Understanding Procedural Search Tasks "in the Wild"

Bogeum Choi
University of North Carolina at Chapel Hill
North Carolina, USA
choiboge@live.unc.edu

Jaime Arguello
University of North Carolina at Chapel Hill
North Carolina, USA
jarguello@unc.edu

Robert Capra
University of North Carolina at Chapel Hill
North Carolina, USA
rcapra@unc.edu

ABSTRACT

People often search online for procedural (i.e., "how-to") knowledge. A procedural search task might involve a do-it-yourself project, cooking a dish, fixing a problem, or learning a new skill. Prior research has studied procedural search tasks from different perspectives: estimating the frequency of procedural searches online, understanding how people acquire procedural knowledge in specific contexts, and developing tools to support procedural search. Less research has aimed at deeply understanding procedural search tasks "in the wild". To bridge this gap, we conducted a survey (N = 128) on Amazon Mechanical Turk. Participants were asked to recall a recent procedural task for which they searched online. Participants were asked open-ended questions about the task itself and their unique situation (e.g., constraints and needs). Additionally, participants provided webpages they found useful in their searches and described the characteristics of the page that made it useful. Finally, they provided useful pieces of information from each selected page and explained what they gained from the information. Using an inductive coding approach, we analyzed participants' responses to gain insights about: (1) procedural task characteristics, (2) goals, (3) constraints, (4) contextual factors, (5) relevance criteria, and (6) gains obtained from useful information. Based on our results, we discuss important implications for future research and system design.

CCS CONCEPTS

• **Information systems → Users and interactive retrieval**.

KEYWORDS

procedural knowledge, procedural search, survey study, qualitative analysis

ACM Reference Format:
Bogeum Choi, Jaime Arguello, and Robert Capra. 2023. Understanding Procedural Search Tasks "in the Wild". In *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '23), March 19–23, 2023, Austin, TX, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3576840.3578302

ACM ISBN 979-8-4007-0035-4/23/03…$15.00
https://doi.org/10.1145/3576840.3578302

1 INTRODUCTION

A *procedural search task* requires gathering information to acquire *procedural knowledge*. Procedural knowledge is defined as "how-to" knowledge about performing a task, and can include knowledge about step-by-step procedures, algorithms, techniques, heuristics, methods, and skills [2]. In this respect, a real-world procedural search task might involve looking for information for a do-it-yourself project, cooking a meal, fixing a broken appliance, or learning a new skill.

Prior work has investigated procedural search tasks from different perspectives. First, research has investigated the frequency with which people search for procedural knowledge online, finding that around 3% of all web search queries have procedural knowledge intent [3, 8, 20, 21]. Second, research has investigated how people acquire procedural knowledge in specific contexts, such as software development, cooking, intelligence analysis, and learning [6, 9–11, 15, 19]. Finally, studies have developed and evaluated novel tools to help people search for procedural knowledge [16, 21, 22]. However, less research has studied real-world procedural search tasks and the dimensions that may affect them, including: contextual factors, constraints, goals, criteria used to judge relevance, and different ways that information can help someone move forward with a procedural task. Understanding these dimensions is an important part of supporting users with procedural search tasks. Our research in this paper aims to bridge this gap.

We report on a crowd-sourced survey (N = 128) that asked participants to describe a recent scenario in which they pursued a procedural task that required them to search for information online. In the first part of the survey, participants were asked to describe the task itself, the context of the task, specific constraints they had, and specific types of information they wanted. In the second part of the survey, participants were asked to provide three webpages that they found useful for the task. For each webpage, they were asked to describe characteristics of the page that made it useful. Finally, for each page, participants were asked to provide (i.e., copy/paste or describe) one to three pieces of information they found useful from the page. For each piece of information, they were asked to describe what they gained from it. Using an inductive coding approach, we performed a qualitative analysis of participants' responses to address the following research questions:

- **RQ1:** At the task level, what types of procedural search tasks did participants pursue and what were important characteristics of their situation that may have influenced their searches? We analyzed procedural search tasks from six different perspectives: (RQ1.1) task category, (RQ1.2) task type, (RQ1.3) task goals, (RQ1.4) task constraints, (RQ1.5) contextual factors, and (RQ1.6) prior needs.

- **RQ2:** At the page level, (RQ2.1) what types of pages did participants find useful and (RQ2.2) what were important criteria used to judge the usefulness of a page?
- **RQ3:** At the information piece level, what did participants gain from useful pieces of information?

In Section 5, we discuss how our analysis of real-world procedural search tasks has important implications for future research and system design.

2 RELATED WORK

Prior research in the field of education provides a basis for defining procedural knowledge. The Anderson and Krathwohl (A&K) taxonomy [2] was developed to help educators define learning objectives for students. In the taxonomy, procedural knowledge is defined as "how to" knowledge about performing a specific task. A&K distinguish procedural knowledge from factual, conceptual, and metacognitive knowledge. Procedural knowledge involves knowledge about step-by-step procedures, algorithms, techniques, heuristics, methods, and skills. Additionally, it includes knowledge about when to apply a specific procedure or skill to a task.

Our research uses this definition of procedural knowledge and builds on three areas of prior work: (1) investigating how people search for procedural knowledge using search engines, (2) understanding the *process* through which people acquire procedural knowledge, and (3) developing tools to support users with procedural search tasks.

2.1 Procedural Search on the Web

Understanding procedural search tasks is important because people often use web search engines to acquire procedural knowledge. Völske et al. [20] analyzed about one billion natural language queries (NLQs) issued to Yandex during a one-year period. NLQs accounted for 4% of all query traffic and a large portion (not specified) were queries of the form "how to [verb] [noun]". Common verbs included 'make', 'cook', 'install', 'build', 'calculate' and 'clean'. Interestingly, many queries ended with terms related to specific constraints (e.g., "at home").

Eickhoff et al. [8] analyzed queries issued to Bing over a one-month period. By analyzing clicks on specific websites, an estimated 3% of search sessions had *knowledge acquisition intent* involving declarative or procedural knowledge. For procedural knowledge queries, characteristic n-grams included "how do", "how to", and "can I", suggesting that procedural searches often involve uncertainty about *feasibility*.

Bailey and Jiang [3] analyzed three-month's worth of Bing search sessions to develop a taxonomy of web search tasks. "Learn how to perform a task" was the 12th most frequent category, accounting for 2% of all sessions. Additionally, procedural search sessions were the 3rd longest (13 queries on average). Weber et al. [21] analyzed a sample of 3,000 queries issued to Yahoo! and found that 2% had "how-to" intent.

Collectively, these studies show that procedural searches account for about 2-3% of searches online and that procedural search is complex. While 3% may seem like a small percentage, in practice, it represents a large number of daily queries.

2.2 Procedural Knowledge Acquisition

Studies have also investigated how people acquire procedural knowledge in different contexts. Ertl [9] studied the effects of prior knowledge and collaboration on procedural knowledge gains. Results found two important trends. First, prior knowledge of relevant concepts resulted in better learning outcomes, suggesting that relevant conceptual knowledge is an important prerequisite to gaining procedural knowledge. Second, collaborative (vs. individual) learning resulted in greater procedural knowledge gains, suggesting that exposure to other people's thought processes and perspectives is beneficial during procedural learning.

Freund et al. [10] studied the information-seeking practices of software engineers in a professional setting and found several important trends. First, participants frequently engaged in search tasks involving procedural knowledge (e.g., learning how to do something, troubleshooting a problem, and finding the right tool/resource to solve a problem). Second, participants reported experiencing challenges related to: (1) information overload, (2) inaccurate or obsolete information, and (3) lack of system support for narrowing the search results. Finally, for highly complex tasks, participants preferred information from people with firsthand experience (e.g., asking a colleague or searching on a forum). Byström and Järvelin [4] also found that highly complex tasks (procedural or otherwise) involve greater use of people as information sources.

Pardi et al. [15] studied search behaviors during different types of procedural search tasks: a cognitive task and a physical task. Results found that participants strongly favored visual content, especially during the physical task. Frummet et al. [11] conducted an in-situ study of people's information needs while cooking. The authors developed a hierarchical taxonomy of information needs that included high-level categories related to specific steps, techniques, and recipes with specific inclusion and exclusion criteria. Urgo et al. [19] conducted a study that compared participants' behaviors during search tasks involving procedural versus factual or conceptual knowledge. During procedural search tasks, participants were more likely to engage in creative processes (e.g., modifying and combining procedures). Choi et al. [6] conducted a survey of U.S. intelligence analysts who routinely use an internal system to search for procedural knowledge. Based on their findings, the authors proposed novel features to alleviate the challenges reported by participants.

2.3 Tools to Support Procedural Search

Prior work has also evaluated tools to support procedural search. Pothirattanachaikul et al. [16] leveraged community Q&A data to develop an algorithm to predict alternative procedures that solve the same problem. For example, "taking a sleeping pill", "doing evening exercises", and "drinking chamomile tea" are alternative ways to "improve sleep". Procedural searchers may benefit from seeing advice by people with experience. Weber et al. [21] developed algorithms to automatically identify "how-to" queries and propose relevant tips mined from Yahoo! Answers. Tips were defined as nuggets of advice that are short, self-contained, actionable, and not obvious. Given a procedural query (e.g., "redesigning a living room"), searchers may benefit from seeing query suggestions about the underlying steps. Yang and Nyberg [22] developed algorithms to suggest queries about the steps of an input procedural query.

3 METHODS

To investigate RQ1-RQ3, we conducted a survey study on Amazon Mechanical Turk (MTurk). The study was approved by our university's institutional review board (IRB). In total, we received 128 valid responses. Participants' ages ranged from 21-71 years old (M = 38.1, $S.D.$ = 10.6). Eighty-two participants identified as male, forty-five as female, and one as non-binary. Participants were asked about their highest level of formal education. Twenty-six participants reported obtaining a high-school degree, fifteen an associate degree, seventy a bachelor's degree, and seventeen a graduate degree. Participants reported a wide range of occupations, including chef, engineer, manager, driver, data analyst, and office administrator. Each participant received US$8.00 for completing the survey.

3.1 Survey Design

First, after completing a short demographics questionnaire, participants were asked to recall a specific procedural task that required them to search for information online:

For the questions below, please think of a recent experience when you went online to look for information about how to do *something. For example, perhaps your task involved completing a do-it-yourself project, cooking a meal, repairing a home appliance, performing a task using specific software, and so on. We will refer to the task you select as your "procedural knowledge task".*

Next, participants were asked four open-ended questions about their task:

- **Q1:** Please describe the procedural knowledge task that you have selected. We want to know about a real-world task where you had to search for how-to information online. What was the problem? What were you trying to do?
- **Q2:** Please describe how this task came about. In other words, what caused you to work on this task? We want to know more about the context of the task.
- **Q3:** When working on procedural tasks in real life, you may have constraints that apply to your situation (e.g. avoiding certain ingredients when cooking). Please describe any constraints that were present in your situation.
- **Q4:** Please describe the information you were hoping to find. If you wanted to find multiple different things, please describe them all. Please provide enough details so that we can understand what you were looking for and why.

Then, participants were asked to provide the URL of three webpages they found useful when searching for information. Participants were instructed to re-search for these pages if needed. For each page, participants were asked two open-ended questions:

- **Q5:** What type of page is this? For example, is it a community Q&A page, Wiki article, forum article, review, blog post or video?
- **Q6:** How did you decide that this webpage was useful? Describe the characteristics of this webpage that were important to you.

Then, for each page, participants were asked to provide (i.e., copy/paste or describe) one to three pieces of information they found useful from the page. For each piece of information, participants were asked:

- **Q7:** How did this piece of information help you with the task? Describe what you gained from this information.

3.2 Data Collection and Quality Control

The survey was conducted over a two-week period in May 2021. During this period, we published MTurk Human Intelligence Tasks (HITs) in batches of nine.[1] We published HITs over a two-week period to help recruit a diverse participant sample. Prior work has found that the time of day and the day of the week influences the characteristics of MTurk workers likely to work on HITs, such as their age, employment status, and geographic location [5].

To help ensure English language proficiency and high-quality responses, we limited our HITs to MTurk workers in the U.S., with more than 100 approved HITs, and greater than 95% acceptance rate. In total, we received 247 responses. Of these, we rejected 94 (38%) HITs for which participants did not take the task seriously (e.g., duplicate responses to many different questions). Among the 153 approved responses, we excluded 25 (16%) that involved non-procedural tasks (e.g., learning about COVID-19 safety regulations). Ultimately, a total of 128 valid responses were included in our analysis, consisting of 128 responses to Q1-Q4, 384 (i.e., 128×3) responses to Q5-Q6, and 1,110 responses to Q7.

3.3 Qualitative Analysis

To investigate RQ1-RQ3, we conducted a qualitative analysis of participants' open-ended responses using an inductive coding approach. At the task level (based on responses to Q1-Q4), we developed codes along six dimensions: (1) task category, (2) task type, (3) task goals, (4) task constraints, (5) contextual factors, and (6) prior needs. At the document level (based on responses to Q6), we developed codes related to relevance criteria used by participants when judging the usefulness of a webpage. Finally, at the information piece level (based on responses to Q7), we developed codes related to gains obtained from a specific piece of information.

For all task- and document-level dimensions, two of the authors redundantly coded 100% of the data. Both authors worked together over multiple rounds of independent and collaborative coding to develop, revise, and apply the coding scheme. Ultimately, any disagreements in the application of the coding scheme were resolved through discussion.

At the information-piece-level, which involved coding a larger set of 1,110 responses, our coding of participants' gains followed a different process. First, one author analyzed the data and developed an initial coding scheme of gain categories. Second, two authors independently coded a subset of the data (about 10%) to test the reliability of the coding scheme. At this point, agreement was not high enough, so both authors worked together on refining the coding scheme. Finally, both authors coded a new subset of the data (about 10%). After this second round, intercoder reliability for all gain categories was acceptable. Cohen's κ was greater than 0.50 across all categories and greater than 0.70 for all but two categories. After establishing the reliability of the coding scheme, one author (re-)coded 100% of the data.

4 RESULTS

In this section, we summarize our results with respect to RQ1-RQ3. Figure 1 summarizes all codes associated with RQ1-RQ3.

[1] Amazon charges an additional fee for batches of 10 or more HITs.

Procedural task				
Task category (RQ1.1)	Goals (RQ1.3)	Constraints (RQ1.4)	Contextual factors (RQ1.5)	Prior needs (RQ1.6)
- Create/build - Fix/repair - Upgrade/update - Maintain - Learn - Solving a chosen problem (SCP) - Solving a given problem (SGP) - Daily how-to	- Objective/specific endpoint - Objective/no specific endpoint - Subjective/ operationalizable - Subjective/not operationalizable - Open-ended	Situation-related: - Need to acquire - Need to work without - Need to work with - Money - Time - Environment - External help Individual-related: - Lack of expertise - Physical capacity	- Prior experience - Expected difficulty - Prior skill - Prior knowledge - Trial and error - Inconsistencies	- Steps/process - Outcome - Input/requirement - Implementation details - Alternatives - Domain Information - Tips - People - Diagnosis

Document		Information piece	
Relevance criteria (RQ2.2)		Gain (RQ3)	
- Visuals - Firsthand knowledge - Credibility - Level of language - Level of detail - Alternatives	- Meeting constraints - Amount of information - Arrangement - Simplicity - Popularity - Lists	- Implementation details - Tips and warnings - Input - Example/idea - Visualize - Outcome	- Confidence - Compatibility

Figure 1: Summary of Codes for RQ1-RQ3.

4.1 RQ1: Task Analysis

Based on participants' responses to Q1-Q4, we analyzed tasks along six dimensions: (RQ1.1) task category, (RQ1.2) task type, (RQ1.3) task goals, (RQ1.4) task constraints, (RQ1.5) contextual factors, and (RQ1.6) prior needs.

4.1.1 RQ1.1: Task Categories. From participants' responses to Q1, we identified eight task categories. The percentages below indicate the number of tasks (out of 128) associated with each category. Tasks were assigned to only one category. Therefore, the percentages below sum to 100%.

Create/Build (45.3%) tasks involved creating or building something new. Examples included making rotisserie chicken using an outdoor fire pit, making a new habitat for a pet snake, and building a raised garden bed.

Fix/Repair (23.4%) tasks involved fixing an issue or resolving a negative situation. Examples included unclogging a sink, getting rid of ants inside a house, and fixing a household appliance.

Upgrade/Update (12.5%) tasks involved making improvements. Different from the previous category, these tasks involved improving a situation in cases where nothing was broken. Examples included disabling automatic updates on a computer, replacing a shower head, and refinishing a wooden deck.

Maintain (4.7%) tasks involved maintenance work. Different from the previous category, these tasks did not involve improving a situation, but rather performing a task to avoid a future problem (e.g., deterioration). Examples included changing the oil of a car, cleaning a car's air filter, and polishing leather boots.

Learn (4.7%) tasks involved developing a new skill. Examples included learning to play guitar and learning to monetize a YouTube channel. These tasks are fairly open-ended—they have no clear endpoint and have success criteria that are difficult to measure.

Solving a Chosen Problem (SCP) (3.9%) tasks involved solving a problem that was chosen by the participant. Different from the previous category, these tasks were less open-ended and had more measurable success criteria. Examples included learning how to pack a backpack, how to solve a puzzle, and how to play a game.

Solving a Given Problem (SGP) (3.1%) tasks involved solving a problem that was *not* chosen by the participant. Compared to fix/repair tasks, these tasks involved temporary issues that did not involve something being broken. Examples included joining a videoconference meeting and opening a garage door during a power outage.

Daily How-to (2.3%) tasks involved performing routine day-to-day tasks. Examples included learning to fold clothes.

4.1.2 RQ1.2: Task Types. From participants' responses to Q1 and Q2, we classified tasks into two types: **planned** and **unplanned**. Planned tasks (69%) were initiated by the participant by choice (e.g., making eggplant Parmesan) and unplanned tasks (31%) were externally imposed on the participant (e.g., fixing Wi-Fi connectivity issues).

We examined the relationship between task category and type. Interestingly, 100% of create/build, upgrade/update, learn, and SCP tasks were planned. Conversely, 100% of fix/repair and SGP tasks were unplanned. Maintain and daily how-to tasks had a combination of planned and unplanned cases—50% of maintain tasks and 67% of daily how-to tasks were unplanned. For example, an unplanned maintain task involved "cleaning a car's air filter after *noticing a drop in gas mileage*".

4.1.3 RQ1.3: Task Goals. Participants' responses to Q1-Q3 provided insights about their goals (i.e., success criteria). We uncovered five goal categories.

Objective/Specific Endpoint (70.3%) goals involved success criteria that can be measured objectively soon after a procedure is completed. Examples included making a deep-dish pizza, replacing a car window, and installing a mailbox according to U.S. Postal Service requirements.

Objective/No Specific Endpoint (7.8%) goals also involved objective and measurable success criteria. However, success could not be measured immediately after a procedure is completed. Examples included building a bird feeder that attracts Orioles, getting rid of pigeons nesting on a rooftop, and fixing a reoccurring problem such as a leaky bathtub. This goal category requires *waiting* to see whether a procedure has the intended effect (e.g., the feeder attracts Orioles).

Subjective/Operationalizable (8.6%) goals involved success criteria that are subjective (i.e., vary across individuals) but have common metrics or heuristics to measure success. Examples included making a diet plan that is backed by scientific evidence, making chicken Francese cheaply, and making easy-to-peel hard boiled eggs. Participants often had goals that involved criteria such as 'effective', 'cheap', and 'easy'. While such criteria are subjective, one can imagine metrics or heuristics to measure success (e.g., easy procedures have fewer steps).

Subjective/Not Operationalizable (7.8%) goals involved success criteria that are subjective and lack metrics or heuristics to measure success. Examples included making a delicious prime rib, planting a beautiful garden, and rearranging a home office to be more stylish. Criteria such as 'delicious', 'beautiful', and 'stylish' are highly subjective (i.e., vary across individuals) *and* lack common metrics or heuristics to measure success.

Open-ended (4.7%) goals are both subjective and have no clear endpoint. All tasks that involved learning a new skill (e.g., learning to play guitar) were classified as open-ended.

4.1.4 RQ1.4: Task Constraints. Participants' responses to Q3 provided insights about specific constraints they had. We uncovered nine constraint categories. Two constraint categories (i.e., lack of expertise and physical capacity) are related to the individual and the others are related to the individual's unique situation.

The codes below—and all codes in the rest of Section 4.1—are not mutually exclusive. For example, 14.1% of participants did not mention any constraints and 29.7% mentioned constraints associated with multiple categories. Therefore, in the rest of Section 4.1, the percentages correspond to the percentage of tasks (out of 128) associated with each category and do not sum to 100%.

Need to Acquire (32.0%): In many cases, participants expressed concern about not having (or not knowing if they had) all the necessary tools/materials required by the task. Many participants were concerned about being able to acquire the necessary tools/materials conveniently, cheaply, or at all. For example, for the task of making homemade sushi, P52 wanted recipes that use tools/materials that can be easily purchased.

Need to Work Without (23.4%): Participants often needed procedures that do not require tools/materials they did not already have. Different from the previous category, these participants were not willing to acquire new tools/materials. For example, P75 needed a recipe to make Spätzle without a Spätzle maker.

Need to Work With (18.0%): Participants also mentioned wanting procedures that involved specific tools/materials they had on hand. For example, P81 wanted a recipe for making lemon macarons because they already had lemons.

Money (16.4%): Participants mentioned budget constraints. Some wanted to spend as little money as possible (P9: building a garden arch cheaply) and others had a specific budget in mind (P42: building a computer on a $2,000 budget).

Lack of Expertise (8.6%): Some participants expressed concern about their lack of prior knowledge, skills, and experience in the task domain. For the task of cooking a burrito, P30 expressed concern about not being an experienced cook.

Time (11.7%): Participants mentioned time constraints. As with money constraints, some participants wanted to spend as little time as possible (P82: pre-sprouting potatoes as quickly as possible) and others had a specific time frame in mind (P69: building a raised garden bed before the end of spring).

Environment (7.0%): Participants mentioned constraints related to their environment. Examples included getting rid of ants using a pet-friendly solution, installing Linux given a specific amount of available disk space, and painting the inside of a house with lots of furniture.

Physical Capacity (4.7%): Some participants mentioned being concerned about their physical capacity to complete the task. For the task of installing a ceiling fan, P25 expressed concern about the fan being too heavy for one person to carry.

External Help (4.7%): A few participants mentioned constraints related to external help—completing the task on their own or with the help of others. Examples included texturing drywall without any assistance and making a garden arch with the help of one other person.

4.1.5 RQ1.5: Contextual Factors. Participants' responses to Q1-Q3 provided insights about the context of the task. Participants commented on how the task came about and their unique situation. Based on their comments, we uncovered six contextual factors.

Prior Experience (19.5%): Participants mentioned having (or not having) prior experience with the task or similar tasks. For example, P92 mention having experience with household remodeling tasks but never installing wall tile.

Expected Difficulty (9.4%): Participants commented on their expectations of the task's difficulty. For example, P63 mentioned knowing that cream of chicken soup is a simple recipe.

Prior Skills (6.3%): Participants commented on having (or not having) relevant skills (e.g., being an expert cook).

Prior Knowledge (5.5%): Participants commented on having (or not having) prior knowledge in the task domain. For the task of fixing an electric oven, P89 mentioned knowing about a few possible causes.

Trial and Error (5.5%): Participants mentioned previous attempts to complete the task. For the task of baking bread, P88 mentioned several failed attempts.

Inconsistencies (1.6%): A few participants mentioned finding inconsistencies in previous searches for the task. For the task of baking a chocolate cake, P19 commented on encountering recipes recommending very different temperatures to bake a cake.

4.1.6 RQ1.6: Prior Needs. Q4 asked participants about the types of information they were hoping to find. We identified nine types of needs.

Steps/Process (33.6%): Most frequently, participants mentioned needing step-by-step instructions. For example, P80 wanted step-by-step instructions on how to build an arcade machine using a Raspberry Pi "from beginning to end".

Outcome (32.8%): Participants also wanted information about the outcome of the task or steps of the task. Participants wanted to know what the outcome of the task should look or taste like, whether a solution would be durable, and ways in which the final product would be useful. For example, for the task of making duck stock, P19 also wanted to see recipes that use duck stock.

Input/Requirements (30.5%): Participants wanted information about tools/materials required by the task. For example, P115 wanted to know what equipment they needed to fix a car's air conditioning system.

Implementation Details (23.4%): Participants wanted detailed information about specific steps of the task. Different from the steps/process category, implementation details were related to specific maneuvers involved in the task rather than high-level descriptions of the steps. For example, participants wanted information about exactly what to do to execute a step (possibly under specific constraints) and the rationale behind specific steps.

Alternatives (13.3%): Some participants wanted to find different ways to approach the task. For example, P32 wanted to find different methods for proofing dough to make deep dish pizza.

Domain Information (10.2%): Participants commented on wanting background information about the task domain. Different from implementation details, this type of information is not necessary to complete the task. For example, for the task of learning to mine cryptocurrency, P14 also wanted information about the history of the process.

Tips (8.6%): Participants commented on wanting advice about things to do or avoid. For the task of baking macarons, P81 wanted "troubleshooting tips for common problems".

People (3.1%): Some participants wanted to find other people in a similar situation. For the task of setting up a VPN, P60 wanted to "find someone on a discussion forum who has previously worked with this equipment."

Diagnosis (2.3%): A few participants wanted information to diagnose a problem. For the task of fixing a flickering tail light on a car, P98 mentioned wanting to find information about possible causes of the problem.

4.2 RQ2: Document Types and Relevance Criteria

Participants were asked to provide three webpages that they found useful during their searches. Next, we report on the types of pages provided by participants and the characteristics that made them useful.

4.2.1 RQ2.1: Document Types. From participants' responses to Q5, we identified six document types. Figure 2 shows the percentage of documents (out of $128 \times 3 = 384$) associated with each category.

Some document types deserve additional explanation. How-to pages (e.g., WikiHow articles) contained step-by-step instructions on how to execute a procedure. Informational articles (e.g., Wikipedia articles) contained background information about the task domain. Social forum pages originated from community Q&A sites (e.g., Reddit). Pages in the "other" category included review articles, product diagrams, and magazine articles.

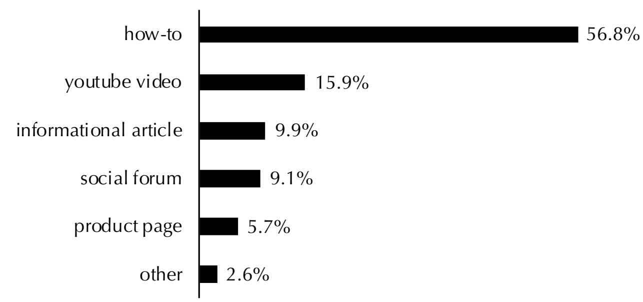

Figure 2: Document Types Percentages.

4.2.2 RQ2.2: Relevance Criteria. From participants' responses to Q6, we uncovered 12 relevance criteria. The percentages below indicate the percentage of pages (out of 128 x 3 = 384) associated with each category. In some cases, participants' responses did not provide insights about specific relevance criteria (e.g., "the page had what I was looking for"). In other cases, participants mentioned relevance criteria related to multiple categories. Therefore, the percentages below do not sum to 100%.

Visuals (25.5%): Most frequently, participants commented on the page having visual information (e.g., images and/or videos). Visual content played three different roles. First, it illustrated how to execute the procedure (e.g., installing a ceiling fan). Second, it helped participants visualize physical components of the task. For the task of removing the headlight of a car, P36 mentioned: "I saw I had the wrong part installed." Third, it helped participants visualize the end product of the task. For the task of baking a casserole, P93 mentioned: "It gave me a general idea of what to expect the casserole to look like."

Firsthand Knowledge (21.4%): Participants commented on the page containing information from people with firsthand experience with the task (or similar tasks). For the task of installing LED lights to the back of a TV, P11 mentioned: "Seeing information posted by people who owned the LEDs was useful."

Credibility (15.6%): Participants commented on the page originating from a credible source. For the task of making a bird feeder, P100 mentioned: "I trust the magazine Birds and Blooms."

Level of Language (15.6%): Participants commented on the page containing information written in simple, easy-to-understand language. For the task of building a computer, P42 commented: "The information was clear and easily digestible."

Level of Detail (15.4%): Participants commented on the page containing detailed (i.e., in-depth) information about aspects of the task. For the task of editing photos using Adobe Lightroom, P28 commented: "[the page] was great at explaining what the different modules are for and in which situations you would use them."

Alternatives (13.3%): Participants commented on the page containing different approaches to the task. For the task of making homemade pizza without special tools, P91 commented: "It gave several options [such as] using a cookie sheet or cast iron pan."

Meeting Constraints (13.0%): Participants commented on the page containing information that was relevant to their unique preferences or constraints. For the task of transforming a patio, P18 mentioned: "I decided this page was useful because of the price of items."

Amount of Information (8.6%): Participants commented on the amount of information on the page. In some cases, participants commented on the brevity of the information (e.g., "It was a short video."). In other cases, participants commented on the vastness of the information (e.g., "[It] provided a lot of information that made it easy to accomplish the task.").

Arrangement (7.6%): Participants commented on how the information was arranged or formatted on the page. For the task of creating a garden arch, P9 mentioned: "It gives the materials in list form and paragraphs of important tips and instructions."

Simplicity (7.6%): Participants commented on the page describing a simple approach to the task. Different from the level of language category, this category relates to the simplicity of the procedure itself instead of the simplicity of the language used in the page. For the task of making deep-fried Oreos, P21 commented: "It had a simple recipe for what I was looking for."

Popularity (5.2%): Participants commented on the page being popular. Participants leveraged different types of evidence to infer a page's popularity, such as the number and quality of reviews, its rating/score, and its view count. For the task of baking bread, P88 commented: "The reviews were good, which told me that this was a decent recipe."

Lists (4.9%): Finally, participants commented on the page having useful *lists* of steps, tips, and required tools/materials. For the task of drilling a hole on a marble countertop, P46 commented: "This page was very useful because it lists the tools."

4.3 RQ3: Information Gains

For each of the three pages provided by participants, they were asked to provide one to three pieces of information they found useful. Additionally, for each piece of information, they were asked about what they gained from the information. From participants' responses to Q7, we identified eight main ways that information helped participants during their task. The percentages below correspond to the percentage of information pieces (out of 1,110) associated with each gain category. Because information pieces could be associated with multiple gain categories, the percentages do not sum to 100%.

Implementation Details (32%): Participants benefited from detailed information about how to perform the task or specific steps. Implementation details included instructions about what to do, when, and how. For the information piece "preheat the oven to 350 degrees F," P19 noted: "I determined the temperature to set my oven."

It may seem obvious that participants gained implementation details from procedural documents. However, we also identified three interesting ways that specific types of information helped participants with implementation. First, participants gained detailed explanations of *how to execute a step*. For the task of replacing caulk in a fiberglass bathtub, P34 said: "A lot of DIY instructions will say things like 'remove the tape' but won't tell you exactly how because they assume it's common knowledge." Second, participants appreciated information that included the *rationale* behind specific choices. For the task of building a raised garden bed, P69 said: "this knowledge [the rationale behind recommended proportions for soil mix] is more useful than just telling me what to add to my mix because it helped me understand why the percentages were there." Finally, participants appreciated information about the *function* of specific components of the task. For the task of unclogging a sink, P1 mentioned: "It gave a rundown of what pipes do what [. . .] it was good to learn how everything works before I started."

Tips and Warnings (21%): Participants mentioned gaining tips and warnings from people with firsthand experience. Tips and warnings refer to advice about things to do and avoid (e.g., common mistakes). Participants gained tips and warnings about the task (e.g., pros and cons of different types of caulk) and specific steps (e.g., not to integrate dry ingredients at first). Tips and warnings were often gained from opinions and judgements by people with experience and were described as "important" and "critical".

Tips and warnings helped participants in different ways. First, they helped participants achieve outcomes with certain desired qualities. For the task of making fried chicken, P17 mentioned: "I knew steam was likely to be an issue [. . .] this tip gave me a way to keep the chicken crispy." Second, tips and warnings highlighted important aspects of the task. Participants found statements such as "the most important thing is to remember is XYZ" as being helpful reminders about where to prioritize their attention. Third, tips and warnings helped participants avoid wasting time, money, or effort. For the task of removing the headlight of a car, P38 encountered information that "stopped me from buying the wrong parts." Finally, tips and warnings helped participants complete the task safely and manage any risks. For the task of making a bird feeder using a plastic bottle, P100 mentioned: "The advice to follow the grooves on the bottle kept me from cutting myself!"

Input (19%): Participants gained information about "inputs" (i.e., tools/materials) required by the task. Participants often gained input information from lists of required tools/materials.

Additionally, input information helped participants in three other ways. First, it helped participants determine where or how to acquire the required tools/materials. For the task of installing a mailbox, P58 mentioned: "It told me that I might be able to rent [vs. buy] an auger to dig the hole." Second, it helped participants learn about alternatives. For the task of replacing a mass airflow sensor in a car, P3 mentioned: "It gave me alternative part numbers that would work with my vehicle if the recommended dealership part was not available." Finally, it helped participants know how to choose the "right" tools/materials. For the task of resurfacing a deck, P59 noted: "It furthered my understanding of treated lumber [and] made me more confident in purchasing the right type for my project.".

Example/Idea (13%): Participants benefited from seeing examples. An example serves as a demonstration (or "instantiation") of a procedure, solution, or outcome. Different from tips and warnings, examples involve people explaining what they did rather than recommending an approach or strategy.

Examples provided three important benefits. First, examples allowed people to *see* a procedure being realized, giving them a clearer idea of what to do and what to expect. For the task of resurfacing a deck, P59 mentioned: "This [gave me] an example of how to repair [the] deck railings." Second, examples provided *new* ideas. This typically happened when participants found examples that illustrated novel approaches to the task. For the task of making duck ramen, P119 referenced someone's comment on a recipe saying that they used chicken eggs instead of duck eggs and mentioned: "I hadn't thought of doing that until I saw the comment." Finally, examples provided inspiration about the range of possible outcomes for the task. This typically happened when participants could see different examples at once. For the task of embroidering towels, P67 encountered a website with different font samples and said: "It gave me an idea of what is available."

Visualize (11%): Participants commented on how visual content helped them visualize different aspects of the task that are difficult to articulate. First, visual content helped participants visualize physical components of the task and their location in the environment. For the task of disassembling a snow blower, P15 mentioned: "This information gave me a visual "jumping-off" point to orient myself." Second, visual content helped participants understand physical movements involved in the task. For the task of texturing drywall, P103 said: "This [cut-in painting technique] is a crucial step in the painting process [and I was] able to see someone do it correctly." Finally, visual content helped participants visualize the outcome of a procedure or intermediate steps. For the task of baking a chocolate cake, P19 mentioned: "I got to know how my butter should look like when I mix the ingredients."

Outcome (11%): Participants mentioned gaining information about: (1) the outcome of the task; (2) the outcome of intermediate steps; and (3) uncertain attributes of the task such as the required time, money, and effort. First, learning about the outcome of a procedure helped participants evaluate their own outcome. For the task of baking bread, P88 mentioned: "It told me what the bread should sound like and feel like when it's finished, which I would have learned a long time after with trial and error by myself." Interestingly, participants also commented that outcome information helped them decide whether to follow a procedure as-is or make modifications. For the task of making barbecue ribs, P118 saw comments about how salty a recipe turned out and decided to use less salt than the recipe suggested. Second, information about intermediate products provided benchmarks that participants could use to determine if they were properly executing a procedure. For the task of installing wall tiles, P92 mentioned: "It showed me how to space the first row of tiles off the floor [...] and how it looks like when it's done correctly vs. incorrectly." Finally, information about uncertain task attributes helped participants set expectations. For the task of making barbecue ribs, P118 mentioned: "It gave me a clear idea [about] the time commitment."

Confidence (10%): Participants gained confidence from certain pieces of information. This happened when information: (1) verified the feasibility of the task; (2) substantiated pre-existing beliefs about the task; or (3) reinforced knowledge about the task gained from a previous source. In all cases, the information helped participants gain confidence and deal with uncertainty. For the task of unclogging a sink, P1 mentioned: "I was happy to learn it's one of the easiest home repair jobs so I was getting confident."

Compatibility (7%): Finally, participants used information to determine if a specific procedure was compatible with their unique preferences or constraints. For the task of getting rid of pigeons nesting on a rooftop, P4 mentioned: "This person had the same problem as me and listed some things they have tried." Similarly, for the task of drilling a hole on a marble countertop, P46 mentioned: "This was very important because my countertop is the exact same thickness."

5 RESEARCH & DESIGN IMPLICATIONS

In this section, we discuss the main implications of our results for future research and system design.

Leveraging our Classification Scheme: Participants reported on real-world procedural tasks that varied along three dimensions: task category, type, and goal. Differences along these three dimensions may lead to differences in searchers' needs, behaviors, and challenges. For instance, in terms of task category, searchers may have different needs during the create/build vs. fix/repair tasks, which were the two most common (70% of all tasks). During create/build tasks, searchers may need more support with the creative process of *ideation* [23]. Conversely, during fix/repair tasks, searchers may need to deal with *hard* constraints unique to their situation. In terms of task type, about 30% of tasks were unplanned. During unplanned tasks, searchers may have less prior knowledge because the task is externally imposed. Finally, in terms of goal type, about 80% of tasks had goals that could be measured objectively. However, some of these had no specific endpoint (e.g., fixing a recurring problem). Such tasks require tracking the performance of a procedure/solution over time. Searchers with such goals may need more information about the outcome of a procedure/solution (e.g., its durability). Future research should leverage our multi-dimensional classification scheme to study how procedural task characteristics impact searchers' needs, behaviors, and challenges.

A (More) Nuanced Classification of Goals: People search with a specific goal in mind. Prior research has distinguished between specific vs. amorphous goals [12, 13]. Specific goals have success criteria that are more objective and measurable, whereas amorphous goals have success criteria that are more subjective and immeasurable. Our analysis uncovered a more nuanced goal classification scheme. Specifically, as mentioned above, objective goals (e.g., fixing a broken item) may vary depending on whether success can (or cannot) be measured *immediately* after a procedure is completed. Other goals have success criteria that are subjective in nature (i.e., will likely vary across individuals). However, some subjective goals (e.g., making a *inexpensive* meal) have sensible metrics to measure success (e.g., total cost) while other subjective goals (e.g., planting a *beautiful* garden) do not. Finally, some goals (e.g., learning a new skill) are completely open-ended, subjective, and lack a specific endpoint.

Constraints: We uncovered nine types of constraints that participants had while working on their tasks. Some constraint categories (e.g., time, money, lack of expertise) are not entirely surprising. However, other categories (e.g., external help, physical capacity) are more unexpected. With respect to constraints, we see two directions for future work. The first research direction involves investigating how users account for constraints in their searches. For example, are users more likely to *explicitly* convey certain types of constraints in their queries? If so, how and are such queries effective? If not, what strategies or workarounds do searchers use to account for such constraints?

The second research direction involves developing novel tools to help searchers account for their constraints. Here, we see three types of tools (from standard to more experimental): (1) faceted filtering, (2) document highlighting, and (3) constraint-based query-by-example. In terms of faceted filtering, systems could enable searchers to filter search results based on their constraints (e.g., recipes for inexperienced cooks). To this end, systems would need to automatically *extract* facet-values from documents (e.g., this recipe takes 40 minutes) and *infer* facet-values that are not explicitly stated (e.g., this recipe is for advanced cooks). Faceted interfaces typically enable users to specify *inclusion* criteria. Our results suggest that interfaces should also enable users to specify *exclusion* criteria. Our results found that 23.4% of tasks had "need to work without" constraints.

In terms of document highlighting, systems could automatically highlight specific types of information within a procedural document. For example, a system could highlight required tools/materials, techniques/skills involved, price references, and temporal references. This feature might enable searchers to judge the usefulness of documents more effectively and efficiently based on their constraints. Indeed, our participants mentioned their constraints when judging document usefulness 13% of the time.

In terms of query-by-example, systems could enable searchers to find related procedures by submitting a procedural document as a query. Additionally, a system could enable users to specify inclusion/exclusion criteria based on the current document (e.g., "find other recipes for the same dish that *use similar tools/ingredients*"). Our participants mentioned documents being useful because they contained alternatives 13.3% of the time. A constraint-based query-by-example feature might help searchers find relevant alternatives in other documents.

Prior work in NLP can provide a starting point to develop these proposed features. Prior work has investigated techniques for automatically populating a procedural knowledge base from semi-structured documents [1, 7, 14, 18, 24] and algorithms to link alternative procedures for the same task [1, 16].

Contextual Factors: We uncovered six contextual factors that impacted how participants approached their task. As expected, participants commented on having (or not having) prior knowledge in the task domain, skills related to the task, or firsthand experience with the task. In our analysis, we *purposely* distinguished between domain knowledge (e.g., knowing the parts of an electric oven), skills (e.g., knowing how to sauté), and experience (e.g., having installed *other* operating systems but not specifically Ubuntu Linux). These three contextual factors might impact searchers differently. For example, searchers lacking domain knowledge may need more background information; searchers lacking specific skill sets may need to see more demonstrations (e.g., videos); and searchers without experience may benefit from comparing/contrasting the current task with other tasks more familiar to them. Future research should further investigate how these factors influence searchers.

Beyond prior knowledge, skills, and experience, we uncovered other contextual factors (expected difficulty, trial-and-error, and inconsistencies) that may impact a searcher's emotional state, engagement, and persistence during search and the task itself. Searchers in such conditions might benefit from information about the *actual* difficulty of the task, as well as the level of trial-and-error and inconsistencies across procedural documents that should be expected.

Prior Needs: As expected, participants frequently commented on wanting step-by-step information on how to complete their task. Interestingly, however, they also frequently commented on wanting information about the outcome of the task and its intermediate steps, the inputs/requirements of the task, and information on how to execute specific steps and their rationales. With respect to prior needs, our results have two major implications. First, resources that host procedural documents (e.g., WikiHow) should encourage authors to include these types of information when writing documents. Systems could even provide authors with a form-based interface that explicitly requests such types of information (e.g., what is the rationale behind this step?).

Second, the document highlighting feature described above could be extended to highlight the types of information that participants needed. For example, the system could highlight steps, input information, outcome information, implementation details, tips and advice, etc. Such a feature might help searchers find the needed information more effectively and efficiently within a procedural document.

Relevance Criteria: Research has studied relevance criteria for decades (see Saracevic [17] for a review). Consistent with prior work, our participants mentioned relevance criteria associated with broad categories proposed by Saracevic [17]: (1) *content* (level of detail, amount of information); (2) *object* (arrangement, lists); (3) *validity* (credibility, firsthand knowledge, popularity); (4) *situational match* (meeting constraints); and (5) cognitive match (level of language, simplicity). In our analysis, we uncovered two relevance criteria that seem uniquely important during procedural search tasks: visuals and alternatives. In terms of visuals, prior work also found a strong demand for visual content during procedural searches [15]. In terms of alternatives, participants described wanting alternatives for several reasons: (1) to have a back-up plan; (2) to choose the most appropriate alternative based on preferences/constraints; (3) to identify steps, tools, or materials that are essential (common across alternatives) or modifiable (different across alternatives); and (4) to combine alternatives into a new procedure. Search systems should consider these relevance criteria when ranking documents and summarizing results on the SERP.

Gains: Our participants reported eight types of gains from pieces of information they found useful (e.g., implementation details, tips and warnings, confidence, etc.). Here, we see two directions for future research. The first research direction involves investigating factors that may impact the types of gains searchers are likely to experience and benefit from. Factors might include the characteristics of the procedural task, the user's unique situation (i.e., constraints and contextual factors), and/or the phase of the task. For example, being able to visualize aspects of the task might be more advantageous during physical (vs. cognitive) tasks; gaining confidence might be more beneficial for searchers lacking prior experience with the task; and gaining inspiration from examples might be particularly important at the beginning of a create/build task.

The second research direction involves automatically linking procedural documents in novel ways. For example, if a document contains step-by-step instructions at a high level, individual steps could be linked to other documents that: (1) describe the step's *inputs*; (2) contain *implementation details* about the step; (3) *visualize* the step; (4) provide *tips* for executing the step; and/or (5) illustrate the *outcome(s)* of the step. From participants comments, it is clear that step-by-step instructions do not necessarily support all types of gains, forcing users to search for other documents. Linking procedural documents in different ways may streamline the search process and improve a searcher's experience.

6 CONCLUSION

To better understand procedural search tasks "in the wild", we conducted a survey on Amazon Mechanical Turk. Our qualitative analysis of survey responses makes several important contributions. First, we developed a classification scheme of procedural search tasks based on three dimensions: category, type, and goal. Future research should leverage our classification scheme to create task scenarios and investigate whether differences along these dimensions influence searchers' needs, behaviors, and challenges. Second, we uncovered a broad range of constraints and contextual factors that searchers must account for during a procedural search task. We proposed novel tools to help searchers account for their unique constraints and contextual factors. Third, we discovered different types of information that participants needed. Besides step-by-step instructions, participants needed background information, information about inputs and outputs, implementation details, alternatives, and advice. Fourth, participants judged the usefulness of documents using a wide range of criteria. Some of these criteria (e.g., visuals, firsthand knowledge, alternatives) seem uniquely important during procedural search tasks. Systems should consider these criteria when ranking and summarizing procedural documents for users. Finally, participants gained a wide range of benefits from useful information. Future studies should leverage our gain categories to investigate their impact on users (e.g., objective and subjective outcomes). Additionally, studies should consider whether their impact *depends* on task characteristics, contextual factors, and/or the phase of the task.

REFERENCES

[1] Eyob N. Alemu and Jianbin Huang. 2020. HealthAid: Extracting domain targeted high precision procedural knowledge from on-line communities. *Information Processing & Management* 57, 6 (2020).

[2] Lorin W Anderson, David R Krathwohl, Peter W Airasian, Kathleen A Cruikshank, Richard E Mayer, Paul R Pintrich, James Raths, and Merlin C Wittrock. 2001. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives, complete edition.*

[3] Peter Bailey and Li Jiang. 2012. User task understanding: a web search engine perspective. https://www.microsoft.com/en-us/research/publication/user-task-understanding-a-web-search-engine-perspective/ Presentation delivered at the NII Shonan: Whole-Session Evaluation of Interactive Information Retrieval Systems workshop. 8-11 October 2012, Shonan, Japan..

[4] Katriina Byström and Kalervo Järvelin. 1995. Task complexity affects information seeking and use. *Information Processing & Management* 31, 2 (1995), 191–213.

[5] Logan S. Casey, Jesse Chandler, Adam Seth Levine, Andrew Proctor, and Dara Z. Strolovitch. 2017. Intertemporal Differences Among MTurk Workers: Time-Based Sample Variations and Implications for Online Data Collection. *SAGE Open* 7, 2 (2017).

[6] Bogeum Choi, Sarah Casteel, Robert Capra, and Jaime Arguello. 2022. Procedural Knowledge Search by Intelligence Analysts. In *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '22)*. ACM, New York, NY, USA, 169–179.

[7] Cuong Xuan Chu, Niket Tandon, and Gerhard Weikum. 2017. Distilling Task Knowledge from How-To Communities. In *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*. International World Wide Web Conferences Steering Committee, 805–814.

[8] Carsten Eickhoff, Jaime Teevan, Ryen White, and Susan Dumais. 2014. Lessons from the Journey: A Query Log Analysis of within-Session Learning. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining (WSDM '14)*. ACM, New York, NY, USA, 223–232.

[9] Bernhard Ertl. 2009. Conceptual and Procedural Knowledge Construction in Computer Supported Collaborative Learning. In *Proceedings of the 9th International Conference on Computer Supported Collaborative Learning (CSCL'09)*. International Society of the Learning Sciences, 137–141.

[10] Luanne Freund, Elaine G. Toms, and Julie Waterhouse. 2005. Modeling the information behaviour of software engineers using a work - task framework. *Proceedings of the American Society for Information Science and Technology* (2005).

[11] Alexander Frummet, David Elsweiler, and Bernd Ludwig. 2022. "What Can I Cook with These Ingredients?" - Understanding Cooking-Related Information Needs in Conversational Search. *ACM Transactions of Information Systems* 40, 4, Article 81 (2022).

[12] Yuelin Li and Nicholas J. Belkin. 2008. A faceted approach to conceptualizing tasks in information seeking. *Information Processing & Management* 44, 6 (2008), 1822–1837.

[13] Jingjing Liu, Chang Liu, and Nicholas Belkin. 2013. Examining the Effects of Task Topic Familiarity on Searchers' Behaviors in Different Task Types. In *Proceedings of the 76th ASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries (ASIST '13)*. American Society for Information Science, USA.

[14] Dena Mujtaba and Nihar Mahapatra. 2019. Recent Trends in Natural Language Understanding for Procedural Knowledge. In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*. 420–424.

[15] Georg Pardi, Yvonne Kammerer, and Peter Gerjets. 2019. Search and Justification Behavior During Multimedia Web Search for Procedural Knowledge. In *Companion Publication of the 10th ACM Conference on Web Science (WebSci '19)*. ACM, New York, NY, USA, 17–20.

[16] Suppanut Pothirattanachaikul, Takehiro Yamamoto, Sumio Fujita, Akira Tajima, and Katsumi Tanaka. 2017. Mining Alternative Actions from Community Q & A Corpus for Task-Oriented Web Search. In *Proceedings of the International Conference on Web Intelligence (WI '17)*. ACM, New York, NY, USA, 607–614.

[17] Tefko Saracevic. 2016. The Notion of Relevance in Information Science: Everybody knows what relevance is. But, what is it really? *Synthesis Lectures on Information Concepts, Retrieval, and Services* 8, 3 (2016), 1–109.

[18] Pol Schumacher, Mirjam Minor, Kirstin Walter, and Ralph Bergmann. 2012. Extraction of Procedural Knowledge from the Web: A Comparison of Two Workflow Extraction Approaches. In *Proceedings of the 21st International Conference on World Wide Web (WWW '12 Companion)*. ACM, New York, NY, USA, 739–747.

[19] Kelsey Urgo and Jaime Arguello. 2022. Understanding the "Pathway" Towards a Searcher's Learning Objective. *ACM Transactions of Information Systems* 40, 4 (2022).

[20] Michael Völske, Pavel Braslavski, Matthias Hagen, Galina Lezina, and Benno Stein. 2015. What Users Ask a Search Engine: Analyzing One Billion Russian Question Queries. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (CIKM '15)*. ACM, New York, NY, USA, 1571–1580.

[21] Ingmar Weber, Antti Ukkonen, and Aris Gionis. 2012. Answers, Not Links: Extracting Tips from Yahoo! Answers to Address How-to Web Queries. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining (WSDM '12)*. ACM, New York, NY, USA, 613–622.

[22] Zi Yang and Eric Nyberg. 2015. Leveraging Procedural Knowledge for Task-Oriented Search *(SIGIR '15)*. ACM, New York, NY, USA, 513–522.

[23] Yinglong Zhang and Robert Capra. 2019. Understanding How People Use Search to Support Their Everyday Creative Tasks. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval (CHIIR '19)*. ACM, 153–162.

[24] Ziqi Zhang, Philip Webster, Victoria Uren, Andrea Varga, and Fabio Ciravegna. 2012. Automatically Extracting Procedural Knowledge from Instructional Texts using Natural Language Processing. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC '12)*. European Language Resources Association (ELRA), Istanbul, Turkey, 520–527.