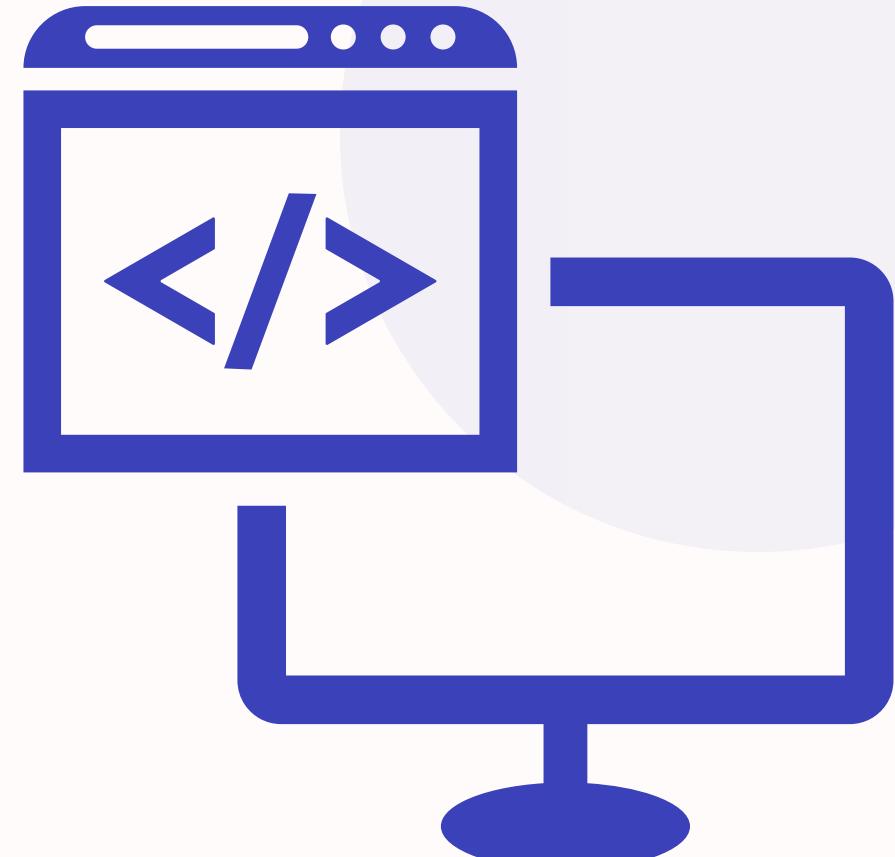


# BÁO CÁO KẾT QUẢ TÍNH TOÁN ĐA PHƯƠNG TIỆN

Lab 1: Dữ liệu dạng Text - Crawler

Giảng viên hướng dẫn: Th.S Đỗ Văn Tiến

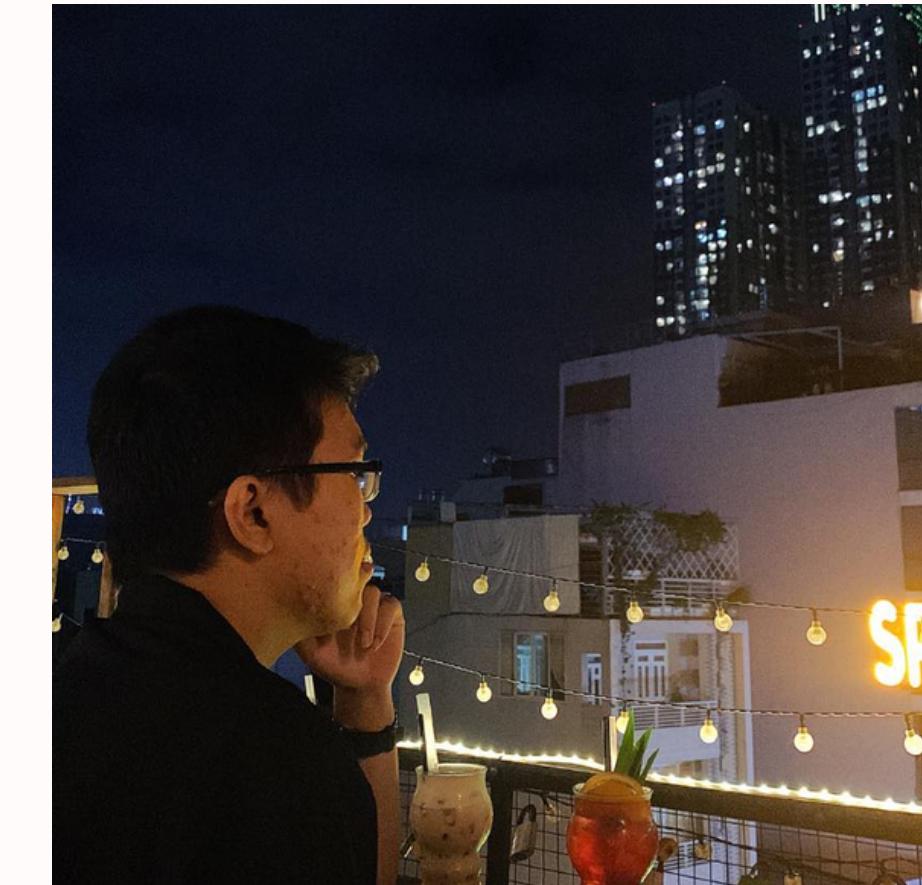


# THÀNH VIÊN



01

Phạm Thiện Bảo  
20521107



02

Lê Nguyễn Tiến Đạt  
20521167

# NỘI DUNG

- 01 VIẾT CRAWLER THU THẬP BÀI BÁO KHOA HỌC CỦA MỘT TÊN TÁC GIẢ BẤT KỲ TRÊN THƯ VIỆN SỐ GOOGLE SCHCOOLAR**
- 02 VIẾT CRAWLER THU THẬP CÁC BÀI VIẾT VÀ BÌNH LUẬN TỪ 1 TRANG FACEBOOK**
- 03 VIẾT CRAWLER THU THẬP CÁC BÀI VIẾT VÀ BÌNH LUẬN TỪ 1 TRANG BÁO ĐIỆN TỬ**
- 04 VIẾT CRAWLER THU THẬP ẢNH LÀ KẾT QUẢ TRẢ VỀ TỪ GOOGLE KHI SỬ DỤNG MỘT TỪ KHÓA**

# THƯ VIỆN SỬ DỤNG



BeautifulSoup



# **CRAWLER THU THẬP BÀI BÁO KHOA HỌC**

# CRAWLER THU THẬP BÀI BÁO KHOA HỌC

## Bước 1: Truy cập đến trang thông tin cá nhân của tác giả

Google Scholar

**Ngan Luu-Thuy Nguyen**

[University of Information Technology](#)  
Email được xác minh tại uit.edu.vn  
[Natural Language Processing](#)

[THEO DÕI](#)

TIÊU ĐỀ	TRÍCH DẪN BỞI	NĂM
<a href="#">Overview of BioNLP'09 shared task on event extraction</a> JD Kim, T Ohta, S Pyysalo, Y Kano, J Tsujii Proceedings of the BioNLP 2009 workshop companion volume for shared task, 1-9	777	2009
<a href="#">The genia event and protein coreference tasks of the BioNLP shared task 2011</a> JD Kim, N Nguyen, Y Wang, J Tsujii, T Takagi, A Yonezawa BMC bioinformatics 13 (11), 1-12	111	2012
<a href="#">UIT-VSFC: Vietnamese students' feedback corpus for sentiment analysis</a> K Van Nguyen, VD Nguyen, PXV Nguyen, TTH Truong, NLT Nguyen 2013 10th international conference on knowledge and systems engineering (KSE ...)	50	2018
<a href="#">Emotion recognition for vietnamese social media text</a> VA Ho, DHC Nguyen, DH Nguyen, LTV Pham, DV Nguyen, KV Nguyen, ... Computational Linguistics: 16th International Conference of the Pacific ...	45	2020
<a href="#">Deep learning versus traditional classifiers on Vietnamese students' feedback corpus</a> PXV Nguyen, TTT Hong, K Van Nguyen, NLT Nguyen 2018 5th NAFOSTED Conference on information and computer science (NICS), 75-80	36	2018
<a href="#">Hate speech detection on vietnamese social media text using the bi-gru-lstm-cnn model</a> T Van Huynh, VD Nguyen, K Van Nguyen, NLT Nguyen, AGT Nguyen arXiv preprint arXiv:1911.03644	35	2019

# CRAWLER THU THẬP BÀI BÁO KHOA HỌC

## Bước 2: Tiến hành mở rộng trang để lấy tất cả class chứa thông tin bài báo

Publication Title	Year	Page Count
machine reading comprehension K Van Nguyen, KV Tran, ST Luu, AGT Nguyen, NLT Nguyen IEEE Access 8, 201404-201417	2008	22
Exploring domain differences for the design of a pronoun resolution system for biomedical text NLT Nguyen, JD Kim Proceedings of the 22nd International Conference on Computational ...	2016	20
Effective attention-based neural architectures for sentence compression with bidirectional long short-term memory NT Tran, VT Luong, NLT Nguyen, MQ Nghiem Proceedings of the 7th Symposium on Information and Communication Technology ...	2012	20
Improving protein coreference resolution by simple semantic classification N Nguyen, JD Kim, M Miwa, T Matsuzaki, J Tsujii BMC bioinformatics 13 (1), 1-12	2019	19
Hate speech detection on vietnamese social media text using the bidirectional-lstm model HTT Do, HD Huynh, K Van Nguyen, NLT Nguyen, AGT Nguyen arXiv preprint arXiv:1911.03648	2011	19
Overview of the protein coreference task in BioNLP shared task 2011 N Nguyen, JD Kim, J Tsujii Proceedings of the BioNLP shared task 2011 workshop, 74-82	2022	18
New Vietnamese corpus for machine reading comprehension of health news articles K Van Nguyen, T Van Huynh, DV Nguyen, AGT Nguyen, NLT Nguyen Transactions on Asian and Low-Resource Language Information Processing 21 (5 ...	2021	18
A large-scale dataset for hate speech detection on vietnamese social media texts ST Luu, KV Nguyen, NLT Nguyen Advances and Trends in Artificial Intelligence. Artificial Intelligence ...	2008	22

# CRAWLER THU THẬP BÀI BÁO KHOA HỌC

Thiết lập cho quá trình mở rộng trang diễn ra tự động

```
--> <div id="gsc_bpf">
    ><button type="button" id="gsc_bpf_more" class="gs_btnPD gs_in_ib gs_btn_flat gs_btn_lrg gs_btn_lsu">...</button>
</div>
</div>
```

```
while browser.find_element(By.ID, 'gsc_bpf_more').is_enabled():
    browser.find_element(By.ID, 'gsc_bpf_more').click()
    sleep(2)
```

# CRAWLER THU THẬP BÀI BÁO KHOA HỌC

## Bước 3: Tìm tất cả các class chứa thông tin bài báo và lấy các thông tin cần thiết của từng bài báo

```
#Tìm tất cả các class chứa thông tin bài báo
all_papers = browser.find_elements(By.CLASS_NAME, "gsc_a_tr")

title,author,year,linkdetail,cited = [],[],[],[],[]

for paper in all_papers:
    title.append(paper.find_element(By.CSS_SELECTOR,"td.gsc_a_t > a").text)
    author.append(paper.find_element(By.CLASS_NAME,"gs_gray").text)
    year.append(paper.find_element(By.CLASS_NAME,"gsc_a_h").text)
    cited.append(paper.find_element(By.CLASS_NAME,"gsc_a_ac").text)
    linkdetail.append(paper.find_element(By.CSS_SELECTOR,"td.gsc_a_t > a").get_attribute("href"))
    sleep(random.randint(1,3))
```

# CRAWLER THU THẬP BÀI BÁO KHOA HỌC

## Bước 4: Lưu tất cả các thông tin từng bài báo vào file csv kết quả

Ttitle	Authors	Year	Cited	Link_detail
0 Overview of BioNL	JD Kim, T Ohta, S Pyysalo, Y Kanc	2009	776	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:KlAtU1dfN6UC">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:KlAtU1dfN6UC</a>
1 The genia event an	JD Kim, N Nguyen, Y Wang, J Tsuj	2012	110	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:UebtZRa9Y70C">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:UebtZRa9Y70C</a>
2 UIT-VSFC: Vietnam	K Van Nguyen, VD Nguyen, PXV N	2018	50	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:qxL8FJ1GzNcC">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:qxL8FJ1GzNcC</a>
3 Emotion recognitic	VA Ho, DHC Nguyen, DH Nguyen,	2020	45	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:hC7cP41nSMkC">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:hC7cP41nSMkC</a>
4 Deep learning vers	PXV Nguyen, TTT Hong, K Van Ng	2018	36	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:4TOpqqG69KYC">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:4TOpqqG69KYC</a>
5 Hate speech detect	T Van Huynh, VD Nguyen, K Van I	2019	35	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:L8Ckcad2t8MC">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:L8Ckcad2t8MC</a>
6 A Vietnamese data	K Nguyen, V Nguyen, A Nguyen, N	2020	31	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:GnP-g6toBAC">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:GnP-g6toBAC</a>
7 Modeling in the tin	N Nguyen, O Strnad, T Klein, D Lu	2021	30	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:D03iK_w7-QYC">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:D03iK_w7-QYC</a>
8 Job prediction: Fro	T Van Huynh, K Van Nguyen, NLT	2020	25	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:mB3voiENLucC">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:mB3voiENLucC</a>
9 Variants of long sh	VD Nguyen, K Van Nguyen, NLT N	2018	25	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:Wp0glr-vW9MC">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:Wp0glr-vW9MC</a>
10 Proceedings of the	JD Kim, S Pyysalo, T Ohta, R Boss	2011	24	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:hqOjcs7Dif8C">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:hqOjcs7Dif8C</a>
11 Filling the gaps bet	Y Kano, N Nguyen, R Saetre, K Yc	2008	23	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:W7OEmFMy1HYC">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:W7OEmFMy1HYC</a>
12 Enhancing lexical-k	K Van Nguyen, KV Tran, ST Luu, A	2020	22	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:YFjsv_pBGBYC">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:YFjsv_pBGBYC</a>
13 Exploring domain c	NLT Nguyen, JD Kim	2008	22	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:roLk4NBRz8UC">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:roLk4NBRz8UC</a>
14 Effective attention	NT Tran, VT Luong, NLT Nguyen,	2016	20	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:IjCSPb-OGe4C">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:IjCSPb-OGe4C</a>
15 Improving protein	N Nguyen, JD Kim, M Miwa, T Ma	2012	20	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:eQOLeE2rZwMC">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:eQOLeE2rZwMC</a>
16 Hate speech detect	HTT Do, HD Huynh, K Van Nguyen	2019	19	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:IWHjjKOFINEC">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:IWHjjKOFINEC</a>
17 Overview of the pr	N Nguyen, JD Kim, J Tsujii	2011	19	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:UeHWp8X0CEIC">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:UeHWp8X0CEIC</a>
18 New Vietnamese c	K Van Nguyen, T Van Huynh, DV I	2022	18	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:r0BpntZqJG4C">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:r0BpntZqJG4C</a>
19 A large-scale datas	ST Luu, KV Nguyen, NLT Nguyen	2021	18	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:70eg2SAElzsC">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;citation_for_view=VoTCbXYAAAAJ:70eg2SAElzsC</a>
20 Overview of BioNL	N Nguyen, JD Kim, J Tsujii	2011	18	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;cstart=20&amp;pagesize=80&amp;citation_for_view=VoTCbXY">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;cstart=20&amp;pagesize=80&amp;citation_for_view=VoTCbXY</a>
21 Constructive and t	LT Nguyen, K Van Nguyen, NLT N	2021	15	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;cstart=20&amp;pagesize=80&amp;citation_for_view=VoTCbXY">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;cstart=20&amp;pagesize=80&amp;citation_for_view=VoTCbXY</a>
22 A transformation r	T Van Dang, VD Nguyen, N Van K	2018	14	<a href="https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;cstart=20&amp;pagesize=80&amp;citation_for_view=VoTCbXY">https://scholar.google.com/citations?view_op=view_citation&amp;hl=en&amp;user=VoTCbXYAAAAJ&amp;cstart=20&amp;pagesize=80&amp;citation_for_view=VoTCbXY</a>

# CRAWLER THU THẬP BÌNH LUẬN TỪ TRANG FACEBOOK

# CRAWLER THU THẬP BÌNH LUẬN TỪ TRANG FACEBOOK

Trang chủ Chính sửa trang cá nhân Tin nhắn Thông báo Chat Tìm bạn bè Trang Nhóm Menu

! Trang này đã được cập nhật thành trải nghiệm Trang mới

Bạn cần dùng máy tính hoặc ứng dụng di động để tiếp tục quản lý Trang này. [Learn More](#)



KTX ĐHQG Confessions

Fanpage confessions

Ký túc xá Đại học Quốc gia TP.HCM

- Link gửi confession:  
[bit.ly/ktxdhqgcf](http://bit.ly/ktxdhqgcf)

Nhắn tin Theo dõi Xem thêm

[Giới thiệu](#) · [Ảnh](#) · [Đang theo dõi](#)

KTX ĐHQG Confessions

#KTX\_Cfs3100  
#BỨC\_XÚC

Chào mọi người, chuyện là phòng mình (3xx, tòa F2) hiện tại chỉ có 2 người ở là mình và 1 thằng nhóc năm nhất. Nó tên PDT, học N\*\*\* B\*\* H\*\* ở N\*\*\* V\*\*.

Không biết mọi người có gặp trường hợp này bao giờ chưa, chứ mình cảm giác bức rúc, khó chịu thật sự. Mình thấy thằng này đúng kiểu không biết điều thật sự. Sáng nào thằng đấy thức dậy là y như rằng mình cũng sẽ thức. Nó cứ như người khổng lồ vậy, làm gì cũng tạo ra tiếng rõ to. Nước uống của phòng thì trước giờ mình...

[Xem thêm](#)

1 giờ · Công khai

33 · Thích · [Bày tỏ cảm xúc](#) · 5 bình luận · Chia sẻ · Toàn bộ tin · Lưu · [Tim hỗ trợ hoặc báo cáo bài viết](#)

## Sử dụng phiên bản mbasic của Facebook:

- + Là phiên bản giao diện đơn giản và gọn nhẹ, hiển thị ít dữ liệu hơn so với phiên bản đầy đủ
- + Không có các tính năng nâng cao như ứng dụng, trò chơi, quảng cáo
- + Trình bày thông tin dễ dàng để xử lý và phân tích

# CRAWLER THU THẬP BÌNH LUẬN TỪ TRANG FACEBOOK

Bước 1: Đăng nhập và truy cập đến trang 'KTX ĐHQG Confessions'

Bước 2: Lấy các đường link truy cập riêng đến từng post trong trang đầu tiên

KTX ĐHQG Confessions

#KTX\_Cfs3100

#BỨC\_XÚC

---  
Chào mọi người, chuyện là phòng mình (3xx, tòa F2) hiện tại chỉ có 2 người ở là mình và 1 thằng nhóc năm nhất. Nó tên PDT, học N\*\*\* B\*\* H\*\* ở N\*\*\* V\*\*.

Không biết mọi người có gặp trường hợp này bao giờ chưa, chứ mình cảm giác bức rúc, khó chịu thật sự. Mình thấy thằng này đúng kiểu không biết điều thật sự. Sáng nào thằng đầy thức dậy là y như rằng mình cũng sẽ thức. Nó cứ như người khổng lồ vậy, làm gì cũng tạo ra tiếng rõ to. Nước uống của phòng thi trước giờ mình...

Xem thêm

1 giờ · Công khai

   33 · Thích · Bày tỏ cảm xúc · 5 bình luận · Chia sẻ · [Toàn bộ tin](#) · Lưu · Tim hỗ trợ hoặc báo cáo bài viết

KTX ĐHQG Confessions

#PHẢN\_ÁNH

---  
#1 4/6/2023 10:38:48

Gửi đến ban quản lý cụm nhà B.A, mong BQL có thể thay máy quét thẻ khác cho tòa BA4. Trong khi các tòa khác chỉ cần cầm thẻ đi ngang qua thì máy đã đọc được thẻ, còn ở đây phải áp sát thẻ vào mới được, rất chi là công kẽm, có khi áp vào mạnh quá cái máy nó bung ra luôn ạ. Em cảm ơn mọi người đã đọc.

---  
#2 4/9/2023 21:48:04

Hẹn chế cho người ngoài phòng mượn đồ nha mọi người, một là họ xu luôn hai là họ quên mình không biết đường đâu mà tìm ấy. 😢😢😢

---  
Xem thêm

5 giờ · Công khai

 6 · Thích · Bày tỏ cảm xúc · 3 bình luận · Chia sẻ · [Toàn bộ tin](#) · Lưu · Tim hỗ trợ hoặc báo cáo bài viết

# CRAWLER THU THẬP BÌNH LUẬN TỪ TRANG FACEBOOK

## Bước 3: Lần lượt duyệt từng bài post để lấy nội dung và bình luận

+ Thực hiện lấy nội dung của bài post

```
elems = browser.find_elements(By.CSS_SELECTOR,"div > p") #Content trong post
content = []
if (elems):
    for elem in elems:
        content.append(elem.text)
else:
    bgs = browser.find_element(By.CSS_SELECTOR,"div > div > div > div > div > div > span") #Content trong background
    content.append(bgs.text)
```

# CRAWLER THU THẬP BÌNH LUẬN TỪ TRANG FACEBOOK

## Bước 3: Lần lượt duyệt từng bài post để lấy nội dung và bình luận

+ Thực hiện lấy bình luận của bài post:

Thông tin bình luận có dạng (người bình luận, nội dung bình luận)

Danh Mai Khang An

#1. Mọi ý kiến đều được xem xét và chấp nhận, nhưng tên quốc gia không viết hoa thì lời lẽ hay như thế nào cũng bị loại trong mắt người đọc!

1 like · Bày tỏ cảm xúc · Phản hồi · Xem thêm · 6 giờ trước

Thanh Nguyen

#1. Cá nhân mình thấy nếu tụi mình sống lạc quan vui vẻ, có cống hiến, hạnh phúc thì đã là một hình thức đấu tranh đúng đắn rồi nhé bạn.

2 like · Bày tỏ cảm xúc · Phản hồi · Xem thêm · 6 giờ trước

Ddawng Nguyen

Lê Bảo đọc cmt cbi kể tội nè

Thích · Bày tỏ cảm xúc · Phản hồi · Xem thêm · 1 giờ trước

Linh Lê

#1 thảo luận cái đặc cầu, viết hoa tên quốc gia đi rồi nói tiếp

1 like · Bày tỏ cảm xúc · Phản hồi · Xem thêm · 6 giờ trước

# CRAWLER THU THẬP BÌNH LUẬN TỪ TRANG FACEBOOK

## Bước 3: Lần lượt duyệt từng bài post để lấy nội dung và bình luận

+ Thực hiện lấy bình luận của bài post:

Thông tin bình luận có dạng (người bình luận, nội dung bình luận)

```
list_cmt = []
name = browser.find_elements(By.CSS_SELECTOR, "div > h3 > a")
#cmt = browser.find_elements(By.CSS_SELECTOR, "div > div.dy")
#cmt = browser.find_elements(By.CSS_SELECTOR, "div > div.dv")
#cmt = browser.find_elements(By.CSS_SELECTOR, "div > div.dz")
if (len(name)):
    pick_id = browser.find_elements(By.XPATH, '//a[contains(@href, "comment/replies")]') #tìm tất cả cmt
    cmt = []
    for id in pick_id:
        takeid= id.get_attribute('href').split('ctoken=')[1].split('&')[0] #lấy id của từng cmt
        text_cmt = browser.find_element(By.XPATH, ('//*[@id=' + takeid.split('_')[1] + ']/div/div[1]')) #lấy phần bình luận
        cmt.append(text_cmt.text)
```

# CRAWLER THU THẬP BÌNH LUẬN TỪ TRANG FACEBOOK

## Bước 4: Lưu kết quả từng bài post vào file txt

--Crawl Post\_Link : [https://mbasic.facebook.com/story.php?story\\_fbid=pfbid02jnPTTz3aoF5BsKEbQEziTwy1pgVKoV5dhFmLRGyijcMRsZLDMuqsrp9w](https://mbasic.facebook.com/story.php?story_fbid=pfbid02jnPTTz3aoF5BsKEbQEziTwy1pgVKoV5dhFmLRGyijcMRsZLDMuqsrp9w)

Content-----

• TỪ THỦ ĐỨC LÊN SÀI GÒN MƯA SML, MƯA KHỦNG, MƯA TRÔI SẠCH KIẾN THỨC BÀI VỞ ĐÃ HỌC ĐỂ THI GIỮA KỲ...

Comment-----

Thành Võ : Tối thứ 7 rảnh rồi đi chơi. Ng tính ko bằng trời độ

Đạt Lyn : Dc có cái t7 hoi mà trời 😊

Phát Đào : Ủa chứ Thủ Đức không phải Sài Gòn ngộ he

Nguyễn Phương Trinh : Ủa chứ Thủ Đức không phải Sài Gòn ngộ he

Thanh Thảo : Nguyễn Nga quên nǎo để chạy deadline kkk

Võ Đình Ngọc Na : Ute that thu

Minh Lê : Nguyễn Như mưa to bự như mình phán hồi chiều 😭

Uyên Hoàng : Trên thiên đìn mở Fest hay j mà anh Lôi đánh đèn ghê quá 🐧

# **CRAWLER THU THẬP BÌNH LUẬN TỪ TRANG BÁO ĐIỆN TỬ**

# CRAWLER THU THẬP BÌNH LUẬN TRANG BÁO ĐIỆN TỬ

Gọi ngay: 19001100 | Đặt vé ngay: vietnamairlines.com

VNEXPRESS | Chủ nhật, 16/4/2023 | Mới nhất | Tin theo khu vực | International

Góc nhìn | Thế giới | Video | Podcasts | Kinh doanh | Khoa học | Giải trí | Thể thao | Pháp luật | Giáo dục | Sức khỏe

**Thủ tướng: TP HCM khó khăn sẽ ảnh hưởng cả nước**

Lãnh đạo Chính phủ đánh giá "đầu tàu" kinh tế TP HCM đang gặp nhiều khó khăn khi chưa phục hồi hoàn toàn sau dịch nhưng đổi mới nhiều thách thức mới.

"TP HCM là đầu tàu kinh tế của đất nước cả về hữu ...

**Hàn Quốc bắn 10 phát cảnh cáo tàu Triều Tiên vượt giới tuyến**

Quân đội Hàn Quốc cho biết đã bắn cảnh cáo 10 phát để ứng phó một tàu tuần tra Triều Tiên vượt qua giới tuyến trên biển giữa hai nước.

**Ngoại trưởng Mỹ khen ngợi ẩm thực Việt Nam**

Ngoại trưởng Mỹ Blinken cho rằng ẩm thực Việt Nam có những hương vị tươi ngon nhất và ông đã được nếm thử trong chuyến thăm ngày 14-16/4.

**Góc nhìn**  
**Công chức tận hiến**  
Gắn nỗ lực phụng sự người dân với thành công nghề nghiệp của mình là chìa khóa cho một bộ máy công chức hiệu quả.

Cẩm Hà

37



Sử dụng ChromeDriver để truy cập vào URL trang báo VnExpress

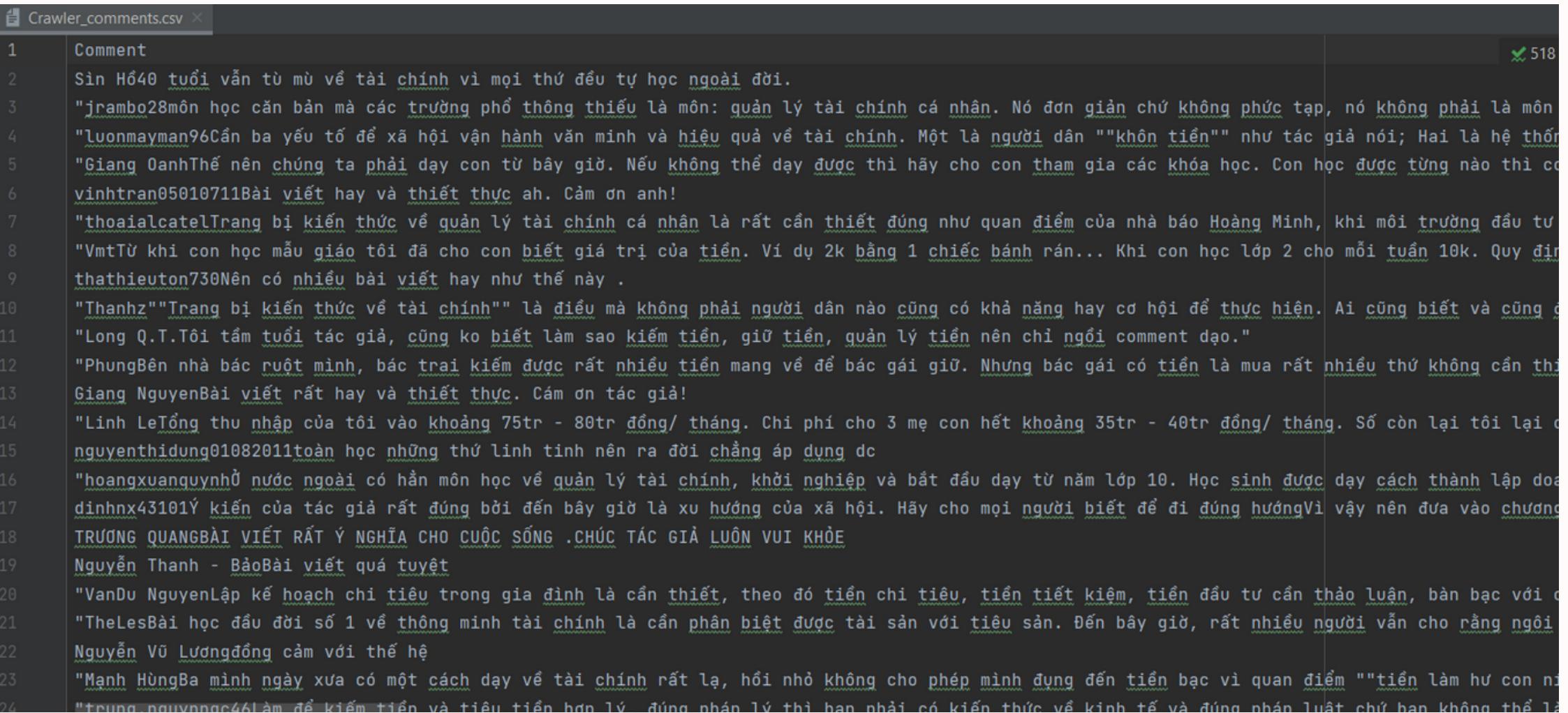


Lấy URL của từng bài báo rồi truy cập vào từng bài

# CRAWLER THU THẬP BÌNH LUẬN TRANG BÁO ĐIỆN TỬ

O3

Truy cập vào từng URL của các bài báo và crawl comment của các bài báo đó



The screenshot shows a CSV file named 'Crawler\_comments.csv' with 24 rows of data. The first column contains user IDs and the second column contains their comments. The comments are in Vietnamese and discuss financial topics like tuition fees, scholarships, and student life.

Comment
Sin Hô40 tuổi vẫn tù mù về tài chính vì mọi thứ đều tự học ngoài đời.
"jrambo28môn học căn bản mà các trường phổ thông thiếu là môn: quản lý tài chính cá nhân. Nó đơn giản chứ không phức tạp, nó không phải là môn "luonmayman96 Cần ba yếu tố để xã hội vận hành văn minh và hiệu quả về tài chính. Một là người dân ""khôn tiền"" như tác giả nói; Hai là hệ thống Giang Danh Thế nên chúng ta phải dạy con từ bây giờ. Nếu không thể dạy được thì hãy cho con tham gia các khóa học. Con học được từng nào thì có vinhtran05010711 Bài viết hay và thiết thực ah. Cảm ơn anh!
"thoaijalcatel Trang bị kiến thức về quản lý tài chính cá nhân là rất cần thiết đúng như quan điểm của nhà báo Hoàng Minh, khi môi trường đầu tư Vmt Từ khi con học mẫu giáo tôi đã cho con biết giá trị của tiền. Ví dụ 2k bằng 1 chiếc bánh rán... Khi con học lớp 2 cho mỗi tuần 10k. Quy định thathieuton730 Nên có nhiều bài viết hay như thế này .
"Thanhz""Trang bị kiến thức về tài chính"" là điều mà không phải người dân nào cũng có khả năng hay cơ hội để thực hiện. Ai cũng biết và cũng cần Long Q.T.Tôi tầm tuổi tác giả, cũng ko biết làm sao kiếm tiền, giữ tiền, quản lý tiền nên chỉ ngồi comment dạo."
"PhungBên nhà bác ruột mình, bác trai kiếm được rất nhiều tiền mang về để bác gái giữ. Nhưng bác gái có tiền là mua rất nhiều thứ không cần thiết Giang NguyenBài viết rất hay và thiết thực. Cám ơn tác giả!
"Linh LeTổng thu nhập của tôi vào khoảng 75tr - 80tr đồng/ tháng. Chi phí cho 3 mẹ con hết khoảng 35tr - 40tr đồng/ tháng. Số còn lại tôi lại c nguyenthidung01082011 toàn học những thứ linh tinh nên ra đời chẳng áp dụng dc
"hoangxuanquynhỞ nước ngoài có hẳn môn học về quản lý tài chính, khởi nghiệp và bắt đầu dạy từ năm lớp 10. Học sinh được dạy cách thành lập doanhnhx43101 Ý kiến của tác giả rất đúng bởi đến bây giờ là xu hướng của xã hội. Hãy cho mọi người biết để đi đúng hướng Vì vậy nên đưa vào chương TRƯỜNG QUANG BÀI VIẾT RẤT Ý NGHĨA CHO CUỘC SỐNG . CHÚC TÁC GIẢ LUÔN VUI KHỎE
Nguyễn Thanh - BảoBài viết quá tuyệt
"VanDu Nguyen Lập kế hoạch chi tiêu trong gia đình là cần thiết, theo đó tiền chi tiêu, tiền tiết kiệm, tiền đầu tư cần thảo luận, bàn bạc với
"TheLesBài học đầu đời số 1 về thông minh tài chính là cần phân biệt được tài sản với tiêu sản. Đến bây giờ, rất nhiều người vẫn cho rằng ngôi Nguyễn Vũ Lươngđồng cảm với thế hệ
"Mạnh HùngBa mình ngày xưa có một cách dạy về tài chính rất lạ, hồi nhỏ không cho phép mình dùng đến tiền bạc vì quan điểm ""tiền làm hư con người""
"trung_nguyenduc441 làm để kiếm tiền và tiêu tiền hơn lý do phản lú thì bạn phải có kiến thức về kinh tế và đúng pháp luật chứ ban không thể làm

O4

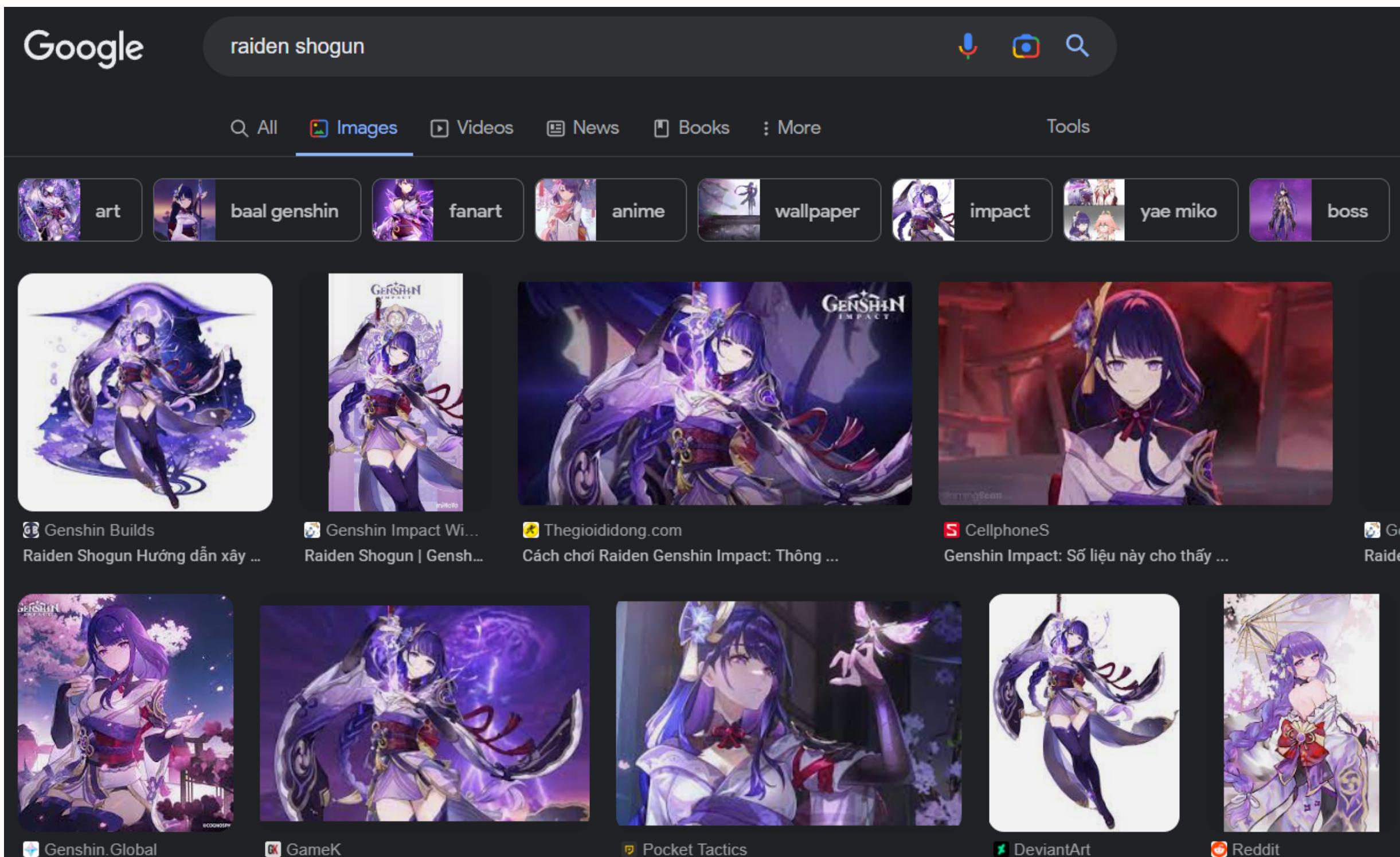
Comment của các bài báo sẽ được lưu lại vào file Crawler\_comments.csv

# CRAWLER THU THẬP ẢNH TỪ GOOGLE



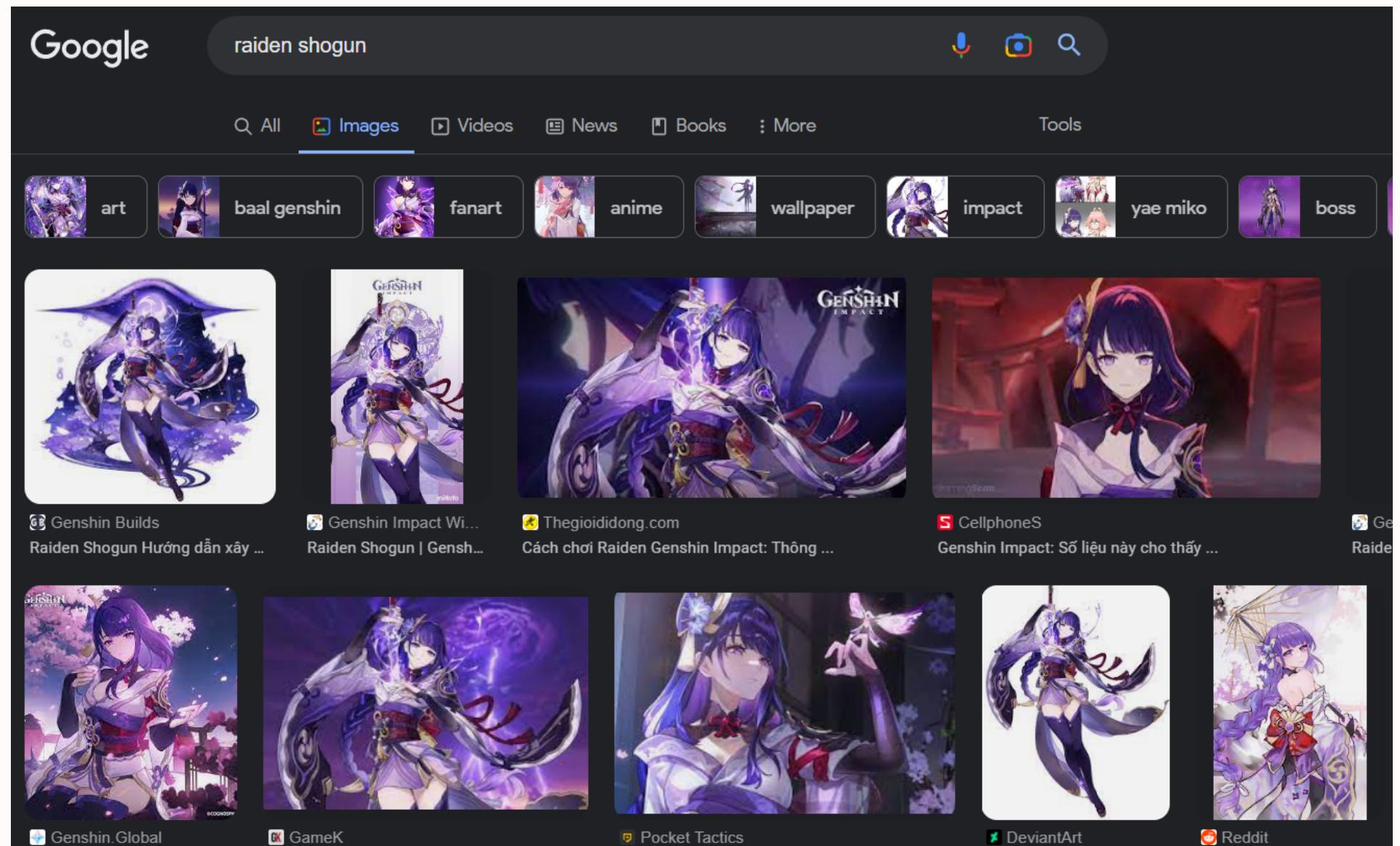
# CRAWLER THU THẬP ẢNH TỪ GOOGLE

## Bước 1: Truy cập vào Google Image qua URL của nó



# CRAWLER THU THẬP ẢNH TỪ GOOGLE

## Bước 2: Nhập từ khóa tìm kiếm và số lượng ảnh mà ta muốn tìm kiếm

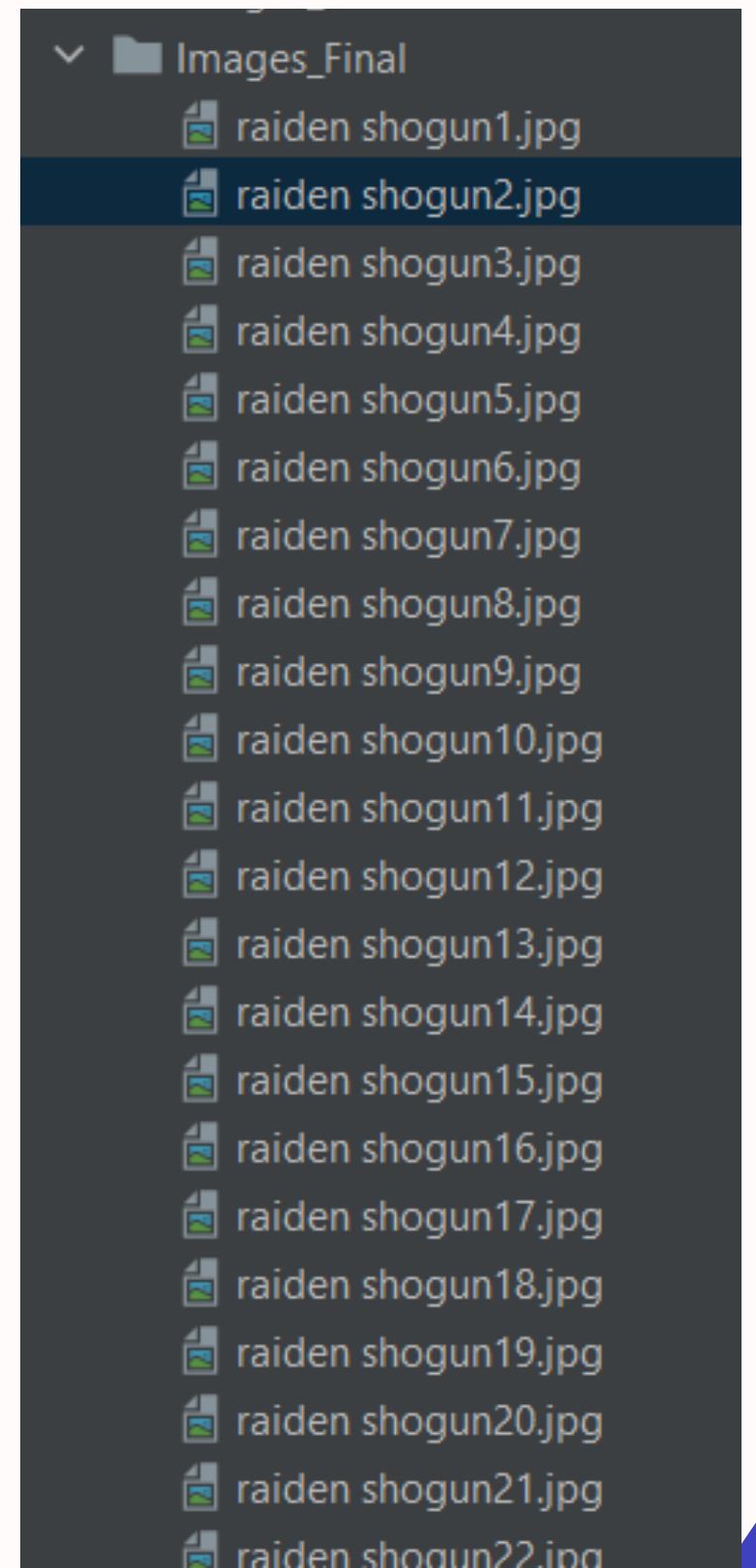
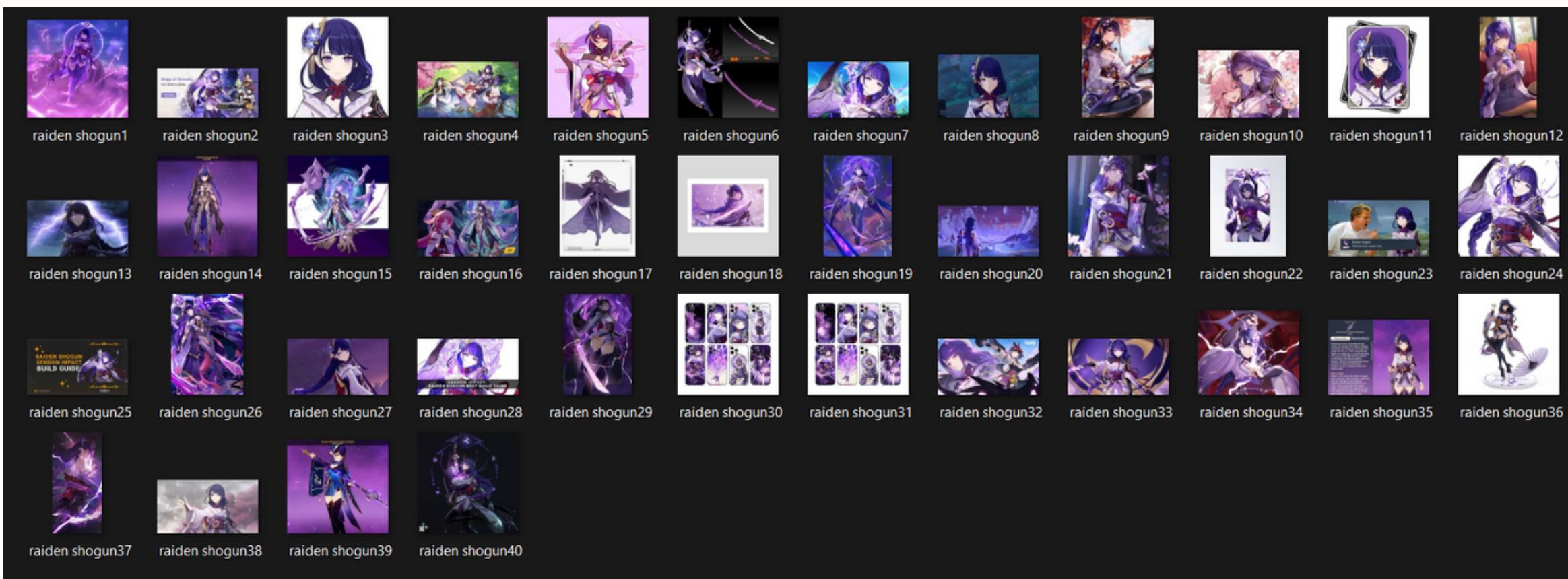


Nhập từ khóa muốn tìm kiếm: **raiden shogun**  
Nhập số lượng ảnh muốn tìm kiếm: **30**  
Tìm kiếm hình ảnh....  
Tìm thấy 30 ảnh  
Bắt đầu tải về...

# CRAWLER THU THẬP ẢNH TỪ GOOGLE

Bước 3: Lấy class rg\_i chứa trong từng thẻ img và tải ảnh xuống folder Image\_Final

class="rg\_i Q4LuWd"



**THANK YOU  
FOR WATCHING**

