CS6200 Project Proposal

Yuting Sun

Background

It is always a challenging task to try out a new cosmetic product. Different people have different skin types and product preference. A product that is good for one person may not be suitable for another. Improper products would waste a lot of money and even cause severe skin problems. Since the COVID-19 pandemic, it's not convinient for people to go to the stores and try the products on themselves. Therefore, it would be helpful if we could build a searching tool for cosmetic products.

Ingredients are key to the property and effectiveness of a cosmetic product. For example, salicylic acid is widely used in treating acnes, titanium dioxide is used for UV absorbance in sunscreen products. Ingredient analysis can be very helpful for people to search for proper products. Some people likes the certain propety of a product, so s/he would propably like a product with similar ingredients. Some people are allergic to a certain ingredient, so it would be necessary to avoid products with the same ingredient. However, most cosmetic products contain more than 30 ingredients, which is difficult for people to read and analysis. It would be a good idea to combine information retrieval techniques with ingredient analysis. This would help customers find out new product that are most suitable for them, as well as getting more insights on cosmetic ingredients.

Solution

In this project, I would like to propose a searching tool for cosmetic products that contains two parts:

1) Search for relevant products based on product type, name, and description.

- The product category (eg. skincare, makeup, etc) can be used as filters for search.
- Product category and brand are indexed as single terms.

  Product name and description are indexed with n-gram overlap (where n=1 to 3) with tf-idf weighting. Stopwords are removed.
- Top-K relevant results can be obtained by BM25 algorithm using python ElasticSearch. The results are by default sorted by relevant scores.

2) Recommend similar product based on ingredient analysis.

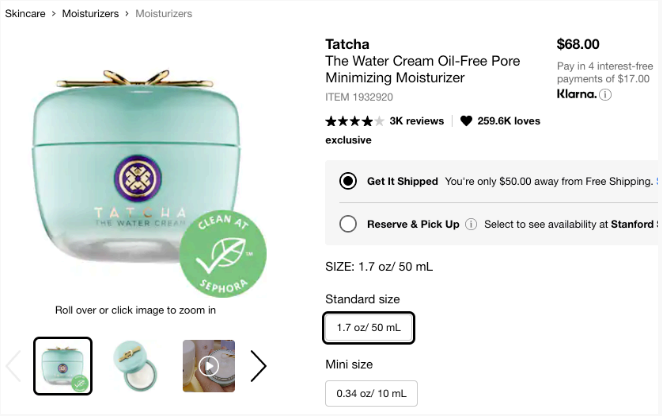
- Ingredients of each product is indexed and vectorized as a single document. Similarity is calculated via cosine score.
- When a user opens a product page, a list of similar products (sorted by similarity) are shown at the right side.

Dataset

The Sephora dataset can be obtained from Kaggle. This dataset contains 9000+ products with product information, detailed description and ratings. More than 7500 products have ingredient information.

Product Information Example:

Basic Info



Skincare › Moisturizers › Moisturizers
Roll over or click image to zoom in

Tatcha
The Water Cream Oil-Free Pore Minimizing Moisturizer
ITEM 1932920

$68.00
Pay in 4 interest-free payments of $17.00
Klarna.

★★★★☆ **3K reviews** | ♥ **259.6K loves**

exclusive

● **Get It Shipped** You're only $50.00 away from Free Shipping.

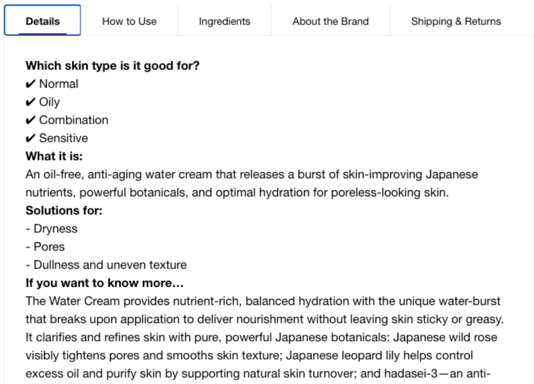
○ **Reserve & Pick Up** Select to see availability at **Stanford**

SIZE: 1.7 oz/ 50 mL

Standard size
1.7 oz/ 50 mL

Mini size
0.34 oz/ 10 mL

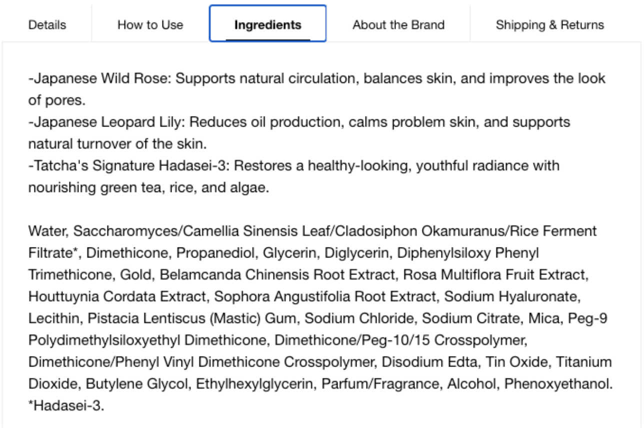
Description



Which skin type is it good for?
✔ Normal
✔ Oily
✔ Combination
✔ Sensitive
What it is:
An oil-free, anti-aging water cream that releases a burst of skin-improving Japanese nutrients, powerful botanicals, and optimal hydration for poreless-looking skin.
Solutions for:
- Dryness
- Pores
- Dullness and uneven texture

If you want to know more...
The Water Cream provides nutrient-rich, balanced hydration with the unique water-burst that breaks upon application to deliver nourishment without leaving skin sticky or greasy. It clarifies and refines skin with pure, powerful Japanese botanicals: Japanese wild rose visibly tightens pores and smooths skin texture; Japanese leopard lily helps control excess oil and purify skin by supporting natural skin turnover; and hadasei-3—an anti-

Ingredients



-Japanese Wild Rose: Supports natural circulation, balances skin, and improves the look of pores.
-Japanese Leopard Lily: Reduces oil production, calms problem skin, and supports natural turnover of the skin.
-Tatcha's Signature Hadasei-3: Restores a healthy-looking, youthful radiance with nourishing green tea, rice, and algae.

Water, Saccharomyces/Camellia Sinensis Leaf/Cladosiphon Okamuranus/Rice Ferment Filtrate*, Dimethicone, Propanediol, Glycerin, Diglycerin, Diphenylsiloxy Phenyl Trimethicone, Gold, Belamcanda Chinensis Root Extract, Rosa Multiflora Fruit Extract, Houttuynia Cordata Extract, Sophora Angustifolia Root Extract, Sodium Hyaluronate, Lecithin, Pistacia Lentiscus (Mastic) Gum, Sodium Chloride, Sodium Citrate, Mica, Peg-9 Polydimethylsiloxyethyl Dimethicone, Dimethicone/Peg-10/15 Crosspolymer, Dimethicone/Phenyl Vinyl Dimethicone Crosspolymer, Disodium Edta, Tin Oxide, Titanium Dioxide, Butylene Glycol, Ethylhexylglycerin, Parfum/Fragrance, Alcohol, Phenoxyethanol. *Hadasei-3.

Future Improvement

The current dataset contains 9k products. If product description and ingredients are considered as different documents, we have 18k documents in total. This might be a bit small for the dataset.

The current dataset contains number of reviews for each product but does not contain review text. The total number of reviews is ~2.7 million. Therefore, we can expand the existing dataset by parsing the review text to conduct better search for relevant product.