

Báo cáo đồ án môn học

Tìm kiếm sản phẩm trên tiki bằng deep learning

Trương Đức Vũ - 18520194

GVHD: TS. Nguyễn Vinh Tiệp

Nguyễn Lâm Quỳnh - 18521326

13/01/2021

Mục lục

1 - Giới thiệu

2 - Cách tiếp cận

3 - Thực nghiệm

4 - Đánh giá - Kết quả

5 - Kết luận

6 - Tham khảo

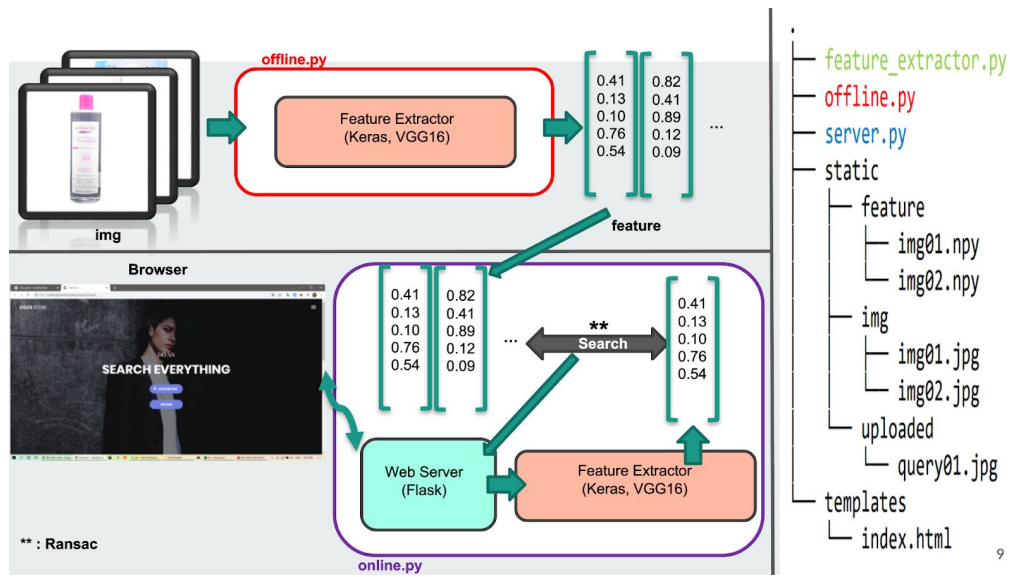
1. Giới thiệu

Thế kỷ 21, nhu cầu mua sắm của con người ngày càng gia tăng. Đặc biệt trong thời đại công nghệ số thì việc mua bán càng ngày càng dễ dàng, người ta không cần phải mất quá nhiều thời gian và công sức để đi đến cửa hàng hoặc trung tâm mua sắm để có thể lựa chọn cho mình một sản phẩm ưng ý. Việc bạn phải làm là có một thiết bị có thể kết nối internet rồi tha hồ mua sắm thỏa thích bằng cách tìm kiếm theo từ khóa. Thế nhưng, đâu phải món đồ nào cũng có thể mô tả bằng từ ngữ được, đặc biệt là trong lĩnh vực thời trang. Thử xem, nếu có một ngày, bạn đang đi đường và thấy một cô gái với một bộ váy cực kì xinh đẹp. Bạn rất muốn có được nó, bạn sẽ làm sao? Nếu là người mà bạn quen biết, bạn có thể lại hỏi người đó để biết xem nên mua ở đâu. Thế nhưng, nếu là người lạ, việc hỏi bộ đồ mà người ta đang mặc sẽ rất ngại đối với cả hai. Và bạn đành ngậm ngùi khi không được sở hữu nó. Đó là động lực thúc đẩy chúng tôi nghiên cứu vấn đề này, tạo ra được một trang web truy vấn bằng hình ảnh. Chỉ bạn chụp hình sản phẩm bạn muốn mua, và upload nó lên trang của chúng tôi, chúng tôi sẽ hiển thị các sản phẩm liên quan với các thông tin về mô tả sản phẩm, giá thành và liên kết để bạn có thể mua nó ngay lập tức.

2. Cách tiếp cận

Bài toán chúng tôi cần giải quyết là Image Search với đầu vào là một tấm ảnh của sản phẩm và đầu ra là hình ảnh của các sản phẩm liên quan, thuộc dạng Similarity Search tức là đi tìm độ giống nhau giữa bức ảnh query và các bức ảnh khác trong dataset, sau đó trả về kết quả dựa trên sự giống nhau từ cao đến thấp. Khó khăn thứ nhất là phải tìm được một cách ‘biểu diễn’ (representation) ảnh tốt dưới dạng các vector để có thể ‘đong đếm’ được sự giống nhau giữa các bức ảnh. Phần này được gọi là Feature Extraction. Nhưng khó khăn lớn hơn là với cả triệu bức ảnh trong dataset, việc tính toán độ giống nhau giữa bức ảnh query và toàn bộ các bức ảnh khác là rất mất thời gian. Dựa trên đó, quy trình chúng tôi xây dựng một image search engine đi theo các bước:

- Thu thập dữ liệu: thu thập trên trang thương mại điện tử tiki.com.vn
- Dùng mạng Deep Learning để rút trích đặc trưng thành features vector
- Xây dựng không gian tìm kiếm
- Đưa ảnh cần truy vấn vào và rút trích đặc trưng tương tự thành feature vector
- So sánh feature vector và không gian tìm kiếm, lựa chọn những vector gần nó nhất
- Xây dựng framework đưa ra màn hình những sản phẩm liên quan bao gồm mô tả, giá và liên kết mua hàng.



Hình 1 : Tổng quan mô hình truy vấn sản phẩm tiki

2.1. Thu thập dữ liệu

Bài toán chúng tôi đề ra yêu cầu phải có lượng dữ liệu lớn, thế nên chúng tôi lựa chọn thu thập các sản phẩm trên tiki. Chúng tôi tìm kiếm được một database có sẵn ở trên github (trong file product.csv). File gồm các mục id, sku, name, url, price, discount, image, seller_id, tiki now, freeship, shocking, under_price, installments, gifts, reviews, rating, info, subcategory của hơn 220.000 sản phẩm. Dựa trên database này chúng tôi tạo ra dữ liệu của mình gồm hai phần. Một là các thư mục chứa hình ảnh của sản phẩm. Chúng tôi sử dụng urllib.request.urlretrieve để có thể download hình ảnh về khi có sẵn đường link của ảnh. Ở đây, chúng tôi chọn đuôi .jpg cho tất cả ảnh. Phần thứ hai bao gồm một file csv cho phép truy cập đến thông tin của sản phẩm với ảnh tương ứng với 4 thuộc tính được lấy từ file product.csv là name,url,price, subcategory và thuộc tính thứ 5 là tên hình ảnh sản phẩm được lưu trong file full_data.csv.

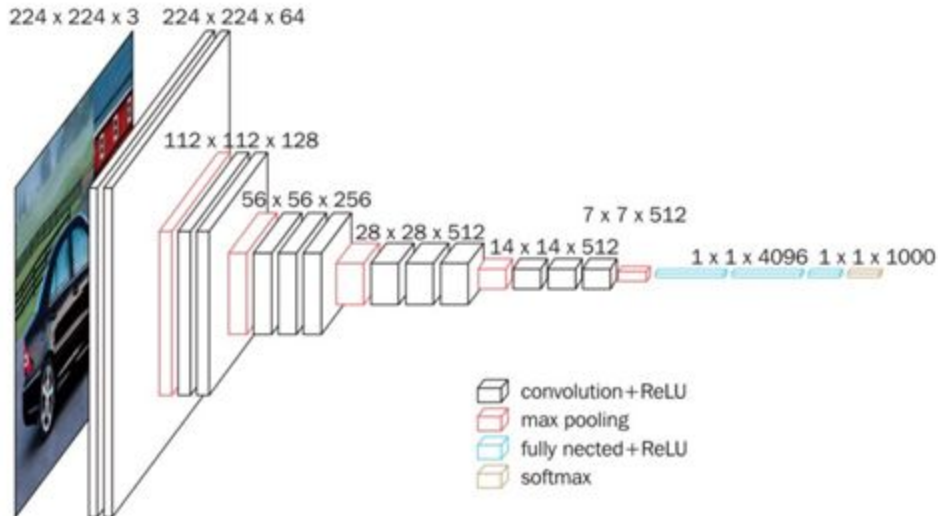
	A	B	C	D	E
1	name_product	link	price	name_img	type
2	Silcot BẮng Trang Ẩiá»fm (l	https://tiki.vn/silcot	33600	1588_1.jpg	1588
3	Bá»™ 3 Há»™p Silcot BẮng	https://tiki.vn/bo-3	101000	1588_2.jpg	1588
4	Silcot BẮng Trang Ẩiá»fm (l	https://tiki.vn/silcot	35000	1588_3.jpg	1588
5	Combo 3 Phim Thá»™m Dá»™	https://tiki.vn/coml	125900	1588_4.jpg	1588
6	BẮng tá»™y trang IPEK 150	https://tiki.vn/bong	31400	1588_5.jpg	1588
7	Khắfn Ắá»™t Cá»™c ẮẮẮ C	https://tiki.vn/khan	23900	1588_6.jpg	1588
8	Phim Thá»™m Dá»™u Jomi N	https://tiki.vn/phim	47000	1588_7.jpg	1588
9	Bá»™ 3 Há»™p Silcot BẮng	https://tiki.vn/bo-3	107000	1588_9.jpg	1588
10	NẮá»™c tá»™y trang Garni	https://tiki.vn/nuoc	148000	1588_10.jpg	1588
11	NẮá»™c Tá»™y Trang TẮ»™	https://tiki.vn/nuoc	59000	1588_11.jpg	1588
12	Combo 3 Giá»™y Thá»™m Di	https://tiki.vn/coml	70000	1588_12.jpg	1588
13	Bá»™ 2 há»™p BẮng Trang	https://tiki.vn/bo-2	76000	1588_13.jpg	1588
14	NẮá»™c tá»™y trang Garni	https://tiki.vn/nuoc	179000	1588_14.jpg	1588
15	TẮ»™i Khắfn NẮ»™n Ắá Nắfn	https://tiki.vn/tui-kl	110000	1588_15.jpg	1588
16	Sá»™a Rắ»™-a Má»™t Dắ»™h C	https://tiki.vn/sua-r	389700	1588_16.jpg	1588
17	NẮá»™c Tá»™y Trang Byph	https://tiki.vn/nuoc	129000	1588_17.jpg	1588
18	NẮá»™c Tá»™y Trang L'ore	https://tiki.vn/nuoc	59000	1588_18.jpg	1588
19	NẮá»™c tá»™y trang Byphi	https://tiki.vn/nuoc	99000	1588_19.jpg	1588

Hình 2: Ví dụ về thông tin của các ảnh

2.2. Dùng deep learning để rút trích đặc trưng

Vấn đề cần giải quyết ở phần này là làm sao cho những sản phẩm đưa ra giống với sản phẩm được truy vấn. Cụ thể là sử dụng các mô hình Convolutional Neural Networks cho Image Classification nổi tiếng, tức đã được trained với các cơ sở dữ liệu lớn và đạt kết quả cao, để tạo ra các feature vector có độ dài như nhau cho mỗi bức ảnh. Cụ thể hơn, đầu ra của layer gần cuối cùng (trước softmax hoặc svm layer) được dùng như là 1 feature tốt .

Đặc điểm dữ liệu của chúng tôi là những sản phẩm với hình ảnh đời thực, chú trọng vào hình dáng và màu sắc. Thế nên để phù hợp với dữ liệu của mình, chúng tôi lựa chọn fc1 layer output của VGG16 đã được đào tạo trên bộ dữ liệu imagenet. Hoàn thành việc rút trích ta sẽ có một features vector với chiều là 220646 x 4096.



Hình 3: Mô hình VGG16

2.3 Xây dựng không gian tìm kiếm

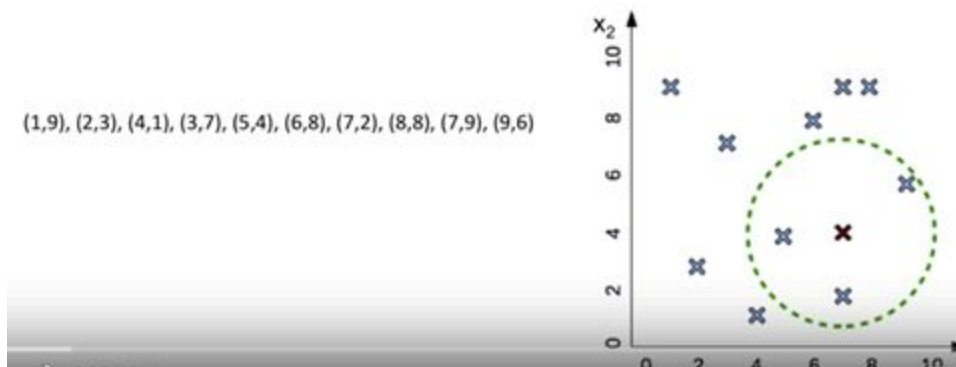
Một search engine tốt đòi hỏi thêm ở chỗ thời gian truy vấn phải nhanh. Với kích thước của bộ dữ liệu nhỏ, chúng ta có thể dùng vét cạn để so sánh giữa vector feature của ảnh truy vấn và features vector của bộ dữ liệu. Nhưng nếu với số lượng bộ dữ liệu lớn, chúng ta không thể nào đi so sánh từng feature vector với nhau được. Bởi vậy, cần có một cách quản lý dữ liệu phù hợp để có thể truy xuất hiệu quả. Hiện nay các kĩ thuật chủ yếu người ta thường sử dụng là Binary Hashing và Nearest Neighbor Search. Ở đây chúng tôi sử dụng cấu trúc Kd-tree một loại của Nearest Neighbor Search để xử lý. Kd-tree là cách tìm kiếm hiệu quả các dữ liệu gần, liên quan nhất trong miền không gian hai chiều và nhiều chiều.

Chúng ta có thể xây dựng Kd- tree bằng cách:

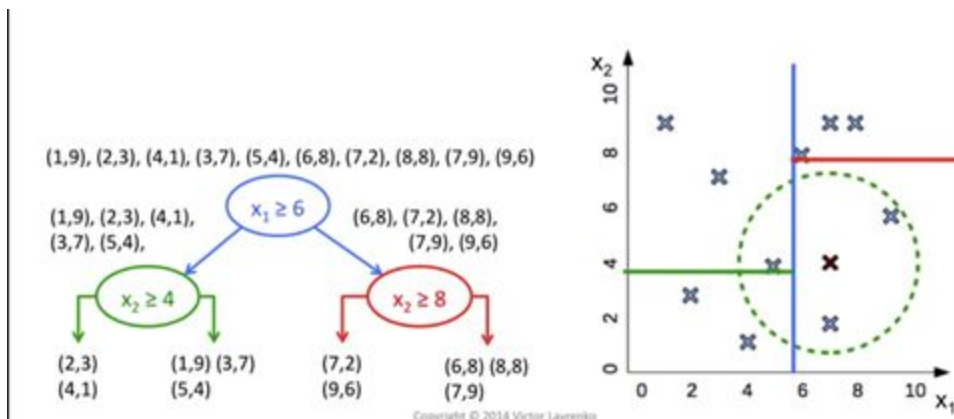
- 1) Chọn một chiều random
- 2) Tìm tọa độ trung bình

3) Chia data theo tọa độ đã chọn

4) Chọn random một chiều nhưng không trùng với những chiều đã chọn trước đó rồi lặp lại (2), (3).



Hình 4: Mô phỏng KD - tree

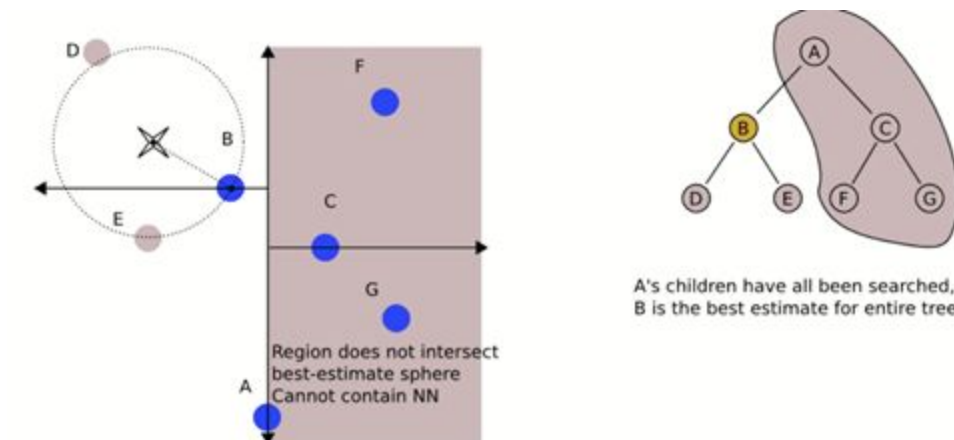


Hình 5: Quá trình tìm kiếm trên KD - tree

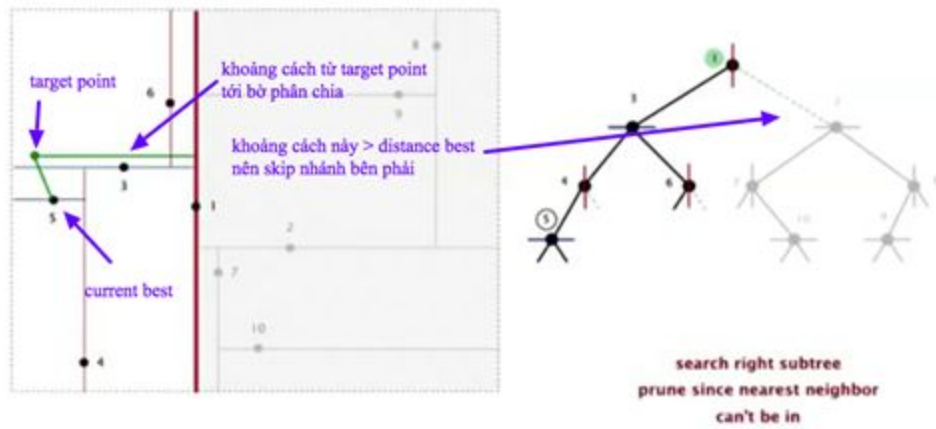
Cách tìm kiếm trên cây từ một điểm cho trước:

1) Bắt đầu từ root node, dịch chuyển qua các nhánh của cây

- 2) Trong khi đi qua các nhánh, thuật toán sẽ lưu lại các node có khoảng cách các gần nhất với target point, được gọi là current best
- 3) Nếu node hiện tại gần target point hơn current best, thì nó trở thành current best
- 4) Trong khi di chuyển nó sẽ kiểm tra xem với điểm current best ở nhánh trái thì khoảng cách từ target point đến current best có gần hơn khoảng cách từ target point tới bờ phân chia hay không. Nếu gần hơn tức là bên nhánh trái đã cho kết quả tốt nhất và ta không cần tìm tiếp bên phải, nếu dài hơn tức là sẽ có một điểm nào đó ở bên phải cho khoảng cách tới target point tốt hơn nên ta phải duyệt qua các nhánh bên phải.
- 5) Với mỗi nhánh, thuật toán hoàn thành cho đến khi chạm tới leaf node.



Hình 5.1



Hình 5.2

3. Thực nghiệm

3.1. Tiền xử lý

Vì số hình ảnh là khá lớn hơn 220.000 hình ảnh các loại, nên để mô hình chạy nhanh hơn, chúng tôi đã giảm lưu lượng bằng cách giảm số chiều từ 4096 xuống còn 512 chiều mỗi hình.

3.2. Ransac

Trong mô hình này, chúng tôi đã sử dụng kỹ thuật ransac để re-ranking lại kết quả truy vấn ra.

Để dễ hiểu, phần tiếp theo sử dụng bài toán cụ thể. Từ cụ thể ta suy ra tổng quát.

- Cho tọa độ của N điểm trên mặt phẳng.
- Giả sử có một số điểm nằm trên đường thẳng, còn các điểm khác là nhiễu (nghĩa là biết chắc có 2 tập hợp con, còn gọi là cluster).
- Hãy tìm phương trình đường thẳng, và cho biết những điểm nào được coi là nằm trên đường thẳng, những điểm nào được coi là nhiễu.
-

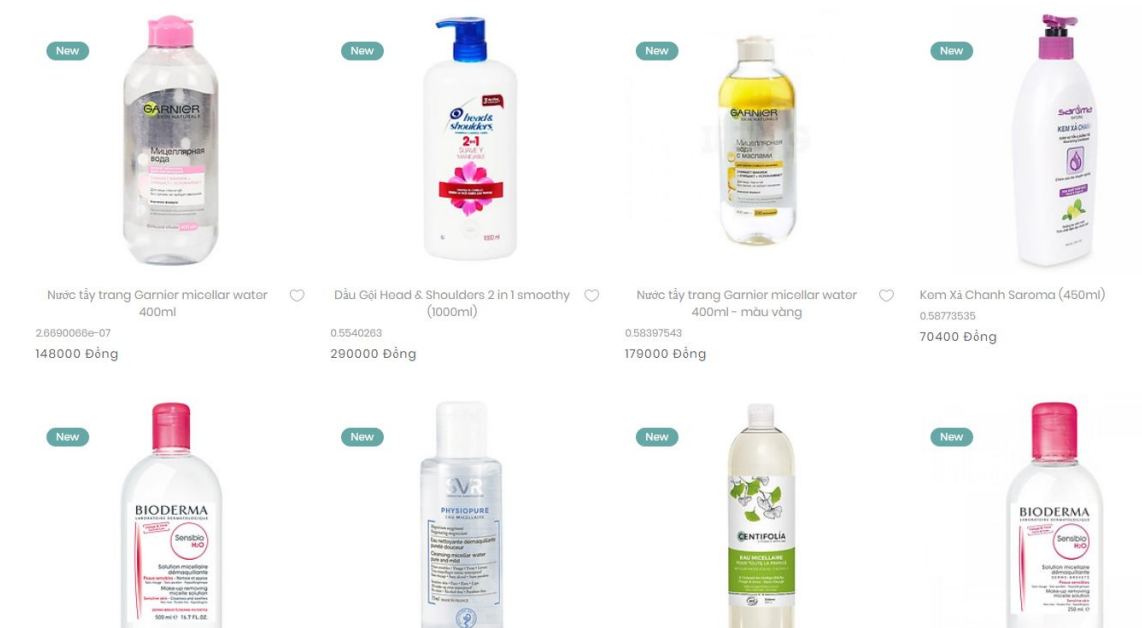


Hình 6: Ví dụ ransac

Ở trong bài làm này, chúng ta đơn giản là so sánh số điểm inlier của ảnh truy vấn và với các ảnh mà chúng ta đã truy vấn ra ở trước đó, nếu ảnh nào có số inlier càng cao thì thứ tự xếp hạng càng cao.

RESULTS

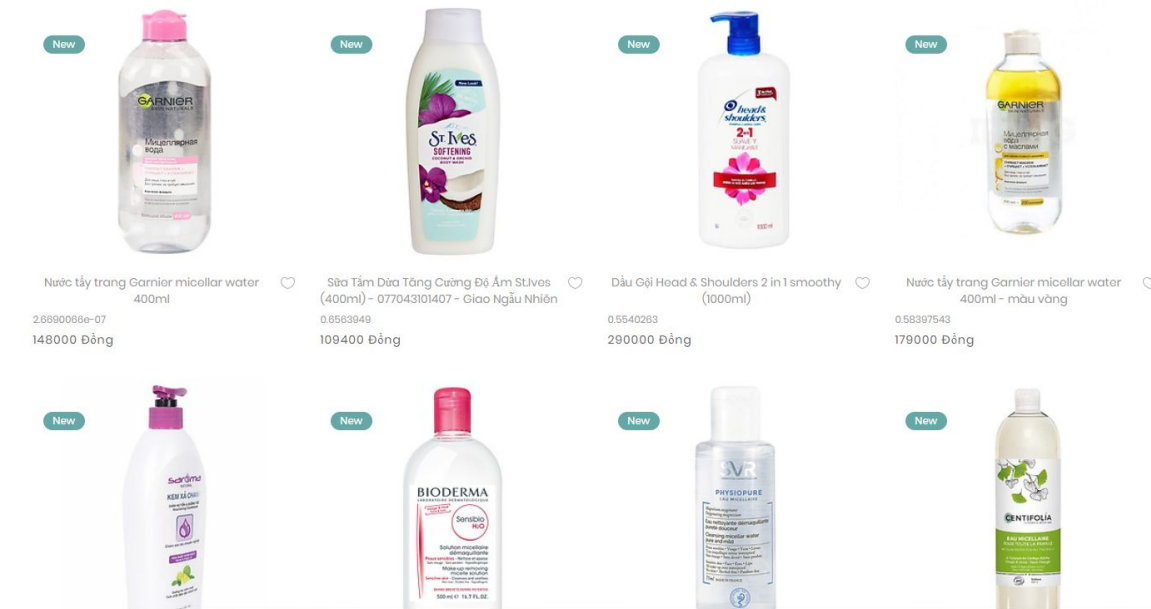
Có 30 kết quả - thời gian tìm kiếm : 4.819089651107788



Không có Ransac

RESULTS

Có 30 kết quả - thời gian tìm kiếm : 8.07412314414978



Có ransac

Ta có thể nhận thấy là kết quả top đầu khi sử dụng và không sử dụng kỹ thuật Ransac là thay đổi không nhiều mà thời gian lại dài hơn trung bình gần 4 giây.

4. Đánh giá - Kết quả

Trong mô hình này chúng tôi sử dụng độ đo **mean average precision (map)** để đánh giá độ chính xác của mô hình truy vấn thông tin này. Để đánh giá, chúng tôi chọn ngẫu nhiên 1000 sản phẩm kèm theo loại sản phẩm và tính toán các thông số mAP@10, mAP@20, mAP@50, mAP@100.

	mAP@10	mAP@20	mAP@50	mAP@100
(%)	64.16	54.96	46.56	39,34

Bảng 1: Kết quả đánh giá

Khi sử dụng ransac, thời gian của mô hình truy vấn thông tin trung bình cần thêm 4 giây để re-ranking. Qua 100 query, chúng tôi ước tính được thời gian trung bình khi sử dụng và không sử dụng ransac như sau:

	Sử dụng ransac	Không sử dụng ransac
giây/hình	8.0376	4.9348

5. Kết luận

Qua đồ án môn học, chúng tôi đã biết cách và xây dựng được một mô hình truy vấn thông tin nói riêng cũng như là một mô hình tìm kiếm các sản phẩm trên tiki. Hơn thế nữa, mô hình này cũng có thể áp dụng vào thực tế một cách thông minh hơn tìm kiếm thông thường. Mô hình có thể tìm kiếm ra chính xác một sản phẩm đang được bày bán trên tiki thông qua một hình ảnh được chụp từ thực tế.

Bên cạnh đó, cũng có một vài hạn chế mà mô hình còn gặp phải. Như thời gian tìm kiếm còn hơi lâu (khoảng 5 giây cho một sản phẩm)

Trong tương lai, chúng tôi dự định sẽ thu thập thêm dữ liệu thường xuyên và cập nhật lên mô hình vì có rất nhiều sản phẩm trên tiki được đăng bán mới mỗi ngày. Bên cạnh đó, chúng tôi cũng sẽ tìm cách tối ưu được hiệu suất cũng như thời gian truy vấn sản phẩm.



6. Tham khảo



[Simple Image Search Engine](#)

[Feature Matching + Homography to find Objects](#)

[Breaking Down Mean Average Precision \(mAP\)](#)

[Tập tành crawl dữ liệu product của Tiki](#)