

Analyzing Web Search Queries of Before and After Purchase on E-Commerce Site

Yuka Kawada¹[0000–0003–0767–5038], Takehiro Yamamoto¹[0000–0003–0601–3139],
Hiroaki Ohshima¹[0000–0002–9492–2246], Yuki Yanagida²[0000–0002–6691–2243],
Makoto P. Kato²[0000–0002–9351–0901], and Sumio Fujita³[0000–0002–1282–386X]

¹ University of Hyogo, Kobe, Japan
ad221019@gsis.u-hyogo.ac.jp, t.yamamoto@sis.u-hyogo.ac.jp,
ohshima@ai.u-hyogo.ac.jp

² University of Tsukuba, Tsukuba, Japan
ynagi2@klis.tsukuba.ac.jp, mpkato@acm.org

³ Yahoo Japan Corporation, Tokyo, Japan
sufujita@yahoo-corp.jp

Abstract. In this study, we investigated how Web search queries change before and after purchasing a product. We focused on the Web searchers who purchased cameras from an e-commerce site. First, we manually classified the words that characteristically appear during pre-purchase and post-purchase searches with cameras. From the manual classification, we found 14 intents. The intents include model number, evaluation, shipping, and accessory, etc. Based on these classified intents, we then analyzed when the words in each intent were used in queries before/after the purchase. Our analyses revealed that the users narrowed the search space with respect to the product as the user nears purchase.

Keywords: Web search · Consumer behavior · Behavior analysis.

1 Introduction

Numerous users search the Web to obtain information about a product. Market research firm Survata conducted a survey to investigate which websites users in the US in searching for products [4]. Survata reported that 35.5% of users utilize Web search as their first step in purchasing a product, and 45.5% of users search the Web when they have not decided which specific product to buy.

Web search queries of users change depending on their interests in the product [3]. Analyzing the changes in Web search queries over time can help us understand the purchasing behavior of users. For example, a user searches the Web issuing the query “*camera types*” to look for suitable cameras for purchase. At this time, the user still has not yet decided which product to purchase. The user also searches the Web issuing the query “*canon kiss x10 price*” when trying to find the price of the Canon Kiss X10 camera model. This time, the user is searching for a specific product and is likely to purchase the same. Understanding the purchasing behavior process based on Web search queries enables

us to support users, by customizing advertisements to be shown to the users or, by presenting search results that are customized to the user’s information needs [1][5].

In this study, we analyze the change in Web search queries before and after purchasing a camera using the Web search logs of a commercial Web search engine and purchase data from an online e-commerce site. We address the following research questions in this study.

RQ1: What types of words are used in Web search queries before and after a purchase?

RQ2: How do the words identified in RQ1 change when users approach a purchase and after the purchase?

2 Data

We used the Web search logs provided by Yahoo! Search and purchase data from Yahoo! Shopping. We used Web search logs for 13 months between October 1, 2016 and October 31, 2017 provided by Yahoo! Search. We used the search logs of users who performed at least 10 searches per month to target users who use Web search engines on a daily basis. As for the purchase data, we analyzed the logs of users who purchased cameras between October 1, 2016 and October 31, 2017 from Yahoo! Shopping. The cameras include digital cameras and digital SLR cameras.

In this study, we focused on cameras since we believe that users who purchase cameras tend to use a Web search engine to gather information before making a purchase. Thus, it is important to understand the search behaviors of users with respect to these products. In this study, we analyzed 5,508 users who bought cameras.

3 Extraction of Characteristic Words Used during Pre- and Post-purchase Queries and their Classification into Query Intents

We investigate the query types before and after purchases by identifying characteristic words and categorizing them into **query intents**. For instance, users search **price** information before buying and **accessories** after purchasing. In this study, we aim to identify the intent of queries issued before and after a purchase.

We prepared product-related queries for the 72 hours before and after a purchase. For pre-purchase queries, we extracted sessions with product-related queries in 30-minute inactivity intervals [2]. Post-purchase queries were also extracted. The list of product names was prepared based on the purchase data and data from kakaku.com⁴.

⁴ <https://kakaku.com/>

We trained a classifier using logistic regression to determine pre-purchase and post-purchase queries, employing tf-idf weighted words. We analyzed queries from 5,508 users related to cameras, resulting in an accuracy and an F1 score of 0.65 and 0.64. Furthermore, we extracted the top 100 words with the largest (absolute) regression coefficient for both positive and negative coefficients, leading to the extraction of 200 words in total.

We then manually classified the characteristic words into query intents, which represent the types of information the users searched for. To prepare query intents, we first refer to the taxonomy from the previous study [6] that analyzed the relationship between search behavior and satisfaction related to product search. By referring to their taxonomy and sampling queries, we first prepared the initial query intents as **category**, **manufacture**, **series**, **model number**, **evaluation**, **sales method**, **store**, **price**, and **function & feature**. Three authors of this paper manually classified the characteristic words into the query intents based on this taxonomy. As a result, we determined 14 query intents.

Table 1 lists these 14 query intents and presents example characteristic words assigned to each query intent prepared in our study. The query intents **accessory**, **shipping**, **how to use**, and **new & used** are the newly prepared query intents based on our study. From this classification process, we obtained the query intents that could be issued after purchasing a product. For example, users search for information regarding product shipping by issuing queries containing words such as “*tracking*”. Moreover, users search for information regarding product usage by issuing queries containing words such as “*treatment*.” These query intents are expected to be issued after a purchase. The queries related to these intents may not be issued on an e-commerce site, but rather on a Web search engine. We believe that these intents could be revealed by analyzing queries on a Web search engine. The identification of these intents suggests that it is important for the system to not only recommend products related to the purchased product (such as accessories) but also to recommend information regarding usage and guarantees to support the users’ purchasing process. In this next section, we will analyze how these query intents occur and change before and after a purchase.

4 Time Series Analysis of Queries Based on Query Intents

In this section, we analyze how users’ Web search queries changed as they were about to make a purchase and as after making a purchase. By understanding these changes, we can understand when the users make a purchase, and what information they search for after the purchase. First, we analyzed the changes in queries on a 24 hours basis. Next, we analyzed the changes in queries on a query basis.

To analyze the queries, we classified them into the query intents prepared in Section 3. A query is classified into a query intent if it contains a word that belongs to the query intent.

Table 1: List of query intents, definitions, and example words to be assigned for each intent prepared in our study.

Query intent	Definition	Words
Category	A type of the product.	digital cameras
Manufacturer	A manufacturer of the product.	Casio
Series	A series name of the product.	ixy
Model number	A model number of the product.	g800
Function & feature	A function or a feature of the product.	pixels
Store	A store name.	Kitamura
Price	Price of the product.	price
Guarantee	Product registration and guarantee.	customer registration
Evaluation	A reputation of the product.	comparison
Sales method	Selling and payment method for the product.	on sale
Accessory	Another product to be used with the product.	sd card
Shipping	Shipping of the product.	tracking
How to use	Information on how to use the product.	treatment
New & used	A condition of the product.	used
Other	Do not fit into any of these categories.	-

4.1 Occurrences of query intents per 24 hours

We analyzed the frequency of query intents at each 24 hours interval in the time period between one week before and one week after making a purchase. Table 2 presents the ratio of the frequency of queries for each query intent. The value in the table is computed based on the ratio of the number of queries in a certain period divided by the total number of queries issued in two weeks. For example, in Table 2 the value of query intent **price** in the interval $(-24h, 0]$ is 0.18. This means that 18% of queries that are classified as **price** were issued in the period between 24 hours before the purchase and immediately before the purchase.

First, we describe the trend of change in queries before the purchase. As the purchase time approaches, the frequency of **category**, **manufacturer**, **series**, and **model number** gradually increases. In addition, the largest proportions of **category**, **manufacturer**, **series**, **model number**, **function & feature**, **store**, **price**, and **evaluation** appear in the time period between 24 hours before the purchase and immediately before purchase. These results suggest that many users perform Web searches in the 24 hours period before purchasing a camera.

Next, we describe the trend of change in queries after the purchase. The highest number of queries from the query intents **accessory** and **how to use** are issued after purchasing cameras. Moreover, the query intent **shipping** is issued the most after a purchase. It is worth noting that products bought from e-commerce are typically shipped immediately, arriving within 24 to 48 hours.

We observe that information regarding **accessory** of the camera is searched just after the purchase $((0h, 24h])$. This indicates that users who purchased a camera search for the related products immediately after the purchase. Information regarding **shipping** of the product $((24h, 48h])$ is then searched, followed by information regarding **how to use** of the product $((48h, 72h])$. These re-

Table 2: Frequency of queries for each intent analyzed in 24 hours intervals between one week before and after the purchase.

Query intent	pre-purchase							post-purchase						
	(-168h, -144h]	(-144h, -120h]	(-120h, -96h]	(-96h, -72h]	(-72h, -48h]	(-48h, -24h]	(-24h, 0h]	(0h, 24h]	(24h, 48h]	(48h, 72h]	(72h, 96h]	(96h, 120h]	(120h, 144h]	(144h, 168h]
Category	0.05	0.05	0.06	0.07	0.09	0.11	0.25	0.08	0.05	0.05	0.04	0.04	0.03	0.03
Manufacturer	0.05	0.05	0.05	0.06	0.07	0.09	0.21	0.10	0.07	0.06	0.05	0.04	0.04	0.04
Series	0.03	0.05	0.05	0.05	0.07	0.10	0.37	0.09	0.05	0.04	0.04	0.02	0.02	0.02
Model number	0.04	0.04	0.05	0.06	0.07	0.09	0.25	0.11	0.07	0.06	0.05	0.04	0.04	0.03
Function & feature	0.07	0.06	0.07	0.07	0.07	0.09	0.16	0.08	0.06	0.06	0.05	0.05	0.05	0.05
Store	0.06	0.06	0.06	0.07	0.08	0.10	0.20	0.08	0.06	0.05	0.05	0.05	0.04	0.04
Price	0.07	0.07	0.06	0.07	0.08	0.09	0.18	0.08	0.06	0.05	0.05	0.05	0.05	0.05
Guarantee	0.05	0.05	0.04	0.06	0.06	0.08	0.27	0.07	0.06	0.08	0.06	0.05	0.05	0.04
Evaluation	0.07	0.07	0.07	0.07	0.08	0.09	0.15	0.08	0.06	0.05	0.05	0.05	0.05	0.05
Sales method	0.06	0.07	0.06	0.06	0.07	0.09	0.17	0.10	0.06	0.05	0.06	0.05	0.05	0.06
Accessory	0.04	0.04	0.04	0.04	0.05	0.06	0.14	0.16	0.10	0.10	0.07	0.06	0.06	0.05
Shipping	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.11	0.19	0.12	0.10	0.08	0.06	0.06
How to use	0.05	0.05	0.05	0.06	0.05	0.06	0.07	0.09	0.10	0.10	0.09	0.08	0.08	0.07
New & used	0.06	0.07	0.07	0.07	0.08	0.10	0.17	0.08	0.06	0.05	0.05	0.05	0.05	0.05

sults suggest that the users search for a variety of information after purchasing a product, and the timing of when this information is searched varies depending on the information type.

4.2 Occurrence of query intents per query

In this section, we analyze the changes on a query basis in a more fine-grained way. First, we extract queries from users who have issued exactly five queries that contain any of the characteristic words presented in Section 3 before and after their purchases. We then analyze the changes in query intents in the issued queries when users approach the purchase and after the purchase.

Table 3 presents the changes in the occurrence of query intents before and after the purchase. For example, the value of the query intent **model number** in the $n = 1$ of “The n -th query pre-purchase” column is 0.30. Therefore, 30% of the queries that contain model names occurred just before the purchase. Similarly, the value of query intents **model number** in the $n = 1$ of “The n -th query pre-purchase” column is 0.21. This shows that 21% of the queries that contain model names occurred just after the purchase. The query intents for which the total frequency of queries was 50 or less were excluded from the analysis (denoted as – in the table) because it was difficult to find reliable trends for these intents.

As for the query intents when the purchase approaches, we can observe that the frequency of some query intents increases as the purchase approaches while that of the some other query intents decreases. The occurrences of query intents **series** and **model number** in issued queries increase as the purchase approaches. Conversely, the occurrence of the query intent **evaluation** gradually decreases as the purchase approaches⁵. These findings suggest that the users first search for information on a broad scope by looking for recommended prod-

⁵ Due to the space limitation, we only report the results related to users who issued five queries. We observed a similar trend with respect to users who issued six queries.

Table 3: Frequency of queries for each query intent for users who conducted five searches related to product purchases before and after the purchase.

Query intent	The n -th query pre-purchase					The n -th query post-purchase				
	n=5	n=4	n=3	n=2	n=1	n=1	n=2	n=3	n=4	n=5
Category	0.18	0.20	0.22	0.21	0.19	0.20	0.21	0.22	0.20	0.17
Manufacturer	0.17	0.18	0.21	0.22	0.22	0.19	0.22	0.21	0.20	0.18
Series	0.10	0.13	0.18	0.25	0.34	0.25	0.25	0.20	0.16	0.14
Model number	0.12	0.15	0.19	0.24	0.30	0.21	0.21	0.21	0.20	0.17
Function & feature	0.21	0.18	0.23	0.20	0.18	0.17	0.21	0.23	0.18	0.21
Store	0.19	0.21	0.19	0.19	0.21	0.22	0.20	0.21	0.19	0.19
Price	0.17	0.19	0.21	0.22	0.22	0.21	0.20	0.21	0.19	0.19
Guarantee	—	—	—	—	—	0.14	0.14	0.20	0.23	0.20
Evaluation	0.22	0.24	0.21	0.19	0.14	0.20	0.19	0.19	0.20	0.21
Sales method	0.21	0.18	0.16	0.21	0.24	0.23	0.19	0.18	0.15	0.25
Accessory	0.19	0.18	0.20	0.21	0.23	0.21	0.20	0.20	0.19	0.20
Shipping	0.19	0.22	0.21	0.21	0.17	0.21	0.22	0.19	0.20	0.18
How to use	0.21	0.27	0.17	0.20	0.15	0.16	0.18	0.20	0.20	0.25
New & used	0.22	0.19	0.18	0.19	0.22	0.15	0.20	0.20	0.23	0.22

ucts and then gradually narrow down the scope by issuing a specific model name of a product in their query.

As for the query intents after purchase, we can observe that the frequencies of **category**, **manufacturer**, **series** and **model number** decrease with increasing time after the purchase. This suggests that users search for information related to a product immediately after making a purchase.

5 Conclusion

In this study, we investigated how words used in Web search queries change over time from the pre-purchase to the post-purchase phase with respect to users who purchased a camera from an e-commerce site.

We classified pre- and post-purchase queries based on distinctive words found through manual categorization using a logistic regression. Our study shows that users search for diverse information both before and after making purchases.

Before making a purchase, **evaluation** searches decreased while **manufacturer** and **series** searches increased, revealing the product selection process. After making a purchase, users search **shipping**, **accessory**, and **how to use** information.

The findings of this study will help in understanding how users reach the point of purchase via Web search and their post-purchase information needs. In the future, we aim to explore search behavior variations based on user characteristics.

Acknowledgment

This work was supported in part by JSPS KAKENHI Grant Numbers JP21H03774, JP21H03775, JP22H03905.

References

1. Ashkan, A., Clarke, C.L.: Characterizing commercial intent. In: Proceedings of the 18th ACM conference on information and knowledge management. pp. 67–76 (2009)
2. Huang, J., Efthimiadis, E.N.: Analyzing and evaluating query reformulation strategies in web search logs. In: Proceedings of the 18th ACM conference on Information and knowledge management. pp. 77–86 (2009)
3. Jansen, B., Simone, S.: Bidding on the buying funnel for sponsored search and keyword advertising. *Journal of Electronic Commerce Research* **12**(1), 1 (2011)
4. Murga, G.: Amazon Takes 49 Percent of Consumers’ First Product Search, But Search Engines Rebound. <https://blog.survata.com/amazon-takes-49-percent-of-consumers-first-product-search-but-search-engines-rebound>
5. Sondhi, P., Sharma, M., Kolari, P., Zhai, C.: A taxonomy of queries for e-commerce search. In: Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. pp. 1245–1248 (2018)
6. Su, N., He, J., Liu, Y., Zhang, M., Ma, S.: User intent, behaviour, and perceived satisfaction in product search. In: Proceedings of the 11th ACM International Conference on Web Search and Data Mining. pp. 547–555 (2018)