifulSoup

- BeautifulSoup là một thư viện Python để phân tích dữ liệu có cấu trúc. Cho phép tương tác với HTML theo cách tương tự như cách tương tác với một trang web bằng các công cụ dành cho nhà phát triển. BeautifulSoup cho thấy một số chức năng trực quan mà có thể sử dụng để khám phá HTML mà mình nhận được.
- BeautifulSoup cung cấp một số phương pháp đơn giản và thành ngữ Pythonic để điều hướng, tìm kiếm và sửa đổi cây phân tích cú pháp như một bộ công cụ để phân tích tài liệu và trích xuất những gì bạn cần. Không cần nhiều mã để viết một ứng dụng.

- BeautifulSoup tự động chuyển đổi tài liệu đến sang Unicode và tài liệu đi sang UTF-8. Bạn không phải suy nghĩ về các mã hóa, trừ khi tài liệu không chỉ định một mã hóa và BeautifulSoup không thể tự động phát hiện một mã hóa. Sau đó, bạn chỉ cần chỉ định mã hóa ban đầu.
- BeautifulSoup nằm trên các trình phân tích cú pháp Python phổ biến như lxml và html5lib, cho phép bạn thử các chiến lược phân tích cú pháp khác nhau hoặc tốc độ giao dịch để linh hoạt.
- Hướng dẫn cài đặt BeautifulSoup:

```
1 pip install requests
2 pip install html5lib
3 pip install bs4
```

### 3.1.3 Selenium

- Selenium là một dự án cho một loạt các công cụ và thư viện kích hoạt và hỗ trợ tự động hóa các trình duyệt web.
- Selenium cung cấp các tiện ích mở rộng để mô phỏng tương tác của người dùng với các trình duyệt, một máy chủ phân phối để mở rộng phân bổ trình duyệt và cơ sở hạ tầng để triển khai đặc tả W3C WebDriver cho phép bạn viết mã có thể hoán đổi cho tất cả các trình duyệt web chính.
- Vì thế, nhóm cũng sử dụng Selenium để thu thập dữ liệu từ web mặc dù nó chủ yếu là một công cụ kiểm tra web tự động, không nhiều về việc thu thập dữ liệu.
- Hướng dẫn cài đặt Selenium:

- Đầu tiên, máy tính cần có python (chúng tôi đang sử dụng Python 3) và pip.

Nếu không có, vui lòng truy cập:

<https://realpython.com/installing-python/> để cài đặt python

<https://www.liquidweb.com/kb/install-pip-windows/> để cài đặt pip.

- Sau đó trong cửa sổ lệnh (hoặc Terminal) gõ:

```
1 pip install selenium
```

- Một việc cần làm nữa là tải webdriver (ở đây chúng tôi sử dụng trình duyệt Google Chrome) (Bạn có thể tải webdriver Google Chrome tương thích với hệ điều hành máy tính của bạn từ đây:  
<https://chromedriver.chromium.org/downloads>

- Sau đó, vui lòng giải nén tệp đã tải xuống và sao chép tệp **chromedriver.exe** giống như tệp này đường dẫn:  
<C:Webdriver//chromedriver.exe>

## 3.2 Công cụ Crawl dữ liệu

### 3.2.1 Format data

- Data:

- Dữ liệu định tính: Khi dữ liệu được trình bày dưới dạng từ ngữ và mô tả đề cập đến văn bản, hình ảnh, video, bản ghi âm, quan sát,...
  - \* Mặc dù bạn có thể quan sát dữ liệu này, nhưng nó chủ quan và do đó, khó phân tích dữ liệu trong nghiên cứu, đặc biệt là để so sánh.
  - \* Ví dụ: Dữ liệu chất lượng đại diện cho mọi thứ mô tả hương vị, kinh nghiệm, kết cấu hoặc ý kiến được coi là dữ liệu chất lượng. Loại dữ liệu này thường được thu thập thông qua các nhóm tập trung, phỏng vấn cá nhân hoặc sử dụng các câu hỏi mở trong các cuộc khảo sát.
- Dữ liệu định lượng: Bất kỳ dữ liệu nào được biểu thị bằng số lượng các số liệu. Loại dữ liệu này có thể được phân biệt thành các loại, được nhóm, đo lường, tính toán hoặc xếp hạng.
  - \* Ví dụ: các câu hỏi như tuổi, thứ hạng, chi phí, chiều dài, cân nặng, điểm số,...

- Phân loại data:

- Quan sát:
  - \* Được chụp trong thời gian thực.
  - \* Không thể tái tạo hoặc lấy lại. Đôi khi được gọi là 'dữ liệu duy nhất'.
  - \* Ví dụ bao gồm chỉ số cảm biến, đo từ xa, kết quả khảo sát, hình ảnh và quan sát của con người.
- Thực nghiệm:
  - \* Dữ liệu từ thiết bị phòng thí nghiệm và trong các điều kiện được kiểm soát.
  - \* Thường có thể tái tạo, nhưng có thể tốn kém để làm như vậy.
  - \* Ví dụ bao gồm trình tự gen, sắc ký đồ, đọc từ trường và quang phổ.

- Mô phỏng:
  - \* Dữ liệu được tạo ra từ các mô hình thử nghiệm nghiên cứu các hệ thống thực tế hoặc lý thuyết.
  - \* Mô hình và siêu dữ liệu trong đó dữ liệu đầu vào quan trọng hơn dữ liệu đầu ra.
  - \* Ví dụ bao gồm các mô hình khí hậu, mô hình kinh tế và kỹ thuật hệ thống.
- Bắt nguồn hoặc biên dịch:
  - \* Kết quả phân tích dữ liệu hoặc tổng hợp từ nhiều nguồn.
  - \* Có thể tái tạo (nhưng rất đắt).
  - \* Ví dụ bao gồm khai thác văn bản và dữ liệu, cơ sở dữ liệu được biên dịch và mô hình 3D.
- Tham chiếu hoặc chuẩn:
  - \* Tập dữ liệu thu thập cố định hoặc không phải trả tiền, thường được đánh giá ngang hàng và thường được xuất bản và sắp xếp.
  - \* Ví dụ bao gồm cơ sở dữ liệu trình tự gen, dữ liệu điều tra dân số, cấu trúc hóa học.
- Định dạng tệp:
  - \* Định dạng tệp nên được chọn để đảm bảo chia sẻ, truy cập lâu dài và bảo quản dữ liệu của bạn.
  - \* Chọn các tiêu chuẩn và định dạng mở để sử dụng lại.
  - \* Nếu bạn đang sử dụng một định dạng khác trong giai đoạn thu thập và phân tích nghiên cứu của mình, hãy đảm bảo bao gồm thông tin trong tài liệu của bạn về các tính năng có thể bị mất khi tệp được chuyển sang định dạng bảo quản, cũng như bất kỳ phần mềm cụ thể nào sẽ cần thiết để xem hoặc làm việc với dữ liệu.

### 3.2.2 JSON

- JavaScript Object Notation (JSON) là một giải pháp thay thế đang thu hút rất nhiều sự chú ý. Điều đầu tiên khi nhắc đến JSON với các nhà phát triển là được thiết kế nhẹ nhàng để có thể dễ dàng đọc, trao đổi dữ liệu và thực thi. Tuy nhiên, đó không phải là lý do duy nhất bạn nên sử dụng JSON cho tích hợp API Restful.
- Các API RESTful phụ thuộc vào việc trao đổi dữ liệu dễ dàng, đáng tin cậy và nhanh chóng. JSON phù hợp với hóa đơn cho từng thuộc tính này. Đây là lý do tại sao sử dụng ngôn ngữ này nhiều sẽ giúp bạn tạo ra các dịch vụ API thú vị.
- Mặc dù CSV là định dạng tệp phổ biến nhất cho dữ liệu “phẳng”, JSON là định dạng tệp phổ biến nhất cho dữ liệu “dạng cây” có khả năng có nhiều lớp, như các nhánh trên cây:

```
1      {
2      [
3          {
4              "id":0,
5              "type":"banana",
6              "amount":12
7          },
8          {
9              "id":1,
10             "type":"apple",
11             "amount":7
12          }
13      ]
14      }
```

- Đối với tệp JSON, bản xem trước tab Dữ liệu sẽ hiển thị một cây tương tác với các nút trong tệp JSON được đính kèm. Bạn có thể nhấp vào các phím riêng lẻ để mở và thu gọn các phần của cây, khám phá cấu trúc của tệp dữ liệu khi bạn tiếp tục. Các tệp JSON không hỗ trợ mô tả hoặc chỉ số cột.

### 3.2.3 Ưu điểm của JSON

- JSON nhanh hơn:  
Cú pháp JSON rất dễ sử dụng. Chúng ta chỉ phải sử dụng như một cú pháp giúp chúng ta dễ dàng phân tích dữ liệu và thực thi dữ liệu nhanh hơn. Vì cú pháp của nó rất nhỏ và trọng lượng nhẹ, đó là lý do mà nó thực thi phản hồi theo cách nhanh hơn. Cách tiếp cận nhẹ của JSON có thể tạo ra những



cải tiến đáng kể trong các API RESTful hoạt động với các hệ thống phức tạp.

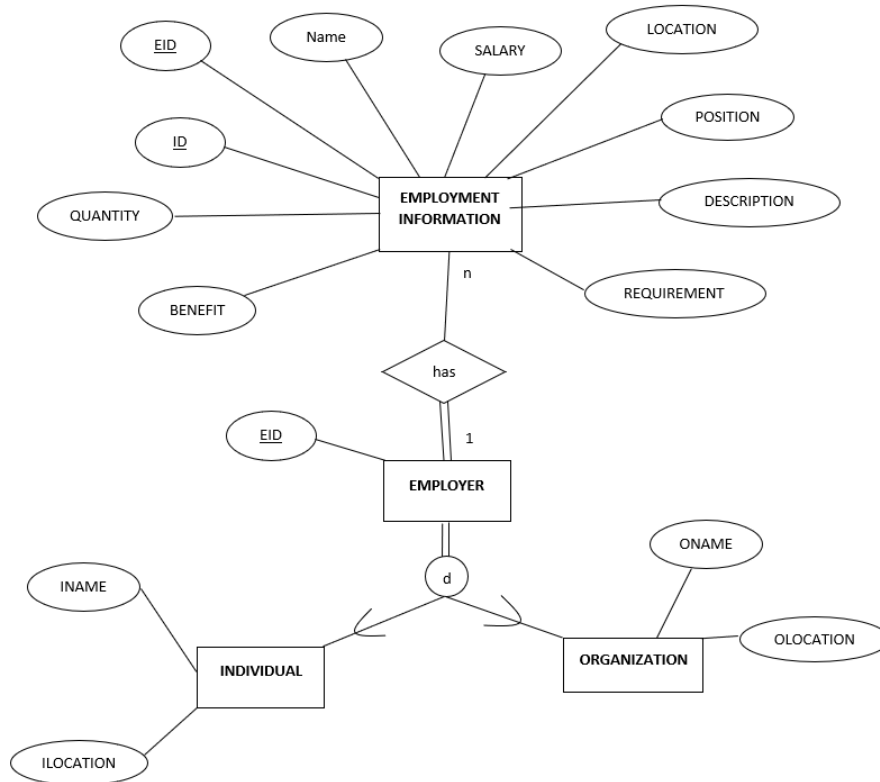
- Phân tích cú pháp máy chủ:  
Phân tích cú pháp phía máy chủ là phần quan trọng mà các nhà phát triển muốn nếu quá trình phân tích cú pháp sẽ nhanh ở phía máy chủ thì chỉ người dùng mới có thể nhận được phản hồi nhanh chóng của phản hồi của họ vì vậy trong trường hợp này phân tích cú pháp phía máy chủ JSON là điểm mạnh mà cho biết chúng tôi sử dụng JSON ở phía máy chủ.
- Hỗ trợ lược đồ:  
Nó có nhiều loại trình duyệt được hỗ trợ tương thích với các hệ điều hành, vì vậy các ứng dụng được tạo bằng mã hóa JSON không đòi hỏi nhiều nỗ lực để làm cho nó tương thích với tất cả các trình duyệt. Trong quá trình phát triển, nhà phát triển nghĩ cho các trình duyệt khác nhau nhưng JSON cung cấp chức năng đó.
- Công cụ chia sẻ dữ liệu:  
JSON là công cụ tốt nhất để chia sẻ dữ liệu ở bất kỳ kích thước nào kể cả âm thanh, video,... Điều này là do JSON lưu trữ dữ liệu trong các mảng nên việc truyền dữ liệu dễ dàng hơn. Vì lý do này, JSON là một định dạng tệp cao cấp cho các API web và để phát triển web.

### 3.2.4 Mô hình hóa data

#### 1. Mô tả:

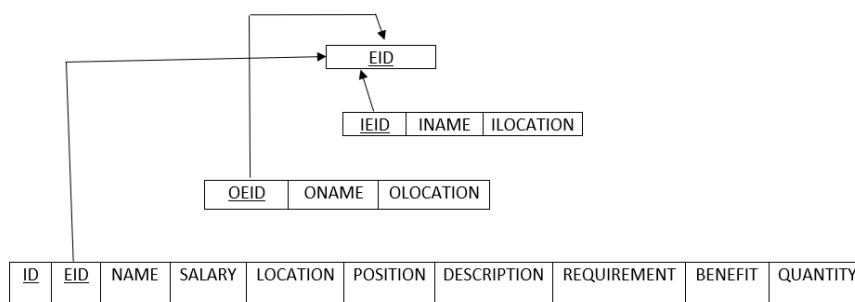
- Bộ dữ liệu thông tin tuyển dụng gồm 2 đối tượng chính là nhà tuyển dụng và thông tin tuyển dụng, trong đó nhà tuyển dụng có thể là tổ chức hoặc cá nhân.
- Tổ chức có thể là một công ty, tập đoàn. Cần lưu trữ các thông tin bao gồm mã nhà tuyển dụng, tên tổ chức, địa chỉ.
- Cá nhân cần lưu trữ các thông tin bao gồm mã nhà tuyển dụng, tên, địa chỉ.
- Về thông tin tuyển dụng, cần lưu trữ các thông tin về việc làm cần tuyển như mã số, tên việc làm, mô tả công việc, mức lương, nơi làm việc, yêu cầu, lợi ích, số lượng...

2. Sơ đồ liên kết thực thể (ER Diagram) cho CSDL theo các mô tả:



**Hình 1:** ER diagram

3. Ánh xạ lược đồ liên kết thực thể sang lược đồ dữ liệu quan hệ:



**Hình 2:** Ánh xạ lược đồ

### 3.2.5 Cách kết nối Mysql sử dụng python

- To update pip in the virtual environment, type the following command:

```
1 pip install -U pip
```

- Type the command for the package you want to install:

- To install the mysqlclient package, type the following command:

```
1 pip install mysqlclient
```

- To install the mysql-connector-python package, type the following command:

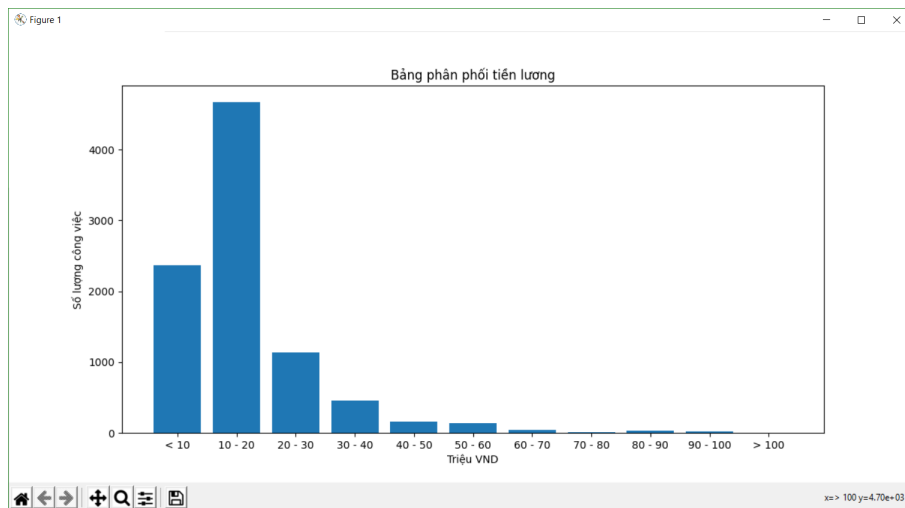
```
1 pip install mysql-connector-python
```

- To install the pymysql package, type the following command:

```
1 pip install pymysql
```

## 4 Phân tích dữ liệu thu thập được từ các trang

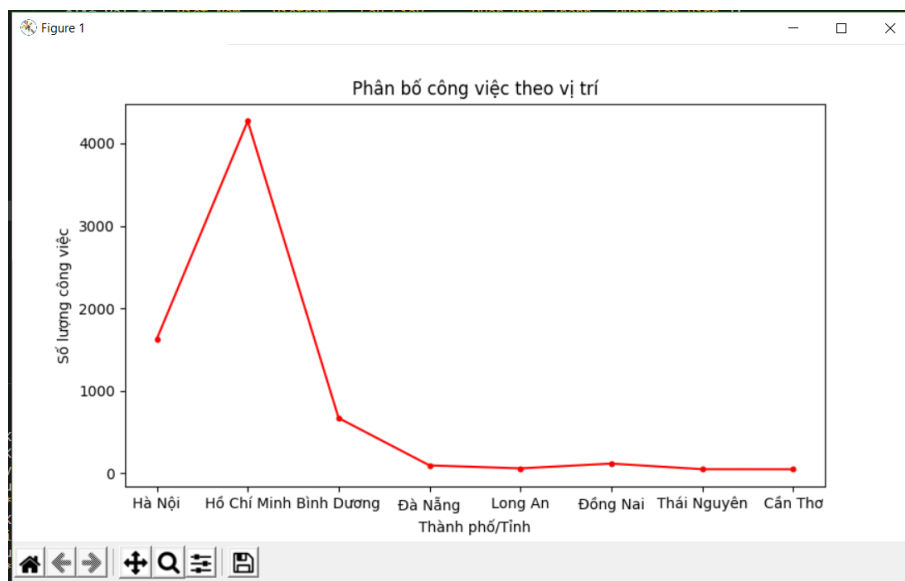
### 4.1 Sơ đồ phân bố các tin tuyển dụng theo giá trị tiền lương



Hình 3: Sơ đồ phân bố tin tuyển dụng theo giá trị tiền lương

- Theo như sơ đồ ở trên, mức lương mà các nhà tuyển dụng đưa ra nằm chủ yếu trong phân khúc từ 10.000.000 VNĐ - 20.000.000 VNĐ, tiếp theo là phân khúc dưới 10.000.000 VNĐ và các phân khúc còn lại thấp hơn hẳn so với hai phân khúc trước.
- Có thể thấy, nhóm đối tượng nằm trong phân khúc tiền lương 10.000.000 VNĐ - 20.000.000 VNĐ là những đối tượng cho tri thức, trình độ chuyên môn nhưng chưa có nhiều kinh nghiệm như: kỹ sư, kế toán, chuyên viên thiết kế, ...
- Bên cạnh đó, nhóm đối tượng nằm trong phân khúc tiền lương dưới 10.000.000 VNĐ tập trung vào các đối tượng không cần trình độ chuyên môn, như bảo mẫu, bảo vệ, bán quán, ...

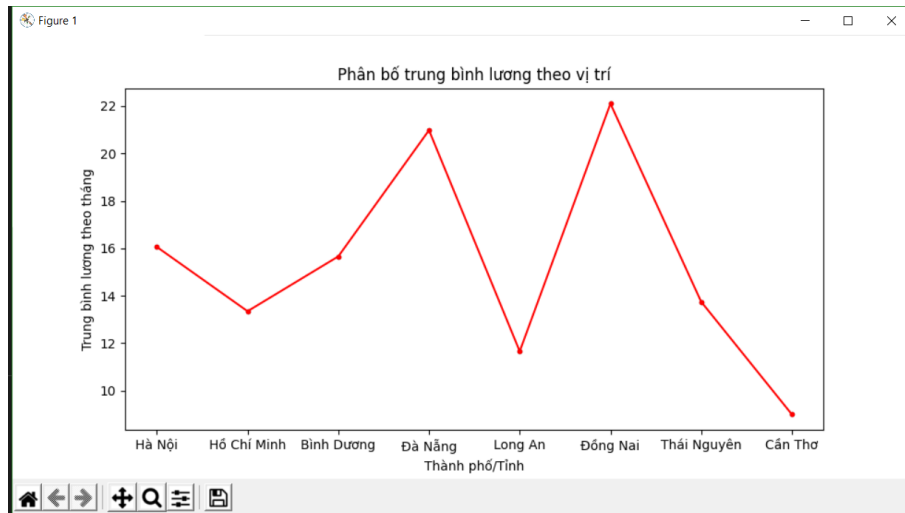
## 4.2 Sơ đồ phân bố các tin tuyển dụng theo vị trí địa lý



Hình 4: Sơ đồ phân bố tin tuyển dụng theo vị trí địa lý

- Theo như sơ đồ ở trên, hai trung tâm kinh tế lớn nhất của cả nước là Thành phố Hồ Chí Minh và Hà Nội có mức tuyển dụng cao nhất, đặc biệt là Thành phố Hồ Chí Minh với hơn 4000 tin tuyển dụng.
- Bên cạnh đó, các trung tâm công nghiệp như Đồng Nai, Bình Dương, Quảng Ninh,... cũng có nhu cầu tìm kiếm nhân lực cho các nhà máy, xí nghiệp sản xuất,...

### 4.3 Sơ đồ phân bố giá trị tiền lương trung bình theo vị trí địa lý



Hình 5: Sơ đồ phân bố tiền lương trung bình theo vị trí địa lý

- Theo như sơ đồ ở trên, hai trung tâm kinh tế lớn nhất của cả nước là Thành phố Hồ Chí Minh và Hà Nội tuy có mức tuyển dụng cao, nhưng mặt bằng lương không cao so với các trung tâm công nghiệp như Đồng Nai, Đà Nẵng,...
- Có thể hiểu đơn giản do đặc thù chuyên môn nghề nghiệp ở các trung tâm công nghiệp mà mặt bằng tiền lương cũng nhỉnh hơn so với các nơi khác.

### 4.4 Các công cụ hỗ trợ phân tích

- Sử dụng ngôn ngữ Python để phân tích dữ liệu.
- Sử dụng module **pandas** từ pip để đọc dữ liệu từ file **xlsx**.
- Sử dụng module **matplotlib.pyplot** để vẽ đồ thị minh họa.

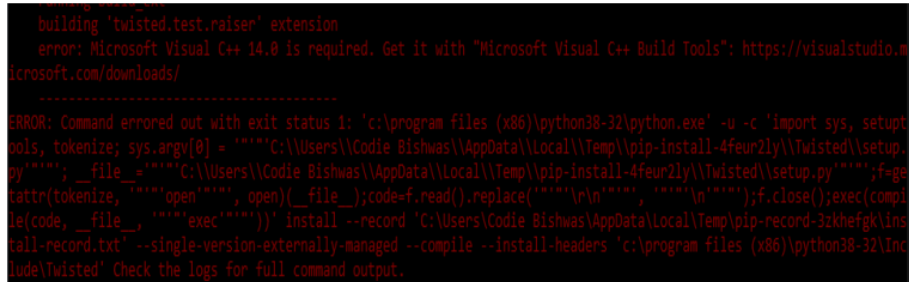
## 5 NHỮNG KHÓ KHĂN GẶP PHẢI

### 5.1 Cài đặt thư viện scrapy của python

- Khi cài đặt thư viện scrapy của python bằng command trên terminal:

```
1 pip install scrapy
```

- Máy có thể bị lỗi như sau:



```
building 'twisted.test.raiser' extension
error: Microsoft Visual C++ 14.0 is required. Get it with "Microsoft Visual C++ Build Tools": https://visualstudio.m
icrosoft.com/downloads/

-----
ERROR: Command errored out with exit status 1: 'c:\program files (x86)\python38-32\python.exe' -u -c 'import sys, setup
tools, tokenize; sys.argv[0] = '''C:\Users\Codie Bishwas\AppData\Local\Temp\pip-install-4feur2ly\Twisted\setup.
py'''; __file__ = '''C:\Users\Codie Bishwas\AppData\Local\Temp\pip-install-4feur2ly\Twisted\setup.py'''; f=get
attr(tokenize, '''open''', open)(__file__); code=f.read().replace(''\n''', '''\n'''); f.close(); exec(compile
(code, __file__, '''exec'''))' install --record 'C:\Users\Codie Bishwas\AppData\Local\Temp\pip-record-3zkhefgk\ins
tall-record.txt' --single-version-externally-managed --compile --install-heads 'c:\program files (x86)\python38-32\Inc
lude\Twisted' Check the logs for full command output.
```

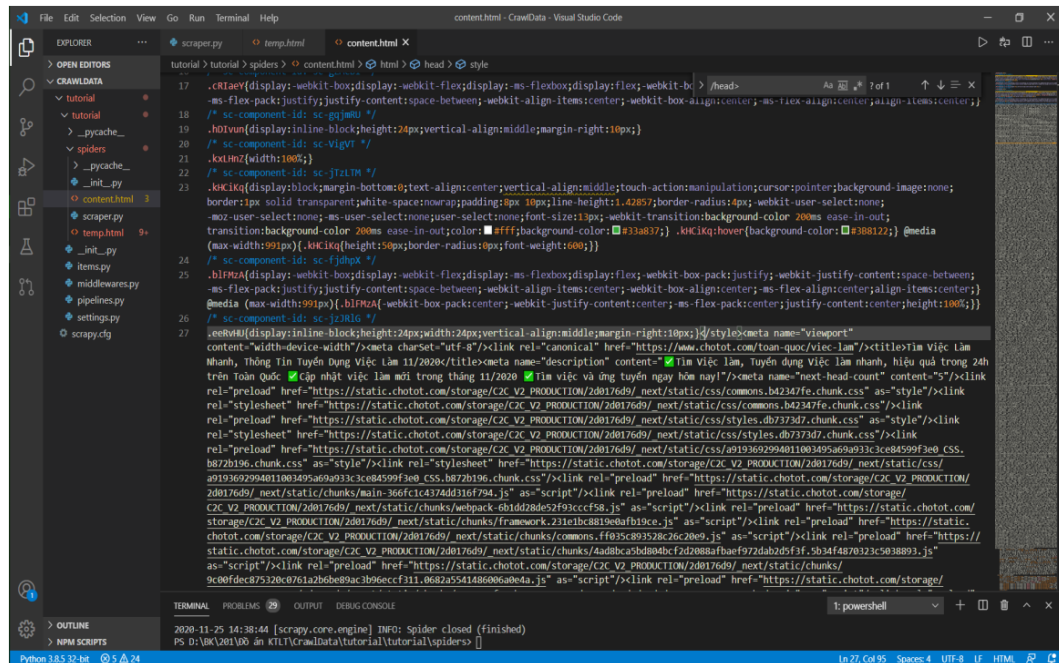
Hình 6: Lỗi khi cài đặt thư viện scrapy của python

- Để khắc phục lỗi này, mở terminal cài đặt Twisted bằng command trên terminal:

```
1 pip install Twisted
```

## 5.2 File HTML bị lỗi dẫn đến việc tìm các tag của file khó khăn

Cụ thể ở trang web <https://chotot.vn/>, khi ta gọi request trả về nội dung của file HTML, ta thu được như hình:



Hình 7: Nội dung file HTML trả về khi gửi request đến trang web

Để xử lý trường hợp này, ta đổi nội dung file sang dạng chuỗi và dùng các method của python để lấy các trường dữ liệu tương ứng.

## 5.3 Trang web yêu cầu đăng nhập để hiển thị thêm thông tin

Cụ thể ở trang web <https://vietnamworks.com/>, khi ta gọi request trả về nội dung của file HTML, web chỉ trả về một số lượng trường dữ liệu nhất định và số lượng link trả về cũng bị hạn chế. Nhóm tìm hiểu thêm về selenium để giải quyết vấn đề này.

## 6 Đề xuất ứng dụng

Thiết kế website giới thiệu việc làm:

- Trang chủ
- Giới thiệu
- Thông tin dịch vụ - việc làm
- Mô tả chi tiết việc làm
- Liên hệ

Các tính năng cơ bản:

- Khách: xem các thông tin public trên trang web, cho phép đăng kí, đăng nhập.
- Thành viên (sau khi đã đăng nhập): cho phép thực hiện một số hàm chức năng cơ bản: thay đổi thông tin cá nhân, mật khẩu, tìm kiếm thông tin tuyển dụng, xem các xu hướng hiện tại.
- Quản trị viên: hiện thực các tính năng quản lý:
  - Quản lý thành viên (xem thông tin, xóa thành viên,...).
  - Tính năng quản lý (xem, thêm, sửa, xóa) các tài nguyên của ứng dụng web như thông tin tuyển dụng, các công việc đang có nhu cầu lớn. . .
  - Hiện thực phân trang hiển thị cho các tính năng quản lý

Thực hiện kiểm tra dữ liệu đầu vào (sử dụng cả kiểm tra bằng javascript (client side) và PHP (server side)).

Tính năng tìm kiếm tài nguyên đơn giản trên trang web.

## 7 LINK PROJECT

Link project: [https://github.com/ThuanNguyen1210/201CO10313\\_BCH-DOAN](https://github.com/ThuanNguyen1210/201CO10313_BCH-DOAN)





## Tài liệu

- [1] Scrapy [online], from: <https://scrapy.org/>, viewed 22/11/2020.
- [2] Selenium [online], from: <https://www.selenium.dev/>, viewed 23/11/2020.
- [3] Beatifulsoup [online], from: <https://pypi.org/project/beautifulsoup4/>, viewed 24/11/2020.