BÁO CÁO THỰC HÀNH

**Môn học: Bảo mật web và ứng dụng**

**Kỳ báo cáo: Buổi 03 (Session 03)**

**Tên chủ đề: RECON**

*GVHD: Ngô Khánh Khoa*

*Ngày báo cáo: 22/04/2024*

1. **THÔNG TIN CHUNG:**

*(Liệt kê tất cả các thành viên trong nhóm)*

Lớp: NT213.O21.ANTN

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Họ và tên** | **MSSV** | **Email** |
| 1 | Phạm Ngọc Thơ | 21522641 | 21522641@gm.uit.edu.vn |

1. **NỘI DUNG THỰC HIỆN:[[1]](#footnote-1)**

|  |  |  |
| --- | --- | --- |
| **STT** | **Công việc** | **Kết quả tự đánh giá** |
| 1 | Kịch bản 01 | 100% |
| 2 | Kịch bản 02 | 100% |
| 3 | Kịch bản 03 | 100% |
| 4 | Kịch bản 04 | 100% |

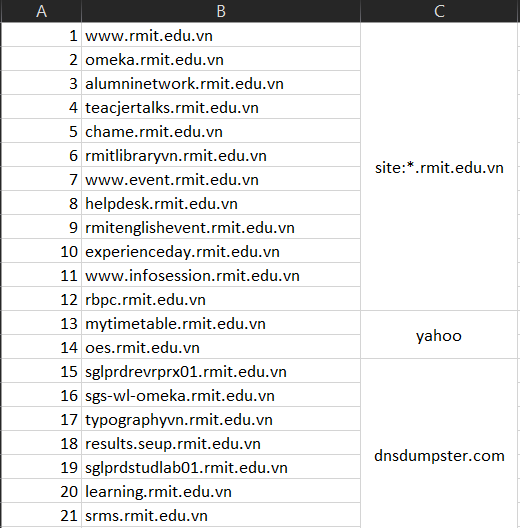
**Phần bên dưới của báo cáo này là tài liệu báo cáo chi tiết của nhóm thực hiện.**

BÁO CÁO CHI TIẾT

## **Subdomain Enumeration:**

1. **Liệt kê thông qua các nguồn trên internet:**

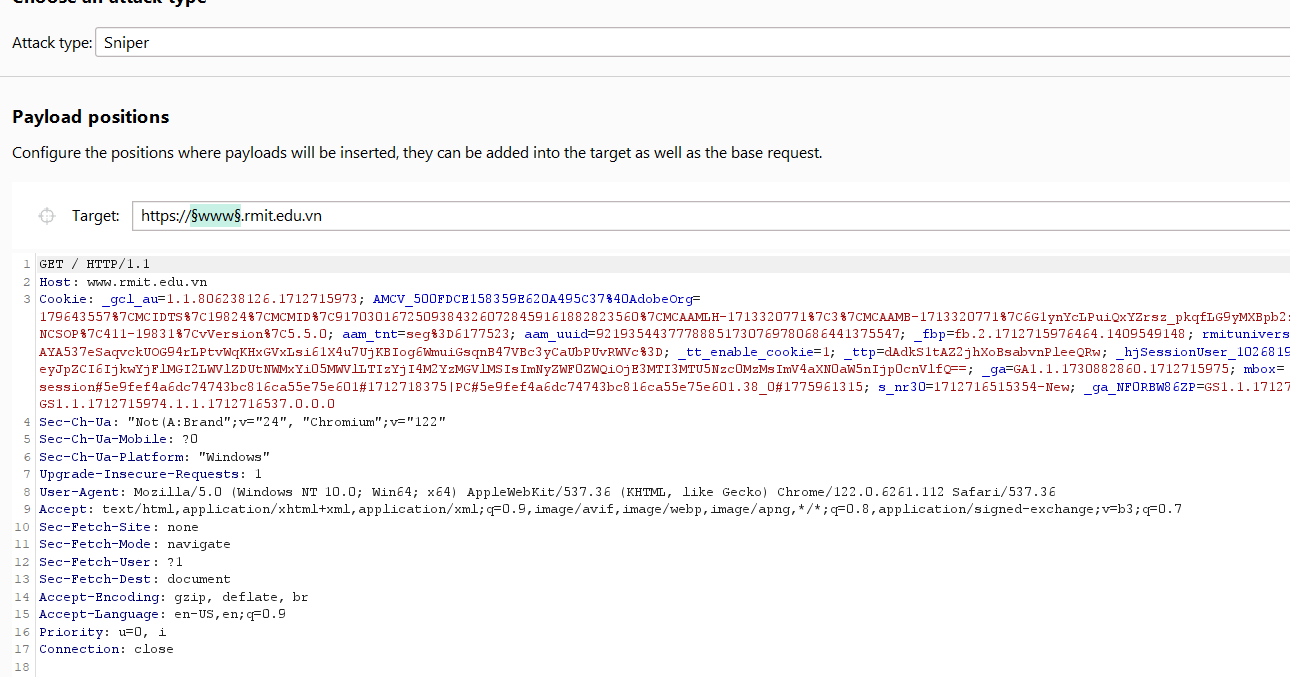
**Bài tập 1:** **Liệt kê ra ít nhất 100 tên miền phụ của rmit.edu.vn, kết quả được lưu trong file csv:**



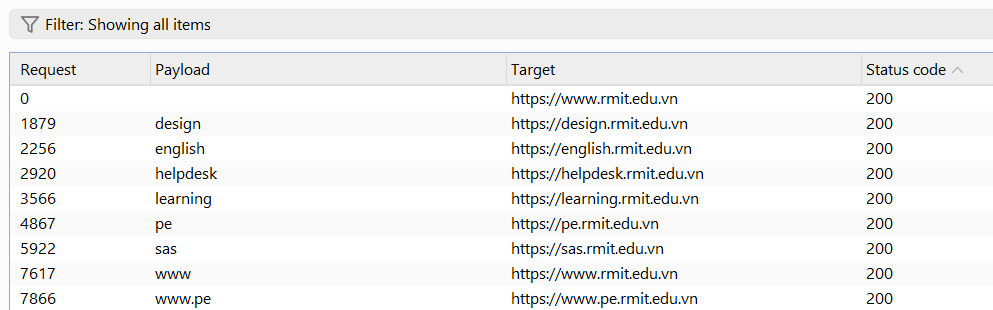
1. **Tìm kiếm chủ động tên miền thông qua kĩ thuật brute-force:**

**Bài tập 2: Dựa vào các tên miền phụ đã tìm kiếm được ở bài tập 1 và các tên miền đã bruteforce được thêm bằng burpsuite intruder. Phân loại các tên miền có kết quả trả về status code 200 và các tên miền có kết quả trả về khác**

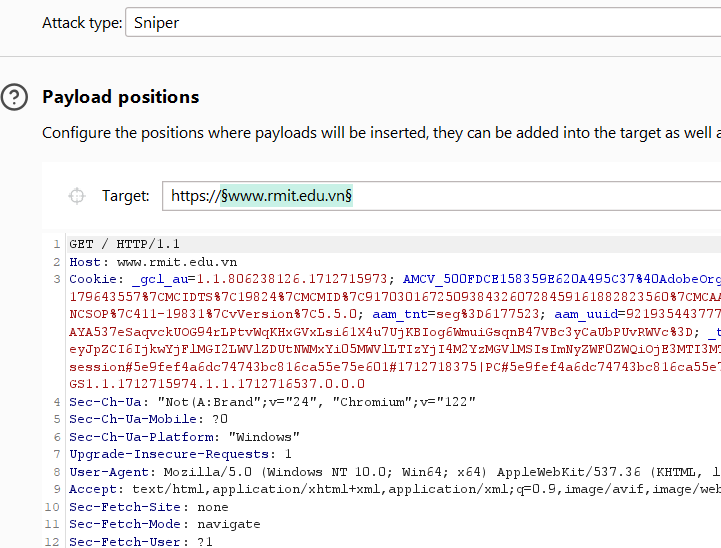
* Bước 1: Truy cập vào trang web muốn tìm tên miền phụ thông qua browser của burpsuite để ghi nhận HTTP history, sau đó send request này vào intruder.
* Bước 2: Thêm subdomain vào url để tìm kiếm, sử dụng nút Add$:

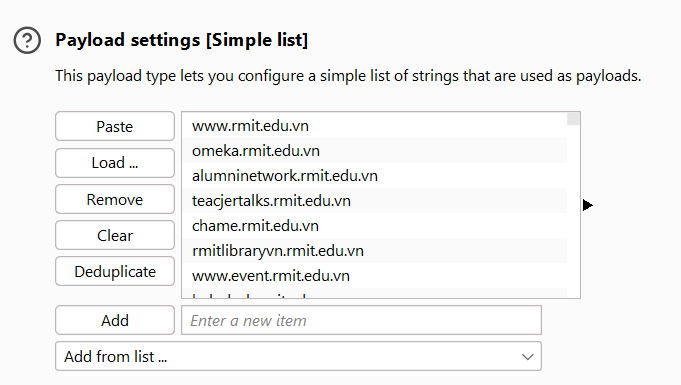


* Bước 3: Qua tab payload, tiến hành load list subdomain vào. Em dùng list subdomain từ link sau để brute force: [subdomain](https://gist.githubusercontent.com/michenriksen/60a5fd10968de2bbe0f6571b6a67e7bb/raw/9f633632ba81dd32ae6219a4c53d1e887aa345c0/subdomains.lst) (gồm 8211 mẫu).
* Bước 4: Start attack và kiểm tra kết quả trả về, có thể tuỳ chỉnh thời gian giữa 2 câu request tại Resource pool để phù hợp với tốc độ trả về của trang web cần tìm kiếm, tránh việc trang web bị ddos. Kết quả tên miền có status code 200 là:

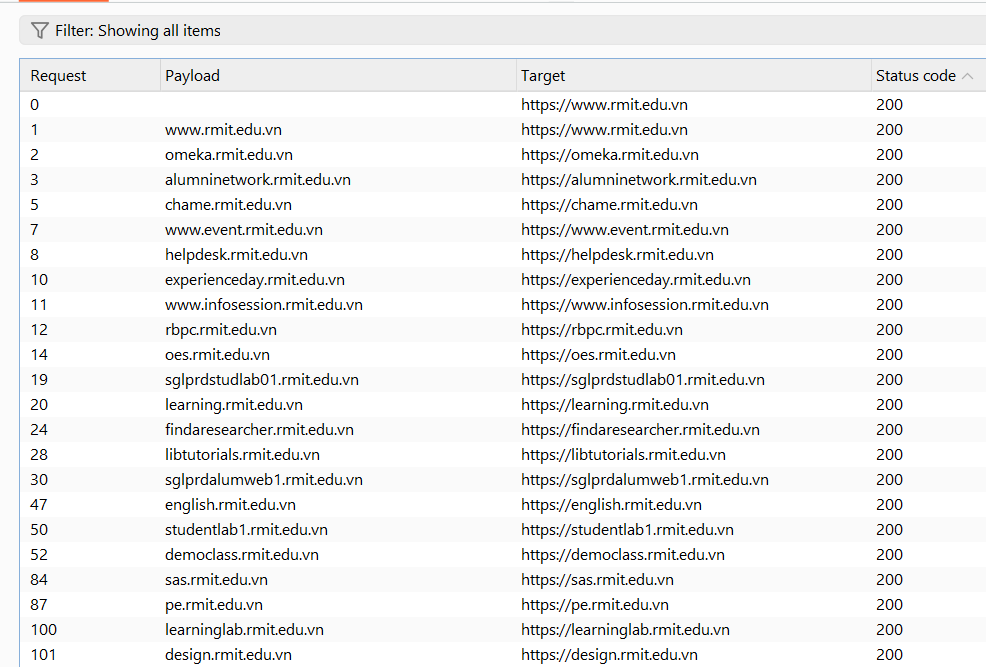


* Tương tự, em sẽ lặp lại 4 bước trên. Lần này em sẽ sử dụng subdomain em đã tìm được ở bài tập 1:





Kết quả:



=> Tên miền do em tìm kiếm trước sẽ có status code 200 nhiều hơn so với việc sử dụng một từ điển tên miền có sẵn trên internet - dù từ điển có số lượng mẫu rất lớn. Qua đó, cho thấy để việc brute force có hiệu quả, nên điều tra để tìm được những thông tin riêng biệt của đối tượng trước khi bắt đầu tấn công.

1. **Host and Port Discovery**
2. **Tìm kiếm các host tương ứng:**

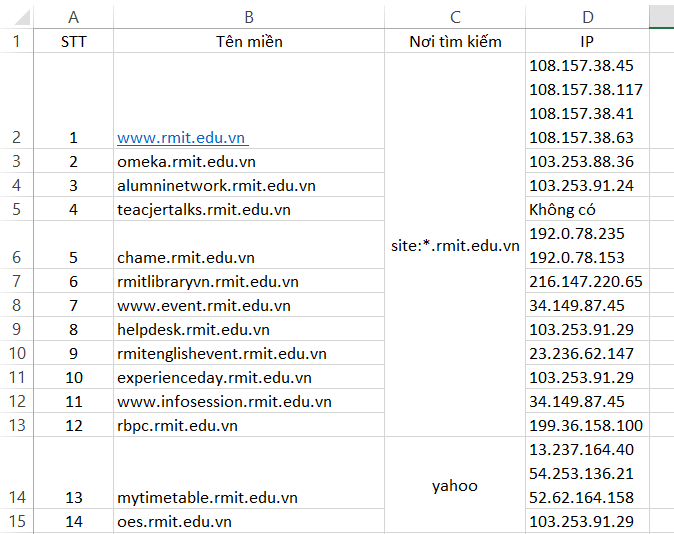
**Bài tập 3: Ghi nhận lại các địa chỉ IP của tên miền phụ tìm được của \*.rmit.edu.vn. Kết quả lưu trong file csv.**

Có thể tìm được IP của tên miền bằng *nslookup,* kết quả đầy đủ em sẽ lưu trong file *baitap.xlsx*:

Để bài tập được nhanh, thay vì làm từng tên miền, em sẽ dùng vòng lặp:

*for /f "tokens=\*" %i in (baitap2.txt) do nslookup %i >> temp.txt*

*baitap2.txt* là tên file chứa các tên miền em tim được ở bài tập 1, *temp.txt* là file lưu kết quả của vòng lặp trên. Để tránh nhầm lẫn với các file khác và để kết quả hiển thị gọn hơn, em sẽ sửa tên và lưu kết quả cuối cùng chung file csv ở bài 1:

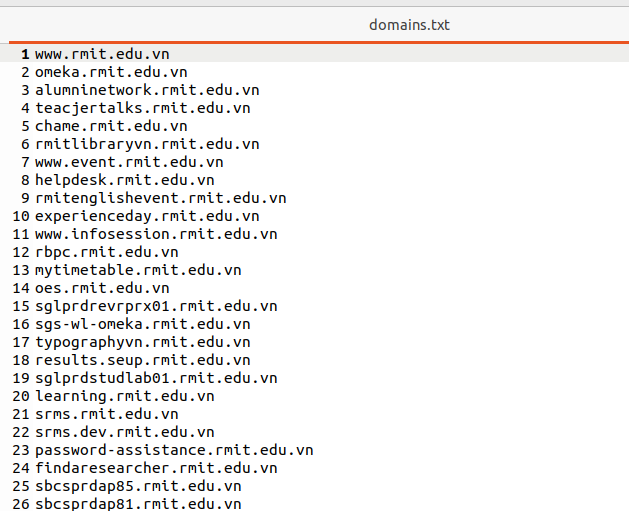


1. **Tìm kiếm port tương ứng:**

**Bài tập 4: Thực hiện scan 1000 port phổ biến trên các danh sách IP tìm được của \*.rmit.edu.vn. Báo cáo kết quả tìm được trong file csv.**

**Danh sách 1000 port tại:** [**https://raw.githubusercontent.com/HeckerBirb/top-nmap-ports-csv/master/top-1000-most-popular-tcp-ports-nmap-sorted.csv**](https://raw.githubusercontent.com/HeckerBirb/top-nmap-ports-csv/master/top-1000-most-popular-tcp-ports-nmap-sorted.csv)

* Em lưu tên miền vào file *domains.txt:*

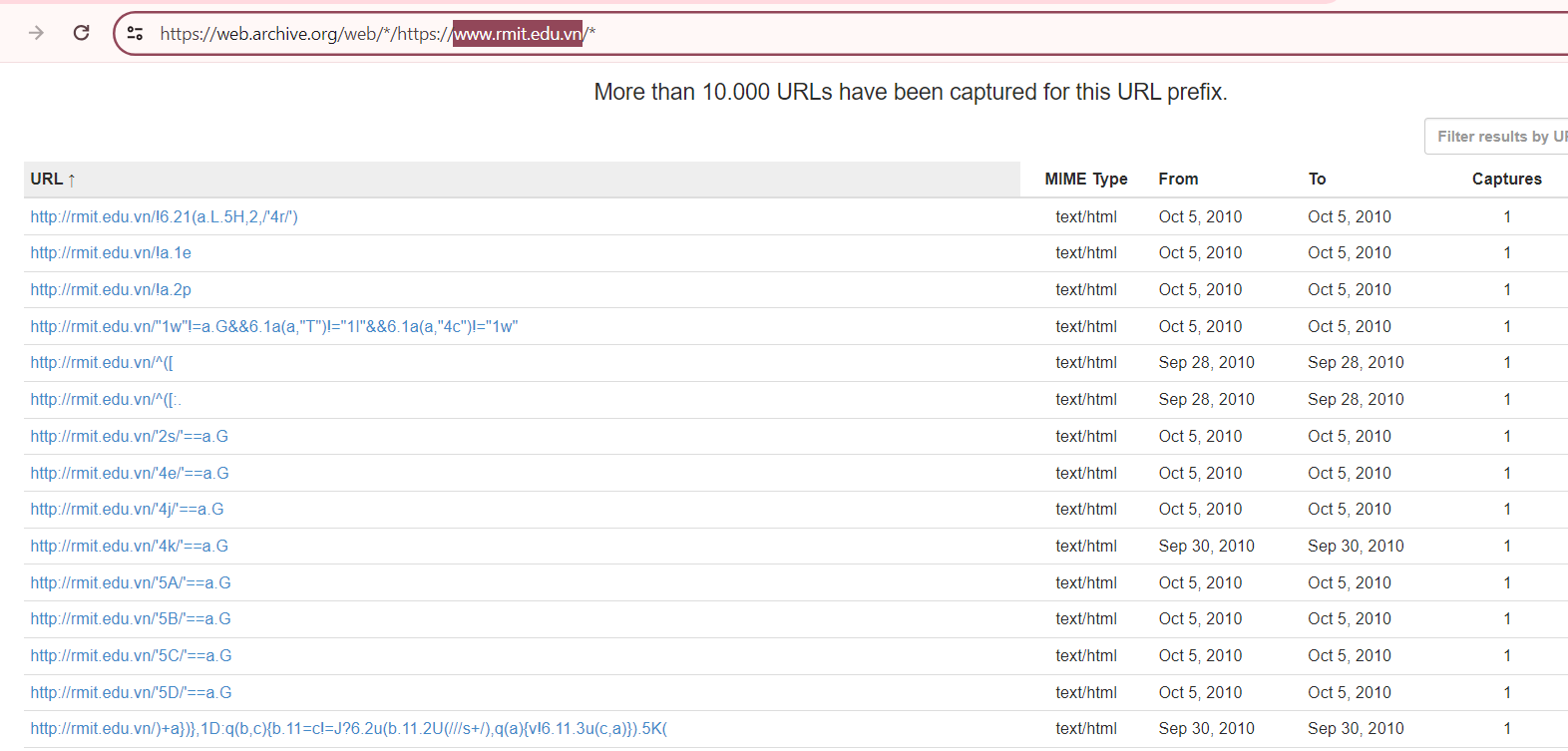


* Để scan tất cả port của tên miền, dùng cú pháp: *nmap -p- <tên miền>*
* Kết quả em sẽ lưu trong file baitap.xlsx.

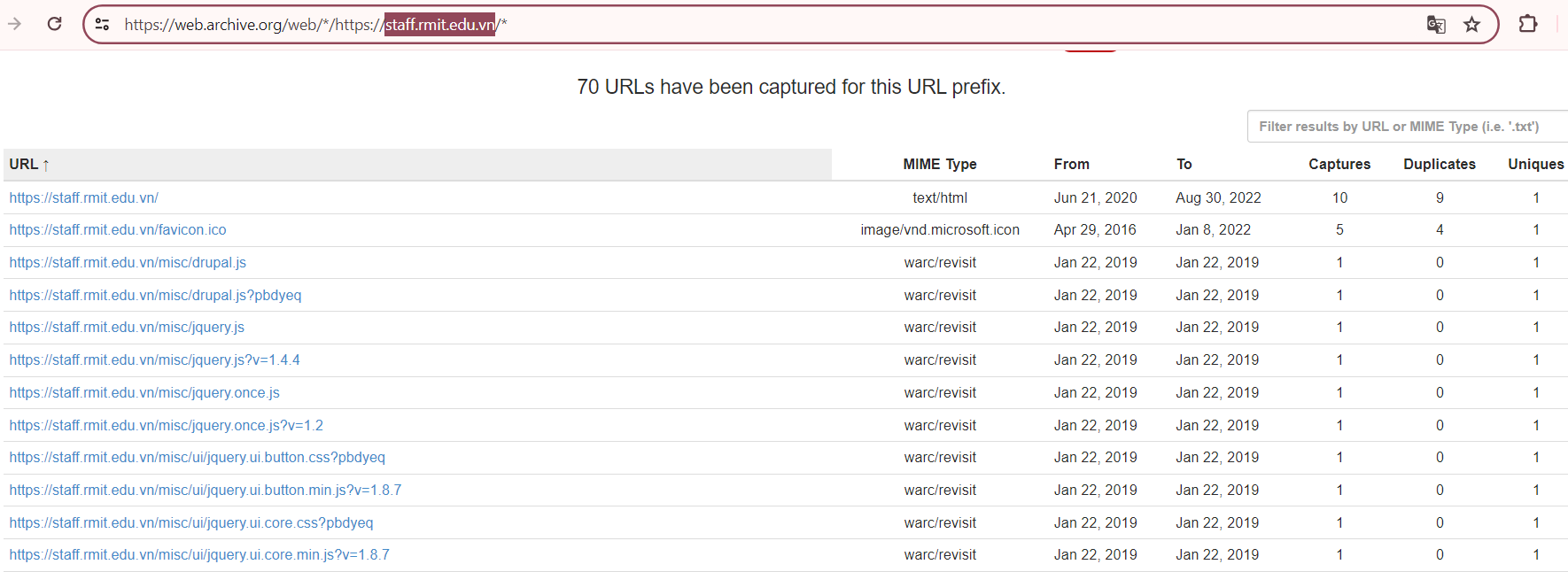
1. **Truy tìm thông tin của website**
2. **Tìm kiếm thông qua Internet Archive:**

**Bài tập 5: Sử dụng https://web.archive.org/ tìm kiếm và ghi nhận lại dữ liệu quá khứ các tên miền phụ không còn tồn tại hiện nay của \*.rmit.edu.vn.**

* Đối với tên miền [www.rmit.edu.vn](http://www.rmit.edu.vn), em tìm được thông tin sau:

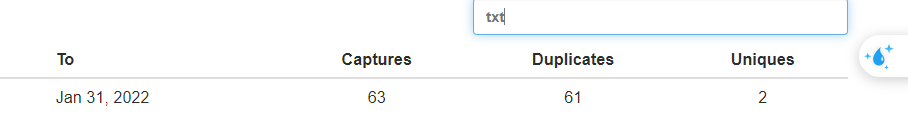


* Tương tự, staff.rmit.edu.vn:



Em lọc kết quả với từ khóa txt thì tìm được 1 kết quả là file *robots.txt*. File "robotx.txt" là một tập tin chứa các chỉ thị cho các trình crawler (chương trình tự động duyệt web) về cách mà họ nên quét và truy cập trang web. Tập tin robots.txt thường được đặt ở thư mục gốc của một trang web và cung cấp các hướng dẫn cho các trình crawler của các công cụ tìm kiếm, như Googlebot, về cách mà họ nên quét trang web đó. Cụ thể, file robots.txt thường chứa các phần như: User-agent, Disallow, Allow, Sitemap.:



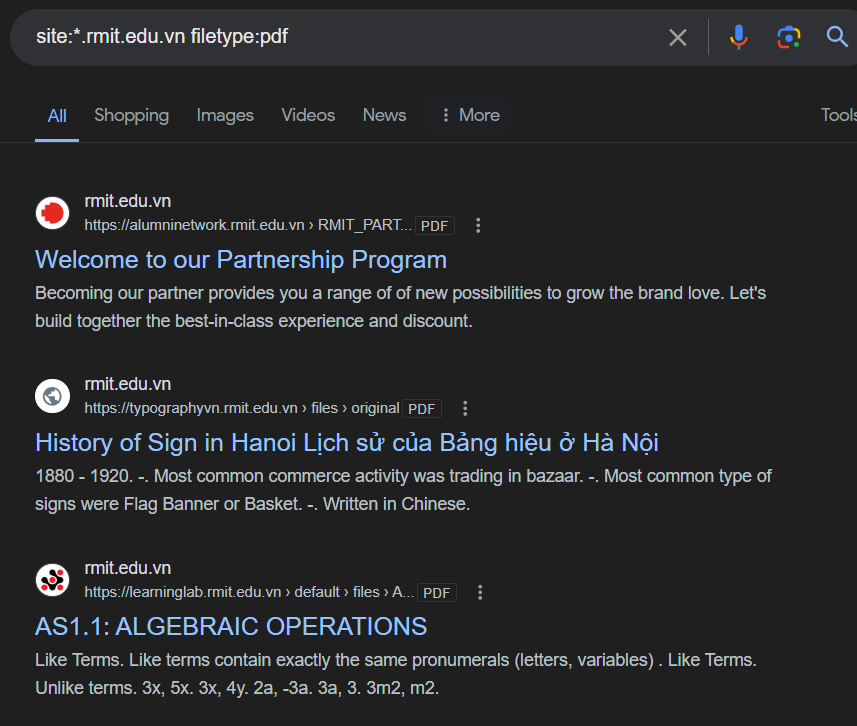


* Ở 2 lần tìm kiếm trên, dựa vào thời gian tồn tại của domain có thể kết luận rằng [www.rmit.edu.vn](http://www.rmit.edu.vn) được sử dụng rất nhiều và có nhiều domain phụ. Còn staff.rmit.edu.vn ít phổ biến hơn và dường như đã ngừng hoạt động từ 30/8/2022.

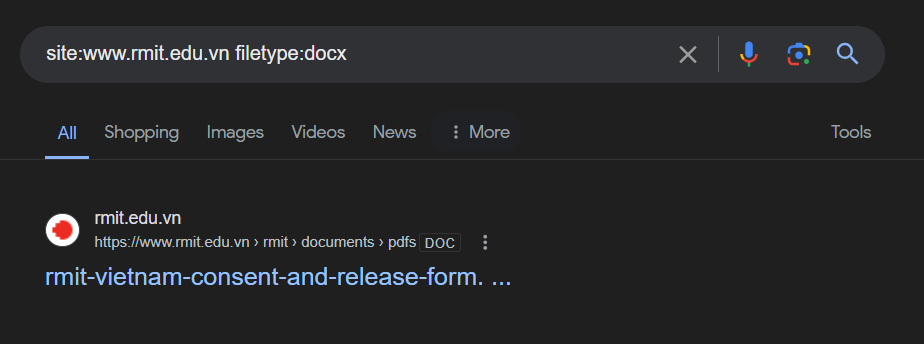
1. **Tìm kiếm thông qua google dork:**

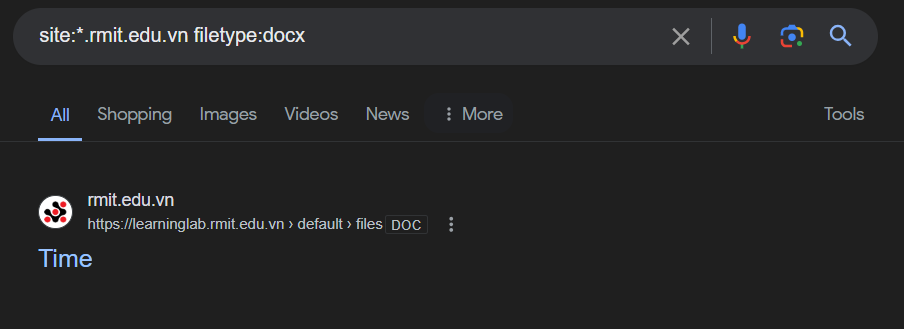
**Bài tập 6: Tìm kiếm các tập tin pdf, excel, word, trên \*.rmit.edu.vn.**

* Sử dụng Google dork để tìm kiếm trên \*.rmit.edu.vn, file pdf:

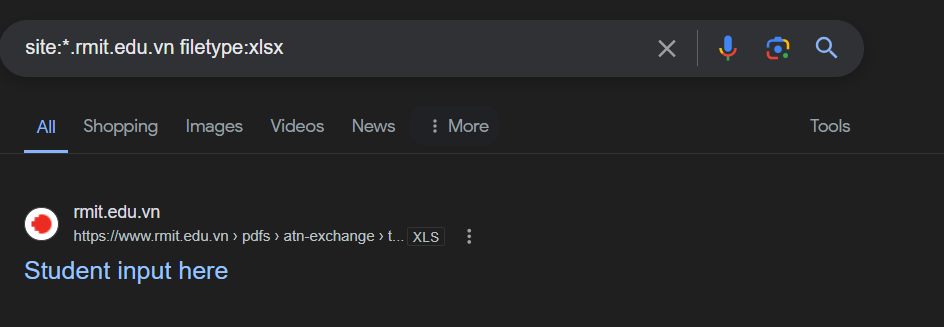


* File word:





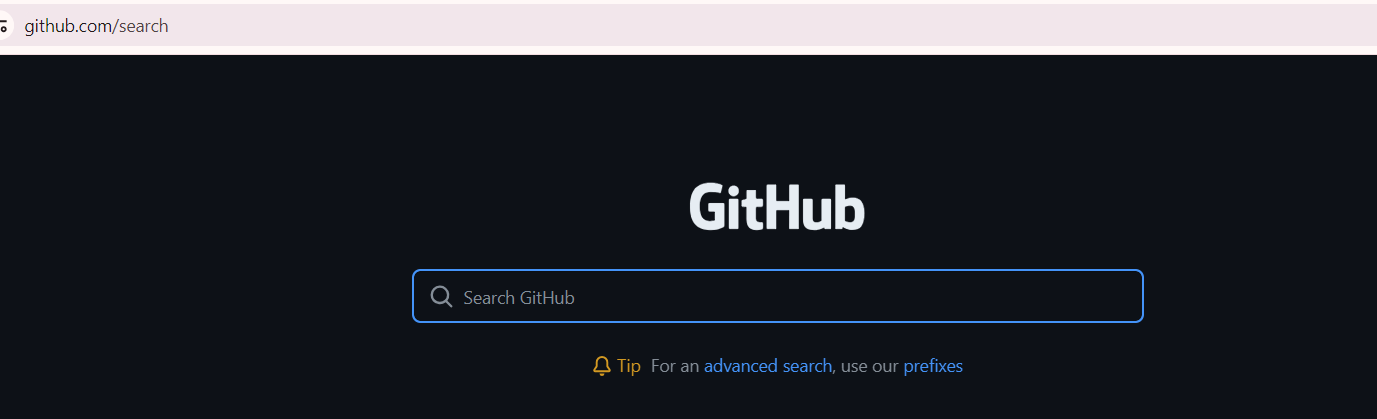
* File excel:



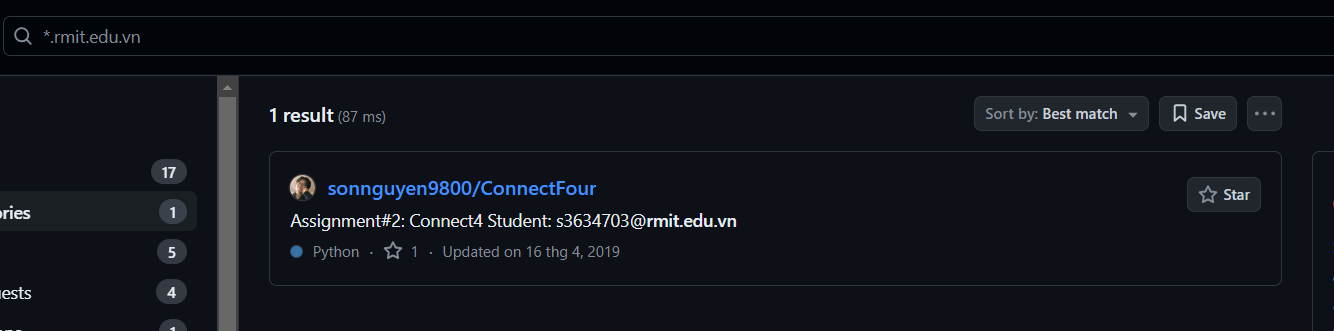
1. **Tìm kiếm thông qua github:**

**Bài tập 7: Ghi nhận một vài thông tin tìm được trên github với domain \*.rmit.edu.vn.**

* Em dùng github search:

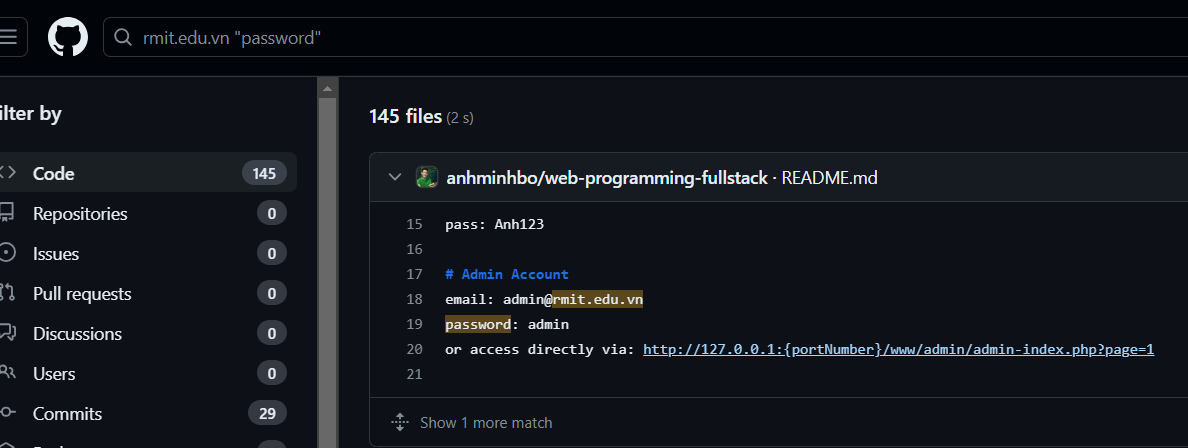


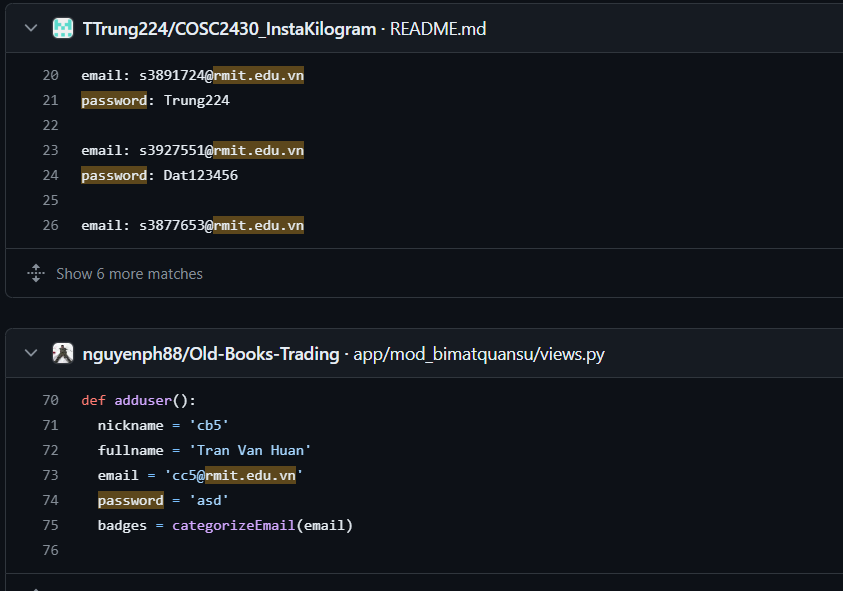
* Theo link bài viết thầy cung cấp, tác giả có đề cập nên sử dụng ký tự đại diện \* để thu được nhiều kết quả hơn, do đó đầu tiên em sẽ tìm kiếm với *\*.rmit.edu.vn*:



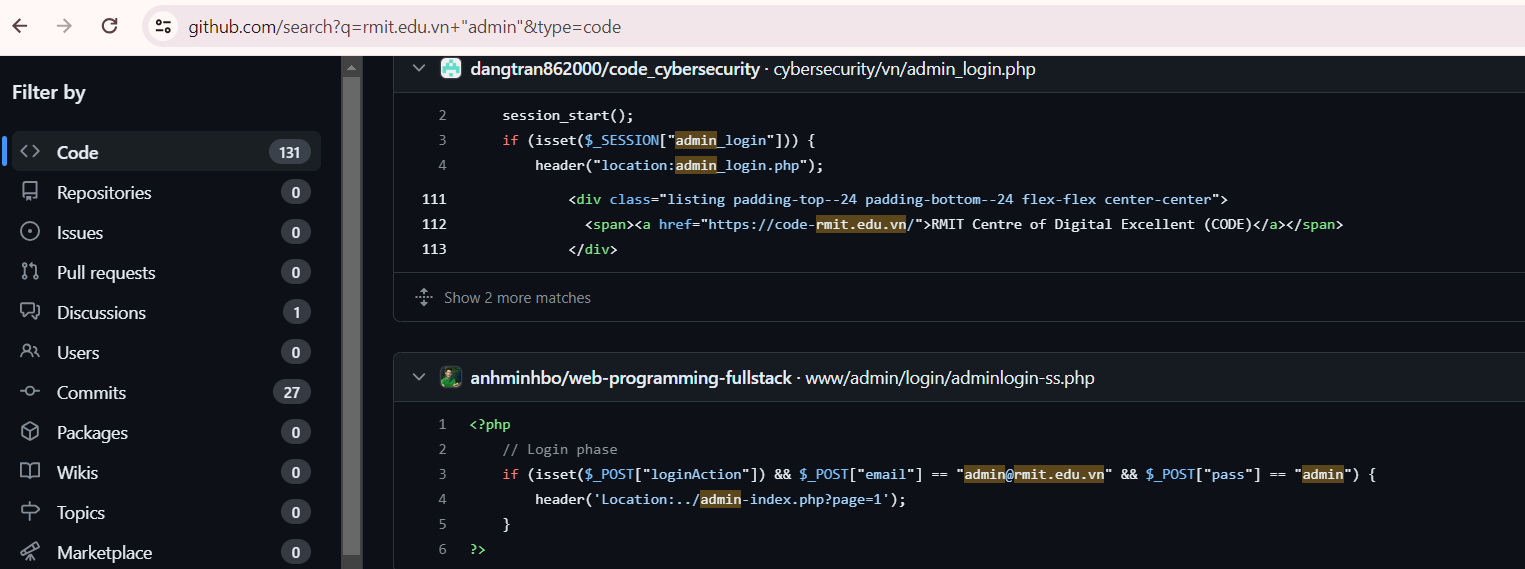
Kết quả ít quá so với mong đợi. Và em chưa tìm được thông tin giá trị.

* Với *rmit.edu.vn “password”*, em tìm được một vài tài khoản như sau:



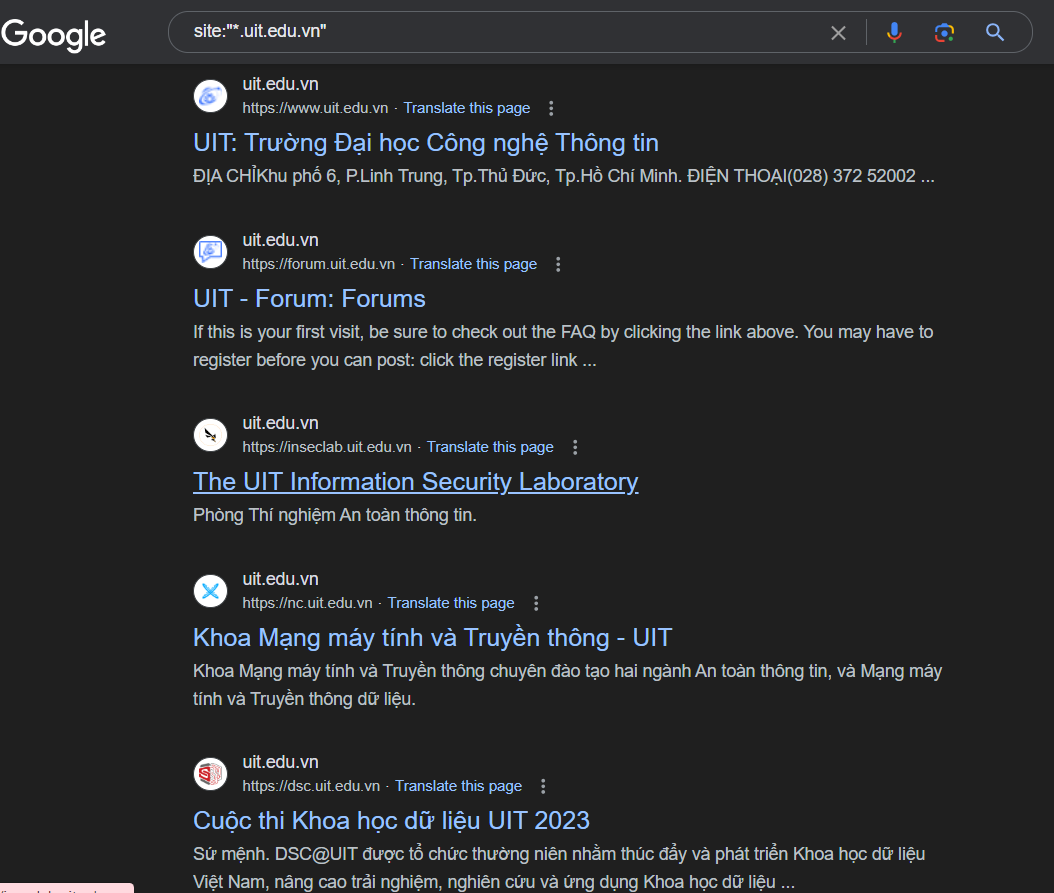


* Tìm kiếm với key “admin” cũng trả về khá nhiều kết quả, dựa vào cách xử lý dữ liệu của chương trình, có thể dựa vào đó để khai thác sâu hơn:



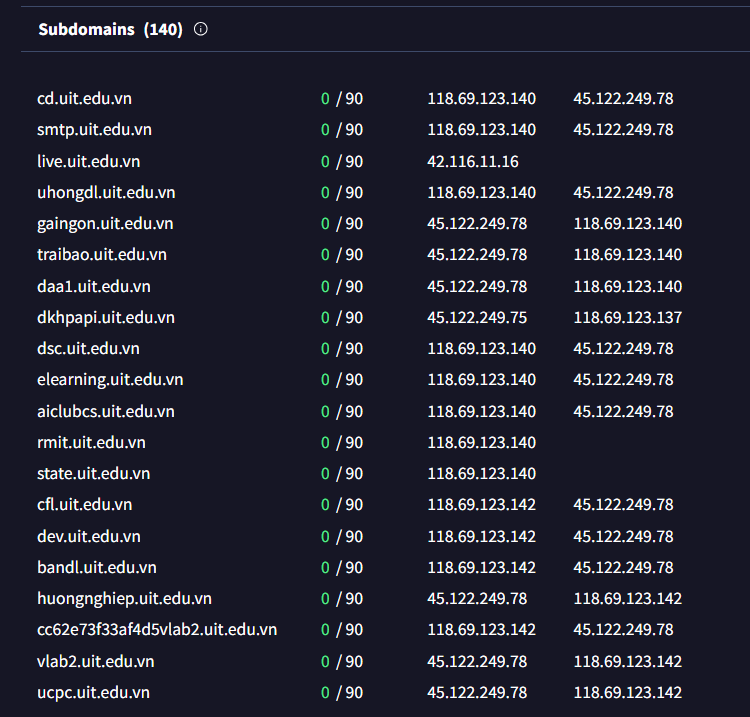
1. **Bài tập thực hành:**

* **Tìm kiếm các tên miền phụ của \*.uit.edu.vn.**
* Tìm kiếm trên google dork được một số kết quả:

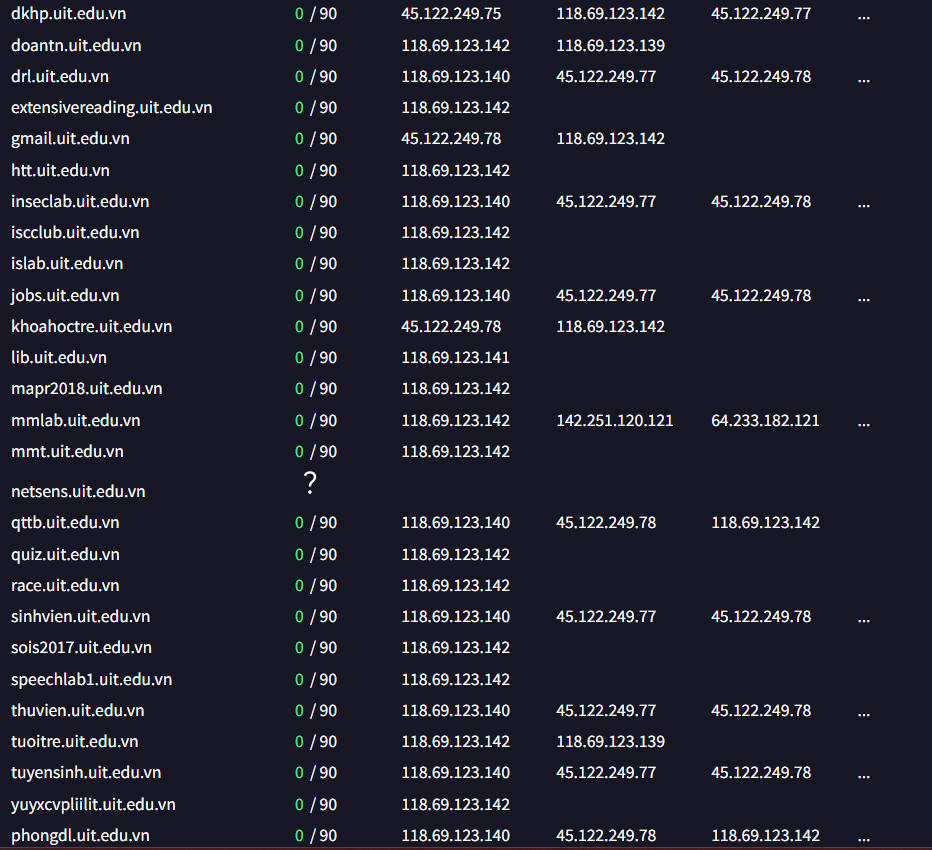
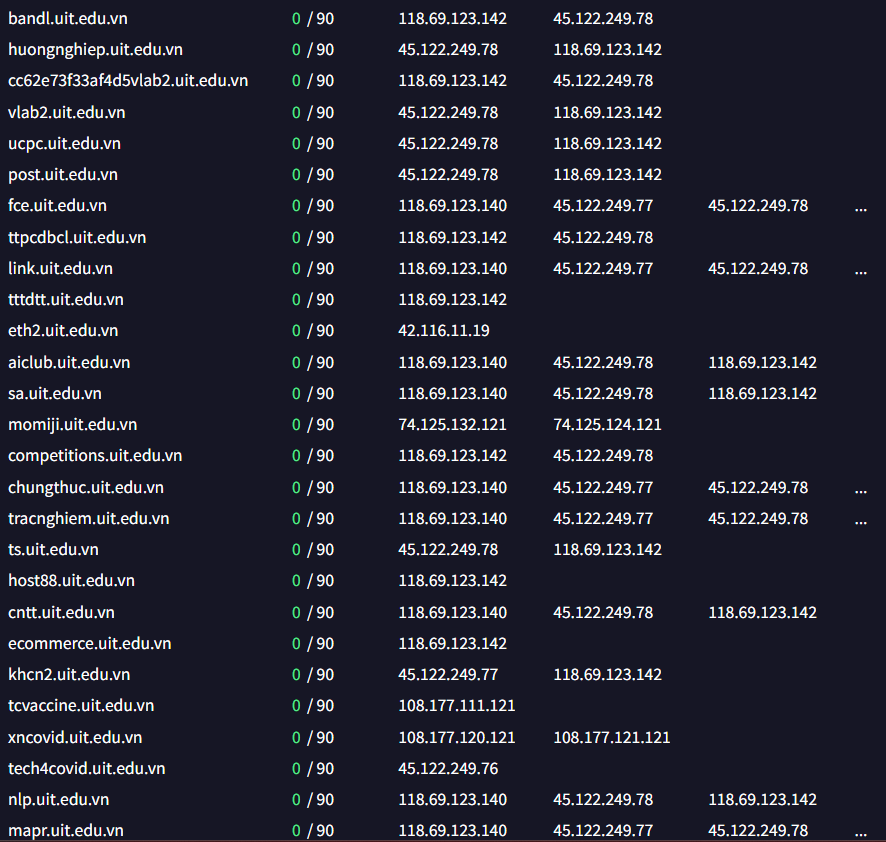


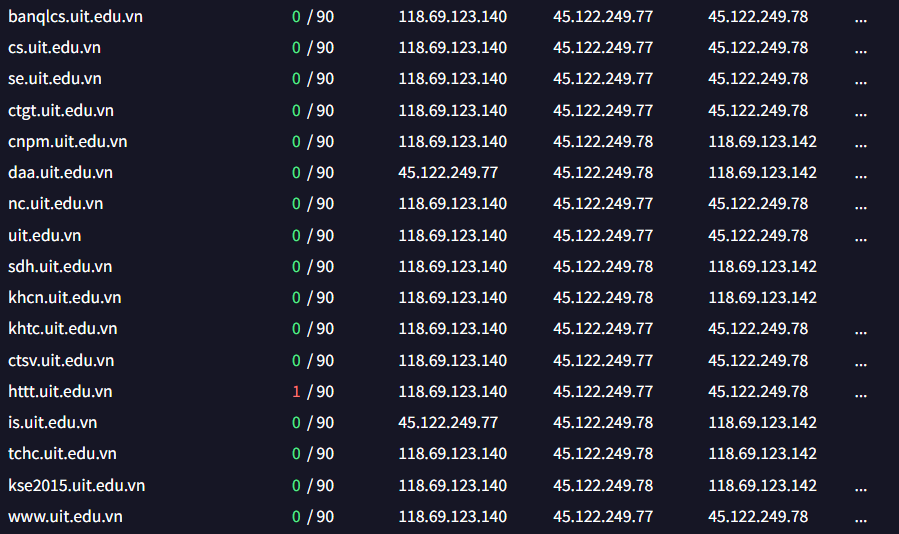
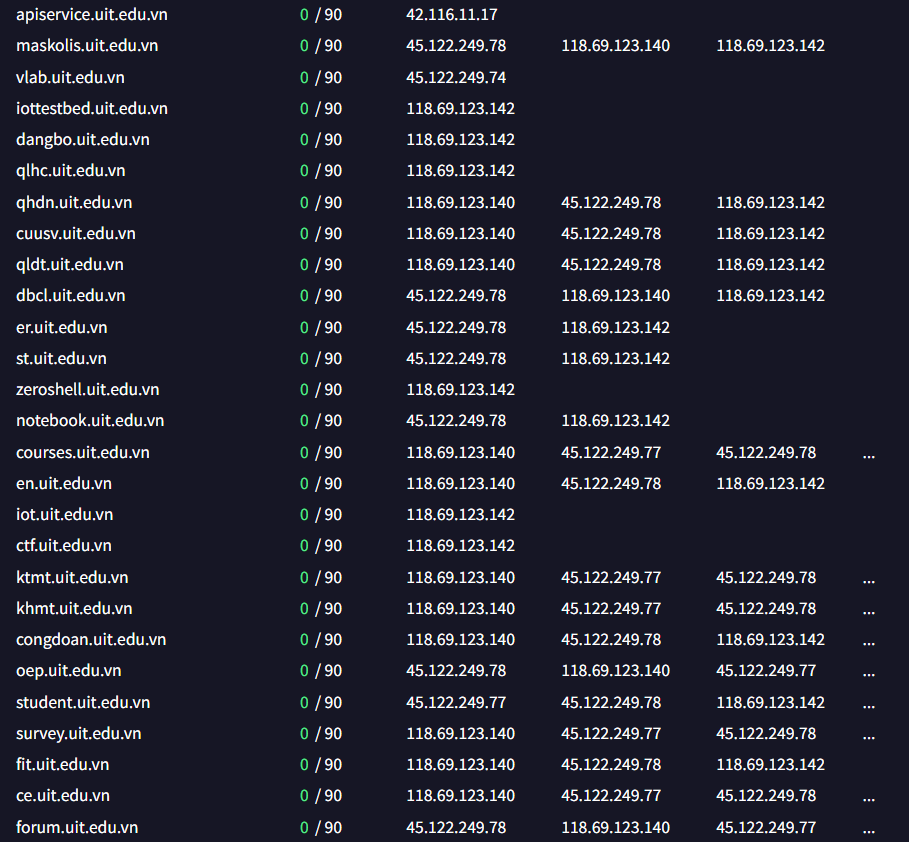
Em tìm được 35 domain. Em lưu vào file *uit\_info.xlsx* để làm tiếp các yêu cầu tiếp theo.

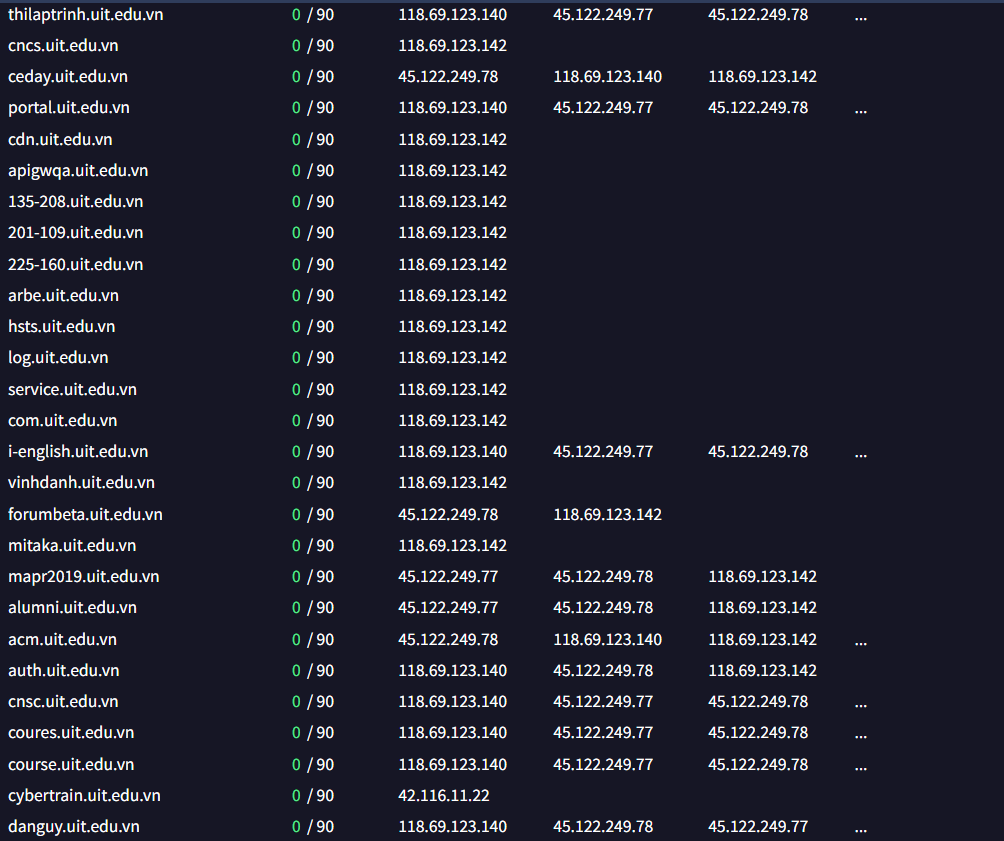
* Ngoài ra, em sử dụng VirusTotal để tìm thì được kết quả nhiều hơn - 140 tên miền phụ:



* **Tìm kiếm các địa chỉ IP thuộc \*.uit.edu.vn và các cổng đang mở tương ứng.**
* Địa chỉ IP tương ứng từng tên miền, em sẽ scan bằng Virus Total:



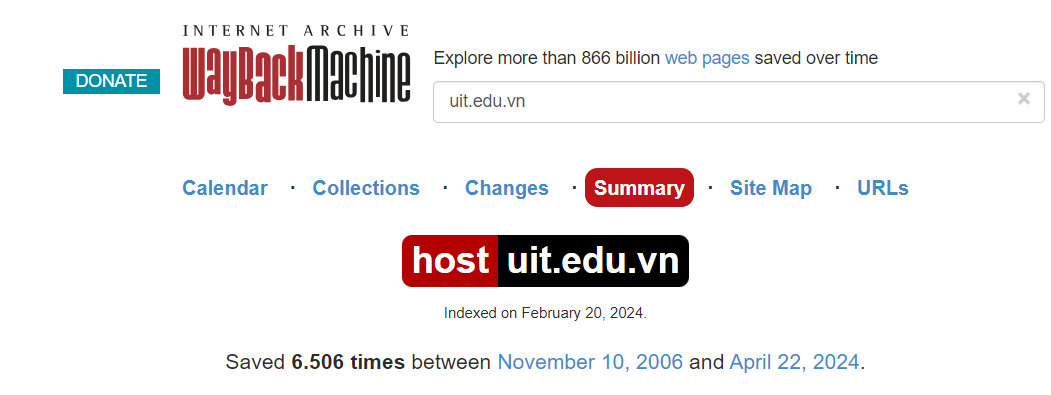




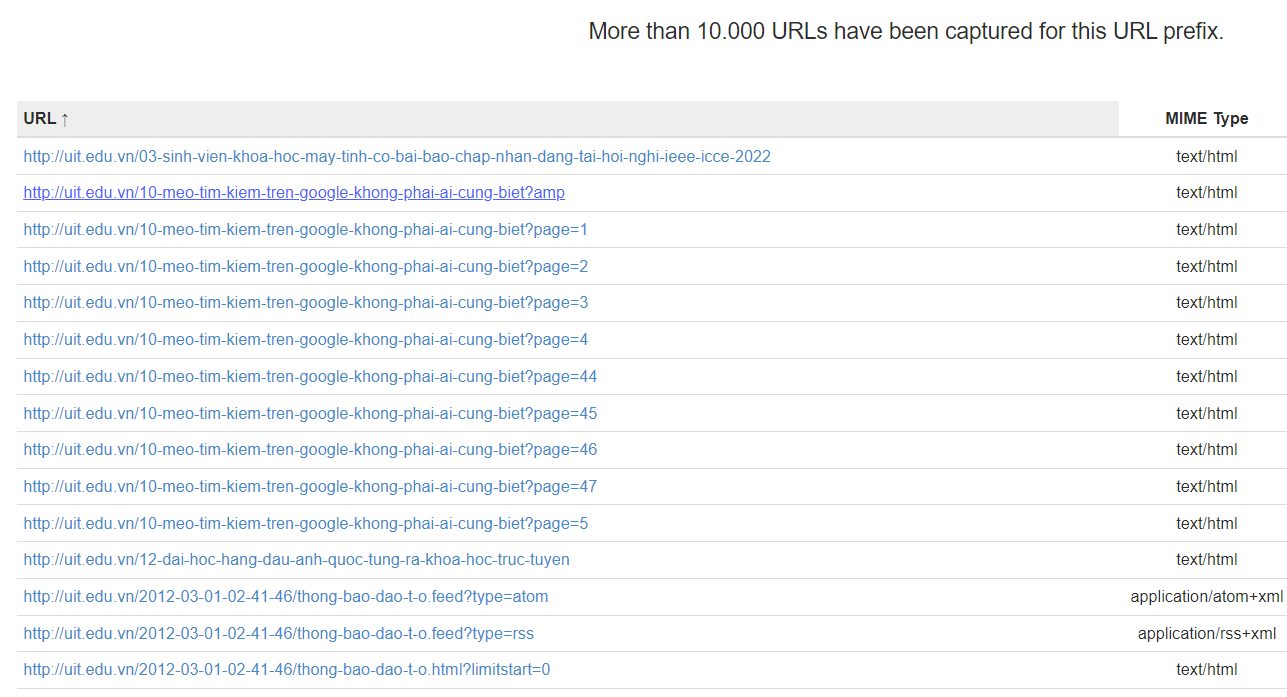
* Các port đang mở tương ứng: dùng nmap để scan, em sẽ scan 34 domain em tìm được trên google dork. Kết quả chi tiết xem tại file *uit\_info.xlsx*.
* **Tìm kiếm các dữ liệu quá khứ của \*.uit.edu.vn:**

Em sử dụng The wayback machine để tìm thông tin.

* Thời gian tìm thấy tên miền đầu tiên là vào năm 2006:

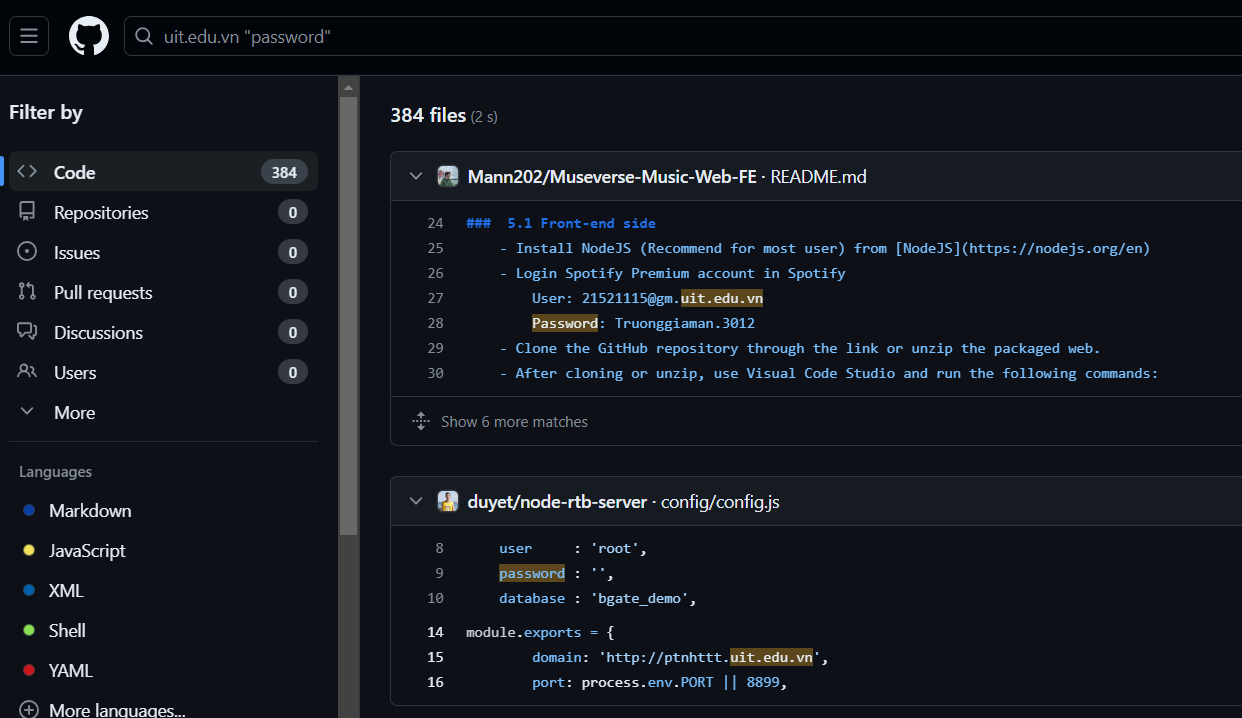


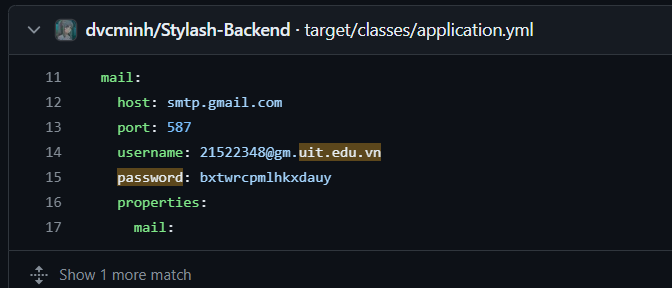
* Một số thông tin về hoạt động của trang web:

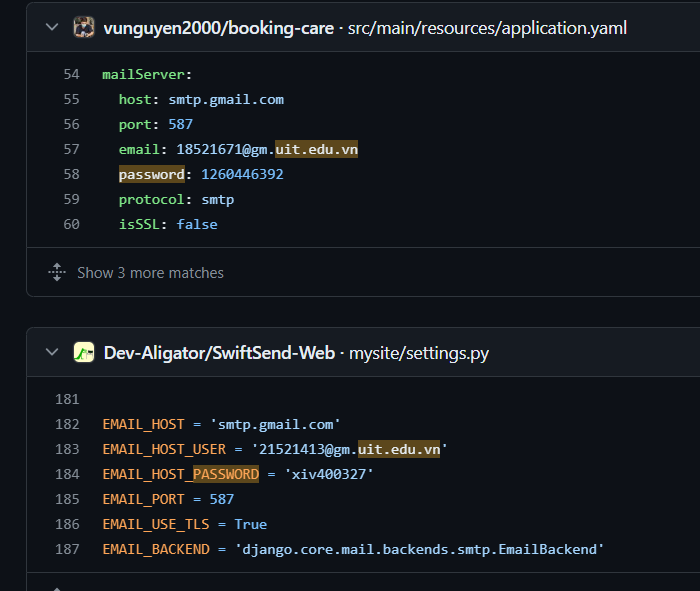


* **Tìm kiếm các dữ liệu nhạy cảm của \*.uit.edu.vn thông qua google dork và github.**

Em tìm được một vài thông tin liên quan đến password. Phần lớn là project của các bạn sinh viên và thông tin admin được đăng kèm:







---

***Sinh viên đọc kỹ yêu cầu trình bày bên dưới trang này***

# **YÊU CẦU CHUNG**

* Sinh viên tìm hiểu và thực hành theo hướng dẫn.
* Nộp báo cáo kết quả chi tiết những việc (**Report**) bạn đã thực hiện, quan sát thấy và kèm ảnh chụp màn hình kết quả (nếu có); giải thích cho quan sát (nếu có).
* Sinh viên báo cáo kết quả thực hiện và nộp bài.

**Báo cáo:**

* File .DOCX và .PDF. Tập trung vào nội dung, không mô tả lý thuyết.
* Nội dung trình bày bằng Font chữ Times New Romans/ hoặc font chữ của mẫu báo cáo này (UTM Neo Sans Intel/UTM Viet Sach)– cỡ chữ 13. Canh đều (Justify) cho văn bản. Canh giữa (Center) cho ảnh chụp.
* Đặt tên theo định dạng: [Mã lớp]-SessionX\_GroupY. (trong đó X là Thứ tự buổi Thực hành, Y là số thứ tự Nhóm Thực hành đã đăng ký với GVHD-TH).

*Ví dụ: [*NT101.K11.ATCL*]-Session1\_Group3.*

* Nếu báo cáo có nhiều file, nén tất cả file vào file .ZIP với cùng tên file báo cáo.
* Không đặt tên đúng định dạng – yêu cầu, sẽ **KHÔNG** chấm điểm bài Lab.
* Nộp file báo cáo trên theo thời gian đã thống nhất tại courses.uit.edu.vn.

**Đánh giá**: Sinh viên hiểu và tự thực hiện được bài thực hành. Khuyến khích:

* Chuẩn bị tốt.
* Có nội dung mở rộng, ứng dụng trong kịch bản phức tạp hơn, có đóng góp xây dựng bài thực hành.

*Bài sao chép, trễ, … sẽ được xử lý tùy mức độ vi phạm.*

**HẾT**

1. Ghi nội dung công việc, các kịch bản trong bài Thực hành [↑](#footnote-ref-1)