Understanding User Search Behavior Across Varying Cognitive Levels

Rishita Kalyani
L3S Research Center
Leibniz Universität Hannover
Hannover, Germany
kalyani@L3S.de

Ujwal Gadiraju
L3S Research Center
Leibniz Universität Hannover
Hannover, Germany
gadiraju@L3S.de

ABSTRACT

The ubiquitous accessibility of the world-wide web has led people to increasingly use web search to learn or acquire new knowledge. Recent research efforts have targeted the optimization of web search to satisfy learning related needs. However, there is little known about how one's search interactions differ across varying cognitive levels that correspond to one's learning. In this paper, we address this knowledge gap by investigating how the search interactions of 150 users vary across 6 search tasks corresponding to distinct cognitive levels. We also analyze how users' knowledge gain varies across the cognitive levels. Our findings suggest that the cognitive learning level of a user in a search session has a significant impact on the user's search behavior and knowledge gain.

Estimating the cognitive level of users during their interactions with search systems will allow us to construct and improve learning experiences for the users. For example, learners can be served content that corresponds to their current cognitive level within their learning process.

CCS CONCEPTS

• Human-centered computing; **• Applied computing → Law, social and behavioral sciences**; **• Information systems**;

ACM Reference Format:
Rishita Kalyani and Ujwal Gadiraju. 2019. Understanding User Search Behavior Across Varying Cognitive Levels. In *30th ACM Conference on Hypertext and Social Media (HT '19), September 17–20, 2019, Hof, Germany.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3342220.3343643

1 INTRODUCTION

One of the most common and frequent uses of the Web is to find information. Whether it is to find answers for a general question or to determine who won the best actor award at the recent Oscars, search is ingrained in our lives. Though it may appear that information seeking type of searches are most common according to classic information retrieval (IR), not all search queries are for information need [2]. Broder categorized search queries broadly into three types - (i) navigational - queries that point to a particular domain or website, (ii) informational - queries that seek information, and (iii) transactional - queries with an intent of completing a transaction. Such categorization revises the basic IR model, but does not help in identifying queries with learning needs, especially over different cognitive learning levels during web search.

Information look-up tasks on the web are often simple and may get completed in a short search session. However, this may differ in learning scenarios. The queries fired can be long and the search tasks required until the entire learning process ends may span over several sessions [4, 15]. The importance of learning as an outcome of web search has been recognized by recent works [26]. Wu et al. predicted the difficulty of search tasks from query and mission-related features [30]. Collins-Thompson et al. investigated the aspects of search interaction which are effective for supporting superior learning outcomes [4]. Yu et al. proposed models for predicting knowledge states and analyzed knowledge gain (KG) of users in informational search sessions [12, 31]. However, there is a lack of understanding of how the search behavior of users and their knowledge gain are affected by the varying cognitive levels of user's learning process.

An ideal search system that can support learning should be able to help users in finding, understanding, analyzing, evaluating and processing documents that contain the information which can provide answers to complex information needs and aid the learning process of users along the way. In order to realize such search systems, it is important to first understand how users' search behavior varies with respect to the cognitive level of the learning process they are embedded in, within a given informational search session. To this end, we present a study that aims to advance the current understanding of how user search behavior and knowledge gain differs across the varying cognitive levels.

Research Questions and Original Contributions

In this study, we investigate learning theory in order to further understand information search. The original Bloom's taxonomy was introduced to serve as a framework for the classification of educational statements which was later revised and restructured by Anderson and Krathwohl [21]. We adopt this revised Bloom's taxonomy for our study, and use it as a guide to design our tasks, measure knowledge gain and user interactions at various cognitive levels. Using quantitative analysis of crowdsourced user interaction data, we aim to answer the following research questions.

> RQ1: How does user knowledge evolve in informational search sessions with respect to the varying cognitive learning levels?

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.
HT '19, September 17–20, 2019, Hof, Germany
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-6885-8/19/09…$15.00
https://doi.org/10.1145/3342220.3343643

We asked users to submit answers for the tasks designed corresponding to each of the cognitive levels. We measured the knowledge gain based on the submitted user responses. Through quantitative analysis of results we showed that users experienced an increase in knowledge for search sessions of varying cognitive complexities. We also found evidence that knowledge gain is impacted by the cognitive complexity of the search session.

RQ2: How is user behavior in informational search sessions affected by the corresponding cognitive learning level?

We collected user data in terms of user interactions while users indulged in search tasks of varying cognitive levels. Our analysis revealed that user interactions are impacted by the cognitive complexity of search sessions; user interactions increase with the increase in cognitive complexity of search session such that it is minimum for the cognitive level corresponding to least complex level in cognitive domain of revised Bloom's taxonomy and maximum for the cognitive level corresponding to most complex level in cognitive domain of revised Bloom's taxonomy.

For the benefit of the community, we publicly release the tasks we created as well as the corresponding user data[1].

2 BACKGROUND AND RELATED WORK

We discuss the background and related literature in three distinct realms: Bloom's taxonomy of learning, information seeking, and the impact of search tasks.

2.1 Bloom's Taxonomy

Bloom developed a taxonomic structure to encourage education at a deeper level as compared to mere fact recalling [1]. The categories of the taxonomic structure were viewed as learning levels. His motivation was to create thinkers in the world and in order to do so, he proposed hierarchical structure of six categories/levels in the taxonomy where the top levels of the structure were more abstract and required higher level of thinking and reasoning in contrast to lower levels. The categories of Bloom's taxonomy from least abstract to most abstract are: *Knowledge, Comprehension, Application, Analysis, Synthesis,* and *Evaluation.* While Bloom's taxonomy became a guidebook in designing educational coursework, many claimed that there aren't any perfect educational taxonomies as most taxonomies have their weaknesses and Bloom's taxonomy was no exception to this rule [3, 21, 22]. The original Bloom's taxonomy has been challenged by many. It has been shown that while all the other categories of Bloom's taxonomy led to an increase in memory, the *Evaluation* level failed to do so [22]. Authors questioned the position and inclusion of *Evaluation* in the educational taxonomy. Anderson and Krathwohl, students of Bloom argued the usability of Bloom's taxonomic structure in educational systems as educators are used to designing the learning objectives in a "subject-description" format where subject would refer to subject matter of the content and description would include an explanation of how to deal with the content [21]. They further illustrated that this "subject-description" format can also be viewed as a "noun-verb" pair. Krathwohl et al. modified the original taxonomy into a 2-dimensional taxonomy, where *Knowledge* formed one of the dimensions and cognitive processes of learning the knowledge formed the second dimension. The revision allowed evaluation for both, the learning outcomes as well as the cognitive process used by learners [27]. To better understand how the knowledge of users in informational search sessions evolves corresponding to the different cognitive levels, we adopted the revised Bloom's taxonomy in our study.

[1]https://sites.google.com/view/searchquestionshypertext19/home

2.2 Information Seeking

Information need often refers to one's underlying motivation to seek the specific type of content [29]. Every individual has his or her own view of the world around them, specific typification that are used to model and explain all the phenomena around them and when one encounters a problem which won't fit in their model, the individual would seek out for more information and remodel knowledge in order to solve the problem and fix the anomaly [24]. Several prior works talk about the relationships between sense-making models and information seeking. Some of them view information seeking as a means to demolish the uncertainty between desired and observed scenarios [5], while some review users' sensemaking approach by transforming users' conceptualization from a noun based knowledge framework to verb based framework [6] . Likewise, there have been discussions where strong relations between information seeking, knowledge and human cognition levels have been displayed for example, through theory on relation between text retrieval and cognitive framework [18] or through a problem solution model [29].

However, in this study we will focus not only on information seeking but also on searching and learning of information. There exists an abundant amount of literature which emphasizes on relations between search, learning, and user or more specifically user behavior in the area of information science [4]. These literature include studies explaining affects of certain additional features like annotation on learning [7], studies exploring knowledge building models [25], studies inspecting learners' behavior in an collaborative system [28], etc. This study will focus on exploring individual learners' behavior online and the relation between learners' behavior and the cognitive learning complexity level.

2.3 Search Tasks

To comprehend the interrelation between search, user, and learning and to inspect the procedure of learning; it is crucial to design search tasks with great caution and clarity. Designing search tasks is a difficult and time consuming problem as it requires specialized knowledge [19]. The modeling is further complicated in the light of previous works which illustrate how variations in search tasks and search task properties can impact searcher behavior [20, 30]. Poorly designed search tasks can often lead to invalid results as users participate in unrealistic searches and depict inadmissible user behavior. This will hence, lead to wastage in resources (time and money). For example, in an attempt to design a complex search task it is not useful when one creates a task where the learner can find the answer from the first Wikipedia page by firing a simple search query.

Tasks can be classified in many ways, based on their type - *e.g., open, factual, navigational, decision-making*, or based on their topic - *e.g., difficulty, urgency, structure, stage.* Within the context of our work, we are interested in classifying the tasks based on their complexity. To do so, we use the revised Bloom's taxonomy to classify the tasks into the six cognitive processes, much like some of the previous studies [19, 20]. However, designing tasks based on Anderson and Krathwohl's taxonomy is complicated as the categories of the taxonomy are not distinct from each other [19]. It also implies that the revised Bloom's taxonomy has learning levels that overlap its boundaries with its immediate top and bottom levels. To study how user behavior varies across the cognitive levels, it is crucial to design the search tasks appropriately; such that they call upon the users to utilize the corresponding cognitive process.

The search tasks created within the scope of this study for each category, were guided by previous literature [8, 23]. In previous studies [13, 19, 20], the setup was such that each participant would answer all of the search tasks corresponding to varying cognitive levels. The search tasks designed were either for all cognitive learning levels of Bloom's taxonomy [19] or for a subset of cognitive learning levels [13, 20]. This would mean that if a learner performs tasks for two different levels, her behavior for other levels would be influenced by the knowledge carried forward from previous levels. To overcome this problem, in our work, we designed distinct search tasks for all of the cognitive learning levels and gathered responses from participants in a between-groups study design. Our experimental results showed distinct user behaviors among different levels of cognitive processes for all of the user interactions discussed in this study. Our study differs from previous works, in that we recognize and consider the fact that the revised Bloom's taxonomy has overlapping levels while allocating tasks to learners.

3 STUDY DESIGN

For this study, we proposed a unique crowdsourcing experimental setup to gather user data and carry out a quantitative analysis. The user data is gathered by deploying different search tasks corresponding to varying cognitive stages of learning. We segregated each level of cognitive learning and accumulated data in the form of logged user-interactions as well as submitted answers.

3.1 Motivation for Proposed Design

The revised Bloom's taxonomy has relaxed constraints between the cognitive processes [21]. This implies that each cognitive process has an overlap in boundaries with its immediate top and bottom cognitive processes in the taxonomic structure. The cognitive processes can be arranged in a hierarchical manner, however, due to its relaxed nature of boundaries and the relation between higher cognitive levels with lower, it can be inferred that if a learner appears for two or more cognitive levels, he is carrying some existing knowledge from previous levels. While this is the motivation in classroom learning scenarios, i.e., to have a smooth transition between cognitive processes, it might create imprecise results while measuring user behavior in online learning.

Jansen et al.'s justification of results where an inverted U relationship is observed between search difficulty and cognitive learning level also sheds some light on the interrelation between the cognitive levels [19]. The authors' justification that the higher level tasks did not call for many searches as users already possessed the knowledge from lower levels and hence, required mere verification shows that cognitive levels are not independent of each other. Further, in a recent study [13], the authors speculate that the reason behind the lack of statistically significant difference in user behavior they studied across cognitive levels was the ascending order in which tasks were provided to users which resulted in users familiarizing themselves with the topic before they reached the higher cognitive levels. Therefore, for our study, we distribute tasks among distinct participants such that each cognitive learning level has unique participants carrying out the search task.

3.2 Experimental Setup

We designed questions for six different search tasks labeled *Remember, Understand, Apply, Analyze, Evaluate,* and *Create*, *i.e.*, each of the six cognitive learning levels of revised Bloom's taxonomy for the domain "Vitamin and Nutrients". A scientific domain like "Vitamin and Nutrients" allowed us to design technical questions on topics that would be unfamiliar to users unless they possessed a scientific background. These search tasks were deployed on a popular crowdsourcing platform called Figure-Eight[2]. Since the cognitive levels are not independent of one another, we assigned only one of the search tasks to a given user. This setup allowed us to gather user data that corresponded to an isolated cognitive level. Any user who tried to attempt one of the search tasks was blocked from attempting any other task in our setup thereafter. This ensured that there was no carry over of knowledge from one cognitive learning level to another. This also ensured that for a higher cognitive level, ideally, the user will have to first familiarize himself with the topic and carry out research instead of a mere fact-verification as observed in some previous studies [19]. We chose a crowdsourcing platform over laboratory experiments in order to scale up the number of participants. We followed the guidelines laid out by previous work while doing so [10].

The title of the crowdsourcing jobs with respect to all tasks on the platform was kept uniform – "Search and Answer" – so that users would not receive any prior indication regarding the type of work that the search tasks demanded. In addition to having a general title, the description of the task too was kept non-specific. A classic description for most of the tasks looked as follows:

> "*In the task you will answer a few questions and use our custom search engine. The topic for questions will be introduced once you click the task link. You can search for answers when you do not know them using our search engine. IMPORTANT: The task requires you to have proficiency in English language*".

These precautions were taken to avoid any bias resulting from reading the title and description [17]. Each task had 30 minutes as maximum allocated time. Workers on Figure-Eight often abandoned tasks for the lack of reward, difficulty, and clarity in the task instructions [14]. To avoid such task abandonment and motivate the workers to complete the tasks, we paid workers at an hourly rate

[2] https://www.figure-eight.com/

of 7,50 USD. Additionally, the workers were given an incentive of a bonus equivalent to 1 USD depending on their performance. Further, certain quality control measures were set on Figure-Eight for all the tasks. A worker was allowed to submit 1 judgment per task, and only Level-3[3] workers were allowed to attempt the task. We limited participation to workers from English-speaking countries, to ensure that workers understand the instructions and questions with complete clarity [9, 11].

Figure 1 provides an overview of the workflow for this study. Users are recruited from Figure-Eight and redirected to an external platform on accepting to participate in our tasks. Users are then provided with further instructions on how to complete the given task along with a small introductory passage describing the importance of "*Vitamin and Nutrients*" in a healthy diet. The instructions informed the user to use `SearchWell`, our customized search engine exclusively for any search related actions. `SearchWell` is built on top of the Bing Web Search API. It uses a WAPS proxy tracker to log and track user activities on the platform including mouse movements, clicks, key presses, URLs visited, time spent on URLs, etc. A user can attempt the task after reading the instructions and upon a valid submission, receives a completion code. In order to receive the monetary reward on Figure-Eight, the user provides the completion code. Care was taken to ensure that users who completed a task were not allowed to participate in another. The validity of submission was determined by the rule that if a user submits a task without carrying out a search and/or the task contains incorrect answers then the submission is automatically rejected. As the aim of our work is to further the understanding of the relation between user, search, and learning online, it advocated that we discard those users who did not enter a single search query. In case of the cognitive learning levels of *Evaluate* and *Create*, due to open-ended nature of the tasks (as described in the next section), any submission without issuing a single search query was rejected. 246 submissions were collected in total. Of these, 150 were accepted and 96 rejected.

3.3 Task Design and Questions

Drawing inspiration from previous works [8, 19–21, 23], we designed task questions that would require the users to call upon the action words corresponding to each of the cognitive learning levels while answering the questions. Verbs or action words reflect the type of action to be carried out on knowledge, for example recollection of a fact, providing a judgment, etc. Table 1 provides an overview of the mapping of each cognitive level of revised Bloom's taxonomy to the words that can be used to design the questions for the chosen cognitive level. The entire set of questions for each level used in this study can be found at this anonymized webpage [4].

We designed 15 questions for *Remember*, 10 questions for *Understand*, 9 questions for both *Apply* and *Analyze*, 2 questions for *Evaluate*, and 1 question for *Create*. The inequality in the number of questions corresponding to the different cognitive learning levels was due to the increasing difficulty of the task as we go higher in the task complexity pyramid. This would mean that having fifteen questions for *Create* level would be unreasonably time-consuming and tedious in comparison to *Remember* for example. To encourage realistic submission goals from the crowdsourced participants, we did not have the same number of questions for all search tasks. Note that during the quantitative analysis, the user data was normalized by the number of questions in each search task.

While designing the questions, care was taken to distinguish between questions of two consecutive levels. Also, for questions corresponding to higher levels, it was made sure that the answers are not available online with simple straightforward search queries.

3.4 Measuring Knowledge Gain

3.4.1 Remember. The *Remember* task consists of three stages: 1) a pre-test for calibrating user knowledge, 2) search session, and 3) a post-session test (where pre-test and post-test consists of same questions, unknown to the participants beforehand). Answers submitted in pre-test can be be viewed as existing knowledge of a user [12]. The pre-test and post-test consist of statements which participants are asked to respond to by selecting one of the options: '*True*', '*False*', or '*I Don't Know*'. Participants are encouraged to respond honestly by pointing out that their accuracy in the pre-test would have no influence in the final monetary reward that they can earn. The second stage, i.e., the search session is where a user is assigned a learning task and can consequently carry out search queries to gain knowledge. We measure the knowledge gain (KG) of the users by comparing the answers submitted in post-test to pre-test. The rules for calculating KG are as follows:

$$
\begin{aligned}
1.\forall x : KG = KG + 1 \iff & \; x = \textit{unknown} \text{ in pre-test } \& \\
& \; x = \textit{correct} \text{ in post-test} \\
2.\forall x : KG = KG + 1 \iff & \; x = \textit{incorrect} \text{ in pre-test } \& \\
& \; x = \textit{correct} \text{ in post-test} \\
3.\forall x : KG = KG + 0 \iff & \; x = \textit{incorrect} \text{ in post-test } \| \\
& \; x = \textit{unknown} \text{ in post-test}
\end{aligned}
$$

where, x represents the correctness of the answer submitted for each question. The answer can either be correct, incorrect, or marked as unknown if '*I Don't Know*' is selected by the participant.

3.4.2 Understand. The *Understand* task has multiple choice questions (MCQs). An answer for a question is considered to be from existing knowledge if the user did not open the search frame before attempting to answer as it implies that the user is answering the question by using his existing knowledge. If the search log indicated that user carried out activities on web before attempting to answer the given question, then the submitted answer for the question is used to calculate the knowledge gain. Users are made aware that there is no benefit for them to use an alternative search engine, apart from the customized one they are directed to use. Thus, we ensure reliability of the user logs. Participants were encouraged to not guess the answers by informing them that the monetary reward is not influenced by their existing knowledge. The knowledge gained with respect to a single question is calculated according to

[3]Level-3 workers on figure-eight are highest Quality workers. It is a group of most experienced, highest accuracy contributors
[4]`https://sites.google.com/view/searchquestionshypertext19/questions`

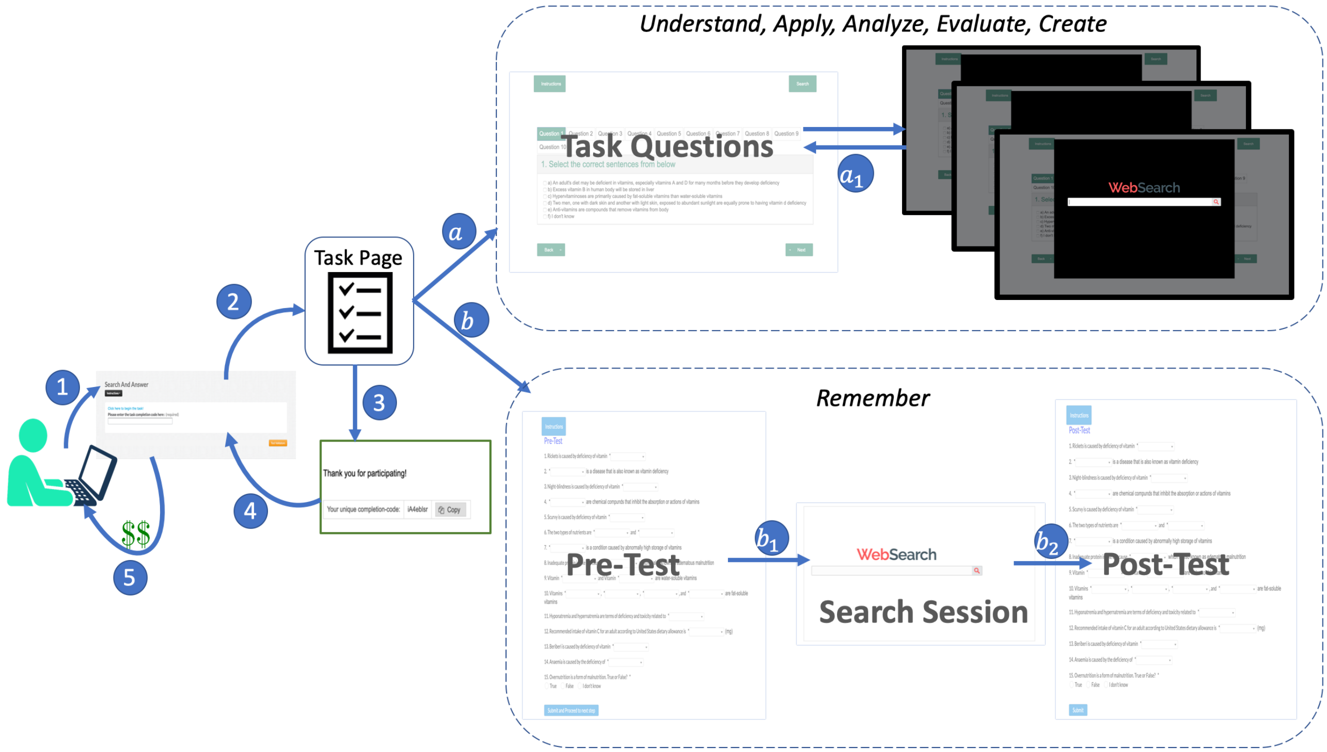

Figure 1: Workflow: 1 Workers are recruited from the Figure-Eight platform. 2 They are then redirected to an external task page, where the task can be accomplished. 3 Workers receive a completion code upon making a valid submission. 4 They copy the completion code and return to platform. 5 Finally, the workers receive their reward on pasting the completion code. Task type a: Workers attempt task questions (a_1) They search from search frame many times during the course of task completion. Task type b: Workers attempt pre-test questions (b_1). They then carry out searches corresponding to the task assigned to them (b_2). Workers finally submit a post-session test.

Table 1: Cognitive learning processes mapped to verbs and typical actions required.

Cognitive Level	Verbs/Action Words	Potential Task
Remember	recognize, recall, repeat, state, define, identify, name, list	Recall information and basic concepts
Understand	classify, summarize, infer, explain, exemplify, identify, locate, recognize, report, select, describe	Explain ideas or concepts
Apply	solve, use, interpret, schedule, execute, implement, demonstrate, operate, sketch	Use the information in new situations
Analyze	differentiate, organize, attribute, relate, compare, contrast, distinguish	Draw connections among ideas
Evaluate	justify, check, critique, weigh, support, judge, defend, argue, appraise	Justify a stand or decision
Create	create, generate, plan, produce, design, construct, assemble, develop, conjecture, formulate, author	Produce new or original work

following rules:

$$1. \forall x : KG = KG + 1 \iff x = correct$$
$$2. \forall x : KG = KG - 1 \iff x = incorrect$$
$$3. \text{Total } KG = 0 \iff |x = incorrect| \geq |x = correct|$$

where, KG refers to knowledge that is gained, and x is the choice selected as one of the answers of multiple correct answers. $|x = incorrect|$ & $|x = correct|$ are the number of incorrectly selected answers and number of correctly selected answers respectively.

3.4.3 Apply. The *Apply* task consists of two types of questions:- MCQs and questions with statements describing a set of events to be ordered in a correct sequence. If the search log indicated web activities while a user was attempting to answer, then the knowledge that is gained is computed. Rules for measuring the KG for MCQs are same as discussed above for *Understand*. However, the rules for ordering or sequencing type questions are as follows:

$$1. KG = KG + 1 \iff \text{submitted sequence is correct}$$
$$2. KG = KG + 0 \text{ for all other cases}$$

where, KG refers to the knowledge that is gained for give question.

3.4.4 Analyze. The *Analyze* task questions ask users to match different properties to different attributes. For example, it may question the user to analyze the given list of nutrients and attribute each nutrient to be either a vitamin or mineral. The rules for calculating the knowledge gain if search log indicated web activities for each question are as follows:

1.$\forall x, a : KG = KG + 1 \iff x.a = correct$

2.$\forall x, a : KG = KG - 1 \iff x.a = incorrect$

3.Total $KG = 0 \iff |x.a = incorrect| \geq |x.a = correct|$

where, KG refers to knowledge that is gained for given question, x refers to answer submitted, a refers to attributes in the question, and $x.a$ refers to answer attributed to certain attribute a which can be either correctly assigned or incorrectly assigned. Both, $|x.a = incorrect|$ and $|x.a = correct|$ indicate the total number of incorrectly attributed answers and total number of correctly attributed answers respectively.

3.4.5 Evaluate and Create. Both, *Evaluate* and *Create* have creative, exploratory questions to which participants are asked to provide open-ended answers. The *Evaluate* task asked users to provide judgment and justify it while *Create* task required users to design a plan. As each of these answers are subjectively dependent upon user's perception, we do not mark them as correct or incorrect through a set of rules and assign a numerical value for increase in knowledge. Authors of this paper manually checked and marked submissions as valid upon encountering complete and comprehensive submissions. User data of valid submissions also indicated search actions.

4 RESULTS

4.1 Knowledge Gain and Cognitive Level

RQ1 was aimed at studying changes in knowledge gain of users in informational search sessions across varying cognitive learning levels. To find the relation between the knowledge of a user and the cognitive complexity of the search task, we asked the users questions on topic "*Vitamin and Nutrients*" corresponding to search tasks of varying cognitive complexity level. We measured the knowledge gained among users based on the answers that they submitted. To explore the relation between KG and cognitive learning levels, we intend to investigate the following hypotheses.

Hypothesis 1.1: Users exhibit measurable changes in knowledge gain for search tasks of varying cognitive learning levels

We carried out crowdsourced experiments and measured KG across 4 cognitive levels of revised Bloom's taxonomy. Of the 150 valid, accepted submissions, we calculated the numeric value of knowledge gain for 100 workers for *Remember, Understand, Apply,* and *Analyze* cognitive levels. Of these 100 workers, 86 workers exhibited an increase in knowledge. A minimum knowledge gain of 6% was exhibited whereas, the maximum knowledge gain recorded was 94.7%. From figure 2, we can observe the distribution of knowledge gained across the first four cognitive levels. Table 2 provides the percentage of users who exhibited a gain in knowledge for the four cognitive levels. We observe that the majority of the users (>86%) who participated in the learning tasks within informational search sessions exhibited knowledge gain, which suggests learning to be an outcome for each of the chosen cognitive level. We found the average knowledge gain to be between 11% to 22% depending upon the cognitive learning level. These findings suggest that users' knowledge evolves across the four cognitive domains *Remember, Understand, Apply,* and *Analyze.*

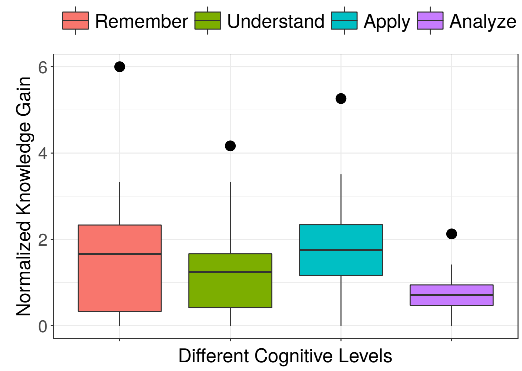

Figure 2: Knowledge Gain across cognitive levels (normalized by max. KG possible at each cognitive level as well as the no. of task questions in each level).

Due to the open-ended nature of *Evaluate* and *Create* tasks, we did not measure the numerical value of knowledge gain, however, we believe that the user behavior in terms of increase in knowledge can be extended to the highest two levels of cognitive learning levels. This is due to the fact that the users of *Evaluate* and *Create* carry out search and spent significant amount of time on web while solving task questions. In addition, looking at search interactions in the following section 4.2, we can say that the search behavior for the *Evaluate* and *Create* tasks were comparable to the remaining four domains, and in many cases as shown in following section more than the the lower four domains. Further, the manual assessment of answers showed that users submitted valid answers. Therefore, we believe that it is safe to assume a gain in knowledge occurred for *Evaluate* and *Create* tasks for users who searched online and submitted valid entries. Hence, we find support for Hypothesis 1.1 with empirical proof for the first four cognitive learning levels.

Table 2: Number of users who exhibited knowledge gain.

Cognitive Level	#Users with KG
Remember	80%
Understand	88%
Apply	84%
Analyze	92%

Hypothesis 1.2: The change in knowledge gain is dependent upon the cognitive learning level of the search task.

In order to determine if the cognitive learning level had any impact upon knowledge gain that is observed for each level, we carried out a one-way between subjects ANOVA for the calculated knowledge gain across the first four cognitive levels. Knowledge gain is calculated as discussed in Section 3.4 and further normalized by the maximum knowledge gain that is possible for each task. This provides us with the percentage of knowledge that is gained. Further, this value is normalized by the number of questions in each search task. A one-way between subjects ANOVA shows that the increase in knowledge gain is affected by the cognitive complexity of the task [$F(4, 100) = 21.88, p < 0.001$]. Post-hoc comparisons using the Tukey-HSD test revealed significant differences in the knowledge gained by users at the $p < 0.01$ level corresponding to *Apply* in comparison to *Analyze*. Results of one-way between subjects ANOVA show that there is a significant impact of the cognitive learning level of a search session on knowledge gain. Hence, we find support for Hypothesis 1.2.

4.2 Search Behavior and Cognitive Level

RQ2 was aimed at studying the relations between a user's search behavior and the cognitive learning level of the search task and investigating whether there is any impact of cognitive learning level on user interactions. We intuitively hypothesize that search behavior in terms of user interactions increase with the increase in cognitive learning complexity of the task. To find a solution for RQ2, we collected data related to user interactions while they performed in search tasks of varying cognitive complexities.

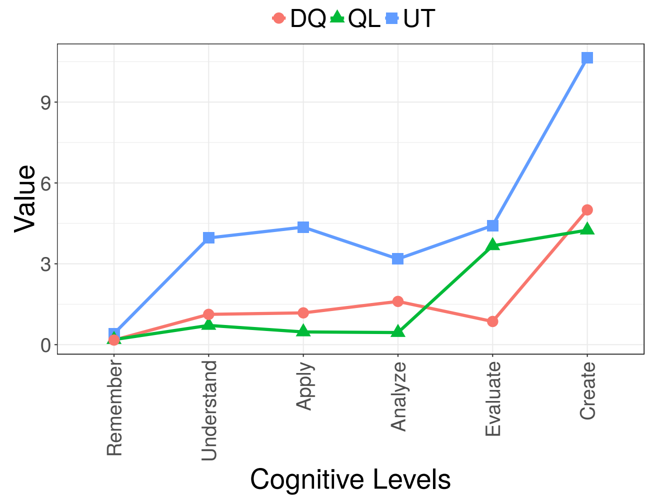

Figure 3: DQ represents the no. of distinct queries per user, QL represents the average query length, and UT represents the average no. of unique terms across the different cognitive levels respectively. All these values are normalized by the no. of task questions in each level.

User Queries

Hypothesis 2.1: Search queries will increase in number with the increase in cognitive learning complexity of the task

We collected a total of 1285 distinct queries across all the cognitive levels. The number of distinct queries (DQ) per user varied from 1 query per user to 36 queries per user. This accumulated data of number of distinct search queries per user was normalized by number of questions in each search task. A one-way between groups ANOVA showed that there were significant statistical differences [$F(6, 150) = 27.68, p < 0.001$] between the number of distinct queries per user for varying cognitive learning levels which implies that the number of search queries is impacted by cognitive level of search task. Post-hoc comparisons using the Tukey-HSD test revealed significant differences in the number of queries entered by users at the $p < 0.01$ level corresponding to *Remember* in comparison with Understand, Apply, Analyze and *Evaluate* as well as *Understand* in comparison with *Apply* and *Analyze*, *Apply* in comparison with *Evaluate* and *Create*, and *Analyze* in comparison with *Evaluate* and *Create*. Similarly, significant differences were revealed at the $p < 0.05$ level corresponding to *Understand* in comparison with *Create*. Figure 3 shows the trend of increase in distinct queries per user as the complexity of the task increases. The number of distinct queries is lowest for *Remember* and highest for *Create*. However, we observe fewer number of distinct queries corresponding to the *Evaluate* tasks in comparison to *Understand, Apply*, and *Analyze*. We believe that this is due to the judgmental nature of questions in the *Evaluate* task. As *Evaluate* is a judgment type task, more effort is spent on finding support or proof for the specific viewpoint of a participant providing an answer. Hence, this may have led users to fire fewer number of distinct queries. Thus, our results lend partial support to Hypothesis 2.1.

Hypothesis 2.2: Query length will increase with the increase in cognitive learning complexity of the task.

We collected and compared average query lengths of queries fired by users while performing search tasks. A one-way between subjects ANOVA showed statistically significant differences [$F(6, 150) = 47.44, p < 0.001$] for the query length across all cognitive levels for user data normalized by the number of questions in each search task. Post-hoc comparisons using the Tukey-HSD test revealed significant differences in the query lengths of queries entered by users at the $p < 0.01$ level corresponding to *Evaluate* in comparison with *Remember, Understand, Apply, Analyze,* and *Create*. It can be seen from Figure 3 that results support the hypothesis partially. The graph shows an upward trend in increase in query length from *Remember* to *Create*, however, for the middle tasks - *Understand, Apply,* and *Analyze*, the average query length decreases slightly such that *Analyze* <*Apply* <*Understand*. We also, analyzed the results for minimum, maximum, first, and last query lengths. For all of these features, one-way between subjects ANOVA results supported that cognitive learning levels had an impact on them. Further, user data for the mentioned features increased from lowest to highest complexity. These results support Hypothesis 2.2 partially.

Hypothesis 2.3: Number of unique terms will increase in number with the increase in cognitive learning complexity of the task.

Analysis of the user interaction data shows that there is a significant effect on number of unique terms in query by the cognitive complexity of the task. While there were a few users who saw search queries with only 1 unique term, by the completion of study, the user with maximum number of unique terms in the entire session had carried out search queries including a total of 100 unique terms. Figure 3 shows the pattern of changes in number of unique terms across cognitive levels. It can be seen from the upward growth in the figure that the total number of unique terms per question are lowest for the search task of lowest complexity and highest for the most complex search task. However, one of the search tasks, *Analyze*, in intermediate levels of revised Bloom's taxonomy falls out of the continuous growth pattern where *Remember <Analyze <Evaluate*. A one-way between subjects ANOVA supported that cognitive complexity of the task affects the number of unique terms observed for the task [$F(6, 150) = 24.59, p < 0.001$] and that there are statistically significant differences in the results across all the cognitive learning levels. Post-hoc comparisons using the Tukey-HSD test revealed significant differences in the number of unique terms entered by users at the $p < 0.01$ level corresponding to *Remember* in comparison with *Understand, Apply, Analyze,* and *Evaluate* as well as *Create* in comparison with *Understand, Apply, Analyze,* and *Evaluate*. These results support Hypothesis 2.3 partially.

Websites and Search Pages

Hypothesis 2.4: Number of websites visited will increase with the increase in cognitive learning complexity of the task

User interactions varied from having an average between 2 to 5 web pages visited depending upon the complexity level of the search task to a maximum of 42 pages visited in the entire search session. A one-way between subjects ANOVA for the user data normalized by number of questions in each search task supported the presence of a statistically significant difference between the total number of web pages visited across the various cognitive learning levels of the search task [$F(6, 150) = 12.49, p < 0.001$] which implies that the total number of web-pages visited by a user is affected by the cognitive learning level of task. Further, this is true for the relation between number of unique web pages visited and cognitive learning levels [$F(6, 150) = 11.33, p < 0.001$]. Post-hoc comparisons using the Tukey-HSD test revealed significant differences in the total number of webpages visited by users at the $p < 0.01$ level corresponding to *Remember* in comparison with *Understand, Apply, Analyze,* and *Evaluate* as well as *Evaluate* in comparison with *Create*. Similarly, significant differences were revealed at the $p < 0.05$ level corresponding to *Create* in comparison with *Understand, Apply,* and *Analyze*. Figure 4 shows a relation between total number of web pages as well as number of distinct web pages visited per user and cognitive learning level. The figure shows that Hypothesis 2.4 is supported partially as the number of web pages visited increases from least complexity level to most complexity level, however, the continuous growth trend does not hold true for intermediate levels.

Hypothesis 2.5: Number of search pages visited will increase with the increase in cognitive learning complexity of the task

We analyzed the search engine results pages (SERPs) consumed by users in search tasks for varying cognitive learning levels. Figure 4 illustrates the observed relation between the number of SERPs visited and cognitive learning levels. Users navigated an average of 3 to 17 distinct search engine result pages during the entire search session. Maximum number of search pages visited by a user was observed to be 39. A one-way between subjects ANOVA showed that the total number of search pages visited is affected by the search complexity of the task [$F(6, 150) = 20.87, p < 0.001$]. The same holds true for the number of distinct SERPs visited [$F(6, 150) = 21.95, p < 0.001$]. Post-hoc comparisons using the Tukey-HSD test revealed significant differences in the total number of SERPs visited by users at the $p < 0.01$ level corresponding to *Remember*

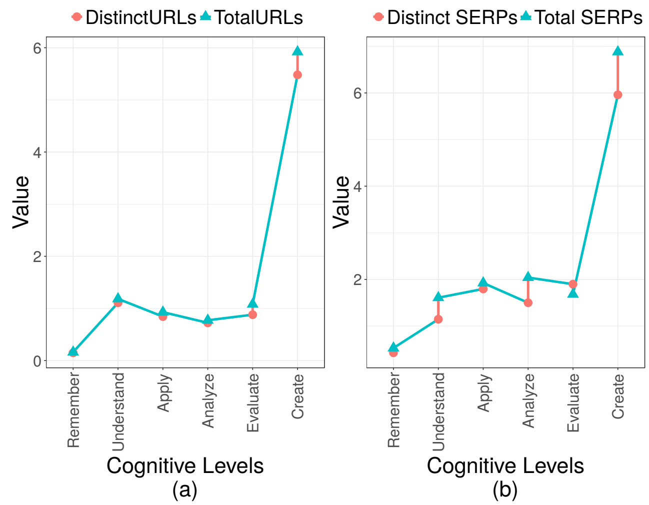

Figure 4: (a). Total & distinct no. of URLs visited across cognitive levels (normalized by no. of task questions in each level), (b). Total & distinct no. of SERPs visited across cognitive levels (normalized by no. of task questions in each level).

in comparison with *Understand, Apply, Analyze,* and *Evaluate* as well as *Create* in comparison with *Understand, Apply, Analyze,* and *Evaluate*. Much like all the other results, the normalized data in Figure 4 shows an upward trend where the search behavior for total number of search pages visited is maximum for the highest complexity task and least for lowest complexity task, however, it does not follow the increasing sequence for intermediate search tasks. Hence, the results support Hypothesis 2.5 partially.

Time Spent Online

We examined the amount of time that the user spent online. This included the total time required to complete the task as well as the active time spent during the search sessions. Active time is the time in which the user actively carried out interactions in the search session.

Hypothesis 2.6: Time spent online will increase with the increase in cognitive complexity of the task

Time spent while users actively interacted with the search engine was logged. This time, called user's active time spent on web across all the cognitive learning levels was measured. When the user's active time, normalized by the number of questions in each search task was analyzed to find the relation it has with the search complexity of the task, we established that the cognitive learning level had a direct impact on it. A one-way between subjects ANOVA shows that there is significant statistical difference [$F(6, 150) = 42.67, p < 0.001$] between the active time and the cognitive complexities of the task. Post-hoc comparisons using the Tukey-HSD test revealed significant differences in the active time spent by users at the $p < 0.01$ level corresponding to *Evaluate* in comparison to *Remember, Understand, Apply, Analyze,* and *Create*. This proved that the active time spent online is impacted by the cognitive level of search task. Figure 5a shows the relation between active time and cognitive complexity

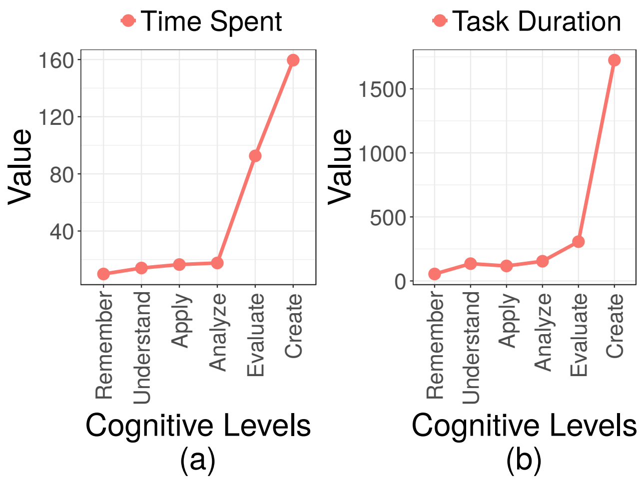

Figure 5: (a). Active time spent in seconds across cognitive levels (normalized by no. of task questions in each level), (b). Total task duration in seconds across cognitive levels (normalized by no. of task questions in each level).

of the task. It gives an upwards trend where *Remember* <*Understand* <*Apply* <*Analyze* <*Evaluate* <*Create*. Hence, hypothesis 2.6 is supported by these results.

We also measured the time taken by user from the moment the task link is opened till the time when the user hits the submit button. We called this time as the total task duration and compared the total time taken to complete search tasks of various cognitive levels. We analyzed the total task duration by normalizing the data with the number of questions in each task. A one-way between subjects ANOVA supports $[F(6, 150) = 10.55, p < 0.001]$ that the total task duration is affected by the cognitive complexity of the search task. Post-hoc comparisons using the Tukey-HSD test revealed significant differences in the total task duration measured for each users at the $p < 0.01$ level corresponding to *Evaluate* in comparison to *Remember, Understand, Apply, Analyze,* and *Create* as well as *Remember* in comparison to *Analyze*. Figure 5b shows the resulting relation of task duration and cognitive learning levels. It can be seen from figure 5b that our hypothesis is partially supported since total task duration increases while moving from least complex task to most complex task, however, the continuous increasing trend is not true for one of the intermediate levels since *Apply* is less than *Understand*. These results show that Hypothesis 2.6 is partially supported.

Results stated to support Hypothesis 2.1, 2.2, 2.3, 2.4, and 2.6 partially support the statement that the "User behavior in terms of user interactions increases with the increase in cognitive learning complexity of a search session". We found that all the user interactions are impacted by the cognitive complexities of the task. Our results further show that search behavior in the form of user interactions is minimum for lowest cognitive complexity level and highest for maximum complexity level. However, the increase in search behavior does not always correspond with increase in cognitive complexity for intermediate levels of revised Bloom's taxonomy.

5 DISCUSSION

Our study design was informed by the revised Bloom's taxonomic structure and explored the relation between users' learning, and their corresponding search interactions. Our study setup which disjoints each cognitive learning search task from the others allowed us to identify the impact of cognitive complexity of a search session on a user's activity and change in knowledge.

In response to **RQ1**, we observed that the knowledge of users on average evolves in the search tasks corresponding to each of the first four cognitive levels. We also found that the average knowledge gain of users across the different cognitive levels was significantly different. Although we did not explicitly study the evolution of knowledge corresponding to the cognitive learning levels of *Evaluate* and *Create*, our findings suggest that users carrying out these tasks also potentially gained some knowledge through the course of the tasks. This is because the user interactions corresponding to users who submitted valid answers in these cognitive levels were similar to those observed in the lower four cognitive levels.

Corresponding to **RQ2**, we found that user interactions across the cognitive learning levels were impacted by the cognitive complexity of the search session. Further, user interactions followed an upward trend from *Remember* to *Create* for all of the observed user interactions, i.e., the search behavior in terms of user interactions is minimum for least complex task and maximum for most complex task. While this was an expected result for RQ2 (intuitively when a learner participates in a more complex task she will require to exert more effort in information-finding), this is the first study that presents quantitative evidence in support of this hypothesis within informational search sessions.

Our findings also show that while the trend for changes in user interactions was upward from least complex to most complex search task, the tasks corresponding to intermediate cognitive levels in the revised Bloom's taxonomy did not strictly follow the continuous growing trend. However, the differences observed between the observed user data for intermediate levels in such scenarios was not significant in all of the cases. We believe that these aberrations were observed due to the fact that revised Bloom's taxonomy is not rigidly bound and the transition from one level to another is a continuous process. Since there are no rigid distinctions between the levels, it is easy for a user interaction feature to fall out of the expected boundary for a particular level.

5.1 Caveats & Limitations

As described earlier we took several measures to ensure reliable participation from our crowdsourced subjects [10]. Although it is challenging to ensure that workers are invested in the experimental tasks, we facilitate legitimate user behavior by using monetary rewards as an incentive for users to learn via searching in the tasks corresponding to each cognitive level.

Due to the study design required to answer the research questions addressed in this paper, we were unable to gather user feedback regarding the relative difficulty of the different tasks corresponding to each cognitive level.

As a result of platform level restrictions on Figure-Eight, workers were constrained to completing their tasks within a maximum duration of 30 minutes (from accepting to participate in our study to submitting their completion codes). This is the case with all crowdsourcing tasks on Figure-Eight. To ensure there was no effect of this restriction on our experimental study, we made sure that all tasks could be completed well within 30 minutes.

Note that to control for Type-I error inflation in our multiple comparisons, we used the Holm-Bonferroni correction for family-wise error rate (FWER) [16] at the significance level of $\alpha < .05$.

6 CONCLUSION AND FUTURE WORK

In this study we gathered and analyzed user data in terms of user interactions and knowledge gain for the cognitive dimension of the revised Bloom's taxonomy. We designed search task questions corresponding to each cognitive learning level and deployed a between-groups experimental study using a crowdsourcing plaform. We ensured that no worker participated in search tasks corresponding to more than one cognitive learning complexity level. We analyzed the users' search interactions to investigate the relationship between user search behavior and the learning levels of the revised Bloom's taxonomy. We demonstrated a relation between the search task complexity and knowledge gain for the first four cognitive levels.

The analysis of user data demonstrated that users' search interactions increased as participants attempted a search task of higher cognitive level of revised Bloom's taxonomy with certain deviations in the intermediate levels. This allowed us to conclude that if a user advances to highest complexity level from lowest, there will definitely be an increase in the observed search behavior. The analysis further indicated that there was statistically significant difference between the cognitive complexity of search interactions across cognitive levels for all of the discussed user interactions. All these results display the importance of the search tasks designed distinctly corresponding to each learning level. These search tasks can be reused to further study the user interactions by search interfaces facilitating learning needs. Thus, our work has important implications on supporting learning oriented search on the Web. Machine learning models can be trained using the user data collected to automatically detect the cognitive complexity of a search session based on user interactions. This can allow for interventions where a learner's needs can be optimized. We will pursue this goal in the imminent future. Furthermore, we will extend our study considering additional topics to analyze topical effects on user interactions across different cognitive learning levels.

REFERENCES

[1] Benjamin S Bloom et al. Taxonomy of educational objectives. vol. 1: Cognitive domain. *New York: McKay*, pages 20–24, 1956.
[2] Andrei Broder. A taxonomy of web search. In *ACM Sigir forum*, volume 36, pages 3–10. ACM, 2002.
[3] Charles C Chan, MS Tsui, Mandy YC Chan, and Joe H Hong. Applying the structure of the observed learning outcomes (solo) taxonomy on student's learning outcomes: An empirical study. *Assessment & Evaluation in Higher Education*, 27(6):511–527, 2002.
[4] Kevyn Collins-Thompson, Preben Hansen, and Claudia Hauff. Search as learning (dagstuhl seminar 17092). In *Dagstuhl Reports*, volume 7. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.
[5] Brenda Dervin. *An overview of sense-making research: Concepts, methods, and results to date.* The Author, 1983.
[6] Brenda Dervin. Sense-making theory and practice: an overview of user interests in knowledge seeking and use. *Journal of knowledge management*, 2(2):36–46, 1998.
[7] Evren Eryilmaz, Jakko van der Pol, Terry Ryan, Philip Martin Clark, and Justin" Mary. Enhancing student knowledge acquisition from online learning conversations. *International Journal of Computer-Supported Collaborative Learning*, 8(1):113–144, Mar 2013.
[8] Chris Ferguson. Using the revised taxonomy to plan and deliver team-taught, integrated, thematic units. *Theory into practice*, 41(4):238–243, 2002.
[9] Ujwal Gadiraju, Ricardo Kawase, Stefan Dietze, and Gianluca Demartini. Understanding malicious behavior in crowdsourcing platforms: The case of online surveys. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1631–1640. ACM, 2015.
[10] Ujwal Gadiraju, Sebastian Möller, Martin Nöllenburg, Dietmar Saupe, Sebastian Egger-Lampl, Daniel Archambault, and Brian Fisher. Crowdsourcing versus the laboratory: Towards human-centered experiments using the crowd. In *Evaluation in the Crowd. Crowdsourcing and Human-Centered Experiments*, pages 6–26. Springer, 2017.
[11] Ujwal Gadiraju, Jie Yang, and Alessandro Bozzon. Clarity is a worthwhile quality: On the role of task clarity in microtask crowdsourcing. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, pages 5–14. ACM, 2017.
[12] Ujwal Gadiraju, Ran Yu, Stefan Dietze, and Peter Holtz. Analyzing knowledge gain of users in informational search sessions on the web. 2018.
[13] Souvick Ghosh, Manasa Rath, and Chirag Shah. Searching as learning: Exploring search behavior and learning outcomes in learning-related tasks. In *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval*, pages 22–31. ACM, 2018.
[14] Lei Han, Kevin Roitero, Ujwal Gadiraju, Cristina Sarasua, Alessandro Checco, Eddy Maddalena, and Gianluca Demartini. All those wasted hours: On task abandonment in crowdsourcing. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 321–329. ACM, 2019.
[15] Ahmed Hassan Awadallah, Ryen W White, Patrick Pantel, Susan T Dumais, and Yi-Min Wang. Supporting complex search tasks. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 829–838. ACM, 2014.
[16] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.
[17] Christoph Hube, Besnik Fetahu, and Ujwal Gadiraju. Understanding and mitigating worker biases in the crowdsourced collection of subjective judgments. 2019.
[18] Peter Ingwersen. Cognitive perspectives of information retrieval interaction: elements of a cognitive ir theory. *Journal of documentation*, 52(1):3–50, 1996.
[19] Bernard J Jansen, Danielle Booth, and Brian Smith. Using the taxonomy of cognitive learning to model online searching. *Information Processing & Management*, 45(6):643–663, 2009.
[20] Diane Kelly, Jaime Arguello, Ashlee Edwards, and Wan-ching Wu. 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*, pages 101–110. ACM, 2015.
[21] David R Krathwohl. A revision of bloom's taxonomy: An overview. *Theory into practice*, 41(4):212–218, 2002.
[22] Seth Kunen, Ronald Cohen, and Robert Solman. A levels-of-processing analysis of bloom's taxonomy. *Journal of Educational Psychology*, 73(2):202, 1981.
[23] Thomas Lord and Sandhya Baviskar. Moving students from information recitation to information understanding-exploiting bloom's taxonomy in creating science questions. *Journal of College Science Teaching*, 36(5):40, 2007.
[24] Alfred Schutz and Thomas Luckmann. *The structures of the life-world*, volume 1. northwestern university press, 1973.
[25] Gerry Stahl. A model of collaborative knowledge-building. In *Fourth international conference of the learning sciences*, volume 10, pages 70–77. Mahwah, NJ: Erlbaum, 2000a, 2000.
[26] Rohail Syed and Kevyn Collins-Thompson. Retrieval algorithms optimized for human learning. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 555–564. ACM, 2017.
[27] Martin Valcke, Bram De Wever, Chang Zhu, and Craig Deed. Supporting active cognitive processing in collaborative groups: The potential of bloom's taxonomy as a labeling tool. *The Internet and Higher Education*, 12(3-4):165–172, 2009.
[28] James Waters and Susan Gasson. Social engagement in an online community of inquiry. *ICIS 2006 Proceedings*, page 47, 2006.
[29] Tom D Wilson. On user studies and information needs. *Journal of documentation*, 37(1):3–15, 1981.
[30] Wan-Ching Wu, Diane Kelly, Ashlee Edwards, and Jaime Arguello. Grannies, tanning beds, tattoos and nascar: Evaluation of search tasks with varying levels of cognitive complexity. In *Proceedings of the 4th Information Interaction in Context Symposium*, pages 254–257. ACM, 2012.
[31] Ran Yu, Ujwal Gadiraju, Peter Holtz, Markus Rokicki, Philipp Kemkes, and Stefan Dietze. Predicting user knowledge gain in informational search sessions. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 75–84. ACM, 2018.