Framework for Assessing Relevancy

Executive Summary

The biggest product discovery challenge that most online retailers face is ensuring their visitors see the most relevant search results for a specific search query on their website . Broadly, three key factors play a role in determining relevancy: (i) Query Intent – interpreting the initial search query, (ii) User Intent – understanding the user's purchase preferences, and (iii) Catalog Enrichment – ensuring the product catalog contains relevant product information. In this article, we will focus on deciphering Query Intent.

I. QUERY INTENT

QUERY INTENT CAN BE CATEGORIZED INTO TWO GROUPS

I. Basic Analysis

These are fundamental steps taken to ensure the query is processed and interpreted correctly

Ex: Synonym Auto-Tagging, Stemming, Stop Words

II. Advanced Analysis

For more complex, multi-word, long tail queries, the following advanced techniques ensure higher relevancy

a. *Phrase Identification*
b. *Feature Extraction*
c. *Dimension Handling*
d. *Specification Handling*
e. *Relational Queries*
f. *Diversity Handling*
g. *Query Clustering*

HOW DOES RELEVANCY WORK?

In order to illustrate how to assess relevancy, let's use the following search query as an example - "The Red Italian Leather L-Shaped Sofa". As mentioned above Relevancy can be assessed using the following methods:

Basic Analysis

Initially, during the query pre-processing phase, your search engine should perform the following sanity checks

Synonyms Auto-Tagging

Does the product catalog contain terms that are synonyms and not exact matches when compared to the initial query?

Ex: "Rouge" and "Red" are synonyms. So Italian Leather Sofas that have been tagged Rouge instead of Red should surface in the search results.

Stemming

The search engine understand that words with the same root should be treated equally.

Ex: "Shaped" and "Shape" are the same when processing the query prior to retrieving search results.

Stop Words

Eliminates any unnecessary prepositions or filler words from the query

Ex: Removes "The" from "The Red Italian Leather L-Shaped Sofa" above

Advanced Analysis : For more complex queries, your search engine should leverage the following advanced techniques

Phrase Identification

Segments the original query into human-readable phrases.

Ex: Combines "Italian" and "Leather" as "Italian Leather"

Feature Extraction

Categorizes the query into appropriate attributes.

Ex: Red→Color, Italian Leather→Material, Sofa→ProductType

Dimension Handing

Identifies and interprets product dimensions in the search query correctly.

Ex: the following queries should be treated the same - "42 TV", "42 in TV" and "42 inch TV" since the search engine should understand that 42', 42in and 42 inches are dimensions of the product type - TV

Specification Handling

For a given query (ex. "iPhone 6s charger"), correctly identifies that '6s' is a specification of product brand 'iPhone' and the product type that the user is searching for is 'charger'.

Relational Queries

For theme based searches like "Dresses for Parties", the ability to assess that "Dress" is the primary product type and that "Parties" is related to "Dress" and is a secondary attribute is important.

Diversity Handling

In order to prevent 'starvation', a term used for describing new arrivals in the product catalog that never appear in search results due to lack of historical performance data, diversity handling ensures that a certain percent of the search results are reserved for surfacing new arrivals.

Query Clustering

For long tail queries with very little search hits or historical data, mapping those queries to head queries (popular queries) with more performance data to ensure better Recall drastically reduces null results.

Ex: "Red Italian Leather L-Shaped Sofa" can be mapped to "Italian Leather Sofa"

II. USER INTENT

After your search engine has successfully captured query intent and identified the list of products most relevant to that query, it is equally important to determine the rank order in which they should appear on the search results page. Ranking should be personalized based on each shopper's individual browsing behavior. Here are techniques to capture user intent accurately:

Demand Forecasting

By analyzing historical purchase data of all shoppers on your website, you can assess what products are most popular for a particular category. These insights can then be used to predict future demand by boosting products that have a higher likelihood of being purchased towards the top of the search results page for new shoppers.

If "Weston" brand Italian Leather sofa purchases have been significant over the past 30 days, then your search engine should give a higher weightage to "Weston" brand couches when users are searching for "Italian Leather sofas".

Understanding User Behavior

Capturing and analyzing shopper click stream data is another valuable source for determining user preference. Some of these signals include product CTR, Add-to-Cart, Price and Purchase data. Utilizing these signals to give more visibility to products based on past shopping behavior can really help increase search conversion

Personalization

62% of consumers feel that online shopping is more personal than in-store and 35% believe online personalization is the most important reason for buying online. Keeping these in mind, its critical to offer a personalized shopping experience to your users. Each shopper will have a unique browse and purchase pattern. Ensuring that they see products based on their personal preference will positively impact conversion.

Ex: if two shoppers search for "Polo Shirts", and one has a brand affinity for "Ralph Lauren" and the other for "Tommy Hilfiger", then they should see different search results where their respective brand preferences are given higher weightage.

III. ENRICHMENT

Often, queries provided by users are unstructured and incomplete. An incomplete query hinders search relevance performance since it does not contain all the product related information necessary to determine intent. In such situations, Enrichment enhances the query by adding critical details that help with information retrieval. Following are some techniques that your search engine should utilize to enhance relevancy:

Automated Thesaurus

Leverages machine learning to automatically identify similar words in a document based on occurrence patterns, frequency, etc. These words can then be appended to the query to increase recall.

Ex: The words "Sofa" and "Couch" in the example above are similar and would be automatically included when retrieving results.

Analyzing Query Logs

Query logs are maintained by each search engine in order to analyze the behavior of shoppers while interacting on a website. These query logs can sift through the browsing patterns of shoppers, identify user preference, and add corresponding terms to the query to increase relevancy.

Ex: If a shopper enters "Red Sofa", the search engine would know that the user is interested in sofas where "Italian Leather" is the material type based on their previous search queries and append it to the query - "Red Sofa Italian Leather"

Wikipedia & External Data Sources

There are several web-based data sources like Wikipedia, ConceptNet and WordNet that contain vast amounts of information that could be utilized to fetch related terms for a specific search query. Crawling these web-based lexical corpuses, structuring the data and identifying semantic relevance can really boost search performance.

IV. GLOBAL SHOPPING TRENDS

In order to truly build a world class product discovery experience for customers, it's imperative for a search engine to consider global trends and analyze data beyond just your website. Following are some critical factors that could significantly impact relevancy

Seasonality

Every calendar year there are certain holidays like Thanksgiving, Christmas, Valentine's Day, etc where overall web traffic and demand for certain products substantially increases.

Ex: During Valentine's Day shoppers tend to prefer purchasing red color items and demand for gift cards surges around Christmas. You can increase search conversion by automatically boosting these high demand products during these periods.

Recent Trends

Often, there are certain topics or products that suddenly start trending on social media resulting in a significant surge in demand. If your catalog has products that are relevant, your search engine should be sophisticated enough to understand such trends and give these products a higher weightage when surfacing search results.

Ex: If a certain athlete or sports team starts playing really well, typically jerseys and products associated with that player are in high demand and should surface higher in search results.

Unexpected Events

Sometimes there are certain unforeseen events that can lead to a sudden change in user behavior. For example, if a certain region forecasts that a hurricane or blizzard might strike in the near future, you'll see an increase in demand for groceries in those regions. So a smart search engine should dynamically promote staple groceries in those regions

Ex: milk, bread, water, etc

CONCLUSION

By following these simple guidelines and providing shoppers with an enhanced customer experience our clients have typically seen a 30% - 40% increase in search conversion. For further information on how to enhance search relevancy, contact sales@unbxd.com.