**Tài liệu mô tả**

**\*\*\*\*\*\*CRAWL DATA**\*\*\*\*\*\*

Dữ liệu được crawl từ trang web http://sis.hust.edu.vn/ModuleProgram/CourseLists.aspx:

Crawl dữ liệu bằng selenium lấy dữ liệu thông qua XPATH

Các thành phần dữ liệu cần crawl trên trang web này

Xpath

//\*[@id="MainContent\_gvCoursesGrid\_DXDataRow0"] #Dữ liệu bảng

//\*[@id="MainContent\_gvCoursesGrid\_DXDRow0"] #Dữ liệu chi tiết

Các thành phần cần tương tác để lấy dữ liệu

Xpath

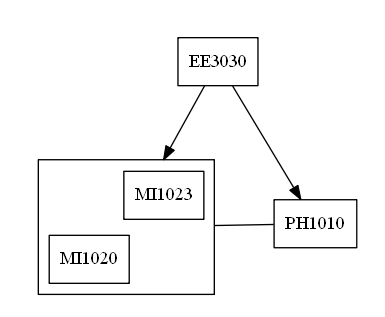
//\*[@id="MainContent\_gvCoursesGrid\_DXDataRow0"]/td[1] #show dữ liệu chi tiết

//\*[@id="MainContent\_gvCoursesGrid\_DXPagerBottom"]/b[5]/img #nút chuyển trang kế tiếp (tại page6 Xpath này sai thứ tự thẻ b nên thay bằng

**\*\*\*\*\*\*VẼ GRAPHVIZ​​**​\*\*\*\*\*\*

Hình được phân tích và vẽ ra theo quy tắc:

* mũi tên biểu diễu quan hệ phụ thuộc
* đường thẳng biểu diễn quan hệ AND
* Các node trong cùng một cụm không được kết nối bởi đường thẳng sẽ là quan hệ OR



Như hình vẽ trên ta có môn EE3030 có 2 phụ thuộc phụ thuộc vào cụm (MI1023 và MI1020) và môn PH1010 ,

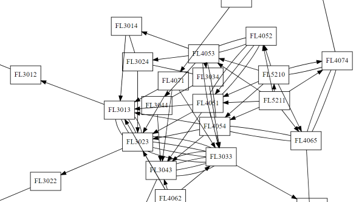
Do cụm (MI1023 và MI1020) và môn PH1010  nối với nhau bởi đường thẳng nên 2 thành phần này tương ứng với AND

Trong cụm (MI1023 và MI1020) 2 môn không có kết nối giữa 2 môn ở trong nên quan hệ này là OR

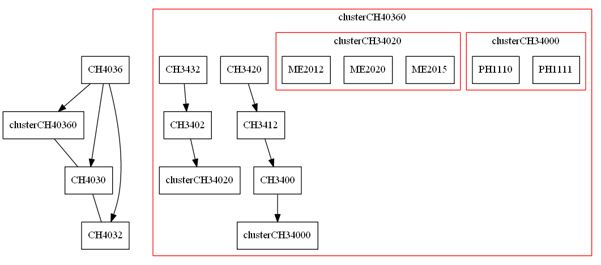
Vậy EE3030 sẽ phụ thuộc tương ứng là (MI1020 OR MI1023) AND PH1010.

Em có vẽ ra quan hệ của các môn còn lại dựa vào tên học phần và dữ liệu mở rộng tương ứng của nó sau đó em vẽ bằng graphviz sau đó lưu vào thư mục dotsource.

Từ những dữ liệu đó em có phân tích và vẽ phụ thuộc chi tiết hơn thì nó sinh ra một số hình vẽ khá khó nhìn như sau:



Em có chỉnh lại bằng cách vẽ riêng rẽ ra và nhìn quan hệ chung thông qua cluster thì được hình dễ đọc hơn như sau:



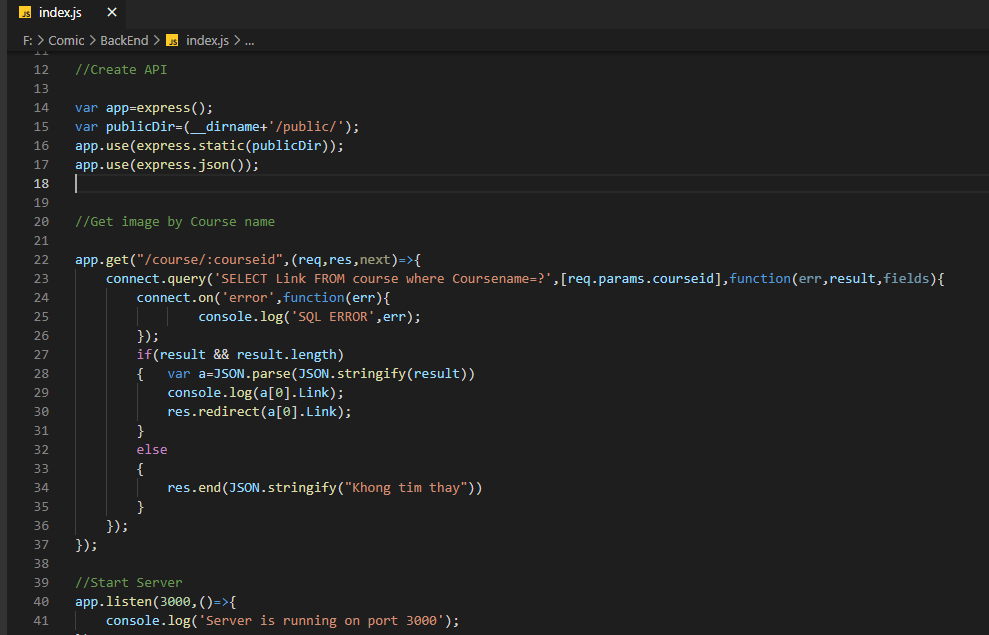
Sau đó em lưu dữ liệu trên vào thư mục dotsourcemoredetail

Link dữ liệu file dot và png của dotsource dotsourcemoredetail:

<https://drive.google.com/drive/folders/138Vy2GEwfflWLX6li10PBNpgtINlo7zg?usp=sharing>

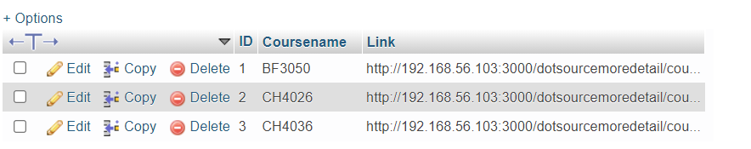
**\*\*\*\*\*\*API​​**​\*\*\*\*\*\*

Em có viết API server bằng nodejs và sử dụng database MySQL

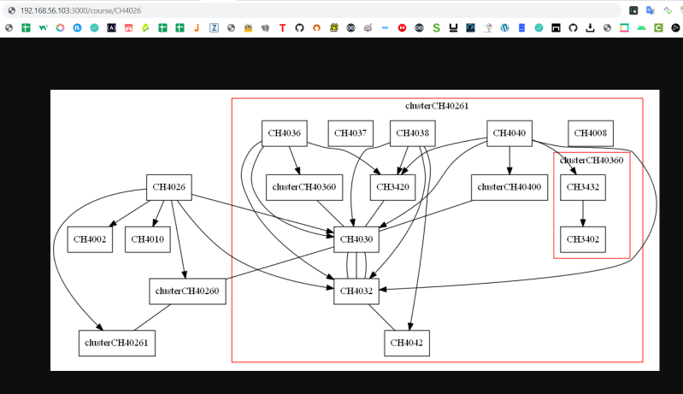


Khi gọi API thì server sẽ truy vấn cơ sở dữ liệu để lấy được đường dẫn tới file ảnh và sẽ chuyển hướng đến API trả lại kết quả file ảnh

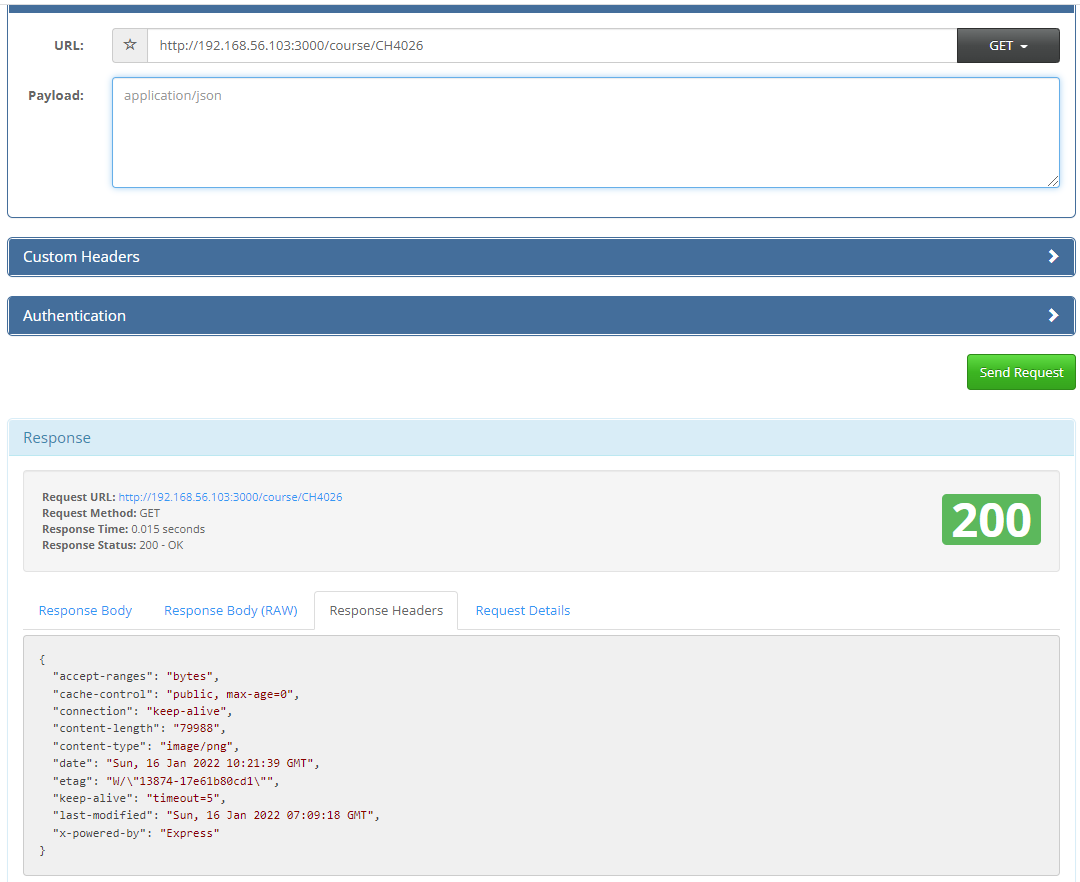
Database sẽ có 1 bảng bao gồm 2 trường đó là tên học phần và đường dẫn để lấy file ảnh từ server lưu trữ file ảnh



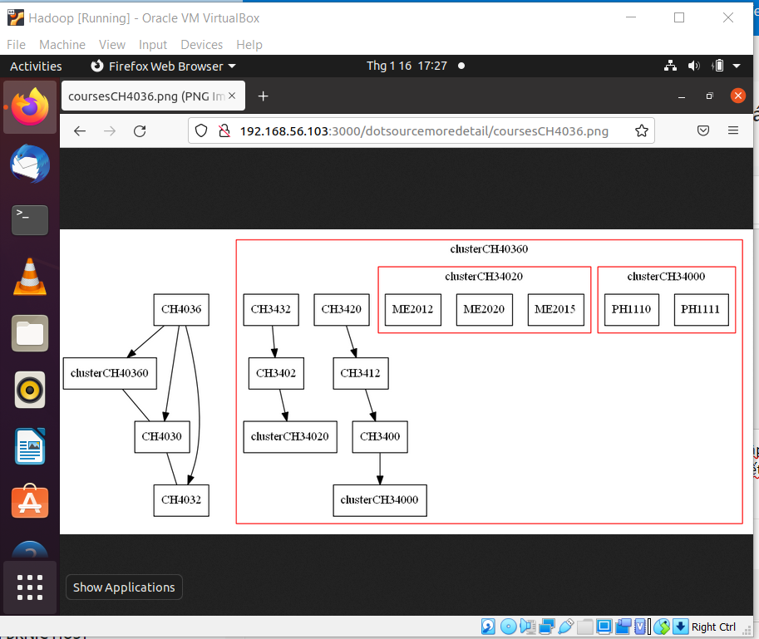
 Đây là kết quả trả về khi truy cập vào <http://192.168.56.103:3000/course/CH4026> (IP [192.168.56.103](http://192.168.56.103:3000/course/CH4026) này là IP  private của 1 adapter của máy em)



Test bằng YARC có thể thấy dữ liệu được truyền theo kiểu binary và có type là png



Do chưa triển khai bằng máy có thể truy cập vào nhờ IP public để truy cập được từ internet nên em có mô phỏng lại bằng cách sử dụng máy ảo để xem dữ liệu có thể nhận được từ máy client không và kết quả là đã nhận được:



Trên đây là những kết quả mà em đã thực hiện thêm, em mong nhận được thêm những lời khuyên từ thầy để hoàn thiện thêm nội dung này.

**\*\*\*\*\*\*Nội dung các file chương trình chính​​**​\*\*\*\*\*\*

**Crawldata.py** crawl dữ liệu và lưu dữ liệu thành 2 file csv là CouseListdata.csv và CourseListdataextend.csv

**graphcourseupdate.py** vẽ graphviz cho môn học dựa vào dữ liệu học phần và dữ liệu mở rộng của 1 môn dữ liệu được lưu vào thư mục dotsource

**graphcoursemoredetail.py** vẽ graphviz cho môn học dựa vào dữ liệu học phần của 1 môn và các môn phụ thuộc và dữ liệu mở rộng tương ứng