Fifty Shades of Pink: Understanding Color in e-commerce using Knowledge Graphs

Lizzie Liang
eBay Inc.
San Jose, CA, USA
lizliang@ebay.com

Sneha Kamath
eBay Inc.
San Jose, CA, USA
snkamath@ebay.com

Petar Ristoski
eBay Inc.
San Jose, CA, USA
pristoski@ebay.com

Qunzhi Zhou
eBay Inc.
San Jose, CA, USA
qunzhou@ebay.com

Zhe Wu
eBay Inc.
San Jose, CA, USA
zwu1@ebay.com

ABSTRACT

The color of the products is one of the most prevalent aspects in many e-commerce domains, and it is one of the decisive purchasing factors. Besides having thousands of color variations and shades, many brands continuously develop proprietary colors and color names to attract more customers. This often leads to color ambiguity (textual and visual), and vocabulary mismatch between buyers and sellers. Therefore, it is crucial for any e-commerce search engine to correctly identify the buyer's color intent and match it to the corresponding product listings. To address this challenge, in this work, we introduce a color query expansion approach using color Knowledge Graphs. We use Knowledge Graphs to unambiguously identify all the colors based on their properties, and the relationships to other colors, which allows us to perform semantic query expansion. Similar expansion concepts could be applied to domains outside of color.

CCS CONCEPTS

• **Applied computing** → **Online shopping**; *Document searching*.

KEYWORDS

Query Expansion, Knowledge Graphs, e-Commerce

ACM Reference Format:
Lizzie Liang, Sneha Kamath, Petar Ristoski, Qunzhi Zhou, and Zhe Wu. 2022. Fifty Shades of Pink: Understanding Color in e-commerce using Knowledge Graphs. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22), October 17–21, 2022, Atlanta, GA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3511808.3557513

1 INTRODUCTION

Traditional query expansion techniques represent queries and documents as a bag-of-words and the goal is to find synonyms and other semantically related terms of the user query in the documents. While effective, lexical match falls short at capturing scenarios that use different terms to describe the same concept. The vocabulary mismatch between query and documents is a critical challenge in search [1]. There is a growing development of the embedding based approach to match a query to its semantic close documents. The embedding based approaches aim at overcoming the vocabulary mismatch challenge. In e-commerce product search, user search queries tend to be concise, and they contain a combination of keywords that represent the attributes of the products. Despite the advance of embedding approaches, they may not be effective at matching all important keywords in a query [3].

To fill in this gap, we proposed a Knowledge Graph (KG) based query expansion approach. Our work consists of two main components. The first component uses a color knowledge graph to identify colors in the standard color palette referenced by search queries. We then find the shades of the colors using a color knowledge graph and expand these colors to their shades. For instance, if the color pink appears in the search query, we believe the implicit semantic match for pink could be its shades, including rose gold and magenta, however, rose gold and magenta should not be matched to pink. In other words, if a user is dedicated to finding a rose gold product, we do not surface other shades of pink products. The second component of our work expands the color in a search query contextually. We learn brand specific colors and expand the colors with brand information. For instance, we expand the search query "pink sneakers" to "nike laser fuchsia sneakers" because laser fuchsia is a color specific to Nike.

We have performed our experimental study across all verticals, including Fashion, Electronics, Homes and Garden, Collectibles etc. Both the offline evaluation and the A/B tests showed a significant recall increase and a decrease in low recall query rate in many verticals.

2 APPROACH

Our approach for color query expansion consists of 3 components -

- knowledge graph
- attribute co-occurrence mining
- query expansion

The color subgraph in eBay KG captures colors' name variations, visual similarity and shade relations. In particular, shade relation is defined as the mappings between arbitrary colors to a small

CIKM '22, October 17–21, 2022, Atlanta, GA, USA

ACM ISBN 978-1-4503-9236-5/22/10.
https://doi.org/10.1145/3511808.3557513

set (16) of predefined palette colors such as red, blue and so on, which are commonly mentioned in search queries. To achieve high coverage, we processed multiple external data sources including public knowledge bases and manufacturer websites to extract color entities. Color similarity and shade relations are computed by an algorithm based on Delta E distances in the LAB color space, which showed around 96% accuracy in human judge evaluations. Currently, there are a total of 10K color entities captured in the graph covering color mentions in around 5% queries and 90% items.

Besides the generic color KG, we mine eBay inventory to identify brand-specific colors, which usually are not well represented outside the e-commerce domain, e.g., Blue Sierra is a color specific to the iPhone 13 Pro phone model by Apple. To do so, we construct a directed weighted aspect-value graph based on aspect-value pairs co-occurrence in listings. We embed the graph using RDF2vec [2], which allows us to quickly identify similar colors (e.g. Graphite is similar to Sierra Blue in the "Cell Phones" category), and related concepts (e.g. Sierra Blue is only related to the brand Apple, and the iPhone 13 Pro and Pro Max models). This allows us to perform better explicit contextual rewrites, conditioned by additional concepts, such as brands, models, product lines, categories etc.

We use the color knowledge graph to identify the color mentions in search queries. Currently the standard 16 palette colors are considered. Once a color mention is identified, we retrieve its shades from the color graph and expand the color in the search query to the shades. In our experience, a color have many shades, although a shade of a color is not a synonym of the color, we argue that surfacing similar products containing the shade of the color in addition to an exact match of the color is reasonable because the standard colors are universal across brands and they are more frequently used when users fail to recognize their uncommon shades.

Furthermore, to achieve contextualized query expansion, we identified brand specific colors with attribute co-occurrence. Once a color mention is identified in a search query, the query is not only rewritten to a form containing the shades of the color, but it can also be expanded to a brand and a color specific to this brand. For instance, the search query "pink sneakers" is rewritten to "magenta sneakers", "rose gold sneakers" and "nike laser fuchsia sneakers", because magenta and rose gold are shades of color pink in the color knowledge graph and laser fuchsia is identified as a specific color of the brand Nike.

3 EVALUATION

We performed an offline and online evaluation for our query expansion approach. In the offline evaluation, a recall relevance analysis was performed on 20,000 color search queries across all verticals. As can be seen in Table 1, Around 44.8% of the queries showed no change in recall, 50.3% of the queries see a recall increase between 0 and 30%, and 4.7% of the queries have a large recall increase greater than 30%. While seeing promising recall increase, as seen in Table 2, 78.3% of the queries have a neutral or positive relevance change when measuring with our relevance model. Within the colors in the standard color palette, black, green, blue, red and white showed a more significant lift than other colors.

Table 1: Color Query Expansion Offline Evaluation - Recall Change

Recall Change%	Perc.	Comments
no change	44.8%	44.8% of the queries showed no recall change.
between 0 and 30%	50.3%	50.3% of the queries showed a up to 30% recall increase.
greater than 30%	4.7%	4.7% of the queries showed a at least 30% recall increase.
negative change	0.2%	0.2% of the queries showed a recall drop.

Table 2: Color Query Expansion Offline Evaluation - Relevance Change

Relevance Change	Perc.	Comments
no change	34.5%	34.5% of the queries have no rel. change.
increase	43.8%	43.8% of the queries have a rel. increase.
decrease	21.7%	21.7% of the queries have a rel. drop.

In addition, we observed this color query expansion technique is more effective in some verticals than others. Electronics & Accessories showed the most gain in low recall query rate with a 14 point drop, followed by Home & Garden and Jewelry & Watches with a 6.8 point and a 5.5 point drop respectively in low recall rate, whereas Collectibles didn't show any decrease in the same metric.

4 CONCLUSION

In this work, we introduced a knowledge graph based query expansion approach. This approach effectively filled in the vocabulary gap in query and document matching. From the experiments, we find that our approach significantly increased the recall and reduced low recall rate search queries. We also provide insights to help practitioners apply similar concepts to domains outside of color.

In future work, we aim to bring this expansion technique to domains outside of color such as material, brand model, and product line etc. Furthermore, we hope to achieve contextualized query expansion by recognizing a term in relation to important entities such as vertical, brand, model, and product line etc.

REFERENCES

[1] Hang Li, Jun Xu, et al. 2014. Semantic matching in search. *Foundations and Trends® in Information Retrieval* 7, 5 (2014), 343–469.
[2] Petar Ristoski and Heiko Paulheim. 2016. Rdf2vec: Rdf graph embeddings for data mining. In *International Semantic Web Conference*. Springer, 498–514.
[3] Fatemeh Sarvi, Nikos Voskarides, Lois Mooiman, Sebastian Schelter, and Maarten de Rijke. 2020. A comparison of supervised learning to match methods for product search. *arXiv preprint arXiv:2007.10296* (2020).