Why People Search for Images using Web Search Engines

Xiaohui Xie
DCST, Tsinghua University
Beijing, China
xiexh_thu@163.com

Yiqun Liu
DCST, Tsinghua University
Beijing, China
yiqunliu@tsinghua.edu.cn

Maarten de Rijke
University of Amsterdam
Amsterdam, The Netherlands
derijke@uva.nl

Jiyin He
CWI
Amsterdam, The Netherlands
j.he@cwi.nl

Min Zhang
DCST, Tsinghua University
Beijing, China
z-m@tsinghua.edu.cn

Shaoping Ma
DCST, Tsinghua University
Beijing, China
msp@tsinghua.edu.cn

ABSTRACT

What are the intents or goals behind human interactions with image search engines? Knowing why people search for images is of major concern to Web image search engines because user satisfaction may vary as intent varies. Previous analyses of image search behavior have mostly been query-based, focusing on what images people search for, rather than intent-based, that is, why people search for images. To date, there is no thorough investigation of how different image search intents affect users' search behavior.

In this paper, we address the following questions: (1) Why do people search for images in text-based Web image search systems? (2) How does image search behavior change with user intent? (3) Can we predict user intent effectively from interactions during the early stages of a search session? To this end, we conduct both a lab-based user study and a commercial search log analysis.

We show that user intents in image search can be grouped into three classes: Explore/Learn, Entertain, and Locate/Acquire. Our lab-based user study reveals different user behavior patterns under these three intents, such as first click time, query reformulation, dwell time and mouse movement on the result page. Based on user interaction features during the early stages of an image search session, that is, before mouse scroll, we develop an intent classifier that is able to achieve promising results for classifying intents into our three intent classes. Given that all features can be obtained online and unobtrusively, the predicted intents can provide guidance for choosing ranking methods immediately after scrolling.

KEYWORDS

Image search; User intent; User behavior

ACM Reference Format:
Xiaohui Xie, Yiqun Liu, Maarten de Rijke, Jiyin He, Min Zhang, and Shaoping Ma. 2018. Why People Search for Images using Web Search Engines. In *Proceedings of WSDM 2018: The Eleventh ACM International Conference on Web Search and Data Mining (WSDM 2018).* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3159652.3159686

WSDM 2018, February 5–9, 2018, Marina Del Rey, CA, USA

ACM ISBN 978-1-4503-5581-0/18/02…$15.00
https://doi.org/10.1145/3159652.3159686

1 INTRODUCTION

Intent is assumed to be the immediate antecedent of behavior [2, 16]. For any information access service, it is important to understand the underlying intent behind user behavior. A widely used Web search intent taxonomy was proposed by Broder [5], based on log-based user studies. In Broder's taxonomy, search intent of Web search users is categorized into three classes: informational, transactional and navigational.

The diversification of online information and online information services that we have seen since the introduction of Broder's search intent taxonomy creates challenges to this taxonomy. Hence, several refinements of this taxonomy have been proposed [37, 38]. However, few efforts have been made towards understanding the intents of Web image search users. In image search engines, the items users search for are images instead of Web pages or online services. And the different result placement and interaction mechanisms of image search also make the search process rather different from general Web search engines [44]. Kofler and Lux [23] supported this view by conducting a user study and showing that without adaptation, user intent taxonomies applied in general web search are not applicable for image search. Similar to general purpose Web search [10, 14], we believe that a thorough understanding or even successful detection of users' image search intent helps search engines provide better image search results and improve their satisfaction. This motivates our first research question:

RQ1: Why do people search for images in text-based Web image search systems?

User behavior data has been successfully adopted to improve general Web searches in result ranking [1, 45], query suggestion [6, 43], query auto completion [21, 26], etc. We therefore believe that understanding user interaction behavior in the multimedia search scenarios will also provide valuable insight into the understanding of user intent. As user behavior varies with search intent, looking into differences in behavior and how such differences relate to search intent will help to improve the performance of image search engines.

Previous work on image search intent understanding usually focuses on the query proposed by users and assumes that the query represents user intent well. However, determining users' search intents or information needs based on the queries they submitted is sometimes rather difficult, as a large proportion of keyword based queries are short, ambiguous or broad [11, 20, 35]. Compared to general (Web) search, queries used in image search on the Web tend to be even shorter [13]. As the same query may come from different search intents, previous work attempts to bridge the **intent gap** between query and the underlying intent through explicit

methods, including query suggestion [46, 47] and result diversification [40, 42]. Besides work that investigates user intent based on query analysis, others look into the relationship between what people search for (query-based) and how they interact with image search engines. For example, Park et al. [33] categorize queries using two orthogonal taxonomies (subject-based and facet-based) and identify important behavioral differences across query types. Unlike previous work, we avoid analyzing query content and, instead, provide a thorough investigation into the whole interaction processes of users. As user behavior is directly affected by search intent, we propose an intent taxonomy based on the differences in user behavior patterns. This motivates our second research question:

RQ2: How does image search behavior change with user intent?

Automatically identifying search intent at an early stage of a search session helps a multimedia retrieval system to rerank its results according to the underlying user intent. Since pagination is usually not (explicitly) supported on image search SERPs, users can view results by scrolling up and down instead of clicking on the "next page" button. In this paper, we define the "early stage" of a search session as search behavior before any scrolling takes place. As user interactions with image search engines through mouse and keyboard can be captured online and unobtrusively, it will be practical to build intent recognition system based on these features if they are effective. This motivates our third research question:

RQ3: Can we effectively predict user intent from user interactions at the early stage of a search session?

To summarize, the main contributions of this work are:

- We propose a new image search intent taxonomy based on an open-coded discussion methodology [38]. The taxonomy includes three classes: Explore/Learn, Entertain, Locate/Acquire. As far as we know, ours is the first work to focus on an image search intent taxonomy in Web search engines. We verify the proposed image search intent taxonomy in two ways: through a user survey involving over 200 people and through a Web image search log analysis. Results show that the taxonomy covers a majority of user intents in Web image search.
- We design 12 tasks within the scope of the taxonomy and perform a lab-based user study to show that users interact with image search engines in different ways with different information needs. Differences are observed in temporal patterns, query reformulation patterns and mouse movement patterns.
- We build and evaluate a user intent recognition system based on the user interactions at the early stage of search sessions and achieve state-of-the-art prediction results.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 presents our user intent taxonomy in image search. Section 4 introduces our user-study settings and reports the findings. Intent prediction and its results are given in Section 5. Finally, Section 6 discusses conclusions and future work.

2 RELATED WORK

2.1 Search intent taxonomies

In the past two decades intent taxonomies for general Web search have been investigated by several researchers. Broder [5] introduced a taxonomy of user intent in text-based Web search using both randomly selected search queries and an analysis of survey data collected from AltaVista users. The taxonomy consists of three categories: Informational, Navigational and Transactional. In [19], session characteristics of these three top-level search intents were examined and used to develop a classification algorithm. Building on Broder's taxonomy, Rose and Levinson [37] introduced a subclassification of the taxonomy to classify intent more precisely. By having humans assign task-type labels to search sessions based on Rose and Levinson's taxonomy, Russell et al. [38] found that it is hard to get sufficient inter-rater agreement on ambiguous search tasks and proposed a new search task taxonomy that contains seven categories. Taxonomies investigated in general Web search mentioned above were found to be *not* applicable in image search [23, 24].

Previous work on an intent taxonomy for image search mostly focused on *what* people search for not *why* they search. Pu [34] classified 1,000 frequent image queries based on a proprietary subject-based categorization scheme. By focusing on whether users were searching for people, location, etc., and on whether the search was about unique instances or non-unique instances, Jansen [18] classified queries based on three non subject-based image query classification schemas. Lux et al. [29] are among the first to investigate the image search intent problem. They categorized user intents into knowledge orientation, mental image, navigation, and transaction; these intents describe search activities in Flickr, a digital photo sharing platform instead of a search engine. After that, a two-dimensional taxonomy was proposed in [7]. However, this work was based on another sharing platform (Pinterest). Neither works were conducted in Web search engines which may face more complex search scenarios. Redi et al. [36] showed that Web image search must deal with images from a wide variety of sources including very poor quality images typically absent in photo sharing platforms. Thus, in this paper, we look more deeply into user intent and build a search intent taxonomy in text-based Web image search systems. As far as we know, this work is among the first to discuss the image search intents in Web search engines.

2.2 User behavior in image search

Several studies analyze the user behavior logs of image search engines [3, 13, 31, 34, 41]. Many features, such as query reformulation patterns, session length, and the number of viewed result pages are recorded and investigated. Compared to (text-based) general Web search, image search leads to shorter queries, tends to be more exploratory, and requires more interactions. Interactions with image search result pages contain abundant implicit user feedback. Previous studies on multimedia search [15, 17, 32] explored user click-through data to bridge user intent gaps for image search. O'Hare et al. [31] proposed a number of implicit relevance feedback features based on additional interactions including hover-through rate, converted-hover rate, and referral page click-through to improve image search result ranking performance. These findings help us to understand how users search for images on the Web, but do not capture variation among types of image search intent.

Although image search tends to be more exploratory, image searches can also be intent-directed [3]. Park et al. [33] analyzed a large-scale query log from Yahoo Image Search to investigate user behavior toward different query types and identified important behavioral differences across them. The major difference between their work and ours is that they tried to link query type based behavior to two of four classes of image search intent proposed by Lux et al. [29], which had been derived from user queries on a

Table 1: Age distribution of participants in our user survey and Chinese internet users according to the 38th statistical report of China internet development [8].

Age	Proportion (Survey participants)	Proportion (Chinese internet users)
(, 20)	28.4%	23.0%
[20, 30)	61.2%	30.4%
[30, 40)	5.7%	24.2%
[40, 50)	4.3%	13.4%
[50,)	0.5%	9.0%

photo sharing platform. Since user behavior in search is heavily dependent on intent [27], it is likely that behavior varies across different search intents. Understanding such differences, and how they relate to image search intent in Web search engines, is the main focus of this paper.

3 USER INTENT TAXONOMY

We start by explaining how we used an open-coded discussion methodology to arrive at our proposed image search intent taxonomy. We then verify the proposed taxonomy through a user survey involving over 200 people and through a Web image search log analysis.

3.1 Establishing an intent taxonomy for image search

To build our intent taxonomy, we conducted both an online survey and a Web search log analysis. We applied the methodology proposed in [38] to generate our criteria to categorize image search intents. In [38], an open-coded discussion was performed by Web research professionals based on 700 anonymized sessions aimed at identifying search task categories in general Web search.

3.1.1 User survey. Besides several basic demographic questions, we ask participants to answer two open-ended questions to describe their most recent image search experience. As shown in [29], interviews with search users can bring us a more comprehensive understanding of search intent.

- Please describe your most recent image search experience with as many as details as possible (e.g., time, place, motivation).
- Please provide all the queries you used in this search (you can look into your search history if necessary).

In order to make our survey more accurate, we suggest participants to check their search history to help recall the experience. We spread our survey through a widely used social platform (Wechat) and paid participants about US$0.50 if they answered the questions seriously. A total of 258 people participated in our survey; after removing noise from the answers (e.g., answers that are too short or not about text-based Web image search), 211 valid cases were kept, which are from 47.9% female users and 52.1% male users. The age distribution of our survey participants is shown in Table 1 together with the age distribution of Chinese internet users [8]. From Table 1, we can observe that the age distribution of our survey is similar to that reported in [8], except that the number of people in their 20s of our survey is much higher.

3.1.2 Search logs. From the search logs of a popular commercial image search engine, Sogou,[1] we sampled 500 search sessions during five days in April 2017. Each session contains consecutive queries produced from a single user within 30 minutes. We removed 25 pornographic search sessions and retained the other 475 sessions to avoid disturbing participants. The number of sessions examined here is similar to [38]. The minimum session length is 1 which means the user only submits one query in this session, while the maximum session length is 29. The average session length is 3.44. We show the distribution of session length in these 475 sessions in Table 2 (column 2). From Table 2, we can observe that short sessions account for a large proportion.

[1] http://sogou.com

Table 2: Distribution of session length of sampled search sessions after filtering pornographic searches and the Fleiss' Kappa [12] for sessions with different session length.

Session length	Proportion	Fleiss' Kappa
1	24.4%	0.419
2	30.1%	0.311
3	13.3%	0.317
4	9.1%	0.463
5	6.9%	0.237
> 5	16.2%	0.408
All	100%	0.375

3.1.3 Categorization criteria. After collecting the online survey data and search logs, a group of three web research professionals was recruited to review all 211 survey data cases and 475 search sessions. The researchers read each session closely and discussed to determine criteria to categorize user intent. Following [38] we performed several iterations of this open-coded discussion. In each iteration, the proposed criteria and corresponding intent taxonomy were fine-tuned to cover as many search sessions as possible and to be easy to state and understand. Based on the results of the discussion, two criteria were proposed to categorize user intent:

Criterion 1 Is the user's search behavior driven by a clear objective?

Criterion 2 Does the user need to download the image for further use after the search process?

We first divided user intent into two groups according to Criterion 1. In some cases, people regard image search as an option for entertainment. They can freely browse the image search results without prior requirements for the images. One can enjoy photos of her/his favorite stars or some humorous images without following someone's account on social media communities, which makes image search an easy way to relax. We label this kind of user intent with "*Entertain*."

In other cases, people may want to find specific images that should meet several requirements they have in mind. We further discussed intent categorization by considering Criterion 2 for such cases. In some cases, people have to download images for further use. They may already have captions for these images. For example, they may want to write a report on the 2016 US presidential elections and need to find an image that shows two presidential candidates in a television debate. We call user intent in this group "*Locate/Acquire*:" people want to find and download images for which they already have some requirements, to complete some tasks.

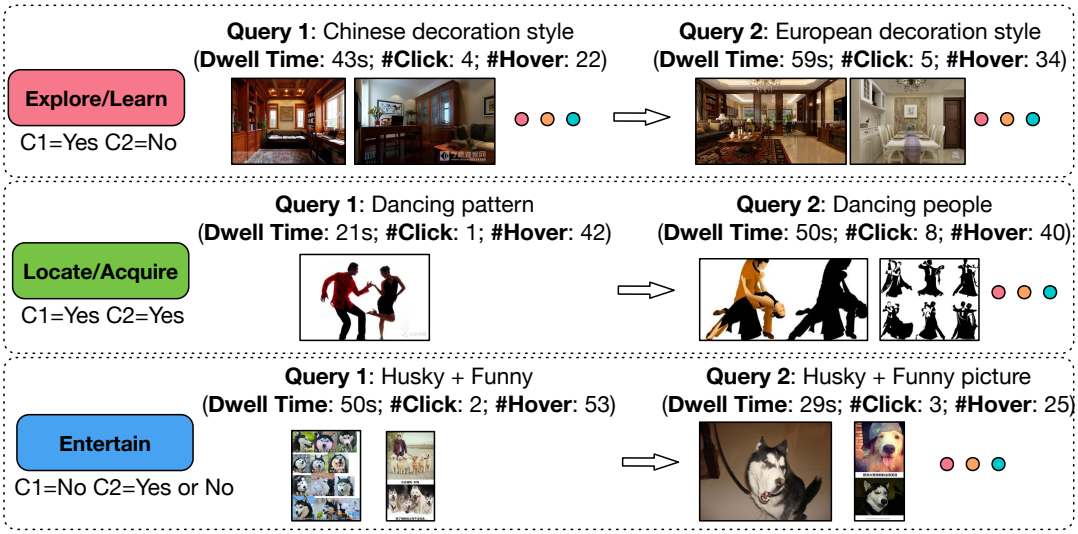


Figure 1: Examples of search queries, user behaviors and clicked images in typical search sessions with different user intents ("Explore/Learn", "Locate/Acquire", "Entertain").

For other tasks, people are capable of satisfying their information need without downloading images. Through image search engines, they can obtain, check or compare information by examining images in result pages only. To be more specific, consider three queries "New Ferrari 458," "a flower with purple stamen" and "decoration style.". The first query may come from the need that people want to know about the appearance of the new Ferrari car. When people want to find out the name of a flower that they saw and remember its major characteristics, they may use the query "a flower with purple stamen." For the third one, people may want to compare different characteristics of different decoration style in order to find the best one for their houses. We call this kind of user intent "*Explore/Learn*:" people expect to obtain informational gain from search result pages and their need can be satisfied without downloading images.

According to Criterion 1 and 2, we group user intents into three categories as follows. We use $C1$ and $C2$ to denote the answer to Criterion 1 and 2, respectively (e.g., $C1$-YES means that the user's search behavior is driven by a clear objective):

(1) **Explore/Learn**. Users want to learn something, confirm or compare information by browsing images. ($C1$-YES, $C2$-NO)
(2) **Locate/Acquire**. Users want to find images for further use. They already have some requirements about these images. ($C1$-YES, $C2$-YES)
(3) **Entertain**. Users want to relax and kill time by freely browsing the image search results. ($C1$-NO, $C2$-YES or NO)

In Fig. 1, we show examples of search queries, user behaviors and clicked images in typical search sessions with different user intents according to the collected Web user behavior log data.

Our intent taxonomy is similar to the Web search task taxonomy proposed by Russell et al. [38]. There are important differences, however. For a start, we do not have their "Navigate" and "Meta" in our intent taxonomy, because people rarely use queries leading to a site or test web sites' capabilities in image search tasks. Furthermore, "Find-Simple" and "Find-Complex" are not in our taxonomy as they are covered by "Explore/Learn." Russell et al. [38]'s "Find-Simple" is a scenario where an information need can be satisfied with a single query and a single result; their "Find-Complex" is a scenario where a user has to search for information that requires multiple searches on related topics, inspect multiple sources, and integrate information across those sources. Finally, we rename Russell et al. [38]'s "Play" to "Entertain" in our intent taxonomy as "Play" in Russel's taxonomy focuses more on transactional needs.

3.2 Verifying the intent taxonomy

To verify our proposed intent taxonomy, we asked three annotators (who are all graduate students majored in computer science and are different from the people who proposed the categorization criteria mentioned in Section 3.1) to manually annotate the search scenarios collected in the survey into our three intent categories (Locate/Acquire, Explore/Learn, Entertain). We provided them with the definitions of each user intent specified in the proposed taxonomy mentioned in Section 3.1. We also gave our annotators two other choices: "Difficult to classify" and "Others." "Others" indicates that the user's intent cannot be fitted into any of the three intent categories and "Difficult to classify" means that the user's intent seems to belong to two or more classes in our proposed taxonomy.

For 203 out of the 211 valid online survey cases our annotators obtain a majority agreement, meaning that at least two raters assign the case into the same category. The Fleiss Kappa score [12] is 0.673 among the three annotators, which constitutes a substantial agreement [25]. The number of cases assigned to "Others" by our three annotators are 1 (0.47%), 2 (0.94%), 5 (2.37%), while the numbers for "Difficult to classify" are 3 (1.41%), 0 (0%), 2 (0.94%), respectively; a closer look at those cases reveals that the descriptions of these cases are vague. Based on these annotation results, we conclude that the proposed image search intent taxonomy covers most of users' actual intents and that the taxonomy is easy to use and apply for annotators.

We employed the same annotators to annotate our search logs. Our annotators were only shown the list of queries, i.e., they were not given hits or clicks. Again, we provided them with the definitions of each user intent in the taxonomy mentioned before. The choices of "Others" and "Difficult to classify" were also given. The Fleiss Kappa score is 0.375 among our three annotators, which constitutes a fair agreement. Also, we list the Fleiss Kappa scores for

different session lengths in Table 2. We can observe that when the session length is around 4, the annotation agreement is highest, leading to moderate agreement. Compared with the substantial agreement reported in survey verification, it seems that by just examining queries, annotators cannot fully capture users' intents. This result echos similar observations by Russell et al. [38]. The result motivated us to further investigate the signals in user behavior that can be applied to distinguish search intents; see Section 4.

Through the user survey involving over 200 people and the Web search log annotations, we show that our proposed taxonomy can cover a majority of user intents and the boundary between different intents is detectable for annotators. The distribution of three intents in user survey is 27% (Explore/Learn), 66% (Locate/Acquire) and 7% (Entertain). In the search log the distribution is 56% (Explore/Learn), 39% (Locate/Acquire) and 5% (Entertain).

In summary, our answer to RQ1 is that user intents in image search on the web can be grouped into three classes: Explore/Learn, Locate/Acquire, and Entertain.

4 IMAGE SEARCH WITH DIFFERENT INTENTS

Armed with our intent taxonomy for image search, we address RQ2, "How does image search behavior change with user intents?" by conducting a lab-based user study. In this user study, we pre-design a set of tasks based on the proposed user intent taxonomy.

4.1 User behavior dataset

4.1.1 User study task. We designed 12 tasks based on the results of the user survey mentioned in Section 3.1.1. Each category of the proposed taxonomy accounts for 4 tasks. Examples of the user study tasks are shown in Table 3.

In order to simulate a realistic image search scenario, for the "Locate/Acquire" tasks, we not only ask participants to complete the search part of the tasks but also ask them to use the images they find to create some multimedia productions (e.g., a slide, a poster, a computer desktop). We provided them with frequently used software to help them make the productions, with several default settings chosen by us, including the text part and background, which guarantees that the participants only need to use the image search engine to complete their tasks. Through this setup, we want to ensure that each participant faces the same task difficulty. For example, in one of the "Locate/Acquire" tasks, participants are asked to make a slide about Harry Potter. We pre-set the theme of the slide as "The movie characters of Harry Potter" and provide the names of three characters on the slide; the participants need to find posters of these three characters and coordinate different posters and the background. For the "Explore/Learn" tasks, we ask participants to verbally answer certain questions related to the query to ensure that the task is done seriously. For example, in one of the "Explore/Learn" tasks, participants are asked to find the name of a flower that has some characteristics. As the scenario assumes that participants already saw the flower, we provide an image of the flower before they search to make sure they have a mental impression. For the "Entertain" tasks, participants can freely browse the image search results to relax. We only pre-set the theme of the task without any further constraints.

4.1.2 Data collection procedure. In the user study, each participant was asked to complete all 12 image search tasks, which were offered in a random order. Compared with collecting data from real search logs, or by browsing plugins, the laboratory user study has

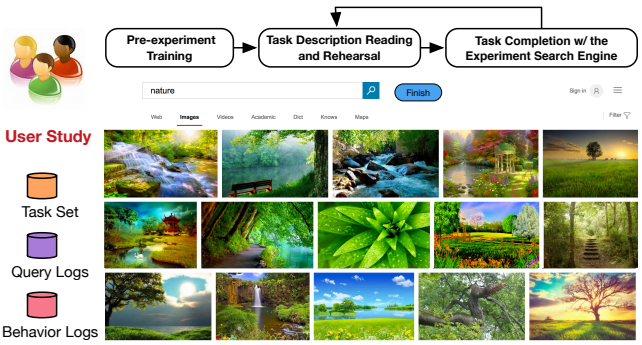


Figure 2: Data collection procedure. We designed our tasks based on the proposed user intent taxonomy. With enrolled participants, we collected query logs and behavior logs.

a smaller scale, but it does allow us to fully control the variabilities in search tasks and information needs.

We recruited 35 undergraduate students, 13 female and 22 male, via email and online social networks, to take part in our user study. The ages of participants range from 18 to 25 and their majors include engineering, humanities and social science, and arts. All participants were familiar with basic usage of Web search engines. It took about an hour and a half to complete the user study. And we paid the participants about US$25 after they completed all the tasks seriously.

To make sure that every participant was familiar with the experimental procedure, an example task was used for demonstration in a Pre-experiment Training stage. In the example task, we asked participants to use the image search engine to learn how to tie a tie and they were required to describe the process in voice after searching. They could browse and click the search results, and adjust the sensitivity of the mouse to the most appropriate level as well. We did not give any further instructions on which results to click on or when to end until participants were familiar with the experimental search engine. After the Pre-experiment training stage, they were asked to complete all 12 image search tasks. For each task, the participants had to go through 2 stages, as shown in Fig. 2. Firstly, the participants should read and memorize the task description (note that the complete task description is provided to the participants). After that, they were required to describe the task in voice without viewing it. Then they can push a "Begin Task" button and will be redirected to an experimental search engine. Like Web image search scenario, they could scroll to move the page up and down, use the mouse to see hover text, and click a thumbnail to view and download the full-size image in the preview page. While no task time limits were imposed, they could stop searching and click the finish button when they thought that the task was completed or no further helpful information would be found.

We injected customized JavaScript into search result pages to log mouse activities on search pages when users perform search tasks. The search system was deployed on a 17-inch LCD monitor with a resolution of 1366×768. The Google Chrome browser was used to display results of search system. A database with full records of the experiments is available online for academic research.[2]

[2] https://tinyurl.com/y8pa8zk6

Table 3: Examples of user study tasks.

Category	Goal	Constraint	Success Criteria
Explore/Learn	Imagine you prepare to renovate a new house. You would like to compare different decoration styles (e.g., Chinese style, Simple European style).	–	Please introduce and compare the characteristics of different decoration styles in voice
Locate/Acquire	Please change the desktop background of this computer.	The background image should have blue sky and forest.	Change the desktop background to the required image.
Entertain	Now take a break, you can browse some posters or photos of your favorite stars.	–	–

4.2 User interaction features

From the query logs and behavior logs, we extracted 28 features that can be grouped into 6 types. We calculated all features both **globally** (i.e., using all data) and **at the early stage of a search session** (i.e., only using data captured before the first mouse scroll). Table 4 lists the complete list of 28 features.

Table 4: The list of 28 features extracted from the logged implicit user interaction with the image search engine. ("*" means that a type of feature can be extracted both in a global view and at the early stage of a search session).

Feature type	Description	#
Dwell time*	The dwell time on the SERP	1
Mouse clicks*	Number of clicks, first click time	2
Mouse hover*	First and longest hover time	2
Mouse movement*	Min, max, mean and median of the mouse movement speed at three directions(original, X-axis, Y-axis), and the mouse movement angle and radian.	20
Query reformulation	Adding and deleting terms, partially change.	3

Temporal information is a widely-used feedback feature in the setting of document relevance [4, 28], and it varies with different search tasks [22]. In this paper, we considered temporal information including dwell time on the SERP, time to first click, time to first hover and longest hover time on the images. We also examined the number of clicks on the SERP as clicks are strongly correlated with relevance and examination [9]. Temporal information is also dependent on image search query types [33].

Mouse movement features, which were explored in previous image search analyses [39], are investigated in this paper as well. Especially, as the placement of image search result is a two-dimensional grid instead of a linear result list, we considered not only the speed of mouse movement in the original direction but also in the X-axis (horizon) and Y-axis (vertical) direction.

Finally, query reformulation patterns in a search session were investigated. Previous studies show that query reformulation patterns vary with search goals [16], they occur frequently on image search platforms [3], and our data also provides evidence to support this, with approximately 76.7% of the sessions involving more than one query. Query reformulations between consecutive queries can be grouped into four categories [19]: adding terms, deleting terms, partial change, and complete change. Because the participants had to search for different items in some tasks, which was caused by the task description we provided, not by the user's cognitive processes, we only compare differences between adding terms, deleting terms, and partial change.

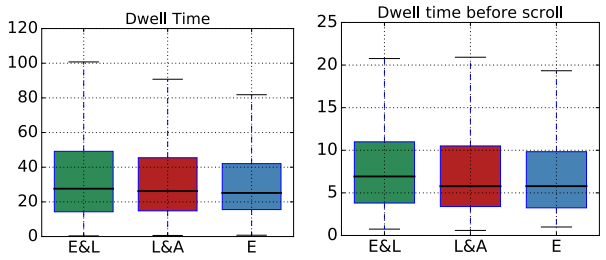
4.3 Statistical analysis

In this subsection, we report the relationship between the extracted features and search intents using box plots. We also perform a series of one-way ANOVA tests and pairwise t-tests to determine the significance. We show the box plots in Fig. 3. The difference in query reformulation patterns across different search intents is plotted in Fig. 4. The results of the ANOVA tests (ANOVA-p) are reported in Fig. 3 and Fig. 4. The results of the pairwise t-tests (p) are discussed in Sections 4.3.1–4.3.5.

4.3.1 Dwell time. From Fig. 3(a) and 3(b), we can observe that the mean dwell time in "Explore/Learn" tasks is longer than in "Entertain" tasks, both globally and at the early stage of a search session ($p < 0.02$ and $p < 0.07$, respectively). And the mean dwell time in "Locate/Acquire" tasks is longer than in "Entertain" tasks globally ($p < 0.05$). Before scrolling, dwell time on the SERP is significantly different between "Explore/Learn" and "Locate/Acquire" ($p < 0.05$). Recall our criteria for categorizing user intent: as users' search behavior in "Explore/Learn" and "Locate/Acquire" tasks is driven by a clear objective, there will exist more confirmation and comparison of image content, which results in more time spent on search engine result pages.

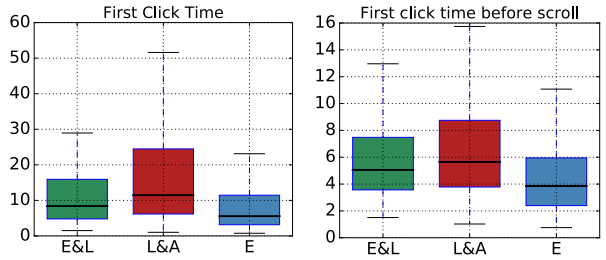
4.3.2 Mouse clicks. The number of clicks shows significant differences between the three intent classes ($p < 0.01$). The mean number of clicks in the three classes follows this relative order: "Explore/Learn" < "Locate/Acquire" < "Entertain." The first time to click is also a useful implicit feedback signal in differentiating different intents. From Fig. 3(c) and 3(d), we can observe that the average first click time of queries driven by the "Locate/Acquire" intent is longer than for queries with another intent, both globally and at the early stage of search (both have $p < 0.001$). The average first click time in "Explore/Learn" tasks is longer than in "Entertain" tasks before scrolling ($p < 0.001$) as well. When performing a "Locate/Acquire" or "Explore/Learn" task, users already have some specific requirements about the images. Image search results are self-contained, so that users do not need to click the document as in general Web search to view the landing page. Instead, they can observe several images before deciding which ones to download or to click to see a larger version. For this reason, users will spend more time on the search result page before the first click.

4.3.3 Mouse hover. Unlike the time to first click, time to first hover shows no significant differences between the three intents. However, the mean longest hover time in "Entertain" tasks is shorter than in "Locate/Acquire" and "Explore/Learn" tasks, both globally ($p < 0.001$) and at the early stage of search ($p < 0.01$), as shown in



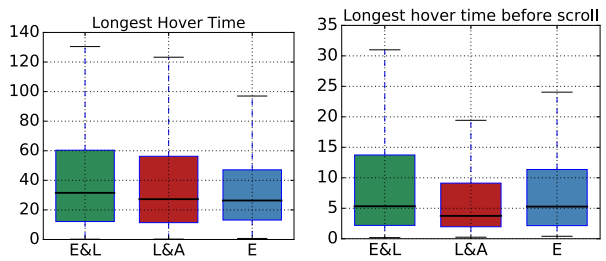
(a) ANOVA-$p < 0.05$

(b) ANOVA-$p < 0.1$



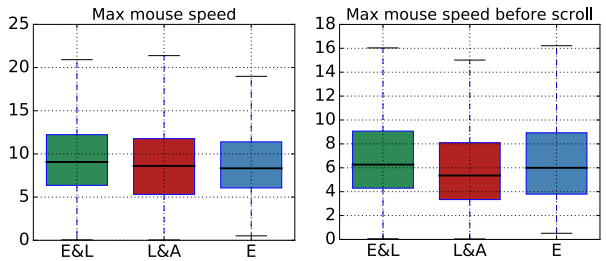
(c) ANOVA-$p < 0.05$

(d) ANOVA-$p < 0.1$



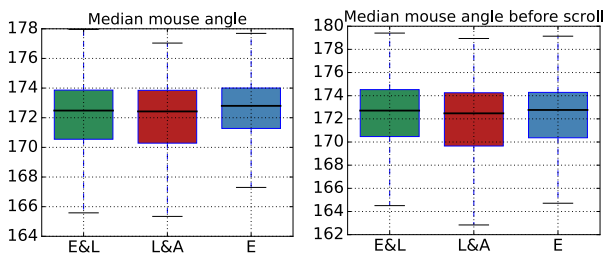
(e) ANOVA-$p < 0.01$

(f) ANOVA-$p < 0.01$



(g) ANOVA-$p < 0.05$

(h) ANOVA-$p < 0.01$



(i) ANOVA-$p < 0.01$

(j) ANOVA-$p < 0.05$

Figure 3: The extracted features on the SERP boxplots for queries with different search intents *globally* (a, c, e, g, i) and *at the early stage of search* (b, d, f, h, j). The unit on the y-axis is second for a, b, c, d, e, f; pixels per second for g, h; and number of degree for i, j. E&L denotes Explore&Learn, L&A denotes Learn&Acquire and E denotes Entertain.

Fig. 3(e) and 3(f), respectively. At the early stage of a search session, the mean longest hover time in "Locate/Acquire" is shorter than in "Explore/Learn" tasks ($p < 0.001$). As hovering on a document can be regarded as a signal that users are inspecting it; our results may be caused because "Explore/Learn" tasks require more complex cognitive processes, hence users may need to compare the image content with the background knowledge and mental impressions of their task.

4.3.4 Mouse movement. The speed of mouse movement varies between intents. As shown in Fig. 3(g) and 3(h), the average maximum speed of mouse movement in "Entertain" tasks is lower than in "Explore/Learn" tasks ($p < 0.02$) and "Locate/Acquire" tasks ($p < 0.08$), globally. And at the early stage of a search session, "Explore/Learn" tasks obtain a higher average max speed than "Locate/Acquire" tasks ($p < 0.001$). Besides in the original direction, the speed of mouse movements also shows significant differences along the X-axis and Y-axis (e.g., "Explore/Learn" tasks receive the highest average mean speed of mouse movement along the Y-axis while "Locate/Acquire" tasks receive the lowest, both globally ($p < 0.001$) and before scrolling ($p < 0.05$)). For the angle of the mouse movement, the median angle of mouse movement between the three intents is significantly different ($p < 0.01$, $p < 0.05$, and $p < 0.01$, respectively), globally. And the differences are also significant between "Locate/Acquire" and "Entertain" ($p < 0.05$), before scrolling. Thus, mouse movement patterns have the potential to help us recognizing image search intents.

4.3.5 Query reformulation. Fig. 4 shows the average number of times a query reformulation occurs, across different search intents. The numbers indicate the number of reformulations per task, on average (globally). We observe that for "Locate/Acquire" tasks, participants tend to refine queries more frequently, which serve as evidence of focused search behavior. In "Locate/Acquire" tasks, users need to find the most appropriate images to create some productions. For example, images used as materials in designing a poster should fit the background and other materials, which means that the user needs to try different styles of images until the poster looks aesthetically acceptable. Thus, "Locate/Acquire" tasks receive a larger number times for query reformulation. In contrast, in "Entertain" tasks, the users' search behavior is not driven by a clear objective. This more exploratory, browsing-like behavior results in a smaller number of query reformulations. We performed a paired two-tailed t-test to verify the significance of the observed differences: $p < 0.01$ for all comparisons except for for differences in adding terms and deleting terms between Explore/Learn and Locate/Acquire ($p < 0.05$).

In summary, through our user study, we collected query logs and behavior logs. Based on these data, we are able to answer *RQ2* by investigating several frequently used implicit signals in general Web search. We find that temporal information and mouse movement patterns are useful in distinguishing search intents. Also, the cognitive process under different intents may result in different types and numbers of times of query reformulation.

5 IMAGE SEARCH INTENT PREDICTION USING EARLY STAGE FEATURES

The features discussed in Section 4.3 show potential in helping us identifying search intent automatically. In this paper, we utilize interaction features at the early stage of a search session to build an early stage user intent recognition system aimed at addressing *RQ3*. We aim to predict intent at the query level. If our features are

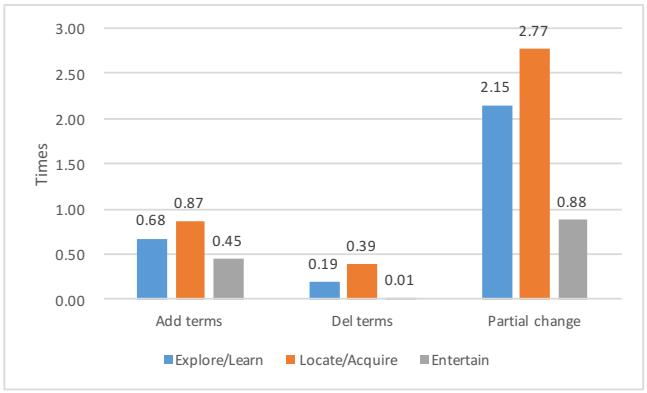


Figure 4: Average number of times of query reformulations per task, across different search intents (ANOVA-$p < 0.01$ for every type of query reformulation).

effective, this system can be practical for an image search engine to rerank its results even before users begin to scrolling, which will likely improve the satisfaction of users.

We compare different combinations of features. Due to a lack of space we do not consider all features described in Table 4 but some natural groupings only. In particular, we use "Time" to denote the combination of dwell time, time to first click and time to first hover. And as the number of clicks and first hover time show no significant difference under different intents at the early stage of a search session, we do not include it in the set of features that we consider. Also, we do not use the query reformulation patterns because we want to predict the user intents at the query level. We concatenate various features into a long feature vector to fuse all features (which is known as "early fusion" of different feature groups [39]). As this task can be treated as a multi-class classification problem, we apply a gradient boosting classifier [30] and perform 10-fold cross validation. We assign the label from the majority class to all the instances to generate a baseline.

The results are shown in Table 5. We can observe that our recognition system with user behavior features outperforms the baseline significantly ($p < 0.001$). And temporal features are more effective than the other two combinations. Fusing all features together achieves better prediction results than only using a single feature group. However, early fusion does not lead to a large increase in performance over the best single group of features. We compared another fusion method, i.e., "late fusion" in which each feature group has its own classifier and the output of all classifiers are combined to obtain a final result. We used the weighted sum of scores for "late fusion," similar to [39]. The results are also shown in Table 5. We see that "late fusion" and "early fusion" perform very similarly. It is worth pointing out that in absolute terms our performance figures are similar to other intent classification tasks considered in the literature, such as [7, 39], even though we use a much sparser signal than [7] and, unlike [39], only use features that can be collected in real-world scenarios.

Finally, concerning *RQ*3, we have found that based on interactive features at the early stages of a search session, we can recover user intents effectively through a combination of temporal features and mouse movement features.

6 CONCLUSION AND FUTURE WORK

In this paper, we proposed a new user intent taxonomy for image search and verified the taxonomy through a user survey involving

Table 5: Classification performance based on the user interactions features in terms of weighted average of F-1 score. Best results are in boldface. (E&L denotes Explore&Learn, L&A denotes Learn&Acquire and E denotes Entertain).

Features	All classes	E&L vs. L&A	E&L vs. E	L&A vs. E
Baseline	0.28	0.44	0.42	0.54
Time (#3)	0.42	0.56	0.56	0.64
Mouse move speed (#12)	0.40	0.55	0.54	0.60
Mouse move angle (#8)	0.38	0.55	0.54	0.61
Late fusion all features	0.44	0.57	0.56	0.64
Early fusion all features	**0.45**	**0.58**	**0.58**	**0.65**

over 200 people. Based on a lab-based user study, we discovered significant differences in user behavior under different intents. Finally, we used these behavioral signals to recover user intents and achieved promising results.

Implications. As user intents can be different in image search scenarios, considering an evaluation metric appropriate for different intents can be beneficial. Also, recommender systems could prioritize showing specific, targeted content to users based on their search goals. Last but not the least, the optimization goal of search engines should be designed according to different search intents. For example, for users with an "Entertain" intent, the goal may be to keep them engaged with the image search engines for as long as possible. This is where our work contributes.

Future work. Interesting directions for future work include investigating the intent prediction beyond the early stage, e.g., by incorporating content features of query and images besides user interaction signals. Moreover, we plan to design strategies to rerank image search results according to different intents, with the aim to improve user satisfaction. Also, a parallel comparison between video search intents and image search intents might be interesting as both belong to multimedia search.

Acknowledgments. This work is supported by the Natural Science Foundation of China (Grant No. 61622208, 61732008, 61532011), National Key Basic Research Program (2015CB358700), Ahold Delhaize, Amsterdam Data Science, the Bloomberg Research Grant program, the Criteo Faculty Research Award program, Elsevier, the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement nr 312827 (VOX-Pol), the Microsoft Research Ph.D. program, the Netherlands Institute for Sound and Vision, the Netherlands Organisation for Scientific Research (NWO) under project nrs. 13675, 612.001.116, CI-14-25, 652.002.001, 612.001.551, 652.001.003, and Yandex. All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.

REFERENCES

[1] Eugene Agichtein, Eric Brill, and Susan Dumais. 2006. Improving web search ranking by incorporating user behavior information. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 19–26.

[2] Icek Ajzen. 2002. Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology* 32 (2002), 665–683.

[3] Paul André, Edward Cutrell, Desney Tan, and Greg Smith. 2009. Designing novel image search interfaces by understanding unique characteristics and usage. *Human-Computer Interaction–INTERACT 2009* (2009), 340–353.

[4] Alexey Borisov, Ilya Markov, Maarten de Rijke, and Pavel Serdyukov. 2016. A context-aware time model for web search. In *SIGIR 2016: 39th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 205–214.

[5] Andrei Broder. 2002. A taxonomy of web search. In *ACM Sigir forum*, Vol. 36. ACM, 3–10.

[6] Huanhuan Cao, Daxin Jiang, Jian Pei, Qi He, Zhen Liao, Enhong Chen, and Hang Li. 2008. Context-aware query suggestion by mining click-through and session data. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 875–883.

[7] Justin Cheng, Caroline Lo, and Jure Leskovec. 2017. Predicting intent using activity logs: How goal specificity and temporal range affect user behavior. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 593–601.

[8] I CNNIC. 2016. The 38th Statistical report on internet development in China. *China Internet Network Information Center (CNNIC), China* (2016).

[9] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. 2008. An experimental comparison of click position-bias models. In *Proceedings of the 2008 international conference on web search and data mining*. ACM, 87–94.

[10] Honghua Kathy Dai, Lingzhi Zhao, Zaiqing Nie, Ji-Rong Wen, Lee Wang, and Ying Li. 2006. Detecting online commercial intention (OCI). In *Proceedings of the 15th international conference on World Wide Web*. ACM, 829–837.

[11] Zhicheng Dou, Ruihua Song, and Ji-Rong Wen. 2007. A large-scale evaluation and analysis of personalized search strategies. In *Proceedings of the 16th international conference on World Wide Web*. ACM, 581–590.

[12] Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin* 76, 5 (1971), 378–382.

[13] Abby Goodrum and Amanda Spink. 1999. Visual information seeking: A study of image queries on the world wide web. In *Proceedings of the ASIS Annual Meeting*, Vol. 36. ERIC, 665–74.

[14] Sha Hu, Zhicheng Dou, Xiaojie Wang, Tetsuya Sakai, and Ji-Rong Wen. 2015. Search result diversification based on hierarchical intents. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 63–72.

[15] Xian-Sheng Hua, Linjun Yang, Jingdong Wang, Jing Wang, Ming Ye, Kuansan Wang, Yong Rui, and Jin Li. 2013. Clickage: Towards bridging semantic and intent gaps via mining click logs of search engines. In *Proceedings of the 21st ACM international conference on Multimedia*. ACM, 243–252.

[16] Bouke Huurnink, Laura Hollink, Wietske Van Den Heuvel, and Maarten de Rijke. 2010. Search behavior of media professionals at an audiovisual archive: A transaction log analysis. *Journal of the Association for Information Science and Technology* 61, 6 (2010), 1180–1197.

[17] Vidit Jain and Manik Varma. 2011. Learning to re-rank: query-dependent image re-ranking using click data. In *Proceedings of the 20th international conference on World wide web*. ACM, 277–286.

[18] Bernard J. Jansen. 2008. Searching for digital images on the web. *Journal of Documentation* 64, 1 (2008), 81–101.

[19] Bernard J. Jansen, Amanda Spink, and Bhuva Narayan. 2007. Query modifications patterns during web searching. In *Information Technology, 2007. ITNG'07. Fourth International Conference on*. IEEE, 439–444.

[20] Bernard J. Jansen, Amanda Spink, and Tefko Saracevic. 2000. Real life, real users, and real needs: A study and analysis of user queries on the web. *Information Processing & Management* 36, 2 (2000), 207–227.

[21] Jyun-Yu Jiang, Yen-Yu Ke, Pao-Yu Chien, and Pu-Jen Cheng. 2014. Learning user reformulation behavior for query auto-completion. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. ACM, 445–454.

[22] Diane Kelly, Jaime Arguello, Ashlee Edwards, and Wan-ching Wu. 2015. Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*. ACM, 101–110.

[23] Christoph Kofler and Mathias Lux. 2009. Dynamic presentation adaptation based on user intent classification. In *Proceedings of the 17th ACM international conference on Multimedia*. ACM, 1117–1118.

[24] Christoph Kofler and Mathias Lux. 2009. An exploratory study on the explicitness of user intentions in digital photo retrieval. In *Proceedings of the 9th International Conference on Knowledge Management and Knowledge Technologies*.

[25] J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 1 (1977), 159–174.

[26] Yanen Li, Anlei Dong, Hongning Wang, Hongbo Deng, Yi Chang, and ChengXiang Zhai. 2014. A two-dimensional click model for query auto-completion. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. ACM, 455–464.

[27] Jingjing Liu, Michael J Cole, Chang Liu, Ralf Bierig, Jacek Gwizdka, Nicholas J Belkin, Jun Zhang, and Xiangmin Zhang. 2010. Search behaviors in different task types. In *Proceedings of the 10th annual joint conference on Digital libraries*. ACM, 69–78.

[28] Yiqun Liu, Xiaohui Xie, Chao Wang, Jian-Yun Nie, Min Zhang, and Shaoping Ma. 2016. Time-aware click model. *ACM Transactions on Information Systems (TOIS)* 35, 3 (2016), Article 16.

[29] Mathias Lux, Christoph Kofler, and Oge Marques. 2010. A classification scheme for user intentions in image search. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*. ACM, 3913–3918.

[30] Llew Mason, Jonathan Baxter, Peter L Bartlett, and Marcus R Frean. 2000. Boosting algorithms as gradient descent. In *Advances in neural information processing systems*. 512–518.

[31] Neil O'Hare, Paloma de Juan, Rossano Schifanella, Yunlong He, Dawei Yin, and Yi Chang. 2016. Leveraging user interaction signals for web image search. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 559–568.

[32] Bing Pan, Helene A. Hembrooke, Geri K. Gay, Laura A. Granka, Matthew K. Feusner, and Jill K. Newman. 2004. The determinants of web page viewing behavior: an eye-tracking study. In *Proceedings of the 2004 symposium on Eye tracking research & applications*. ACM, 147–154.

[33] Jaimie Y. Park, Neil O'Hare, Rossano Schifanella, Alejandro Jaimes, and Chin-Wan Chung. 2015. A large-scale study of user image search behavior on the web. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 985–994.

[34] Hsiao-Tieh Pu. 2005. A comparative analysis of web image and textual queries. *Online Information Review* 29, 5 (2005), 457–467.

[35] Davood Rafiei, Krishna Bharat, and Anand Shukla. 2010. Diversifying web search results. In *Proceedings of the 19th international conference on World wide web*. ACM, 781–790.

[36] Miriam Redi, Frank Z. Liu, and Neil O'Hare. 2017. Bridging the aesthetic gap: The wild beauty of web imagery. In *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*. ACM, 242–250.

[37] Daniel E. Rose and Danny Levinson. 2004. Understanding user goals in web search. In *Proceedings of the 13th international conference on World Wide Web*. ACM, 13–19.

[38] Daniel M Russell, Diane Tang, Melanie Kellar, and Robin Jeffries. 2009. Task behaviors during web search: The difficulty of assigning labels. In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on*. IEEE, 1–5.

[39] Mohammad Soleymani, Michael Riegler, and Pål Halvorsen. 2017. Multimodal analysis of image search intent: Intent recognition in image search from user behavior and visual content. In *ACM International Conference on Multimedia Retrieval*. 251–259.

[40] Bilyana Taneva, Mouna Kacimi, and Gerhard Weikum. 2010. Gathering and ranking photos of named entities with high precision, high recall, and diversity. In *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 431–440.

[41] Dian Tjondronegoro, Amanda Spink, and Bernard J. Jansen. 2009. A study and comparison of multimedia Web searching: 1997–2006. *Journal of the American Society for Information Science and Technology* 60, 9 (2009), 1756–1768.

[42] Reinier H. van Leuken, Lluis Garcia, Ximena Olivares, and Roelof van Zwol. 2009. Visual diversification of image search results. In *Proceedings of the 18th international conference on World wide web*. ACM, 341–350.

[43] Kuansan Wang, Nikolas Gloy, and Xiaolong Li. 2010. Inferring search behaviors using partially observable markov (pom) model. In *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 211–220.

[44] Xiaohui Xie, Yiqun Liu, Xiaochuan Wang, Meng Wang, Zhijing Wu, Yingying Wu, Min Zhang, and Shaoping Ma. 2017. Investigating examination behavior of image search users. In *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM.

[45] Danqing Xu, Yiqun Liu, Min Zhang, Shaoping Ma, and Liyun Ru. 2012. Incorporating revisiting behaviors into click models. In *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 303–312.

[46] Jinxi Xu and W. Bruce Croft. 1996. Query expansion using local and global document analysis. In *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 4–11.

[47] Zheng-Jun Zha, Linjun Yang, Tao Mei, Meng Wang, Zengfu Wang, Tat-Seng Chua, and Xian-Sheng Hua. 2010. Visual query suggestion: Towards capturing user intent in internet image search. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 6, 3 (2010), 13.