UNBXD

The Science & Art
of Optimizing for Low
Recall & Zero Search
Result Queries

Content

Background

Internet retailers face various challenges in product discovery due to scale. Their challenges are higher as the number of shoppers visiting the sites keep growing and the various ways these shoppers search for products. Moreover, this issue is further exacerbated due to ever increasing catalog sizes. This paper presents approach used by Unbxd that helps to optimize large variety of search queries that result in very low recall or even worse zero results.

Why zero search result queries are important?

Ecommerce retailers have many search queries for which their current systems aren't able to retrieve any products or very few products (low recall). Even for systems that claim to have context understanding, on an average, we have found that 1 to 2 of every 10 search queries results in zero product results. This implies 10-20% of all search queries aren't returning any product results, thus a massive problem, as it not only translates to a bad customer experience, but also in a significant loss of revenue to the retailer.

For example, let's consider an internet retailer which has 50K - 100K products in the catalog with annual revenue in the range of $1bn-$5bn. Based on our estimates, site will have an estimated 1-2mn unique search queries per month, out of which 10-20% can be zero results search queries. To estimate the potential revenue impact of optimizing zero results search queries, let's look at the data below in exhibit 1:
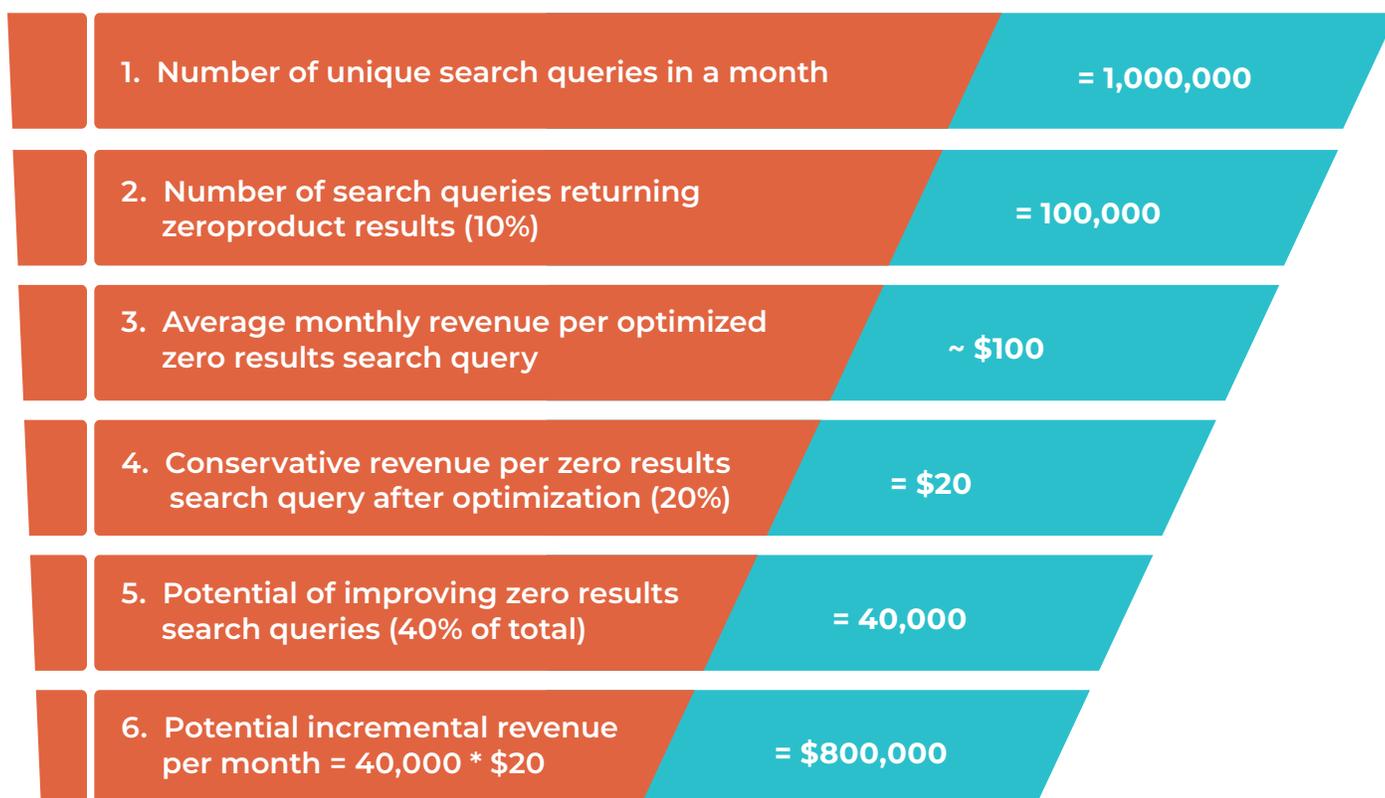
1. Number of unique search queries in a month	= 1,000,000
2. Number of search queries returning zeroproduct results (10%)	= 100,000
3. Average monthly revenue per optimized zero results search query	~ $100
4. Conservative revenue per zero results search query after optimization (20%)	= $20
5. Potential of improving zero results search queries (40% of total)	= 40,000
6. Potential incremental revenue per month = 40,000 * $20	= $800,000

Exhibit 1: Potentially $800K incremental revenue per month could be realized, if 40% of the total zero result search queries are optimized.

Implies, that the company even the size of one-tenth the size elucidated in the exhibit has the potential to realise $80K incremental revenue per month. Clearly, the lacunae due to the zero search result queries cannot be ignored.

Zero result search queries are trickiest to solve for, as the solution lies in making the search systems extremely context sensitive. The ROI is low here as the level of expertise required to solve this problem is quite high, and the incremental revenue might not be able to match it up even for the largest companies. To further qualify, there will be significant investment required in terms of upfront capital expenditure in infrastructure, hiring resources, software maintenance and technology R&D, which translates into dedication just to this focus area and making it a business priority.
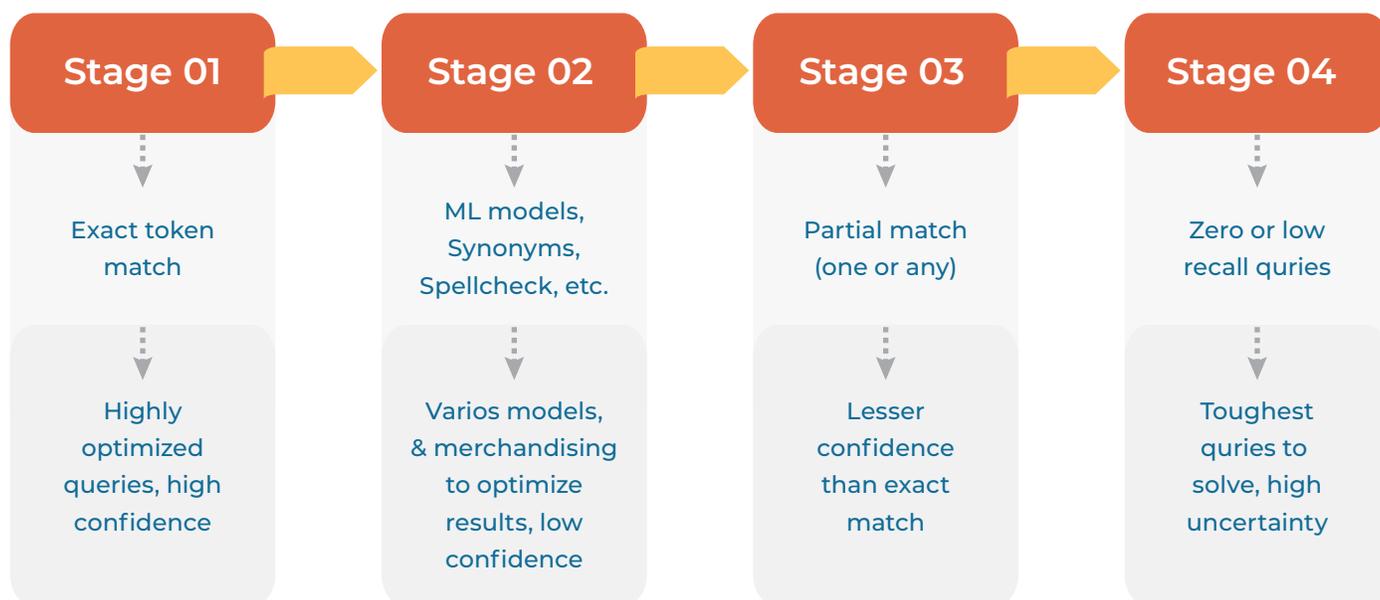
But before we make any judgements, it is important to understand that how search queries are processed and how that process can lead to fixing the low recall and zero search result queries.

How search queries are processed?

Generally, search systems at ecommerce retailers leverage Apache Solr or ElasticSearch as the core search framework underneath. In the search systems, an incoming search query goes through the query pipeline which comprises of 4-stages, where each stage is capable of handling a certain level of complexity in the search query. The four stages for search queries being processed in a search stack are:

▶ **Stage 1:** Token exact match by the search engine. The confidence that query will be resulting in optimized results is high. Also to maximize the potential, companies may setup merchandising on the high traffic queries through custom boosts and sorts, or build landing pages that suits the use case.

▶ **Stage 2:** If there isn't an exact match, AI/ML based models like Learning to Rank for Information Retrieval (LeToR), Named Entity Recognition (NER) are used to build the context and improve results set. Based on algorithm outputs, further spell checks, synonyms or filters may get applied, depending on the algorithms pipeline in the search stack. Some retailers may directly apply query processing models based on synonyms, spell check etc., based on the domain knowledge. Merchandising activities like boosting products from a certain brand or filtering low inventory products may be applied to ensure a consistent and optimal experience for the shoppers.

▶ **Stage 3:** If models and approaches in stage 2 fail to find any product results, retailers may perform a partial match on the search query (where one or any words may match). Some combination of advanced AI models may be applied again to ensure the optimal recall and precision. The confidence is lower than exact match for these set of queries.

▶ **Stage 4:** If search stack and any of the models deployed aren't capable enough of figuring out the intent of the shopper, then system ends up responding with zero or 1-2 irrelevant products in the results. Due to this, shopper feels they've reached a dead-end and probably they are on the wrong site. How Unbxd can help in alleviating this problem, forms the basis of discussion of the next sections.

Stage 01	Stage 02	Stage 03	Stage 04
Exact token match	ML models, Synonyms, Spellcheck, etc.	Partial match (one or any)	Zero or low recall quries
Highly optimized queries, high confidence	Varios models, & merchandising to optimize results, low confidence	Lesser confidence than exact match	Toughest quries to solve, high uncertainty

For an eCommerce company, on an average top 5-10% search queries drive **80%** of the search traffic and revenue on the website.

Optimizing for the top search queries and keywords is necessary. The top keywords are relatively easier to solve for, as there are a lot of data points regarding purchase and browse history based on which the optimization can be done.

However, the remaining search queries, which form the majority, are driving just **20%** of traffic and revenue. Among these search queries, there is a sizeable fraction of search queries where there aren't any product results. As there are no search results, optimizing for these search queries will require a lot of in-built context sensitivity in the search system.

The maths against fixing zero search result queries!

Further, the ROI for the eCommerce enterprise to solve for zero results search queries is low, as the level of expertise required to make sense from these queries is quite high. Top data science professionals are battling with the hardest NLP problems and hiring them will cost a fortune. Continuously identifying patterns from zero or low recall queries requires dedicated investment and effort. The incremental revenue might not be able to match it up even for the largest companies. We did a simple analysis to showcase the amount of investment that would need to be made for fixing these low recall and zero result queries.

Professional	Number of resources	Yearly salary per person	Total
Data scientist	2	$100K-150K	200,000
NLP expert - PhD from Ivy	2	$250K-500K	600,000
Data engineer	2	100,000	200,000
Platform	2	100,000	200,000

Total salary = $1.2mn

Monthly salary = $100K

Incremental profit per month after optimizing search queries from Exhibit 1, assuming a 10% net profit = $800K*0.10 = $80K

Clearly, expenses are more than what incremental revenue, the work on fixing the zero search result queries brings. So, it becomes important for us to understand what at first place causes these low recall and zero result search queries despite the fact the product many times is available in the catalog.

Why this problem arises?

A zero results query is a search request that returns zero products from a search system. While the most important driver for a zero results query is that there were no matching products - however if one drills down - there are many patterns in search queries that may result in zero products. For e.g.

Some of the patterns leading to zero results queries are:

1. Either keyword is not available in product catalog, or relevant products are not carried:

a. A user searches for Hatch-a-mole and client doesn't carry it. Client would like a vendor to redirect to toys or show toys. Having the intelligence to do more than just showing a zero product result with an objective of keeping the user on the site is important.

b. A user searches for "Red Party Dress" and client doesn't carry any red party dresses, however there are a few red dresses in the catalog.

c. A user searches for "Rolex" and client doesn't carry any Rolex watches, however client does carry watches of many similar luxury watch brands. Client might want to show watches of other luxury brands they carry, rather than a zero results page. Client may not know how to solve for this problem.

d. A user searches for "Rolex" and client doesn't carry watches of Rolex brand and any other luxury watch brands. Client might still want to show watches of brands which are superior, though not luxury. It might still be better than a zero results page. This mechanism is called "graceful fallback" that client may not know how to build for. Even though the results aren't something the shopper is looking for, search still responds with shopper to see something. Here, capability of search engine to identify that "Rolex" is a brand of watches and within the luxury niche, allow for the capability to integrate a fallback in case there are no products for the niche shopper might be looking in.

2. Filtering 'out of stock' products

a. Most retailers filter out of stock products from their catalog. There can be a lot of zero search results experiences due to this, including numeric queries. Provisioning for a "sold out page" when products are out of stock, and then recommending similar products may be a way to solve for this.

3. There are a huge number of types of search queries, and not all can be optimally handled. Certain query types which may not be resolved in client's search platform:

a. Thematic search queries such as "Valentines day jewelry"

b. Feature search queries, in case feature is not carried in the catalog

c. Brand search queries, in case the brand is not carried

d. Numeric queries such as "20' tents"

e. Major spelling mistakes that ordinary spell checkers won't catch for example "blck jns"

f. Synonyms not configured for products with lower prices and order values

> The challenge is the inability of a site search solution to:
>
> **a.**   <u>identify</u> these gaps,
>
> **b.**   <u>deploy</u> intelligent <u>AI-based models</u> that can identify, plug these gaps and
>
> **c.**   ensure that this <u>fix works every time</u> a situation like this arises
>
> **d.**   <u>automatically</u> without any manual intervention

And if you have been able to spot and implement such solution, that is great! But if you are still wondering what could be done to fix the overall shopper experience on your eCommerce portal, go ahead and read more.
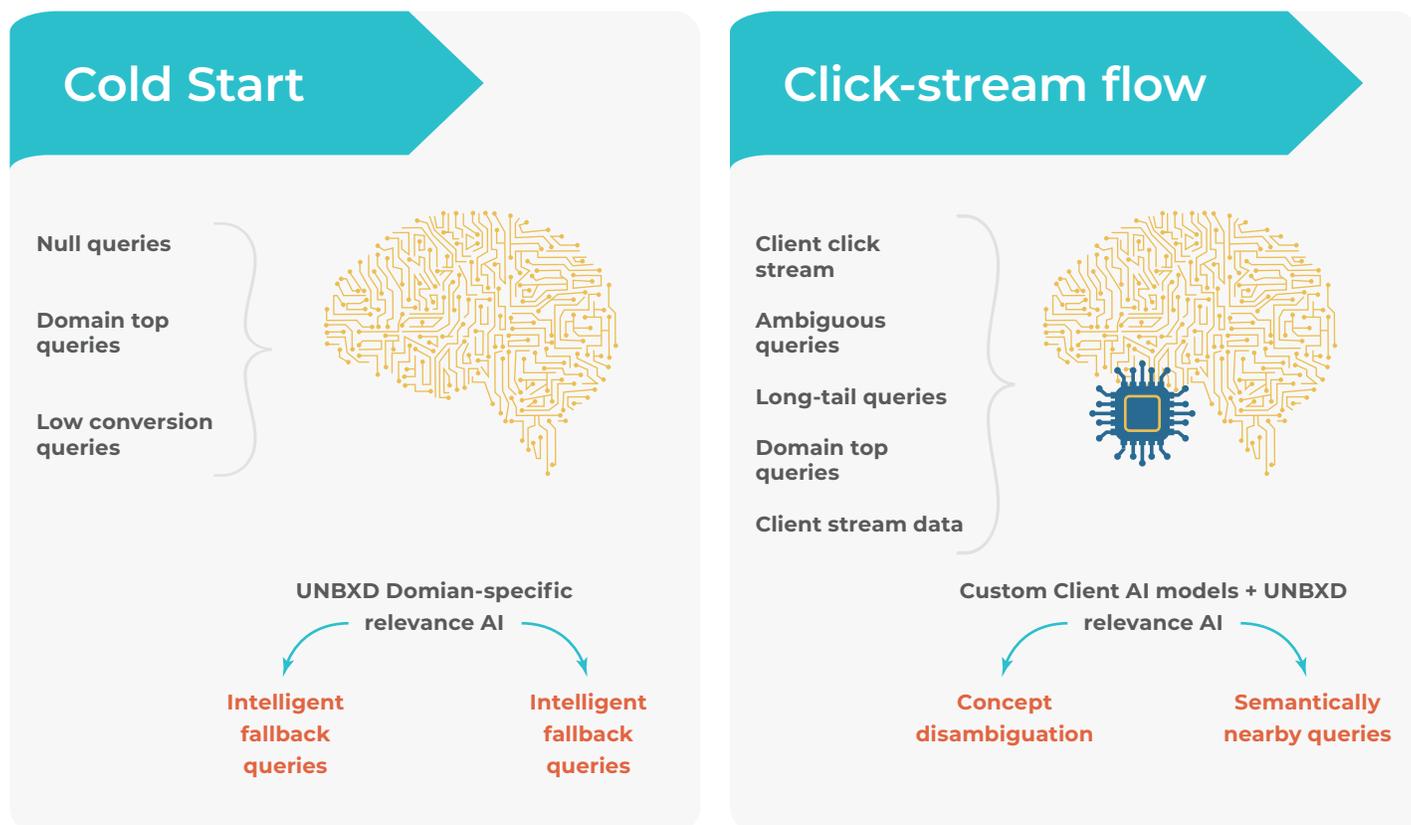
The Unbxd Approach

Unbxd is helping internet retailers in solving Stage 4 problems by addressing the search queries with very low recall or return zero products. Unbxd solutions include:

1.   An ensemble of machine learning models optimized to solve for Learning to Rank (LeToR), Named Entity Recognition (NER), Query Classification, Query Category Prediction, etc.

2.   Ability to process clickstream data and rank popular products.

3.   Automated facets and filters, based on popularity and relevance.

4.   Advanced machine learning algorithms to rank products based on inventory, freshness, diversity, seasons etc. and capability to personalize results on 1:1 basis

Unbxd is able to identify patterns among different types of zero result queries, and find context through its intelligent algorithms. Different techniques that are leveraged during this process are:

1.   Deep Learning, Language Model and Session based Query Reformulations

2.   Language Models using query-click (user session) data and perform automation

      a.   Smart reformulation of phrases and synonyms

      b.   Intelligent fallback queries

      c.   Concept disambiguation

      d.   Spell correction

      e.   Phonetic similarity

Cold Start

Null queries

Domain top
queries

Low conversion
queries

**UNBXD Domian-specific
relevance AI**

**Intelligent
fallback
queries**

**Intelligent
fallback
queries**

Click-stream flow

Client click
stream

Ambiguous
queries

Long-tail queries

Domain top
queries

Client stream data

**Custom Client AI models + UNBXD
relevance AI**

**Concept
disambiguation**

**Semantically
nearby queries**

Unbxd system in turn leverages user behavior/clickstream data, product catalog data as shown in the images in order to train these algorithms and models.

The end objective is to make systems more efficient and apt at finding the results for search queries or identify the fallback if there is a zero result for a particular query. Unbxd aims to tune relevancy into search results produced and providing a seamless shopping experience for the online shoppers. But the eCommerce players because of the abundance of false-promising solutions available in the market are sceptical about how to go around solving for this challenge, that mars their reputation and experience for their shoppers. And they find their own teams too stretched out to fix this "**seemingly last mile problem**".

Conclusion

Solving for zero or low recall search queries is a huge business problem for internet retailers, as this directly translates to bad shopper experience and loss in sales. However this is not the number one problem they are looking to solve because it doesn't cater to the top 20% of the search queries that get the most of the revenue. What businesses generally forget is one bad experience of zero search results and that customer is lost forever. Here, a shopper has already made the effort of visiting your site and searching, however there were no products returned. This is money left on the table. Even if a fraction of those searches could be optimized, there would be a direct revenue impact and gratification for your shopper.

Currently, search providers or in-house technology teams don't want to focus on this problem, as it requires dedicated effort and ROI may not be commensurate. As an example, it might be 3x effort for 1/3rd of the incremental revenue when compared to standard optimizations of search. The problem is extremely challenging as intent needs to be understood by identifying alternative context from a zero results search query.

After implementing Unbxd solution, our clients have been able to solve up to half of their zero or low recall search queries with a significant revenue impact.

*To understand how can Unbxd help in optimizing the search queries that are very low recall or are returning zero product results, and the potential to which the improvement can be seen, please contact us at **sales@unbxd.com***