

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/326133502>

# A Taxonomy of Queries for E-commerce Search

Conference Paper · June 2018

DOI: 10.1145/3209978.3210152

CITATIONS

6

READS

1,332

4 authors, including:



**Parikshit Sondhi**

Snap inc.

18 PUBLICATIONS 209 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



E-Commerce Search [View project](#)



Query Log Analysis [View project](#)

# A Taxonomy of Queries for E-commerce Search

Parikshit Sondhi\*

WalmartLabs  
Sunnyvale, CA, USA  
parikshit@neulogic.ai

Pranam Kolari

WalmartLabs  
Sunnyvale, CA, USA  
pkolari@walmartlabs.com

Mohit Sharma\*

WalmartLabs  
Sunnyvale, CA, USA  
mohit.sharma@walmartlabs.com

ChengXiang Zhai

University of Illinois Urbana Champaign  
Urbana, IL, USA  
czhai@illinois.edu

## ABSTRACT

Understanding the search tasks and search behavior of users is necessary for optimizing search engine results. While much work has been done on understanding the users in Web search, little knowledge is available about the search tasks and behavior of users in the E-Commerce (E-Com) search applications. **In this paper, we share the first empirical study of the queries and search behavior of users in E-Com search by analyzing search log from a major E-Com search engine.** The analysis results show that E-Com queries can be categorized into five categories, each with distinctive search behaviors: (1) Shallow Exploration Queries are short vague queries that a user may use initially in exploring the product space. (2) Targeted Purchase Queries are queries used by users to purchase items that they are generally familiar with, thus without much decision making. (3) Major-Item Shopping Queries are used by users to shop for a major item which is often relatively expensive and thus requires some serious exploration, but typically in a limited scope of choices. (4) Minor-Item Shopping Queries are used by users to shop for minor items that are generally not very expensive, but still require some exploration of choices. (5) Hard-Choice Shopping Queries are used by users who want to deeply explore all the candidate products before finalizing the choice often appropriate when multiple products must be carefully compared with each other. These five categories form a taxonomy for E-Com queries and can shed light on how we may develop customized search technologies for each type of search queries to improve search engine utility.

## CCS CONCEPTS

• Information systems → Query log analysis; • Applied computing → Online shopping;

## KEYWORDS

E-com search, query taxonomy, search log analysis

\*First two authors contributed equally. Work was done when Parikshit was at WalmartLabs

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

SIGIR '18, July 8–12, 2018, Ann Arbor, MI, USA

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-5657-2/18/07...\$15.00

<https://doi.org/10.1145/3209978.3210152>

## ACM Reference Format:

Parikshit Sondhi\*, Mohit Sharma\*, Pranam Kolari, and ChengXiang Zhai. 2018. A Taxonomy of Queries for E-commerce Search. In *SIGIR '18: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, July 8–12, 2018, Ann Arbor, MI, USA*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3209978.3210152>

## 1 INTRODUCTION

Understanding the search tasks and search behavior of users is necessary for optimizing search engine results; this is especially important if the search engine is to personalize the search results for each individual user [4]. For example, the Web query taxonomy suggested by Broder [1], which classifies Web queries into three categories (i.e., informational, navigational, and transactional), has revealed important differences between Web search tasks and traditional library search tasks. Thus enabling researchers to realize that traditional retrieval models were limited to serving informational queries, and helping open up new directions in information retrieval research.

Although much work has been done on understanding users in Web search [1, 5], little knowledge is available about the search tasks and user behavior in the E-Com search applications, despite their increasing importance [3]. To bridge this knowledge gap, we conduct the first empirical study of E-Com search queries, focused on formulating a taxonomy that distinguishes the various types of E-Com search tasks. Our goal is to analyze a large scale commercial E-Com search engine<sup>1</sup> log, in an attempt to discover user behavior patterns that represent distinct search tasks, and naturally lead to a coherent query taxonomy. Considering the significant impact of Broder's work [1], we believe that a meaningful taxonomy of E-Com queries would also help develop a better understanding of E-Com related information retrieval problems, and stimulate research on novel technologies, especially related to comprehensive search task support.

One approach to constructing an E-Com query taxonomy, is to organize the queries in the same product category hierarchy, that is used for organizing E-com catalogs (e.g. Home, Electronics, Apparel etc.). However such an organization provides limited value, as it is "imposed" based on subjective decisions of human curators, not reflective of underlying E-com search tasks. In fact our analysis reveals that: 1) queries exhibiting similar user behavior patterns are often associated with multiple product-categories, and 2) a single product category may be associated with queries exhibiting diverse user behavior patterns.

<sup>1</sup><http://www.walmart.com>

Our goal instead is to “discover” a natural clustering of queries, such that each cluster represents a distinct E-com search task. To this end, we first define a query representation using 15 user behavior features obtained from our search log. Next, we propose a *cluster stability metric*, for tuning suitable clustering algorithm hyper-parameters, leading to “stable” query clusters across different temporally randomly sampled data-sets. Our analysis shows that query clusters generated in this manner naturally map to various E-com search tasks. We ultimately obtain 5 temporally stable clusters. Specifically, they correspond to the following five categories of the E-Com query taxonomy: (1) *Shallow Exploration Queries* are short vague queries that a user may use initially in exploring the product space. (2) *Targeted Purchase Queries* are queries used by users to purchase items that they are generally familiar with, thus without much decision making. (3) *Major-Item Shopping Queries* are used by users to shop for items which are often relatively expensive and thus requires involved exploration, but typically in a limited scope of choices. (4) *Minor-Item Shopping Queries* are used by users to shop for minor items that are generally not very expensive, but still require some exploration of choices. (5) *Hard-Choice Shopping Queries* are used by users who want to deeply explore all the candidate products before finalizing the choice, often appropriate when product constraints are hard to express and multiple products must be carefully evaluated. This taxonomy can be used to infer a user’s intent or task and thus enables a search engine to customize its search algorithm toward the specific needs of a particular user.

## 2 DATASET

We collected data for our experiments from search logs of a major E-Com search engine. In order to create this dataset, we randomly sub-sampled users’ search sessions from the logs for a week in each month from March, 2017 to November, 2017. For the purpose of this analysis, we restricted to the sessions where a user issued at-least two queries and clicked on at-least one product. Also, to eliminate anomalous sessions resulting from scrapers, we ignored sessions where the number of clicks was more than 100. After pre-processing, we ultimately collected 3 Million unique queries spanning over 9 months of user search sessions from March 2017 to November 2017.

For each query, we then constructed features using the following query and session level properties.

### 2.1 Query-specific properties

This group comprises of properties that are solely computed based on a query and a user’s interaction with the query. The query-specific properties used in this work are as follow:

- (i) Impressions or  $\text{imp}(q)$ : number of times query  $q$  was issued in a week sampled from a particular month.
- (ii) Clicks( $q$ ),  $\text{atc}(q)$ ,  $\text{order}(q)$ : total number of clicks, add-to-carts and orders received for query  $q$  in a week.
- (iii) Click rates, add-to-cart ratio and order ratio are computed as follows:

$$\begin{aligned} \text{ctr}(q) &= \frac{\text{clicks}(q)}{\text{imp}(q)}, \\ \text{atcr}(q) &= \frac{\text{atc}(q)}{\text{clicks}(q)}, \\ \text{ordr}(q) &= \frac{\text{order}(q)}{\text{atc}(q)}. \end{aligned}$$

- (iv) Tokens: number of tokens in the query.
- (v) Browsed pages: average number of pages that are browsed for the query.
- (vi) Revenue: revenue attributed to the search query in a week.
- (vii) Clicked items: number of items clicked for the search query on the results page.

### 2.2 Session-specific properties

These properties are computed after considering all the queries issued in a session. The session-specific properties used in this work are as follow:

- (i) Browsed pages: average number of pages that are browsed in a session having the target query.
- (ii) Co-occurring queries: average number of distinct queries that co-occur with the target query in a session.
- (iii) Clicked items: number of items clicked in a session having the target query.

Note that we did not use any content based features, since such a representation would result in clusters that group queries with similar words, but not necessarily similar user behavior.

## 3 METHOD

We used the  $k$ -Means clustering algorithm [2] for generating the clusters. A key technical challenge is to determine the suitable value of  $k$ , such that it results in meaningful query clusters. To solve this challenge, we propose a *metric to measure the stability of a cluster* across all the months under consideration.

Specifically, let  $t$  denote the month in which the data was collected,  $i$  the cluster index, and  $c(i, t, k)$  the cluster center of  $i$ th cluster for month  $t$  generated using a particular value of  $k$ , where  $i \leq k$ . Cluster indices were assigned in descending order of their size, i.e. the largest cluster in any clustering was assigned the index 1, while the smallest was assigned index  $k$ . Such an assignment is appropriate since we would not expect relative cluster sizes associated with distinct query categories to vary dramatically across different clusterings.

We first applied  $k$ -Means clustering for the data collected in each month for different values of hyper-parameter  $k$ , and then evaluated the different clusterings associated with each  $k$  using our *Cluster Stability Metric* ( $S_k$ ) that measures the goodness of clustering as average of  $d_{ik}$  for all values of  $i$  for a particular  $k$ , as shown below:

$$S_k = \frac{1}{k} \sum_{i=1}^k d_{ik}. \quad (1)$$

$$d_{ik} = \frac{2}{|T|(|T|-1)} \sum_{t_p \in T} \sum_{t_q \in T, t_p \neq t_q} \|c(i, t_p, k) - c(i, t_q, k)\|_2, \quad (2)$$

where ( $d_{ik}$ ) is the average pairwise distance between the  $i$ th clusters for different values of  $t$ ,  $T$  represents the set of all the months under consideration,  $t_p$  denotes month  $p$  and  $t_q$  represents month  $q$ .

Our underlying assumption is that query clusters representing prominent search tasks (or query categories) are likely to be stable over different  $t$ s. Thus in a good clustering scheme, all clusters would be stable, and we would expect the value of  $S_k$  to be low.

**Table 1: Cluster Stability Metric ( $S_k$ ) for different values of  $k$** 

$k$	$S_k$
4	1504.401
5	895.305
7	4781.696
9	3424.999
10	1227.422

## 4 RESULTS

### 4.1 Stable cluster discovery

We applied standard scaling to the query attributes prior to applying the  $k$ -Means over the queries. Table 1 shows the values of  $S_k$  for different values of  $k$  over the collected data. As can be seen from the table, for  $k = 5$  we achieve the lowest value of  $S_k$  and this is expected as there may not exist a large number of significantly different patterns in eCommerce search. It is interesting to note that the value of  $S_k$  decreases for higher values of  $k$  ( $k > 7$ ) but nevertheless remains significantly higher than  $k = 5$ . Next, we will analyze the clusters discovered for the month of March, 2017 to derive insights about the queries present in those clusters.

### 4.2 Stable cluster analysis

Table 3 shows the mean and the standard deviation of size and centroids of the clusters across different months in 2017 for  $k = 5$ . We will also briefly discuss appropriate search strategies that can be applied to each cluster whenever possible.

**4.2.1 Shallow Exploration Queries.** This cluster comprises short and broad queries as evidenced by the average token length of 2.9 (Table 3) and its top queries (Table 2). We also observe the lowest click, add-to-cart, and order metrics. A review of queries suggests shoppers use these queries to start their exploration and get inspiration for re-formulating more precise queries after looking at the results. This intuition is further reinforced by the lack of deep pagination for these queries. The users when submitting these queries, are merely browsing the top few products without clicking or paginating. Finally queries in this cluster are associated with a diverse set of product categories, e.g., Electronics, Home Decor, Beauty, Toys, Pets, etc. However, we only observe limited queries from Grocery and Apparel segments, which are more prominent in *Targeted Purchase* and *Hard-Choice Shopping* clusters.

Due to their inherent ambiguity, to improve user experience for these queries, the E-com system, must return a diverse enough collection of products and navigation options, so the user can easily understand the exploration space involved, and follow appropriate navigation links or formulate suitable queries.

**4.2.2 Major-Item Shopping Queries.** This cluster comprises of queries with highest click rates, but comparatively much lower add to cart and order metrics. A look at the queries suggests that queries are dominated by expensive product categories like electronics and furniture. Users typically don't paginate too deep for such queries, suggesting limited choices, but do need time and detailed product information to make a purchase decision.

Beyond ensuring search quality (i.e. relevance of results), improving user satisfaction metrics for these queries would require detailed information on individual items to be provided on product

pages, along with convenient tools for users to compare various candidate products.

**4.2.3 Targeted Purchase Queries.** This cluster is distinguished by its very high click, add-to-cart and conversion metrics. It is dominated by grocery queries associated with inexpensive products that don't require deep decision making. These are also products that users typically purchase repeatedly. While the search engine catalog may contain several relevant products for these queries, users typically may not have a strong preference of one over the other and are not willing to spend too much time making comparisons and deciding. This intuition is further strengthened by the lowest normalized query pagination score (1.000) as shown in Table 3.

Most search engines use user engagement and product popularity as a strong ranking signals. Consequently, queries in this cluster are most likely to suffer from cold-start problem, where new relevant products are never exposed to the users and users don't care to paginate sufficiently to discover them. For such queries, either the search engine must incorporate explore-exploit techniques to introduce a churn in the returned results, or only a small product inventory needs to be maintained.

**4.2.4 Minor-Item Shopping Queries.** This cluster is dominated by queries for inexpensive durable items associated with the Kitchen / Bathroom Appliance and Home Decor categories. These queries have moderate click, add-to-cart, and order metrics. They also have slightly higher pagination metric. A review of queries suggests that users typically do care about certain stylistic/aesthetic aspects of these products, and therefore explore more products and spend more time in making decision compared to *Targeted Purchase Queries*.

**4.2.5 Hard-Choice Shopping Queries.** This cluster is dominated by Apparel, Accessories and Home decor related queries where: 1) a user typically has hundreds of relevant products to choose from, 2) users care significantly about the product being purchased and are willing to spend time in exploring, and 3) it is hard to express exact user requirements or particularly expected style attributes through keyword queries or filters. Thus users end up paginating deep into search results to find suitable products, as evidenced by the high *browsed pages/query* value (5.248) shown in Table 3.

These queries would benefit from user interfaces that allow users to easily navigate a large product space, review several products, while giving them the ability to keep track of products they found interesting (for example via favorite lists). These queries would also benefit the most from abilities that allow users to express information needs that are not easy to express via keywords (example search by image, show similar products etc.).

**4.2.6 Overall Taxonomy.** The five categories of queries form a meaningful taxonomy for classifying queries in E-Com search as illustrated in Figure 1, where we show the following decision process: if a user knows well about what items to buy, the query would likely fall into "Targeted Purchase." If a user doesn't know well, then the user would have to explore. In such a case, if the user cannot even specify a specific query, the query would be very short and fall into "Shallow Exploration." Otherwise, we may further distinguish a case where the relevance criteria are hard to specify and careful comparison between alternative choices must be made, i.e., "Hard-choice shopping" where the user often has to explore deeply

Table 2: Top queries for each cluster

Shallow Exploration	Major-Item Shopping	Targeted Purchase	Minor-Item Shopping	Hard-Choice Shopping
tires	nintendo switch	toilet paper	toothbrush holder	tv
lamps	tv stand	paper towels	dish towels	pioneer woman
bike	curtains	baby wipes	broom and dust pan	bedding comforter set
ipad	laptop	toothpaste	kitchen utensils	women shoes
easter	desk	laundry detergent	knife set	decorative pillows
storage	bikes	easter candy	spatula	pioneer woman clearance
apple watch	microwave	paper plates	cooking utensils	iphone 6s plus case
lego	nintendo switch console	wipes	oven mitts	toddler girl clothes
storage bins	dresser	cat litter	accent pillows set of 4	engagement rings
phones	patio furniture	peanut butter	utensil holder	wedding rings
makeup	mattress	cereal	pot holders	outdoor pillows
shampoo	tablets	body wash	spices	funko pop
dog food	futon	sugar	salt and pepper shaker set	comforter set
candy	vacuum	deodorant	plunger	iphone se case
water	tv stands	womens socks	pizza pan	womens sandals

Table 3: Size and centroids of the clusters found in different months in 2017.

Type	Size*	Click rate*	Add-to-cart ratio*	Order ratio*	Tokens	Browsed pages/query*
Shallow Exploration	15.323 ± 3.444	1.000 ± 0.000	1.0 ± 0.000	1.000 ± 0.000	2.903 ± 0.006	1.107 ± 0.001
Major-Item Shopping	10.512 ± 0.934	3.509 ± 0.018	4.865 ± 0.000	2.800 ± 0.000	3.476 ± 0.006	1.092 ± 0.000
Targeted Purchase	4.319 ± 0.680	3.352 ± 0.000	12.000 ± 0.001	25.900 ± 0.000	3.460 ± 0.012	1.000 ± 0.000
Minor-Item Shopping	3.079 ± 0.097	1.843 ± 0.000	3.500 ± 0.000	2.433 ± 0.000	3.258 ± 0.004	1.209 ± 0.001
Hard-Choice Shopping	1.000 ± 0.289	1.490 ± 0.000	3.481 ± 0.000	2.733 ± 0.000	3.536 ± 0.000	5.248 ± 0.060

\* The values in the column are normalized by dividing them by the lowest value in the column.

by viewing more result pages due to the difficulty in completely specifying the relevance criteria, and also add more products to the cart for a thorough comparison. In the remaining cases, the user would be able to specify relevance criteria, and thus would not need to explore deeply, though the user would still have to explore alternative choices before finalizing a purchasing decision, and here we can further distinguish “Major-item shopping” from “Minor-item shopping”. In the former situation, the relevance of search results is likely high (due to the fact that there aren’t many choices), so clickthrough rates are high; in contrast, in the latter situation, the relevance may not necessarily be high due to too many choices, thus the clickthrough rate tends to be lower.

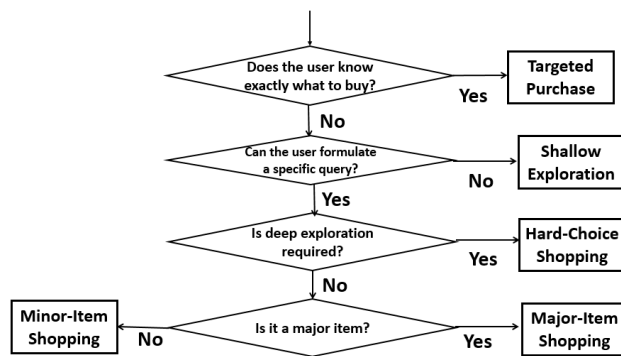


Figure 1: Illustration of Taxonomy of E-Com Queries

## 5 DISCUSSION

We conducted an empirical analysis of an E-Com search log data set and derived a taxonomy for E-Com queries which includes five categories: Shallow Exploration, Targeted Purchase, Hard-Choice

Shopping, Major-Item Shopping and Minor-Item Shopping. They not only allow us to better understand users of E-Com search engines, but also provide a basis for optimizing E-Com search results via personalization and adaptive search task support, on which we have also offered some preliminary thoughts.

Our proposed taxonomy also has interesting parallels with traditional consumer product categories found in economics literature. For example, our Targeted query category aligns well with the notion of *convenience products*<sup>2</sup>, which are generally low priced consumable products that customers buy frequently without investing too much decision making effort. On the other hand, our Hard-Choice category represents queries often used to look for products in the *heterogenous shopping product*<sup>3</sup> category, where every product is fundamentally different from the others, and users’ spend considerable time making decisions.

## REFERENCES

- [1] Andrei Broder. 2002. A Taxonomy of Web Search. *SIGIR Forum* 36, 2 (Sept. 2002), 3–10. <https://doi.org/10.1145/792550.792552>
- [2] Tapas Kanungo, David M Mount, Nathan S Netanyahu, Christine D Piatko, Ruth Silverman, and Angela Y Wu. 2002. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE transactions on pattern analysis and machine intelligence* 24, 7 (2002), 881–892.
- [3] Shubhra Kanti Karmaker Santu, Parikshit Sondhi, and ChengXiang Zhai. 2017. On Application of Learning to Rank for E-Commerce Search. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*. ACM, New York, NY, USA, 475–484. <https://doi.org/10.1145/3077136.3080838>
- [4] Jaime Teevan, Susan T. Dumais, and Eric Horvitz. 2010. Potential for Personalization. *ACM Trans. Comput.-Hum. Interact.* 17, 1, Article 4 (April 2010), 31 pages. <https://doi.org/10.1145/1721831.1721835>
- [5] Ryen W. White and Steven M. Drucker. 2007. Investigating Behavioral Variability in Web Search. In *Proceedings of the 16th International Conference on World Wide Web (WWW '07)*. ACM, New York, NY, USA, 21–30. <https://doi.org/10.1145/1242572.1242576>

<sup>2</sup>[https://en.wikipedia.org/wiki/Final\\_good#Convenience\\_goods](https://en.wikipedia.org/wiki/Final_good#Convenience_goods)

<sup>3</sup><https://www.marketing91.com/shopping-products/>