Library Progress International
Vol.44 No. 3, Jul-Dec 2024: P.13726-13733

Print version ISSN 0970 1052
Online version ISSN 2320 317X

Original Article

Available online at www.bpasjournals.com

Multi-modal Retrieval Augmented Generation for Product Query

Quang-Vinh Dang

British University Vietnam, Vietnam
[0000−0002−3877−8024]
vinh.dq4@buv.edu.vn

How to cite this article: Quang-Vinh Dang (2024). Multi-modal Retrieval Augmented Generation for Product Query. *Library Progress International,* 44(3), 13726-13733.

ABSTRACT

Product Query, when an e-commerce website needs to return a product for a user's query, is essential for any e-commerce system. Traditional search systems only consider the user query in the text and then try to match the search queries with the products' descriptions in the database. Some recent image search systems utilize deep learning methods to match the image query with the products' images in the database. However, none combine text and images in a single query. This kind of search is common in modern daily life as a user can take a photo easily with their smartphone and provide a short text description then try to search for a product. In this paper we consider a multi-modal retrieval augmented generation (RAG) to provide a product query system that allows users to search simultaneously by image and text. Our system will provide a better experience and improve the performance of e-commerce websites.

Keywords: RAG, GenAI, Product Query, Multi Modal

1. Introduction

Product search is an essential task for any e-commerce website. Traditionally, the e-commerce systems allow users to search for products by inputting some texts [11], as displayed in Figure 1.

Recently, some visual search systems [3] allow users to search by image. The system will allow a user to upload a photo and then search for products that are similar to this photo as displayed in Figure 2. The task is usually done via image embedding [4]. The process is visualized in Figure 3.

Both text-based search and image-based search play important roles in the development of the current e-commerce system. However, the recent rise of the multi-modal information retrieval [9] due to the popularity of smartphones requires e-commerce systems to support the multi-modal search, which is currently not available.

In this paper we present a multi-modal search based on multi-modal retrieval augmented generation (RAG), a well-known technique in developing LLM-based AI systems today. The system will enhance the user experience.

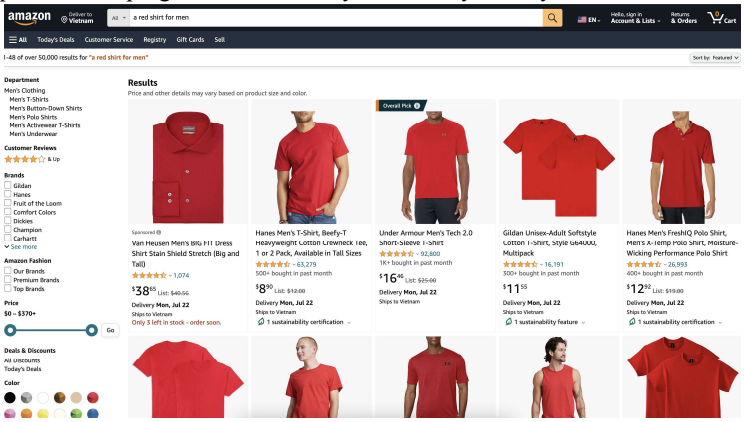

Fig.1. Amazon allows users to search by text.

Fig.2. Amazon visual search [3]

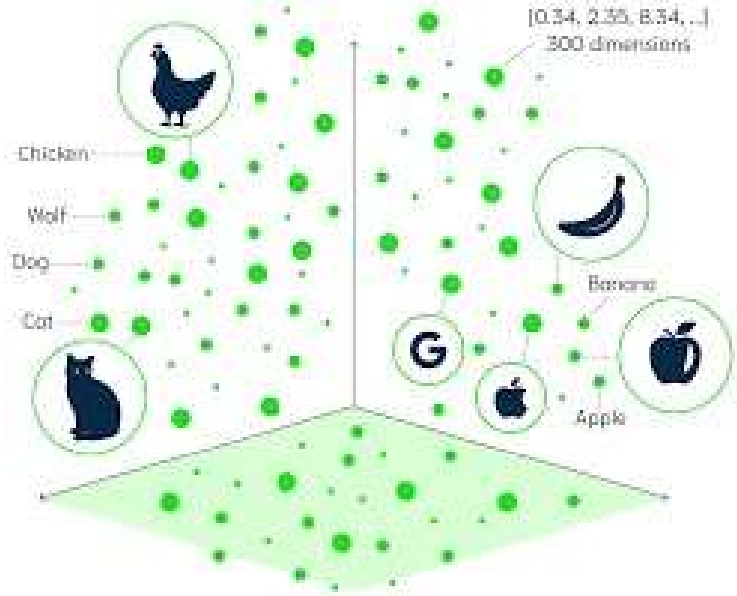

Fig.3. Image Similarity Search via Embedding

2. Related Works

Text search is one of the foundation problems in computer science [1] which has been studied for many years [19]. In industry, elasticsearch [7] is considered one of the most successful search applications. The most common technique to use in text search is the inverted indexing [7] which indexes the word by documents, as visualized in Figure 4.

For image search, the search is done via calculating the similarity in an embedding space such as VGG16 or VGG19 [12]. The architecture of VGG19 is presented in Figure 5.

Other embedding methods have been studied before the deep learning era [10,13]. Their ideas are presented in Figure 6 and 7.

3. Methods

In this paper we presented the method of multi-modal RAG. We used the ChromaDB vector database[1].

We used OpenCLIP for image embedding [5,16,14,2]. OpenCLIP is an opensource implementation of CLIP [15] that can embed both text and image into one space.

[1] https://www.trychroma.com/

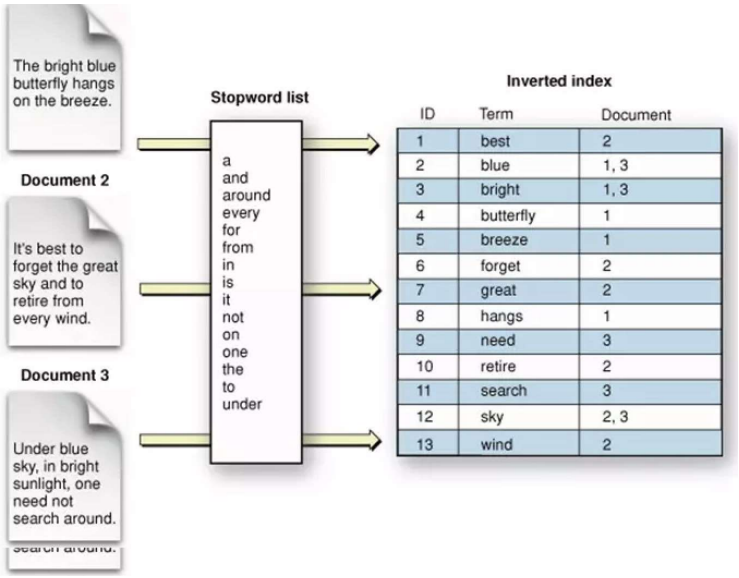

Fig.4. Inverted index for search

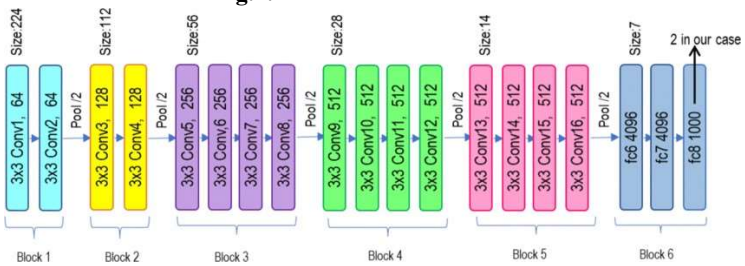

Fig.5. VGG19 architecture [8]

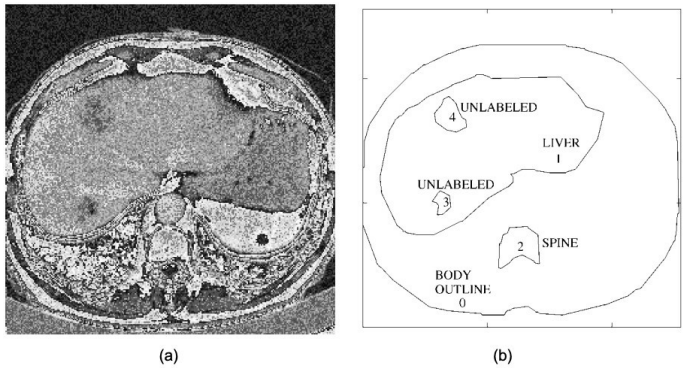

Fig.6. Image embedding by labelled parts [13]

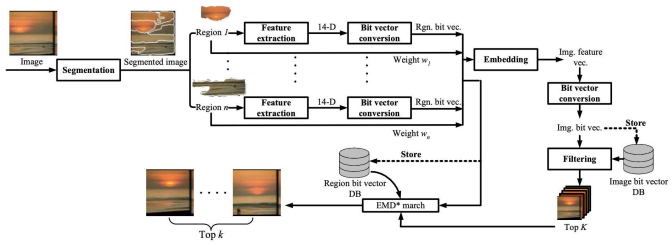

Fig.7. Image embedding by feature extraction [10]

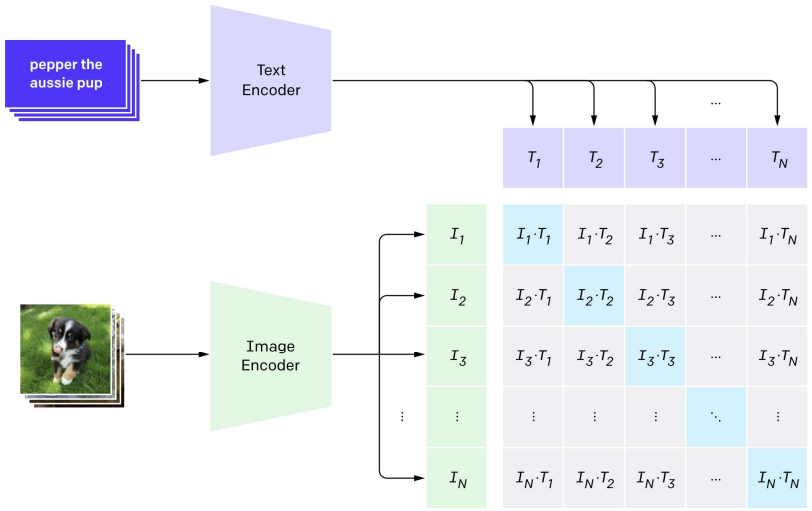

Fig.8. CLIP model

CLIP used the proxy training task: given an image, predict which out of a set of 32,768 randomly sampled text snippets, was paired with it in our dataset. By doing so, we can learn the pair text-image and provide a good learning example.

4. Experimental Results and Evaluation

We used the FashioPedia dataset [6,17,18].

According to the authors, "Fashionpedia is a dataset which consists of two parts: (1) an ontology built by fashion experts containing 27 main apparel categories, 19 apparel parts, 294 fine-grained attributes and their relationships; (2) a dataset with 48k everyday and celebrity event fashion images annotated with segmentation masks and their associated per-mask fine-grained attributes, built upon the Fashionpedia ontology."[2]. A sample image from the dataset is displayed in Figure 9.

We tested the system by some sample queries like "Blue shirt". The system returns an image of a blue shirt and also provide some recommendations about how to wear the blue shirt. It indeed improves the user experience in using the e-commerce website.

Styling a blue shirt can be both fun and versatile ! Here are some ideas inspired by the images you
provided :

[2] https://fashionpedia.github.io/home/index.html

Fig.9. An image from FashionPedia [6]

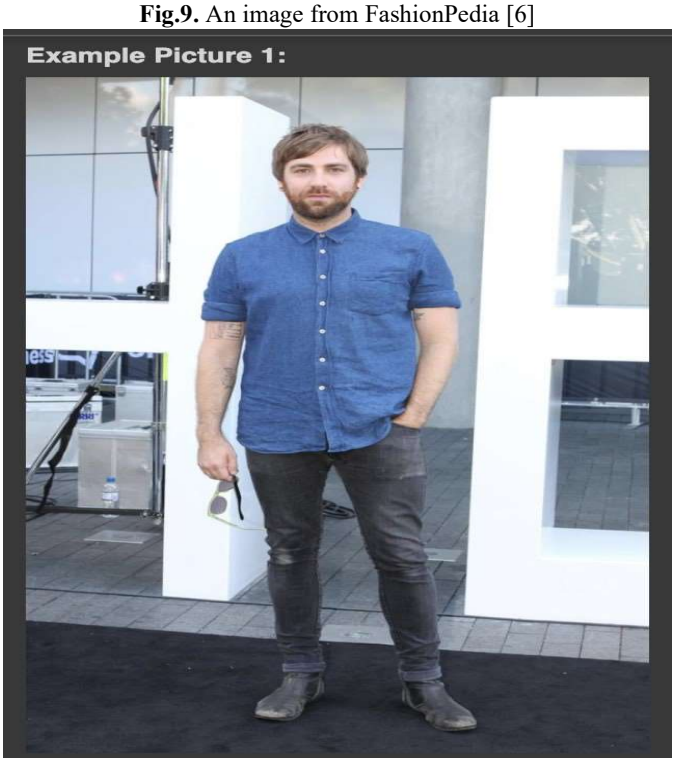

Fig.10. A blue shirt returned by the system

For a Casual Look :

1. Denim on Denim: Like the f i r s t image , pair your blue shirt with dark jeans . This creates a cohesive and stylish look . Rolling up the sleeves adds a relaxed vibe .
2. Layering : Throw on a casual jacket or a cardigan over your blue shirt . This can add depth and texture to your outfit .

3. Footwear : Opt for casual shoes like sneakers or boots .
The boots in the f i r s t image add a rugged touch to the outfit .
For a Chic Look :
1. Leggings : As seen in the second image , pairing a blue shirt with black leggings creates a sleek and comfortable look . This is perfect for a casual day out
or even a relaxed evening .
2. Heels : Elevate the look with a pair of heels . The black heels in the second image add a touch of elegance and sophistication .
3. Accessories : Add some statement jewelry or a stylish belt to accentuate your waist and add some f l a i r to your outfit .
For a Smart−Casual Look :
1. Chinos or Trousers : Swap out jeans for chinos or tailored trousers . This can instantly make your blue shirt look more polished .
2. Blazer : Adding a blazer can transform your blue shirt into a smart−casual ensemble suitable for work or a semi−formal event .
3. Loafers or Dress Shoes : Complete the look with a pair of loafers or dress shoes for a refined touch .
For a Sporty Look :
1. Athleisure : Pair your blue shirt with joggers or athletic shorts . This is perfect for a sporty and comfortable look .
2. Sneakers : Finish off with a pair of trendy sneakers to keep the look casual and sporty .
Color Coordination :
Neutrals : Blue pairs well with neutral colors like black , white , and grey . These combinations are timeless and versatile .
Contrasts : For a bolder look , pair your blue shirt with contrasting colors like mustard yellow , burgundy , or even a bright red .
Accessories :
Sunglasses : As seen in the f i r s t image , carrying a pair of stylish sunglasses can add a cool factor to your outfit .
Bags : A sleek backpack or a crossbody bag can complement your look while being functional .
Remember, the key to styling is to feel comfortable and confident in what you wear . Mix and match these ideas to create your unique style !

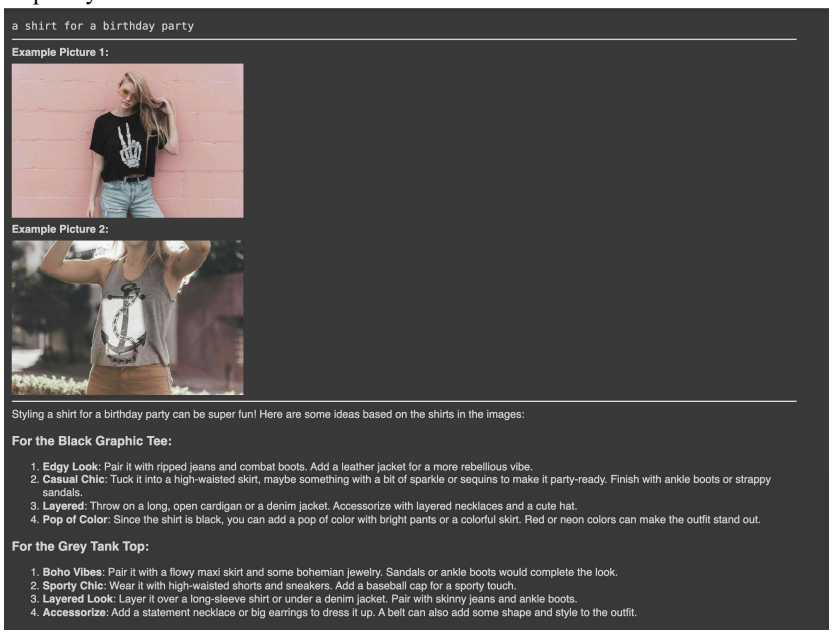

Fig.11. A search for properties not in the database.

We can even search for products with properties that are not explicitly described as in Figure 11.

5. Conclusion

In this paper we consider the multi-modal RAG system that process the image and the text simultaneously. The system will enhance the user experience in ecommerce website and improve the performance of the system. We will integrate other modal to our system to improve their performance in the future.

References

1. Berberich, K., Bedathur, S., Neumann, T., Weikum, G.: A time machine for textsearch. In: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. pp. 519–526 (2007)
2. Cherti, M., Beaumont, R., Wightman, R., Wortsman, M., Ilharco, G., Gordon, C.,Schuhmann, C., Schmidt, L., Jitsev, J.: Reproducible scaling laws for contrastive language-image learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2818–2829 (2023)
3. Du, M., Ramisa, A., KC, A.K., Chanda, S., Wang, M., Rajesh, N., Li, S., Hu, Y.,Zhou, T., Lakshminarayana, N., et al.: Amazon shop the look: A visual search system for fashion and home. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 2822–2830 (2022)
4. Gordo, A., Almaz´an, J., Revaud, J., Larlus, D.: Deep image retrieval: Learningglobal representations for image search. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14. pp. 241–257. Springer (2016)
5. Ilharco, G., Wortsman, M., Wightman, R., Gordon, C., Carlini, N., Taori, R., Dave,A., Shankar, V., Namkoong, H., Miller, J., Hajishirzi, H., Farhadi, A., Schmidt, L.: Openclip (Jul 2021). https://doi.org/10.5281/zenodo.5143773, https://doi.org/ 10.5281/zenodo.5143773, if you use this software, please cite it as below.
6. Jia, M., Shi, M., Sirotenko, M., Cui, Y., Cardie, C., Hariharan, B., Adam, H.,Belongie, S.: Fashionpedia: Ontology, segmentation, and an attribute localization dataset. In: European Conference on Computer Vision (ECCV) (2020)
7. Kathare, N., Reddy, O.V., Prabhu, V.: A comprehensive study of elasticsearch.International Journal of Science and Research (IJSR) (2020)
8. Khattar, A., Quadri, S.: Generalization of convolutional network to domain adaptation network for classification of disaster images on twitter. Multimedia Tools and Applications **81**(21), 30437–30464 (2022)
9. Luo, M., Gokhale, T., Varshney, N., Yang, Y., Baral, C.: Multimodal informationretrieval. In: Advances in Multimodal Information Retrieval and Generation, pp. 35–91. Springer (2024)
10. Lv, Q., Charikar, M., Li, K.: Image similarity search with compact data structures.In: Proceedings of the thirteenth ACM international conference on Information and knowledge management. pp. 208–217 (2004)
11. Maitra, S., Sahoo, L., Tiwary, K.: Study, analysis and comparison between amazona10 and a11 search algorithm. Journal of Computer Science Research **4**(4), 1–6 (2022)
12. Mascarenhas, S., Agarwal, M.: A comparison between vgg16, vgg19 and resnet50 architecture frameworks for image classification. In: 2021 International conference on disruptive technologies for multi-disciplinary research and applications (CENT-
CON). vol. 1, pp. 96–99. IEEE (2021)
13. Petrakis, E.G.M., Faloutsos, A.: Similarity searching in medical image databases.IEEE transactions on knowledge and data engineering **9**(3), 435–447 (1997)
14. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G.,Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision. In: ICML (2021)
15. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G.,Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PMLR (2021)
16. Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C.W., Wightman, R., Cherti,M., Coombes, T., Katta, A., Mullis, C., Wortsman, M., Schramowski, P., Kundurthy, S.R., Crowson, K., Schmidt, L., Kaczmarczyk, R., Jitsev, J.: LAION-5b: An open large-scale dataset for training next generation image-text models. In: Thirty-sixth Conference on Neural Information Processing Systems Datasets and
Benchmarks Track (2022), https://openreview.net/forum?id=M3Y74vmsMcY
17. Shi, M., Belongie, S., Cardie, C.: Fashionpedia-taste: A dataset towards explaininghuman fashion taste. arXiv preprint arXiv:2305.02307 (2023)
18. Shi, M., Cardie, C., Belongie, S.: Fashionpedia-ads: Do your favorite advertisementsreveal your fashion taste? arXiv preprint arXiv:2305.02360 (2023)
19. Tenopir, C.: Full text database retrieval performance. Online Review **9**(2), 149–164

(1985)