Improving Ecommerce Search Experience

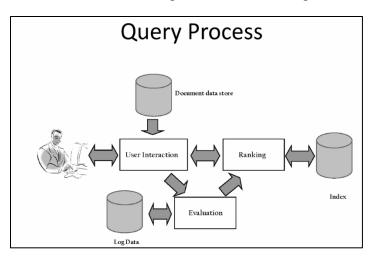
Grania Machado, Yuehua Hu and Ryan Siegel

***** Task and its importance to users

The goal of the task is to explore how search is implemented in a specific ecommerce store environment. We want to understand how the information retrieval (IR) tasks of the application can give relevant results to each user query. We will be ranking and evaluating user queries previously unseen by the search engine. Here, each query returns more than one correct result in the application, hence we need to understand which relevant results are to enhance user experience. The ecommerce store contains a variety of products that the users can explore through an internal company website or mobile application. The design of the store is to allow users to discover products that fit their interests. The ecommerce store accomplishes this by providing a straightforward way for users to lookup a product, comparatively shop for products, and filter for certain prices based on the user's award balance. This award balance is associated with the user's awards that are given to them by the company that employs them. This does not limit the user's interaction with the store. Overall, the goal of the experience is like public facing ecommerce stores, like Amazon.

The way an ecommerce store works is that the user enters a product name to search, and products with matching names are displayed to the user. However, the user may often not know the exact name of the product or may not know the correct spelling of the product. Often users may be looking for a type of product but may be undecided about the exact product of their interest. To start, the relevant result for a search is going to be all products that have matching correct spelling of product within a search query. Brands might have specialized product names depending on the popularity and marketing of their product. In such case, even though the user is looking for that

product, but search will not be able to return it as relevant result since it will not contain any words from the query. Furthermore, since the application will contain correctly spelled product names, a misspelled query from the user should use past search history and similarity measures to find and return relevant results for each query. Additionally, we also give relevant product can suggestions related to the search query. The user can select and apply filters to their query search so that the returned



results are more relevant to the user's query. Filters help focus the results under product type which makes it easier for the IR task to return more relevant results for the search query.

In figure 1, we can see that the user would input query in search box of the application, our search engine would retrieve product results with matching keywords in them, these results are then ranked using evaluation metrics according to their relevancy, and the indexed product list is returned to output screen for users.

***** What do queries look like?

Queries are actual query searches from the period of April 1st to April 25th. They are sourced from users who are leveraging the WorkHuman (WH) to look for products that they might be interested in. Queries would be generated from the search bar, which is a feature that is accessible from anywhere in the store.

We are electing to source query logs from an internal, WH data source. We do not know the user who has written the query or any personal identifiable information about them. These queries are leveraged on an ecommerce search engine which has products from suppliers around the world. Unknown to the user, the product catalog is not as expansive as Amazon, but is still professionally curated and contains products dependent on suppliers to the WH store. There are also particular constraints based on the country for instance. This sort of a priori knowledge can be applied by the user and is something we might have to keep in mind when exploring what queries are made.

In terms of the query's content. The search terms typically are one to four words in length. We will be looking at queries that are about this size. The queries are written in English and the store front is also English. The queries can contain brand names which naturally fall outside the English dictionary. This might cause suboptimal results depending on whether the brand is a product feature that is indexed for the search engine to look for. This will be explored further in what results we see as well as the relevance of said results.

There are currently no query suggestions that might nudge users to 'correct' their query in any way. As for query substitution or query expansion, that is not currently known. In any case, the user will write a query and assume that products relevant to that query will be returned.

A theory we have in relation to brand names is that the user search query, and potential subsequent searches, could help formulate the user's narrative. If the user uses specific brand names associated with a product area, that could be an indication of their subject matter expertise in relation to the product. For instance, if the user were to search golf club, they could retrieve different results than if they searched "sand wedge" or "Titleist golf club", the latter queries indicating that the user has experience with golf.

We will also be looking at atypical queries to explore how the WH search engine handles something more complex, i.e., if a person is looking for a coffee mug, what if they wrote "vacuum coffee mug"? Would the vacuum be ignored? We are looking to understand how dynamic the search engine is and if such irrelevant terms might be discounted or not. We will also look to see what happens if a person types in a query that is not in English (i.e., in Spanish for instance). How would the search engine handle that?

Relevance of results to queries

Depending on the queries that user enters, user would be able to find out the most relevant product that he/she is looking for from the list of results. Users will be able to pick the product that they want. For example, if a user entered the query "Coffee mug" then the results would contain all the coffee mugs that the e-commerce store is currently selling. Those coffee mugs would be distinct sizes, color and other specifications enabling the users to find the most relevant coffee mug that they want to buy. Also, users can add more specific requirements to the queries, things like price range or category filtering. Users can select the price range based on their needs. After their selection on the price range, the returned results would be more relevant than the ones without filtering. Depending on the queries they entered, users can select the categories that they are interested in to get some more relevant results. There is other filtering exist in the ecommerce stores, for example the results from the queries can be viewed as the "price low to high" or "price high to low." This is different than the price range filtering since price range filtering return the selection of products or results based on user's specific range but the "price low to high" or "price high to low" would simply sorting the results by the price regardless of the filtering that users entered.

The number of relevant results for each query would be dependent on the user's need. We are expecting to see there is at least one result showing per query. The number of relevant results would depend on the queries that users entered, if the query contains a general information or a brand name of the product that users wanted to search, then the number of relevant results can be over 100 from showing all the products specific to that brand name. In that case, users can also limit the number of results showing from the website to see the top ranked products based on their selection of limitation numbers. In other cases, if the users entered a specific product name, then the number of results for that query would be limited because user have entered a specific product that they trying to search. Distinct query searches are more likely to have fewer results, but ideally better matches for what the user needs. However, if the query has terms not usually found together like "vacuum coffee mug," we find no results. For that matter, when we did a cursory view of results when utilizing Spanish, the search engine did not make any translation and again no results were retrieved.

Organization of results

Our results will be a product list where the product names or product description contain keywords from query search. However, some results are more relevant for the search query and need to be displayed at the top of search results as they have a higher clickthrough chance by user. In our ecommerce store it is our goal to provide users with products that are of high relevance to the user query to provide an efficient and smooth experience to users. Given the importance of user behavior, it is highly likely that the user will look through the top products more closely as compared to results on next page. Thus, it is important that the ranked results list contain products which have higher chance of selection by user and satisfy user's query search correctly. For this task of efficiently displaying results with higher relevance towards the top of search results, the search results should be organized as a ranked list.

A ranked result list is the process during which indexing evaluates a distinct ranking score for each product item that matches a query, and then sorts the results from most to least relevant (i.e., in

descending score order). We rank product search results by calculating relevance score for evaluation metrics based on a series of ranking factors. The higher the score, the higher the result will be in the result list.

The product scores must be calculated and compared very rapidly to determine the ranked order of products that are in the query result's output. For this purpose, we need to have performance optimization components. Performance optimization involves the design of ranking algorithms and the associated indexes to decrease response time and increase query throughput.

Since the ranked result list emphasizes the quality of the top-ranked documents, rather than the whole product list, it is more appropriate for our search queries in the ecommerce store.

***** Evaluation metrics

We explored the different options based on what we imagine the experience would be from the user perspective. Would it matter that the user sees a relevant product first or last? In the browser experience we expect that users would view a minimum of 9 products because there are a variety of products that could fit the user's needs. The format of results is 4 rows of 3 results across. However, the first page contains approximately 9 full results. In addition, the sorting algorithm for the store, based on our observations, did not necessarily gather the best products at the top. User preferences would also dictate what could be most important and not necessarily the contents of the product as 'likeness to user taste' is a hard concept to define here.

Given the *general* rank of the results is important, i.e., having the most relevant product in slot 200 of 500 would not be conducive to a strong user experience, we narrowed down several different measures. The evaluation metrics to Mean Reciprocal Rank (MRR), Discount Cumulative Gain (DCG) and Mean Average Precision (MAP). We will dive into each of these and understand which of these evaluation metrics would best fit our task.

Mean Reciprocal Rank aims to consider whether a given document is relevant and then the given position of the document. A document closer to the top of the results list will be less penalized for a high rank than if the document is lower in a results list. This is attractive for understanding a query at a granular level if there is a singular item that would perfectly satisfy what a user is looking for. However, given that there could be several relevant products, it is important that there is choice among the selection of products retrieved. MRR fails to account for the number of products retrieved that could be relevant. In other words, if we keyed in on MRR as the evaluation measure, if 10 products were retrieved, we would rather have 5 relevant products in positions 6, 7, 8, 9, 10 than have a relevant product in position 1 while the other products were irrelevant. Not considering the diversity of products is a negative and therefore we should not leverage MRR as the evaluation metric.

Discount Cumulative gain is an effectiveness measure that also puts importance on the position of the relevant document within a search result. DCG and Ideal DCG will penalize a document if it is in the wrong position. Products that are returned from a search query can range in relevance, which underlies DCGs (Discount Cumulative Gain) objective, however, this can be a problem if we optimize towards strength of relevance for assorted products. When searching for products a user's taste can bias results in ways that could not be uncovered within an annotation. In this sense, it is also important to note that there could be significant changes in relevance user to user.

Therefore, ideal DCG and DCG itself might provide results that could be too noisy. The model could be overfitted to the preferences of the annotator and therefore performance could suffer.

Mean Average Precision is the last performance measure to be considered (not including Geometric MAP). MAP considers the position of the product as well as its relevance via precision. An advantage here is that there is a relevant metric that can ingest several relevant documents. The relevance judgement is binary in that a product is or is not relevant and quality of relevance does not matter. This is useful since the retrieved items must match a user's preferences. Deciding the **quality or level** of a product's relevance towards a user's preferences is a complicated task, better left to the user to decide the relevance level than that of the model. This makes the annotation easier as well as we can judge relevance without worrying about a given user's narrative and how closely the product fits that. The second advantage is that a list of many relevant products within the top products list and having a non-relevant product mixed in will not penalize the results as strongly if we utilize Ideal Discount Cumulative Gain.

Therefore, even though DCG provides a range of relevance, important for products, we feel that having a binary classification that is afforded in **Mean Average Precision is critical to offering a list of relevant products to users without overmanaging what we expect them to like.**

***** Modeling approaches for solving this task

We will be using the Pseudo relevance feedback for solving the task by creating a logistic regression model utilizing the words as features within the query, title, brand name, 'hidden tags' generated from the indexing that could take place, and product description to help predict the relevance score. In this development setting, we would be tuning the model on annotated searches to understand what is relevant based on the queries entered. The predictions would then be evaluated from Mean Average Precision since this takes in a binary classification for relevance scores (1 or 0 for relevance by the annotators).

Then during the pre-processing part, we will be conducting pre-processing of the text. It is necessary to spend a large amount of time go into observing the text and operating pre-processing techniques to make it suitable for modeling. In some cases, users might not be aware of the spelling of the words for the product or brand name he/she is looking for, in that case spelling corrections will be taking place. We need to first understand the product name, brand name, product description and English words to form the vocabulary list. Then within a search query, it will look for what is written in the query, if there are no matches at all to a product name, brand name within Levenshtein distance of 1, a spelling correction method will be conducted through the noisy channel model. Then this process will be repeating for the product description and English words if there are no matches to the query. But there are cases where it may be better to accept the misspelling words as the intended words rather than attempt to find the intended word from the vocabulary. Therefore, the corpus will be established including the brand names, product names, English words, and misspellings words to be corrected as well. Techniques from natural language processing will also be used for this part. Things like removing stop-words and option to stem words and use lemmas.

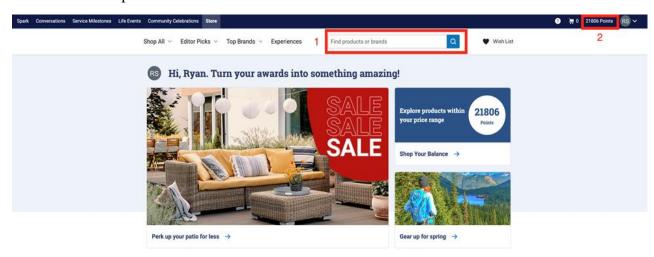
The data collected and annotated would include the user id, time the user entered the query, the product list with name and description and the search query itself. The main part of the data will be search queries since we care more about the information from the queries that what product or brand name that users are interested and intending to search.

Example:

Introduction to what the ecommerce experience is for users

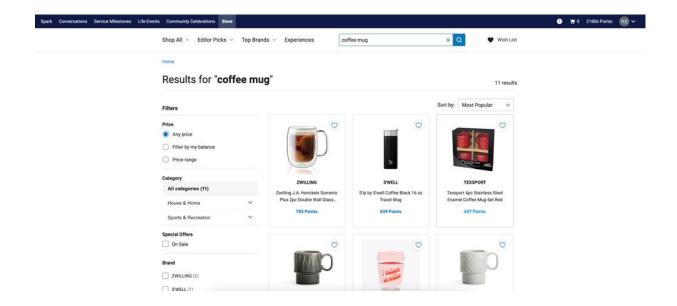


Preview of the top half of the store front.





- 1. Users can enter queries here on the desktop browser. The mobile experience is more compliance
- 2. Users have a limited amount of currency equivalent points (in this example, 20 points = \$1).



Here are the results for a coffee mug. The user can see the ways to sort the products on the left. The initial results are sorted by 'popularity' which is determined by the number of products bought within a country.

* Sample queries, narratives, and relevance judgments

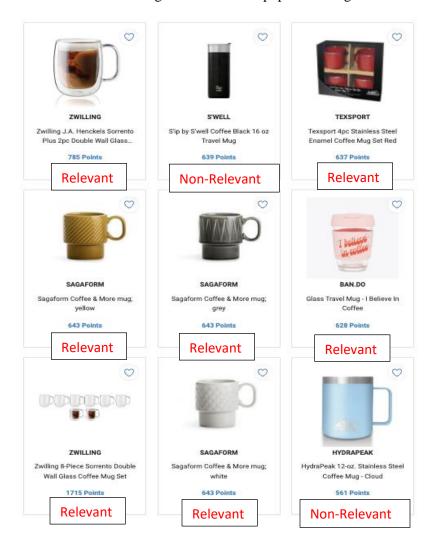
These queries that we enter and the resulting items that come up, without additional sorting, are what we will determine to be relevant. Below is a table of the queries and their average precision. We then found the MAP to be 0.6821. See each query for term searched, narrative and relevancy judgements to determine what was relevant or not relevant for the query.

Query	Average Precision for the query
1. coffee drinking mug	0.826
2. Dog Crate	0.703
3. iRobot	0.945
4. apartment cleaning vacuum	0.97
5. cappachino	no results
6. magnetic airpod	0.872
7. pandora dangle charm	1
8. ergonomic gaming chair	1
9. 14k gold earrings	0.874
10. Bed Bath and Beyond	no results
11. outdoor chairs	0.787
12. ray twin skillet	1
13. Solo Stove	1
14. Bedroom furniture	1
15. electric bike	0.771
16. kids golf clubs	no results
17. weber smoker	0.5
18. metallic chain saw	0.297
19. tree swings	0
20. quick ice maker	0.38
21. otterbox iphone 11 case	1
22. lowes	0
23. cordless lawn	1
24. reflections by design earrings green	1
25. Plush Lounge Chair	0.442
26. Poolside Picnic table	0.73
27. ladies gold watch	1
28. kendra scott necklace	1
29. electric heat lamp	0.833
30. Cuisinart food processor	0.533

Mean Average Precision 0.6821 1. Query: "coffee drinking mug"

Narrative: The user is probably either looking for a single coffee mug or a set of coffee mugs to purchase Relevance judgments:

- a. The user will find results showing coffee mugs within their preferred price range and their favorite brands to be relevant.
- b. Also, the design of coffee mugs will play an important part in user relevance.
- c. Depending on the specific need of the user, the user may find certain results like travel coffee mugs to be non-relevant.
- d. The material of the coffee mug is another important criterion in judging relevance since insulated materials on coffee mugs would be more popular amongst users.



Relevancy (top 9):

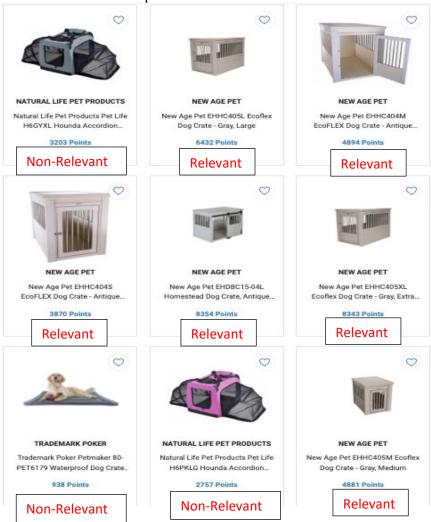
1. R 2. N 3. R 4. R 5. R 6. R 7. R 8. R 9. N

2. Query: "Dog Crate"

Narrative: The user is looking for an indoor place for their dog, although any animal could use this to provide confinement for reasons of safety, security for the pet, house training, prevention of destructive behavior, and/or travel.

Relevance judgments:

- a. The user will find results showing dog crates over other pet carriers to be relevant.
- b. Results showing dog crates from popular pet brands are most relevant.
- c. Prices proportional to the size and material of the crate will be a factor in the relevance decision.
- d. If the dimensions of the crate are suitable for the size of the pet, they have and accommodate the need and comfort of the pet then the result would be deemed relevant.



Relevancy (top 9):

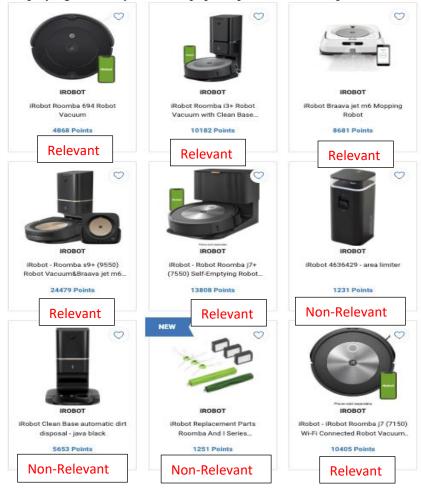
1. N 2. R 3. R 4. R 5. R 6. R 7. N 8. N 9. R

3. Query: "iRobot"

Narrative: The user is looking for an automated way to clean their living space or small office via a vacuum cleaner. The brand iRobot is sought out because this is a well-known brand that has robotic vacuum cleaners. Here, the user is browsing for products by iRobot rather than looking for a specific product from the brand.

Relevancy Judgements:

- a. Since the user is primarily looking for products from iRobot, relevant results are products from iRobot brand.
- b. Considering iRobot is a popular vacuum cleaner brand, the most recently launched efficient vacuum cleaners would be of high relevance to the user search query.
- c. iRobot product with multifunctionality features and proportional price that matches their launch in the market and usability would be a relevance judgment feature.
- d. Battery life of the vacuum cleaner is an important feature in relevance since a cleaner robot with higher battery life can give better cleaning service.
- e. Displaying the newly launched popular products at the top of the search list would be more relevant.



Relevancy (top 9):

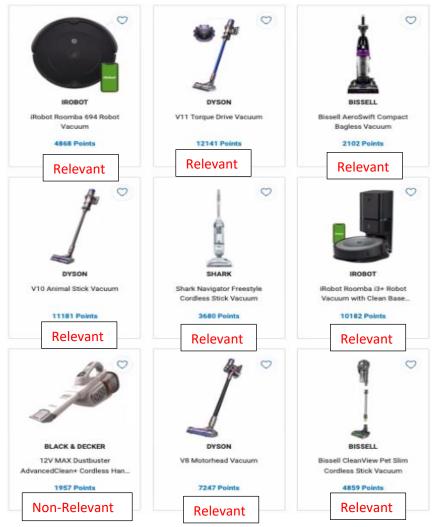
1.R 2. R 3. R 4. R 5. R 6. N 7. N 8. N 9. R

4. Query: "apartment cleaning vacuum"

Narrative: The user is looking for an electronic cleaning product for their living space or small office via a vacuum cleaner. The user is not looking for a specific brand product. Any good vacuum cleaner that can accomplish the task of cleaning with ease and efficiency within a reasonable time would be relevant.

Relevancy Judgements:

- a. While the ease of robot vacuum cleaners is a desirable feature amongst many, many users prefer handheld vacuum cleaners for their living space since there are high-power ones as well as people can use vacuum cleaners on their scheduled time.
- b. Cordless vacuum cleaners offer ease of use and thus a higher price on these would still make them relevant.
- c. Power consumption and battery life for cordless vacuum is an important relevant feature.
- d. Dust blowers and car vacuum cleaners would not be relevant in this case.



Relevancy (top 9):

1. R 2. R 3. R 4. R 5. R 6. R 7. N 8. R 9. R

5. Query: "cappachino"

Narrative: The user is looking for a personal-sized cappuccino machine. Here, the user misspelled "cappachino" by mistake. Ideally, this is an appliance that can be placed on a kitchen counter and can be used to make several types of coffee.

Relevancy Judgements:

- a. The variety of drinks dispensable by the machine and ease of use will be the primary judging criteria for relevance.
- b. Quality of drinks is an extremely desirable feature to avid coffee drinking users and hence the most sought relevant feature.
- c. Convenient design, ease of cleaning the machine, and compact size are relevant features for this search query.
- d. Customizable settings and price flexibility is also a deciding factor for relevance.
- e. The cappuccino maker should be for personal-sized cups, not commercial grade.

Your search did not match any of our products.

- · Try a more general search which will bring you to more related items.
- · Check if there are any spelling mistakes.
- · Limit your search to one or two words.
- · Try to clear or remove any applied filters.
- · Try applying different variations of filters.

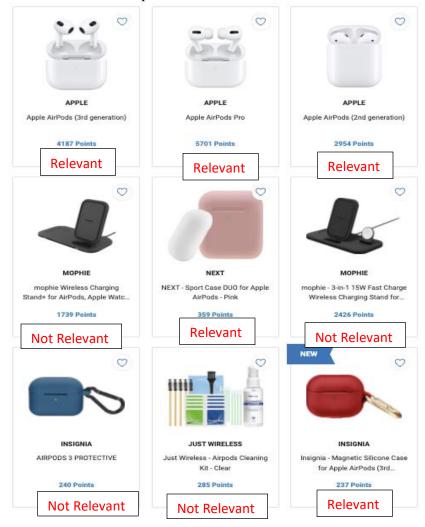
Unfortunately, the current system does not account for spelling correction and hence returned no results for the query.

6. Query: "magnetic airpod"

Narrative: The user is looking for apple wireless earbuds that have magnetic holding cases. AirPods are the signature name of the Apple wireless earbuds product. The user is looking to purchase AirPods that are compatible with their apple phone so compatibility and brand matter.

Relevancy Judgements:

- a. Since compatibility of electronic accessories is important for use, it is extremely important that the product results can pair successfully with current user devices like their apple phone.
- b. The generation of the Apple AirPods is compatible with all phones hence all generations results are relevant; however, the latest and most popular models should be displayed on top.
- c. Accessories of the air pod are not relevant, i.e., protective AirPod cases would be irrelevant
- d. Other wireless earbud products would not be relevant as well



Relevancy (top 9):

1. R 2. R 3. R 4. N 5. R 6. N 7. N 8. N 9. R

7. Query: "pandora dangle charm"

Narrative: Pandora is a specific brand name and therefore the user trusts the brand and is browsing for their specific product. A dangle charm is a specific type of bracelet accessory that is used by people who have pandora bracelets. Together, the product is specific to pandora, but there is a collection of different charms that a person can attach to their Pandora bracelet.

Relevancy Judgements:

- a. Results with dangling charm products with superior design and material are relevant.
- b. Although the user is looking for pandora dangle charms, dangling charms from other brands with exceptional design would also be relevant to the user.
- c. Results with products from varied materials like silver, gold, sterling silver, etc. are relevant.
- d. Pandora music is not relevant.
- e. Small jewelry items like earrings or pendants are not relevant.

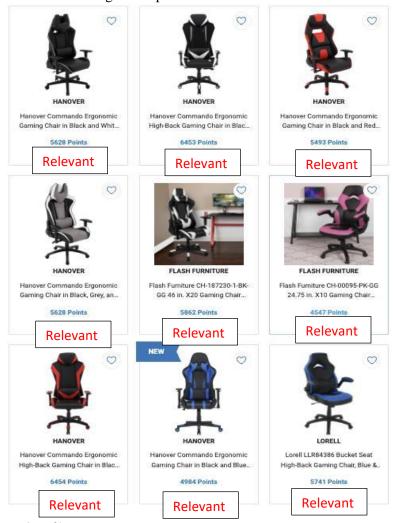


8. Query: "ergonomic gaming chair"

Narrative: The user is looking for a gaming chair to pair with their computer gaming setup. The preference is that the gaming chairs have extra comfort and are suited for extended periods of sitting. An ergonomic chair should have a lumbar adjustment (both height and depth) so each user can get the proper fit to support the inward curve of the lower back.

Relevancy Judgements:

- a. Office chairs would not be suitable since gaming chairs are constructed very differently with better lumbar adjustment.
- b. Good construction of ergonomic support and sturdy material would be of high relevance to the user.
- c. Since the user is looking for a specially built gaming chair, a pricey chair with good specifications would still be relevant.
- d. Lumbar back support from the chair is relevant.
- e. Chairs not meeting these specifications are not relevant.



Relevancy (top 9):

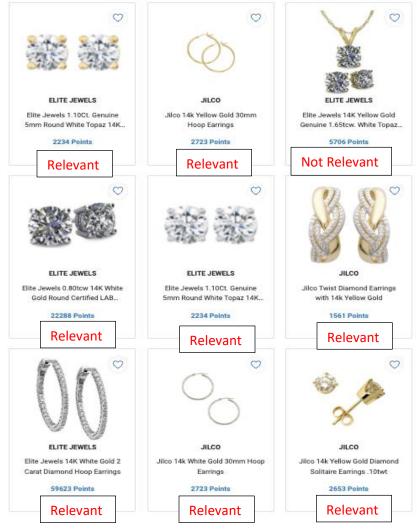
1. R 2. R 3. R 4. R 5. R 6. R 7. R 8. R 9. R

9. Query: "14k gold earrings"

Narrative: The user is looking for earrings of a certain gold amount. 14k stands for 14 karats and designates the amount of gold found within the given jewelry item.

Relevancy Judgements:

- a. The amount of gold in the jewelry is relevant. Therefore, jewelry products with a higher or lower percentage of gold are not what the user would deem relevant.
- b. The user is looking for a specific kind of jewelry item earrings, other items like bracelets or necklaces would be irrelevant.
- c. The brand is not a differentiator.
- d. The size of the jewelry is not important.
- e. Since gold is a pricey metal alloy, and the user is looking for 14k instead of 24k earrings, which means they are looking for a beautiful gold earrings pair that suits their spending budget.



Relevancy (top 9):

1. R 2. R 3. N 4. R 5. R 6. R 7. R 8. R 9. R

10. Query: "bed bath and beyond"

Narrative: The user is looking for retail products for furnishing and decorating their home. Bed Bath and Beyond is a specific brand that has home goods.

Relevancy Judgements:

- a. User is not specifically looking for a bath or bed, hence these items would not be relevant.
- b. Resultant products can be accessories for home décor, bathroom, bedroom, or just items needed in a household.
- c. Home good items would be relevant to the user.
- d. Quality and variety of products are relevant since Bed Bath and Beyond specializes in a wide variety of products at a low-key price range.

Your search did not match any of our products.

- · Try a more general search which will bring you to more related items.
- · Check if there are any spelling mistakes.
- · Limit your search to one or two words.
- · Try to clear or remove any applied filters.
- · Try applying different variations of filters.

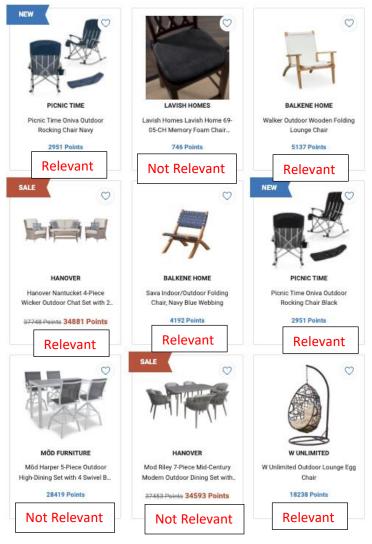
Unfortunately, the current system does not have products from this retail brand and hence returned no results for the query.

11. Query: "outdoor chairs"

Narrative: The user is looking for outdoor chairs. Patio chairs would be another colloquial term.

Relevancy Judgements:

- a. As the user is looking for outdoor chairs, they must be suitable for outside weather like waterproof, easy to clean and maintain to be relevant to users.
- b. Outdoor chairs would be used for relaxing, hence arched back support would make them relevant.
- c. Kitchen chairs are not relevant as they cannot be used outdoors.
- d. If the chair is labeled a patio chair, picnic chair, or deck chair, it is relevant.
- e. Sets of chairs are relevant as well as individual chairs.



Relevancy (top 9):

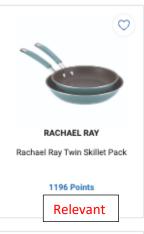
1. R 2. N 3. R 4. R 5. R 6. R 7. N 8. N 9. R

12. Query: "ray twin skillet"

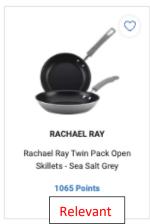
Narrative: The user is looking for a type of skillet specific to a brand. In this case, the brand is likely to be Rachel Ray. Twin skillets of any type would be relevant. Twin seems to represent a 'two pack' of two different-sized pans.

Relevancy Judgements:

- a. Since the user is looking for Rachel Ray brand skillets, skillets from other brands are not relevant.
- b. Any double pack of Rachel Ray skillet is relevant.
- c. There is higher relevance if returned result products are Rachel Ray brand skillets.
- d. Other types of pans such as frying pans or sauté pans are not relevant.
- e. The quality of coating, handle and edges of the pan also decide their relevance in result list.









Relevancy:

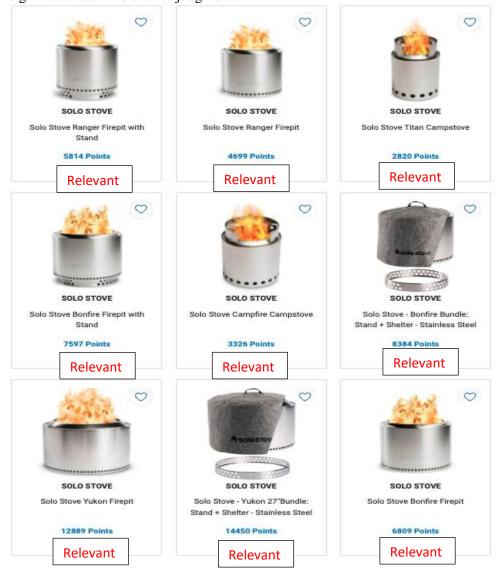
1. R 2. R 3. R 4. R

13. Query: "Solo Stove"

Narrative: The user is looking for a firepit to improve the camping experience. Here Solo Stove turns out to be one of the brand names for this type of product.

Relevancy Judgements:

- a. The user is looking specifically from the brand Solo Stove. Relevant results are products from the Solo Stove brand.
- b. Since solo stoves have a wide price range, the desired price would be relevant.
- c. The size of the firepit is different, therefore depending on the size of the product users are interested in, the size would be important.
- d. Ease of starting a fire and specifications for maintaining a steady flame would be relevant features.
- e. As there is a certain risk involved around fire, the safety features of each product would play a significant role in relevance judgment.



Relevancy (top 9):

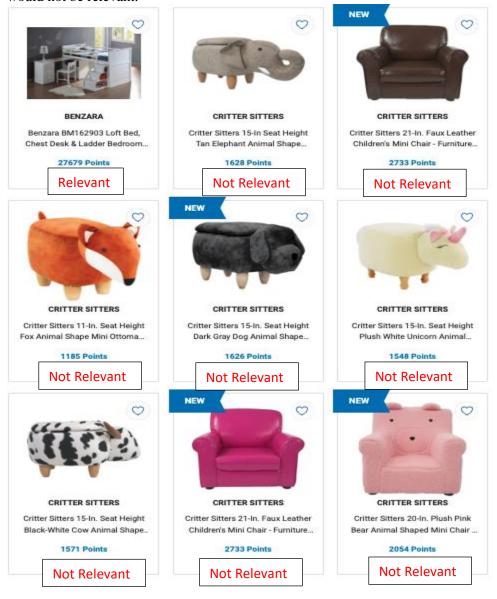
1. R 2. R 3. R 4. R 5. R 6. R 7. R 8. R 9. R

14. Query: "Bedroom furniture"

Narrative: User is looking for furniture in a bedroom which usually includes bed, wardrobes, dressers, chests, nightstands, armoires, vanities, and mirrors

Relevancy Judgements:

- a. User is not looking for specific type of product of bedroom furniture. So, any results of furniture that could be placed in bedroom would be relevant.
- b. Price is important for users to consider as well since the quality of the bedroom furniture would be based on the price sometime.
- c. Based on the user's preference of their bedroom decoration style, the design of the bedroom furniture would also play a key role in judging the relevance.
- d. Other types of furniture such as kids' furniture, outdoor furniture, patio furniture, picnic furniture would not be relevant.



Relevancy (top 9):

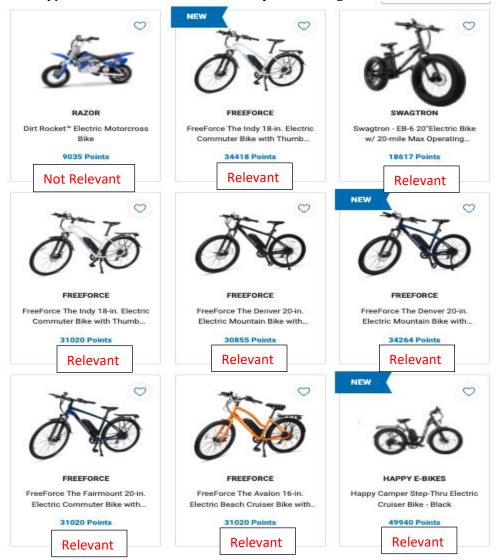
1. R 2. N 3. N 4. N 5. N 6. N 7. N 8. N 9. N

15. Query: "electric bike"

Narrative: The user is looking for an electric bicycle which is an electric chargeable bicycle with an integrated electric motor used to assist propulsion.

Relevancy Judgements:

- a. The user is not looking for a specific brand of electric bike. So, any brand products with electric motor will be relevant.
- b. Since fancy electric bikes will have more features, they will be expected to have higher prices but will still be relevant.
- c. The material and the size of the bike will be a criterion in judging relevance.
- d. Battery life supported by the electric bikes will be an important feature in relevance since it is extremely important for longer ride times.
- e. Other types of bikes, like dirt-bikes, motorcycles, although electric, are not relevant.



Relevancy (top 9):

1. N 2. R 3. R 4. R 5. R 6. R 7. R 8. R 9. R

16. Query: "kids golf clubs"

Narrative: The user is looking for a child's set of golf clubs. Typically, these are cheaper, have less clubs, and are meant for training. The expectations would be that the golf clubs would be suitable for 5-13 years old children. They would be smaller.

Relevancy Judgements:

- a. The size of the clubs is the most important feature for relevance since it should be suitable for use for kids.
- b. Since kids' sizes are variable according to age groups, a variety of products in results would be a criterion in relevance judgment.
- c. The number of clubs in the set is important since a complete set is required for golfing.
- d. The clubs are for a starting kit for a kid, so the material being cheap is not an issue.
- e. Specialized clubs in golf, like a putter, sand wedge, and driver would be desirable along with a 6, 7, 8, 9 iron.

Your search did not match any of our products.

- · Try a more general search which will bring you to more related items.
- · Check if there are any spelling mistakes.
- · Limit your search to one or two words.
- · Try to clear or remove any applied filters.
- · Try applying different variations of filters.

Unfortunately, the current system does not have products matching this query search and hence returned no results for the query.

17. Query: "weber smoker"

Narrative: User is looking for a smoker for a Barbecue supplement. The user is specifically looking for a Weber brand for this type of product.

Relevancy Judgements:

- a. Size of the smoker will be a factor for judging relevance based on user's need and their space availability.
- b. The fuel efficiency of this product will play a key role in determining relevance.
- c. Boxes of smokers might not be relevant but the smoker itself would be relevant.
- d. Smokers made by other brands would not be relevant results for this query.





Relevancy:

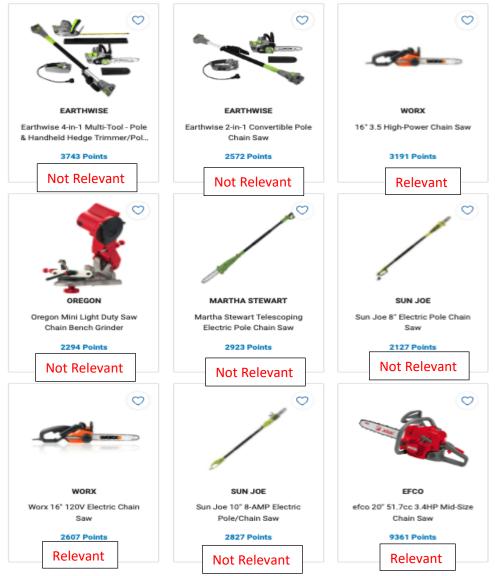
1. N 2. R

18. Query: "metallic chain saw"

Narrative: Chain saw is portable gasoline, electric or battery-powered saw that cuts with a set of teeth attached to a rotating chain driven along a guide bar. The user is looking for chain saws for activities such as tree felling, pruning, cutting firebreaks in wildland fire suppression, and harvesting of firewood.

Relevancy Judgements:

- a. Distinct types of chain saw from the product list, such as gasoline-powered and electric-powered are both relevant if they have metallic teeth.
- b. A chain saw combined with another multi-tool might not be relevant.
- c. As the size of a chain saw being able to meet the purpose of the user is imperative, it will be a factor in judging relevance.
- d. Products showing chain or saw are not relevant since the product in query search is chain saw.



Relevancy (top 9):

1. N 2. N 3. R 4. N 5. N 6. N 7. R 8. N 9. R

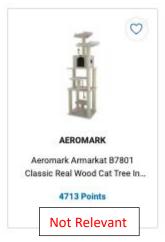
19. Query: "tree swings"

Narrative: The user is looking for tree swings with chains or rope, and hooks that can improve their outdoor experience. It is composed of a single rope or chain attached to a high tree branch, along with a seat, which is typically a wooden plank or tire. The hammock could be an alternative name for this query.

Relevancy Judgements:

- a. As a tree swing is suspended in the air from a tree branch, its sturdiness plays a key role in relevance.
- b. The safety features and materials play the most important part in relevance here.
- c. Plastic trees, décor trees, and cat play trees are not relevant.
- d. Quality and comfort of materials will be an important feature of judging relevance.





Relevancy:

1. N 2. N

20. Query: "quick ice maker"

Narrative: The user is looking for an icemaker, ice generator or ice machine may refer to a consumer device for making ice; a stand-alone appliance for making ice. Ice machine usually refers to the stand-alone appliance.

Relevancy Judgements:

- a. Appliances that can quickly make and dispense ice are of relevance here.
- b. Quantity of ice made; quality of ice produced will also determine the relevance of the product.
- c. Ease of removing ice, storing it, and accessibility to cleaning the appliance are other relevant features
- d. Ice cream maker and coffee maker are not relevant.
- e. The design, custom features and specifications of the ice maker machine will play an important part in user relevance.



Relevancy (top 9):

1. N 2. N 3. N 4. R 5. R 6. N 7. R 8. N 9. R

21. Query: "otterbox iphone 11 case"

Narrative: The user is looking for a specific brand Otter Box which makes mobile accessories, phone cases, and outdoor gear for iPhones. In this query, the user is looking for iPhone 11 phone case accessory specifically.

Relevancy Judgements:

- a. Any types of phone cases for iPhone 11 from Otter Box would be highly relevant.
- b. Since the user is specifically looking for Otter Box products, phone cases from other brands will not be relevant as the user is looking for certain features or quality offered by Otter Box.
- c. Results displaying iPhone 11 will not be relevant.
- d. The design, material, and quality of the phone accessory for iPhone 11 are relevant.





Relevancy:

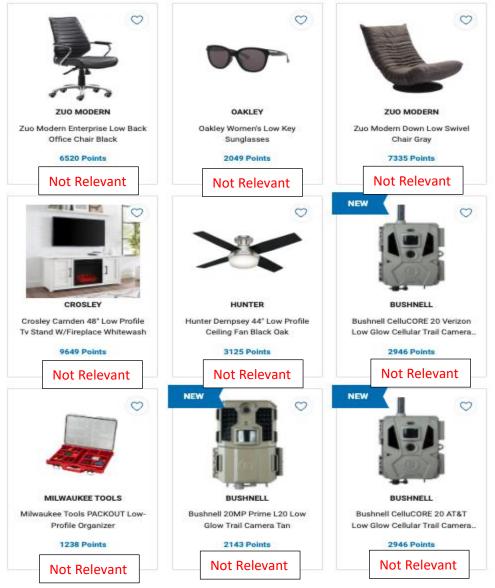
1. R 2. R

22. Query: "lowes"

Narrative: The user is looking for items that are power tools sold at Lowes retail store. Lowes is a brand that is associated with selling home improvement tools and appliances. The items should be oriented to DIY projects. Power tool kits would be the most desirable.

Relevancy Judgements:

- a. Power tools such as automatic hand drills would be relevant.
- b. Items that are associated with brands that could be found within Lowes would be relevant.
- c. Fashion items like jewelry, sunglasses, clothes, and other such products would not fall into the category of home improvement.
- d. Raw materials, like wood or fencing would be relevant.
- e. Products, although from the home improvement category but not from Lowes, are not relevant.



Relevancy (top 9):

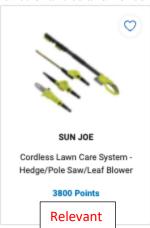
1. N 2. N 3. N 4. N 5. N 6. N 7. N 8. N 9. N

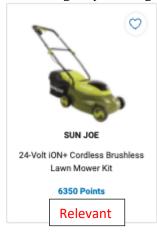
23. Query: "cordless lawn"

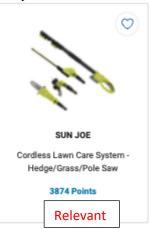
Narrative: The user is not looking for a specific gardening tool. Thus, they are looking for appliances that can do the chopping, trimming, or mowing lawn grass to ensure that the grass is well-groomed and grows at an even, appropriate height. They are looking for tools to maintain their lawn.

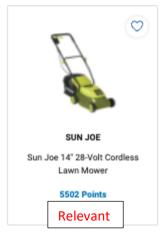
Relevancy Judgements:

- a. The user is looking for a hassle-free way of maintaining their lawn by looking for cordless electronic appliances, hence appliances that meet both these criteria are relevant.
- b. Lawn care appliances that offer multiple functionalities are relevant since the user is not looking for a specific product of lawn care system.
- c. Cordless lawn mower will be most relevant to the user query.
- d. Power consumption and battery life are important relevant judgment features.
- e. Some products come with advanced features and modern technology which offer higher functionalities and hence are on the higher price range, thus price is a relevance feature too.









Relevancy:

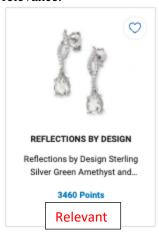
1. R 2. R 3. R 4. R

24. Query: "reflections by design earrings green"

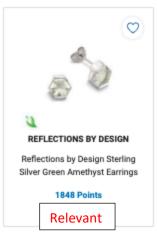
Narrative: The user is looking for a specific brand name Reflections by Design and the product earrings in green color embellishments. Reflections by Design is a brand known to the user, and the user likes its products.

Relevancy Judgements:

- a. Since the user is looking for a jewelry product, resultant products from the specific brand are highly relevant and preferred by the user.
- b. The user is looking for earrings with green embellishment, hence, products highlighting these features are of user relevance.
- c. Jewelry products with a color other than a shade of green are not relevant.
- d. The design and quality of the earrings will be important for judging relevance.
- e. As studded high-end jewelry items are costlier, price will be another criterion for user to consider relevance.









Relevancy:

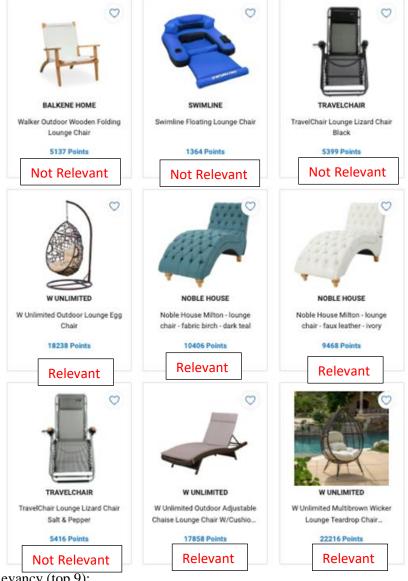
1. R 2. R 3. R 4. N

25. Query: "Plush Lounge Chair"

Narrative: Lounge chairs are typically found within the bedroom or living room. A chair should be comfortable for sitting in for an extended period. The chair should be something in which it is possible to fall asleep in.

Relevancy Judgments:

- a. As the user is looking for a plush lounge chair, the material of the cushion on the chair is important. It should be comfortable and soft.
- The suitability of the chair for use indoors and to add to the furniture collection makes it a relevant
- The length of the chair should be accommodating for an average person to lie down comfortably, this decides the relevance of the product.
- Deck chairs, outdoor chairs, poolside lounge chairs, picnic lounge chairs are not relevant.



Relevancy (top 9):

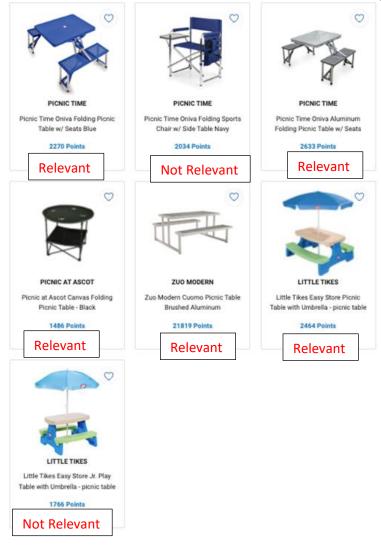
2. N 1. N 3. N 4. R 5. R 6. R 7. N 8. R 9. R

26. Query: "Poolside Picnic table"

Narrative: The picnic table should be able to seat adults comfortably. The idea is to get something in which a meal can be eaten at the table outdoors hence it should be easy to carry, dismantle, and clean. The picnic table can come with an umbrella, shade, and chairs.

Relevancy Judgments:

- a. The poolside picnic table should be foldable or easily dismantled to be taken out and set up.
- b. If the table is a set that has chairs with it, then it would have higher user relevance.
- c. The material and quality of the tables holds relevance as it should be strong enough to hold food items, drinks, and bags.
- d. Accessories like cup holders or umbrellas make the product more interesting to buy.
- e. If the dimensions of the table are more suitable to the user needs, then it should be deemed relevant.



Relevancy:

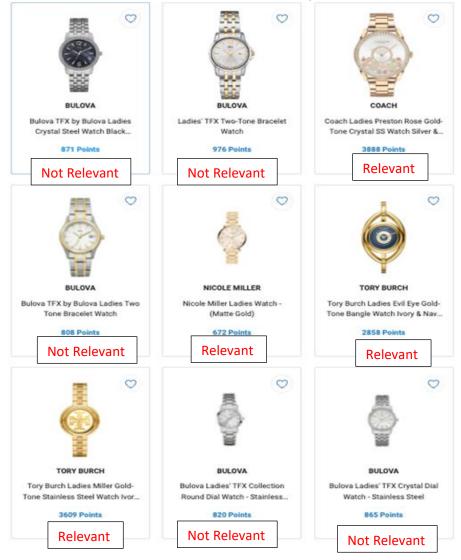
1. R 2. N 3. R 4. R 5. R 6. R 7. N

27. Query: "ladies gold watch"

Narrative: The watch will be a more classically styled, feminine watch made of gold. Shiny finish, high-quality gold, and brand names are important. The watch should be more like a bracelet or jewelry and look regal rather than having practical use outside telling time.

Relevancy Judgements:

- a. High-quality gold watches with impeccable design and style are most relevant.
- b. Diamonds on the watch would be excessive and therefore undesirable.
- c. Gold watches are costly and hence such an investment should show top, popular brand products at the top of the list to be highly relevant.
- d. Since it is a fancy gold watch, it is desired to be analog.
- e. Digital watches would not be relevant.
- f. Watches made from other materials like rubber, stainless steel would not be relevant.



Relevancy (top 9):

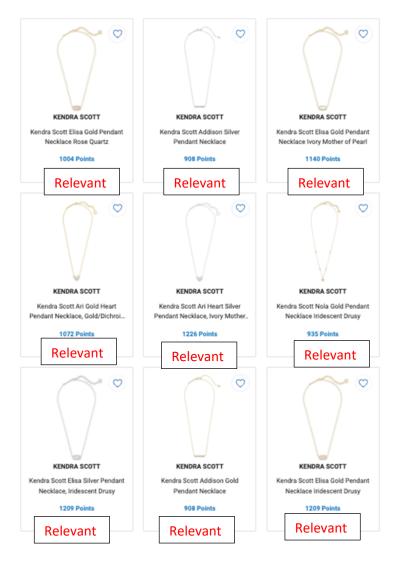
1. N 2. N 3. R 4. N 5. R 6. R 7. R 8. N 9. N

28. Query: "kendra scott necklace"

Narrative: The user is looking for a specific brand of necklace. Karat, or gold content, of the necklace, is not relevant. Kendra Scott brand is a style that is desirable.

Relevancy Judgements:

- a. Necklaces associated with Kendra Scott brand and have excellent design and style are highly relevant.
- b. The material of the necklace is not important.
- c. Necklaces that are not from Kendra Scott brand would not be relevant.
- d. Fashion jewelry that is gold-plated or silver-plated is relevant.



Relevancy (top 9):

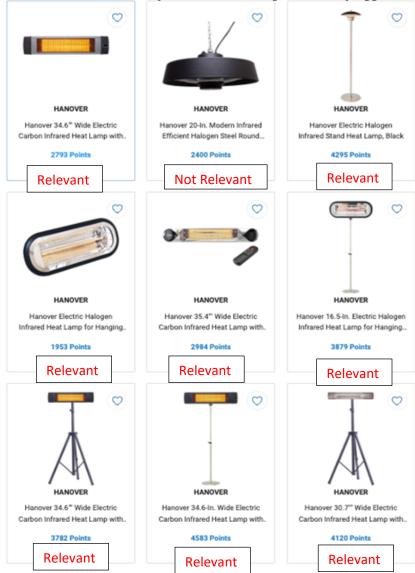
1. R 2. R 3. R 4. R 5. R 6. R 7. R 8. R 9. R

29. Query: "electric heat lamp"

Narrative: The heat lamp would be for outdoor use. It should be a singular unit that can stand on its own. The height of the heat lamp is important so it can warm people underneath it or surrounding it.

Relevancy Judgements:

- a. The heat lamp needs to be standing higher than the average adult or at waist level to be able to provide heat effectively, hence the height of the lamp is a relevant feature.
- b. The heat lamp should be able to warm multiple people at once, hence its effectivity is a major user relevance judgement factor.
- c. Power consumption and features like heat adjustment make significant relevance features.
- d. It should be suited for outdoor use, therefore an indoor heat lamp or one for pets would not be relevant.
- e. If it is corded, the lamp should have a long cord to be plugged in.



Relevancy (top 9):

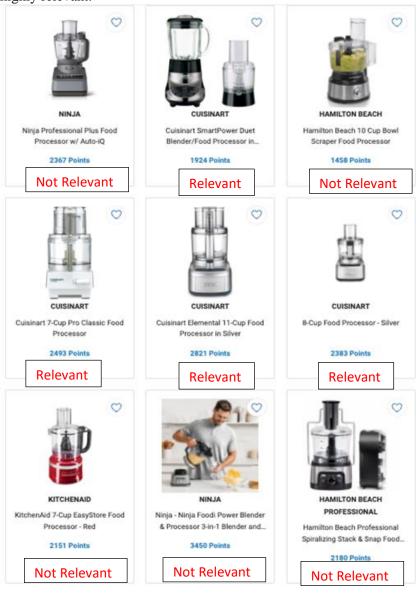
1. R 2. N 3. R 4. R 5. R 6. R 7. R 8. R 9. R

30. Query: "Cuisinart food processor"

Narrative: The user is looking for a Cuisinart food processor to sit on a kitchen counter. It is preferable that it can handle frozen foods and vegetables. It will need to be durable. It will need to have the high capacity to be able to break up a few vegetables or fruit.

Relevancy Judgements:

- a. Food processors of Cuisinart brand that are popular and provide multi-feature selection are relevant.
- b. A food processor that cannot handle more than a small serving of food material would be irrelevant.
- c. Simple clean-up is also desirable, having a dishwasher safe container would be best.
- d. If the food processor has different mode to generate different consistency of processed food, it is highly relevant.



Relevancy (top 9):

1. N 2. R 3. N 4. R 5. R 6. N 7. N 8. N 9. N