An Integrated Model of Task, Information Needs, Sources and Uncertainty to Design Task-Aware Search Systems

Shawon Sarkar, Chirag Shah
{ss288,chirags}@uw.edu
Information School, University of Washington
Seattle, WA, USA

ABSTRACT

The varieties of information seeking behavior encompass a range of practices and constructs such as the realization of an information need, selecting the nature of information, as well as information sources. Most of the past works have studied various constructs of the information seeking process, i.e., information, information need, and information sources individually. However, a person forms and re-forms his or her information seeking strategy based on continually shifting values of these dimensions associated with information seeking. This preliminary study conducted a survey with 15 search scenarios and multiple-choice characteristics completed by 114 Amazon's Mechanical Turk workers to find out more about how these constructs play a role in people's preferences regarding information seeking strategies. The study took an exploratory and inferential research approach to investigate how different forms of information and information needs might lead to different information sources by building binary classification models. The results show that the choice of sources can be predicted (with 80% accuracy) if the information need, representation, and form of information are apparent.

CCS CONCEPTS

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

KEYWORDS

Information seeking; tasks and goals; simulation of information interaction; information search context; user interaction; information need; information sources

ACM Reference Format:
Shawon Sarkar, Chirag Shah. 2021. An Integrated Model of Task, Information Needs, Sources and Uncertainty to Design Task-Aware Search Systems. In *Proceedings of the 2021 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR '21), July 11, 2021, Virtual Event, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3471158.3472231

1 INTRODUCTION

Information seeking is often a purposive effort to acquire information in response to a need or gap in seekers' knowledge [14, 62]. An information seeking process can be characterized in three ways – user-based (e.g., why people seek information), content-based (e.g., what type of information they ask), and based on the types of systems or services or sources and methods (where and how people seek information, what modes or types of sources and channels they use) [52]. However, each information seeking process takes place in its own context created by the above-mentioned interrelated constructs and their interactions, affecting seekers' information needs and subsequent actions in a cyclic pattern, thus making the information seeking a holistic process [52]. Furthermore, information seekers do not always make decisions about why, how, what, and where to seek information by considering these possible constructs in isolation. Instead, information seeking is a holistic process motivated by some tasks where an information seeker makes a set of decisions based on and influenced by all of the possible constructs and various dimensions of each of these constructs [53]. In other words, a person forms (and even re-forms) their information seeking strategy based on continually shifting values of several dimensions associated with the broader task that triggers the seeking process in the first place. A holistic view of information seeking allows a better understanding of how all these seemingly disparate yet interrelated constructs interact with one another and how their interactions structure the process of information seeking within its context (e.g., [22, 25, 52]).

Thus, to support information seekers in task completion and problem-solving, information systems need to apply a holistic perspective that encompasses not only the request an individual is making (what) but also understands and utilizes the intention behind the request and the larger goal (why), the strategies the seekers use to engage with information (how), and problems (barriers) they face at every strategic step while providing information [54, 55]. As an initial step towards task-based intelligent search assistance and to find out more about how these constructs mentioned above play a role in people's general information seeking strategies, the study reported here takes an exploratory research approach by considering the interactions of various aspects of the information seeking process. It aims to understand how different types of tasks trigger various types of information needs triggered by different fundamental human needs (e.g., social, emotional, cognitive) that may lead to different methods of seeking information from different sources. In other words, given answers to the "what" (nature and form of information) and "why" (information need) questions, the study seeks to predict the response for the "where" (information source) question in an information seeking process.

One particular contribution of this study is that it decomposes the information seeking process into information needs, sources,

ICTIR '21, July 11, 2021, Virtual Event, Canada

ACM ISBN 978-1-4503-8611-1/21/07…$15.00
https://doi.org/10.1145/3471158.3472231

methods to access sources, representation of information (e.g., information as knowledge), and the interpretation or form of those representations (e.g., fact, advice, opinion). Although many studies have acknowledged the holistic nature of information seeking, understanding how all these components come together and influence people's task completion is still challenging work. Only a few have approached the seeking process holistically (i.e., [12, 26, 58, 59]) and primarily concentrated on selected elements. In the past, most studies either only focused on one component or a few aspects of one element (e.g., one type of source or a few information needs) and did not consider using multiple sources and channels within one information seeking process [28]. Morris et al. [39] studied individual and social information seeking, respectively but with a limited scope. Byström [11] analyzed the relationship between types of information and types of sources and found out that the effects of task complexity made experts a good information source than other people and all types of documentary sources. Generally, the research was focused on information seeking in one type of task, such as everyday information seeking [50], or searching for health information [23]. Furthermore, the findings of the current study have strengthened the notion that depending on the information need to fulfill a particular task, individuals prefer to choose their information strategy and sources.

The rest of the paper is organized as follows. Section 2 provides an overview of all the complementary aspects of the information seeking process and our rationale behind the study's constructed framework. We then point out that there is a lack of research that connects these modalities by studying online information seeking in a task context. This leads to our research questions and two studies we conducted, described in Section 3. Sections 4 and 5 report the findings of the studies. Next, we conclude the paper by listing the contributions and limitations of this work and giving pointers for future studies.

2 BACKGROUND AND RELATED WORKS

According to Kuhlthau [30], information seeking involves fitting information with what one already knows and extending that knowledge to create new perspectives. Past studies have identified the following fundamental factors involved in individuals' information seeking: information/content/material, information needs, sources of information, and methods to access information sources. This section covers several areas of related work and discusses how our work relates to and extends prior work. Then, more importantly, different dimensions for each of these constructs are identified. From the theoretical perspectives, terms such as information seeking, tasks, and information needs are contentious and difficult to agree on among researchers; therefore, we defined the concepts for this study's purposes and reported the findings.

2.1 Constructs of Information Seeking: Information

The term "information" is typically used to represent various overlapping and contradictory concepts, and the meaning is varied based on different philosophical traditions. Two dimensions of information construct have been presented in the existing literature:

- the representation/manifestation or expression of information such as
  - communicated knowledge perceived by the cognitive state of mind or change in existing knowledge structure [6, 8, 19, 38],
  - sensory stimuli [36, 43],
  - the process of gaining knowledge [9, 34, 38, 42],
  - the process of communication [56],
  - social construct [13, 18, 43],
  - and objects conveying information [9, 38, 43]
- the interpretation (form or nature of information) of above representation (e.g., [9, 37, 51]) such as
  - opinion,
  - advice,
  - fact,
  - social or emotional support

Each of these dimensions of the information construct can play a more significant role than the other depending on the information seeking context and based on other critical attributes of a seeking process such as the source of information, the seeker, and the problematic situation that triggers the information need. Therefore, one can interpret "information" as something meaningful that can add or change existing human knowledge structure, a form of a tangible object, or a form of a thinking process based on what seekers define or seek information in a particular context or situation.

2.2 Constructs of Information Seeking: Information Need

People are driven to act in a certain way to fulfill their information needs. Thus, to define "information need," it is important to consider individuals' perceptions and translations of their information needs. In other words, how they choose, formulate, and express their needs, how they perceive the situation that caused the information need, and the way they seek and use information. Based on the information use, the characteristics of information needs are answers, uncertainty reduction, gaps [14]. Affective aspects such as motivations and expectations may also influence the evaluation of information and its needs [28].

Therefore, understanding the seeker's motivations for asking a question could provide a general framework for conceptualizing different contexts and situations of information needs that drive people to seek information from different sources. Existing studies identified various motivational factors behind information needs (e.g., [16, 27, 40, 65]). Zhang [65] categorizes motivations into three broad factors – *cognitive*, *social* and *emotional*. Katz et al. [27] categorized several basic human needs into five groups:

- needs for strengthening information, knowledge, and understanding (cognitive factors),
- needs for strengthening aesthetics, pleasures and emotional experience (personality factors),
- needs for strengthening credibility, confidence, stability (judgment of information values and relevance)
- needs for strengthening contact with family, friends and the world (situational and social and environmental factors)
- needs for escape or tension release (affective factors)

Broadly, the motivation behind information need could be *cognitive* (seeking relevant information, opinions, or advice for making decisions), *affective* (looking for social and emotional support), social (identifying with others and gaining a sense of belonging), or *personal* (finding support for one's own values) [16, 27, 65].

2.3 Constructs of Information Seeking: Information Source

Individuals utilize various sources available to them while fulfilling their information needs, such as friends and families, books, and websites. An information source is a repository that stores and provides knowledge or information [17, 64]. Sources of information can be impersonal or non-human such as online sources, physical documents, various information retrieval systems, or interpersonal or human-related such as friends and colleagues on various online or offline channels [1]. Several studies on information source selections and preferences have also found that different kinds of information sources are preferred in different situations, in different everyday life and work-related task contexts (e.g., [12, 31, 45, 46, 50]). Moreover, there are various conditions, including seekers' previous experience with a source, accessibility of the source, and the format and content of the information available [33], by which information seekers value and select information sources. Anderson et al. [2] found that people perceived information sources that are easier to use as more accessible and choose them frequently regardless of the quality of the information they expect to obtain. On the other hand, Ashford [3] found out that the source quality is more important than accessibility while thinking in a cost-benefit framework. Besides, time plays an influential role in selecting the type of source [60]. Fisher and Naumer [20] identified several preference criteria for information sources in their study, such as trustworthiness, contact, access or convenience, inexpensiveness, and ease of use.

2.4 Constructs of Information Seeking: Methods to Access Information Source

Although information sources and the methods or channels to access the sources are often synonymous in many existing studies [12, 14], these two concepts are distinct [1]. From existing research, the study identifies two dimensions of the method of accessing information sources and adds another dimension of mediation involved in accessing the sources:

- Physical-electronic dimension - the use of a physical or electronic medium for information transfer [1]
- Synchronous-asynchronous dimension - the synchronicity of communication or connection [1]
- Mediation-No mediation - involvement of any mediated entity

According to Xu et al. [64], channels are the modes of communication through which content is delivered from an information source to seekers, such as face-to-face, phone, or e-mail. Various cues [63] shapes individuals' perceptions of information sources.

To summarize, based on the literature, the current study attempts to depict a multi-dimensional approach to look at information seeking as the product of the relationships among its significant constructs – the type of information (what), information need (why), information source (where), methods (how), and various dimensions of these four aspects to create a framework of information seeking to understand people's information seeking behavior as well as to create better support, services, and systems to enhance this vital activity. Invariably in the research process, this study will make certain assumptions about the information seeking, the information being sought, the methods and sources being used, and the possible solutions in terms of useful or relevant information and the use or sense derived from such information.

3 METHODOLOGY

To understand the relationships among various information needs, forms of information, and information sources, we conducted an exploratory survey of people's information seeking behavior in everyday life. The following research questions guided our study design, data collection, and analysis techniques.

- **RQ1**: What is the relationship between information forms, information needs, information sources, and how are the sources accessed for tasks?
- **RQ2**: Is it possible to predict how different forms of information and information needs might lead to different sources of seeking information for a particular task?

Among the RQs above, RQ1 focuses on identifying the relationships among information needs and information types at different points of search interactions, and RQ2 focuses on modeling the information source use patterns in varying information need scenarios.

3.1 Study Design

To understand how people choose to use information resources, we conducted a survey where we asked people to hypothesize a situation where their information search process significantly impacted their decisions or actions. The survey had 15 unique information scenarios, each representing a particular information need and the required form of information to fulfill that need. Based on Katz et al. [27], Choi and Shah [16], Zhang [65], and Oh [40], information need was divided into four categories – cognitive, personal, social, and affective:

- Cognitive needs – finding relevant information in immediate surroundings, society, and the world; seeking advice or opinions for making decisions; learning or self-education through acquiring information; gaining a sense of security through knowledge
- Personal integrative needs – finding support for one's own values; gaining insight into one's own life; experiencing empathy with problems of others
- Social integrative needs – gaining a sense of belonging; finding a basis for conversation and social interaction; feeling connected with other people
- Affective needs – looking for social and emotional support for personal issues; looking for social and emotional support for someone (e.g., family, friends, and so on.); looking for attainment on personal thoughts or Ideas

Based on theories proposed by Marchionini [37], Buckland [9], and Rulke et al. [44], the representation of information was categorized into three: object, knowledge, and social or sensory stimuli contextual. Moreover, the study borrowed various forms of information from Shah and Kitzie [51] and combined them into four broad categories – advice, opinion, fact, and social or emotional support. Combining all the aspects gave 15 unique combinations, and based on that, we had 15 scenarios. Table 1 presents a short overview of the survey scenario design.

3.2 Developing Scenario

Each of the 15 task scenarios was a unique combination of these three dimensions of two aspects of information seeking – information needs and information. Based on Borlund [7]'s concept of simulated task scenarios, which stated that the scenarios must be realistic to participants, that is, they must fit the participants' real-life situations, we constructed each scenario from various ordinary everyday lives situations. As the observed variations in research findings could be interpreted or explained differently based on the well-known effect of task (here, scenario) complexity (for example, web searching was popular in fact-finding and exploratory tasks), we maintained a consistent task complexity throughout all the seeking scenarios.

For each scenario, there were two sets of information sources. The first set of options included human or interpersonal sources and methods to access those sources. The second set included all impersonal sources, including online and offline information retrieval systems. There were no methods to access these sources because that was implied in their names (see Table 1). There was no particular order in the scenario assignment. The following section describes each of the 15 scenarios.

Although participants' background might affect their responses and the validity of this study's results, we decided not to collect demographic or personal data from the participants. We intended to understand the preferences of sources in general, independent of any contextual factor. However, the unknown demographics and actual information needs of the MTurk workers who participated as well as the artificiality of the task scenarios might raise some questions about how much the survey responses were truly being answered in the context. Therefore, We consulted two expert researchers who specialized in interactive information seeking and retrieval while creating the scenarios to check whether each scenario represents the stated cognitive needs, personal needs, social, affect-driven needs, and forms of information. Furthermore, we performed several rigorous pilot surveys and evaluations with five participants from different backgrounds to verify the effectiveness of these scenarios. We asked for feedback on the descriptions and language of the scenarios and whether they can relate to the task described in the scenarios in an iterative process.

3.2.1 Scenarios for the Survey. **Scenario 1: Cognitive need, Informative object, Fact** Suppose you and your friend went to the grocery store together to buy produce. You are discussing the attributes of different vegetables and fruits; taste, nutrition, etc. Your friend is explaining that the skin of "Red Delicious" apples is thinner than the skin of "Granny Smith" apples. You are not fully convinced of this and you want to find out whether it is true. Where and how do you look for information?

Scenario 2: Cognitive need, Knowledge, Fat Suppose you are just about to start your first semester. You have decided to buy a new laptop instead of taking your 5-year-old computer to the college. There are different types of laptops with various specifications and you cannot decide which laptop you would buy a Mac or Windows? Would you prefer a 4 GB Ram or an 8 GB Ram? You have decided to look for some facts on how to compare different laptops. Where and how do you look for information?

Scenario 3: Cognitive need, Knowledge, Opinion As a student, suppose you have to create a poster presentation for your social science class. You have done presentations with PowerPoint but have never presented with a poster. Since this is your first poster, you are most concerned about how such a document should look. You want to find some opinions concerning the best ways to organize a poster layout so that you are prepared to create your project. Where and how do you look for information?

Scenario 4: Cognitive need, Knowledge, Advice This summer, you are traveling to Australia. You have never been there before and there are many tourist attractions you would like to visit. Therefore, you want information to help you plan your journey. You have set aside one month for the trip and hope to see as much of the country as you can. As you are unfamiliar with the territory, you would like to know about the places before going there. Where and how do you look for information?

Scenario 5: Personal need, Social or Sensory stimuli, Social or emotional support Suppose you have a -3.0 power in both eyes and you wear eyeglasses regularly. In your recent visit to your ophthalmologist, your doctor prescribed contact lenses for clearer vision. You have never used contact lenses before, so you would like to know about overall experience of wearing them - are they easy to wear or painful? Based on your findings you would like to decide whether to wear the lenses. Where and how do you look for information?

Scenario 6: Personal need, Social or Sensory stimuli, Fact Suppose you recently read a book on health, genetic disorders, and diet, and it has influenced the way you think about your health and lifestyle. You have learned that there are some diseases like cancer or diabetes can be caused by the combination of mutations of inherited genes, lifestyle choices, and your environment. Many rare diseases and conditions usually develop when an individual is born with a mutated gene. After reading the book, you want to know if a rare disease or common health conditions run in your family so that you can make precautionary changes in your existing diet or lifestyle. Where and how do you look for information?

Scenario 7: Personal need, Knowledge, Opinion You have followed the news and debate about legalizing marijuana in certain US states. In your personal opinion, alcohol and cigarette smoking are much more dangerous than marijuana.Now you are curious to know how others feel about this issue. Where and how do you look for information?

Scenario 8: Personal need, Knowledge, Fact Suppose your father is returning home tomorrow from the hospital after a prolonged illness. Tomorrow also happens to be his 60th birthday. On this special occasion, you want to make your grandmother's old chicken broth recipe to surprise your father.However, you do not

Table 1: Outline of the Survey Scenario

Given aspects in each scenario	The participants would then have to provide the following aspects for each scenario
What information do you need? Why do you need this information?	What sources do you use while seeking the information? How will you access the source of information?
Example scenario: Suppose your father is returning home tomorrow from the hospital after a prolonged illness. Tomorrow also happens to be his 60th birthday. On this special occasion, you want to make your grandmother's old chicken broth recipe to surprise your father. However, you do not know the recipe and you need to find it out. Where and how do you look for information?	**Example Sources:** **Interpersonal:** Friend, Family, Colleague,... **Access Methods:** Face to Face, Phone, Online chatting,... **Impersonal:** Book/Manual, Search Engine, Social media, pre-posted Forum Entries,...

know the recipe and you need to find it out. Where and how do you look for information?

Scenario 9: Social need, Knowledge, Fact You will be attending asocial gathering this evening. It is a birthday party for a friend being held at a local restaurant. You do not know many of the guests in attendance. You thought you could facilitate conversations with new people if you were up-to-date on some recent topics of interest. You have decided to look into a wide expanse of events since you do not know the other guests' interests and backgrounds. Where and how do you look for information?

Scenario 10: Social need, Social or Sensory stimuli, Social or emotional support Your cousin, a senior in college, said that one of her friends started to smoke. You fear your cousin might begin smoking in the near future and decide to educate her about the risks, so you have to find some information on what could happen if she starts smoking. Where and how do you look for information?

Scenario 11: Affective need, Informative object, Social or emotional support Imagine your family members are out of town for a family wedding in California. You were unable to go because of a work commitment. It is now a few days since they have gone and you are missing them very much, especially when you are at home all alone. You are feeling very sad and in order to feel better you would like to find something to cheer you up. Where and how do you look for it?

Scenario 12: Social need, Social or sensory stimuli, Advice A friend of yours has an appointment to get a flu shot tomorrow, but she has a cold. She is debating whether to keep her appointment. She asked for your advice. You would like to help her by finding out what is generally recommended for people in her situation. Where and how do you look for information?

Scenario 13: Affective need, Knowledge, opinion Suppose you have inherited a large sum of money left by your recently deceased uncle. You are not sure about what to do with this money. You are considering investing it in the stock market as a bond or corporate stocks. However, you are unaware of stock market trends and lack the knowledge required to make a sound judgment on what to do with your inheritance. You would like to find information to help you pick the best type of financial instrument for the investment. Where and how do you look for information?

Scenario 14: Affective need, Knowledge, Fact After graduating as a veterinary doctor, you are just about to start a career in the veterinary industry. You are concerned about how to work towards your retirement.Therefore, you want to know what you should expect from your industry. Where and how do you look for information?

Scenario 15: Affective need, Knowledge, Social or sensory stimuli Your best friend is getting married and asked you to be in their wedding party. You have been friends since the 5th grade. You feel touched, excited, and also a little stressed. You have to plan the shower and bachelor(ette) and you want them to both be great and memorable for your friend. You have no prior experience and no idea where to begin. Therefore, you need information on how to become a perfect head of the wedding party. Where and how do you look for information?

3.3 Study Procedure and Data Collection

We conducted the data collection survey on Amazon's Mechanical Turk (hereafter MTurk). MTurk provides a crowd-sourcing platform for individuals or businesses to post human intelligent tasks (HIT) for MTurk users from around the world. Furthermore, existing studies (e.g., [10, 35]) have found out that MTurk can be used to obtain high-quality data inexpensively and rapidly. MTurk's demographically diverse participants could provide reliable and good quality data like other traditional survey collection methods such as standard Internet samples and typical American college samples [10, 35]. Therefore, since this study's primary interest is exploration, the instant available and easily accessible large participant pool on MTurk was appropriate. We also took several additional measures to ensure data quality and reliability; for example, we recruited the expert participants, offered a higher compensation rate, and maintained a task length workable to all. Several general attention-check questions were also included in various survey parts to check participants' attention and randomness of their answers.

Responses from 114 MTurk workers who were at least 18 years and from the USA were collected. In the 30 minutes lengthy questionnaire, the scenarios were presented to the workers. The participants had to choose where and how they would search for that information from an exhaustive list of information sources. They

could pick any source from any source type (impersonal or interpersonal) for each scenario. All participants responded to all 15 scenarios. Typically, the majority of the respondents took 15 to 30 minutes to complete the survey. Their recorded responses had been downloaded and quantitatively analyzed. The final dataset contains participants' source selections in 1,710 scenarios.

4 DATA ANALYSIS AND RESULTS

To address the **RQ 1** (*the relationships among information forms, information needs, information sources, and how the sources are accessed*), the study takes an exploratory approach to analyze the survey responses (see Tables 2 and 3). It should be noted that multiple or no answers for a scenario were possible.

The descriptive statistics show that most participants chose to use the search engine for scenarios 1, 2, 3, 4 (see Section 3.2.1 for descriptions of the scenarios). The second-largest group of participants chose to consult a person, primarily friends and professionals. For scenarios formulated on personal needs, information as fact, opinion, and social support (5, 6, 7, and 8), most respondents selected interpersonal sources over impersonal sources. While for scenario 5, the popularity of friends and search engine were very close (80 and 83 times respectively), participants preferred to ask family members in scenarios 6 and 8 and friends in scenario 7. Respondents chose interpersonal sources 87 (family), 76 (friend), and 111 (family) times in scenarios 6, 7, and 8. The second most popular choice for these three scenarios is the search engine (impersonal source), which they chose 59, 69, and 43 times respectively. Interestingly, there is a large gap between the first most popular and the second most popular source choice in those scenarios. Most respondents chose to consult an impersonal source in socially-driven scenarios (9, 10, 12). The popular choices for these two scenarios were closely distributed among consulting friends (scenario 9), searching directly on the professional website, reading newspapers, news-media sites, and pre-posted entries on social media. In the scenarios based on affective needs (11, 13, 14, and 15), respondents mostly opted for impersonal sources, especially web search, except for scenario 11, where they chose interpersonal sources (friends) over other sources. Here we also found a mixture of impersonal and interpersonal sources they have chosen to use, for example, pre-posted entries on social media, colleagues, and so on.

From the analysis of people's preferences of methods to access interpersonal sources, it is evident that people preferred to talk to familiar sources such as friends and families directly in person, face to face. When consulting strangers (e.g., customer services, dealers), on the other hand, they also liked to write messages indirectly or to post online (see Table 4).

The question of predicting the answer for the "where to seek information" **RQ2** (predicting information sources from forms of information and information needs) has been molded into a multi-class classification problem. The goal is to examine whether it is possible to predict participants' choice of information sources from the known information need and the nature and form of the required information to fulfill the need. We conducted traditional machine learning classification experiments with an information need, information type, information form, and methods to access sources as independent features and source type as classification labels to address the question. After preparing the data for predictive analysis, we had 4592 total data points.

Table 2: Use of Interpersonal Sources

Scenario	Friend	Family	Colleague	Stranger
1	23	20	2	26
2	60	33	21	31
3	50	18	48	11
4	28	24	5	33
5	80	54	21	20
6	9	87	2	13
7	76	42	21	18
8	3	111	0	2
9	49	17	10	10
10	17	19	2	18
11	84	37	2	2
12	19	16	5	27
13	23	32	15	47
14	12	8	48	29
15	72	54	6	14
Total	605 (35.38%)	572 (33.45%)	208 (12.16%)	301 (17.60%)

Since the data were categorical and the target variable had more than two types of sources, all the features and target variables were encoded using the *One Hot Encoder*, and *Label Encoder* methods [41]. After encoding the data, training, and a testing sample have been created based on 70% and 30% of the original data set grouped by the users, respectively. As the dataset was small and moderately unbalanced, to mitigate the bias towards the majority class (as it would have a more extensive influence on the final loss value), we added weights to cross-entropy losses corresponding to different classes to even out the data bias. Furthermore, to make the data balanced, we increased the frequency of minority classes by using up-sampling clustering techniques to make the data balanced. Generally, up-sampling is preferred when the overall data size is small.

We conducted multiple experiments using several Bagging, Boosting, and Voting models [5]. These ensemble models can handle small and unbalanced datasets very well. In particular, we built Decision Tree [47], Random Forest [32], Extra Trees [57], AdaBoost [24], Gradient Boosting [21], Voting [5] and XGBoost [15] models and evaluated them all on the same training, testing and validation sets. We compared our models against two baseline models – the most frequent model and the stratified random model. We reported the performance of each model in Table 5. To define the Voting classifier using multiple classification algorithms, we used the predictions of logistic regression, regression trees, and support vector machines together. Finally, the Voting classifier averaged the predictions of the sub-models and calculated the accuracy.

Table 5 showed that Random Forest and Extra Trees performed better than other models and could predict the sources with 80% accuracy. To evaluate the performance of our models, we calculated an accuracy score for each model using the mean average

Table 3: Use of Impersonal Sources

Scenario	Book	Offline Catalog	News-paper	Any Object	E-copies	Online Data-base	Google Scholar	Online Catalog	Social media Pre-posted Forum Entries	Websites	News media sites	Search Engine	TV channels	Radio
1	10	3	2	0	12	16	11	4	15	5	2	92	0	0
2	0	6	7	0	7	15	4	1	32	35	1	92	0	0
3	13	2	1	6	11	8	5	3	24	14	1	86	0	0
4	18	4	8	8	3	9	0	2	47	59	7	97	0	1
5	3	1	3	2	4	2	1	0	51	17	4	83	0	0
6	5	2	4	5	16	7	6	4	15	20	2	59	0	0
7	3	1	11	1	10	1	4	1	67	8	26	69	3	2
8	13	2	1	7	3	2	0	2	10	4	0	43	0	0
9	1	0	31	3	4	2	0	1	38	3	67	59	25	6
10	10	0	11	5	36	5	10	0	25	29	13	84	1	0
11	6	0	6	1	1	1	1	0	26	5	5	39	31	4
12	1	2	3	4	16	3	5	1	20	34	7	98	0	0
13	10	2	9	2	11	2	2	0	29	56	13	77	2	0
14	4	1	7	1	11	9	1	0	37	44	3	90	0	0
15	11	2	14	1	1	2	0	0	68	16	3	100	2	0
Total	108 (6.32%)	28 (1.64%)	118 (6.78%)	46 (2.69%)	146 (8.54%)	84 (4.91%)	50 (2.92%)	19 (1.11%)	504 (29.47%)	349 (20.42%)	154 (9.01%)	1168 (68.30%)	64 (3.74%)	13 (0.76%)

Table 4: Use of Channels to access Interpersonal Sources

Methods	Friend	Family	Colleague	Stranger
Face to Face	282	253	144	98
Phone	108	184	15	22
Texting	132	97	22	2
Online Chatting	59	18	4	25
Email	15	16	18	10
Posting on online forums	9	4	5	144
Mail/Letter	0	0	0	0

Table 5: : Accuracy on the data set for several algorithms: DecisionTree (DCT), Random Forest (RF), Extra Trees (ET), AdaBoost (ADA), Gradient Boosting (GB), Voting, and XGBoost. Baselines are a most frequent (MFQ) and stratified random (STR) baseline. Best performers in each column are boldfaced. Significant values indicate whether the predictor is significantly better than its baseline (*=p<.05).

Classifier	Information Source
Decision Tree	78.48%*
Random Forest	**80.0%***
Extra Trees	**80.0%**
AdaBoost	66.67%
Gradient Boosting	78.45%
Voting	78.46%*
XGBoost	75.96%
Most Frequent	48.5%
Stratified Random	50.0%

percentage error subtracted from 100%. Furthermore, for the Random Forest, we generated a single tree as shown in Figure 1. As

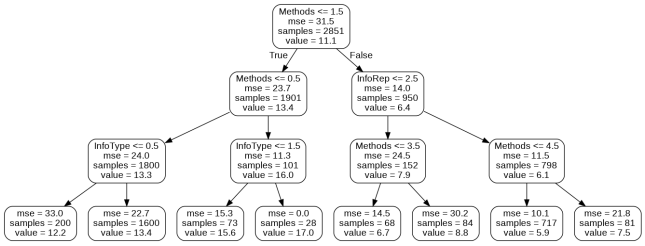

Figure 1: Random Forest tree.

Random Forest is a combination of a subset of the features at each node of the tree, it would be possible to check the accuracy of the predictions by examining a tree. Also, Random Forests can achieve high accuracy without the risk of overfitting or underfitting data.

5 DISCUSSION

From the descriptive analysis of the data, some patterns have emerged among the information needs, information required, and participants' choices of sources and methods to obtain the information. Some of them are new, and some solidify existing research. First, although respondents chose friends and family most of the time as information sources, overall, they consistently opted to use all four types of interpersonal sources in all scenarios (except, scenario 8 where they only chose "family"). However, their choice of impersonal sources was limited only to few types of sources – web search engines, pre-posted entries on social media, professional Websites, online news media sites, electronic copies of journals, magazines, and lastly, newspapers and books or manuals. The popularity of online information sources, especially pre-posted entries of social media and online forums to communicate with strangers, is a reminder of the changing characteristics of human information seeking behaviors. People prefer to use online information sources

more and more, which indicates a rise in collective and collaborative information seeking behavior. Second, although participants had to choose at least one source, almost all preferred to use multiple sources of information and both impersonal and interpersonal for all scenarios. This suggests that many people usually prefer not to seek information only in one way or one time. In the information seeking process, people usually tend to use multiple information sources in various ways, and it is not a linear process; instead, it is a repetitive circular process. However, for the recipe task, people only opted for the interpersonal source, as this task can only be solved by asking interpersonal sources, as participants had to search for their grandmother's chicken broth recipe.

The predictive analysis of the data also produced some insightful observations. The accuracy scores of all classification models are relatively high for the information source, which indicates that it is possible to predict people's choice of information sources if the motivation or need behind the problem situation and the nature of the information they seek to find are known. From the accuracy scores and the descriptive analysis, it is visible that respondents did not show any personal preferences or fondness for any particular source type. Instead, they preferred to choose the source based on the information need-based scenarios. Certain assumptions could be made to explain the behavior. For scenarios that were considered personal and intimate, the participants mainly chose interpersonal sources, especially friends and families who could be trusted or who could provide more contextual information. For scenarios with cognitive needs, respondents opted to use impersonal sources, mostly web searches, pre-posted entries on social forums, or professional Websites. This suggests that the user's motivations and expectations behind information seeking play significant roles in deciding information sources and methods to access them.

Moreover, although people prefer to use web search engines to find information primarily because of their easy accessibility and ability to provide a vast amount of information in a short time, the study shows that people use other information sources as well. People chose to use a book, newspaper, or other physical information sources depending on the information need. Therefore, it can be concluded that the user's motivations and expectations behind information seeking play essential roles in deciding information sources.

Thus the findings of the current study have strengthened the notion that depending on the information need to fulfill a particular task, individuals prefer to choose their information strategy and sources. It confirms the findings of several other studies(e.g., [48, 49, 61]) that while a high proportion of current internet activities involved searching and browsing for answers and gathering information to understand a topic better, users often turned to other people during critical activities, asking for their help or opinion. The findings also highlight participants' preference for multiple sources of information and both impersonal and interpersonal. Thus, this study reported here emphasizes the learning aspect of the seeking process, as outlined by Kuhlthau [29] and Bates [4], and practical design considerations for helping users interact with information and learn.

While there has been growing interest in this area as reflected in the scholarly outcomes and events over the past few years, there is a lack of task-based search and intelligent assistance, which motivates this study to take a holistic view on the information seeking process in order to connect these complementary aspects of individualized and social search processes to a seeker's context and the task at hand. Thus, this study has contributed to this line of investigation further, which has great potential value and implications for personalizing search and recommender systems to suggest types of information and resources in various formats (e.g., people, retrieval databases, document, or query) based on their needs and preferred information forms at task levels and going beyond the query and document only suggestions.

6 CONCLUSION

This study's overall objective is exploratory and to observe information seeking based on the previously identified four aspects of information seeking behavior and various dimensions of them. The study explores how different types of information and information needs may lead to different sources and channels of seeking information. By giving the participants to choose from an extensive list of information sources for each of the 15 scenarios, the study assumed that they would prefer the resource they think is the most useful for similar scenarios in real life. The study displays that when people decide to use a particular source, their decisions are heavily influenced by the type of information they are looking for and the kind of information needs they have.

There are some limitations to this study. The findings are based on an online survey conducted on a crowdsourcing platform, thus limiting the responses and direct communication between the researchers and the respondents. The scenarios were simulated and imposed on the participants, and they had to hypothesize the scenarios, not actually perform the searches. Therefore, the responses were not self-motivated. The study only collected their selection choices and not explanations behind their choices.

Moreover, as with all self-reported data, users' self-reported surveys on information seeking are also open to biases. What participants say they would do in a scenario is not necessarily the same as they would do if they actually searched for the information. Although the findings support that a group of people prefers to use different information sources based on their information needs and their choices of information sources can be predicted from their need for a particular type of information, the exact needs may be interpreted differently based on people's actual situational contexts.

Although there are some limitations to this study, this study produces meaningful patterns regarding individuals' information needs, sources, and methods. It is crucial to note that people prefer to use multiple sources depending on the type of need, probably to receive affirmation or other opinions on the same topic from interpersonal and impersonal sources. The study also shows that it is possible to predict people's choice of information sources based on their information needs and preferred information forms. The findings could provide foundations toward developing personalized search, recommender, or intelligent systems, which can then recommend various sources of information to seekers based on their information needs and preferred form of information. Thus, this study hopes to contribute to the information seeking research and provide a foundation for further research along these lines. Furthermore, in this study, we did not collect and analyze demographic

or personal data. In the future, we would investigate how these data could help determine to which degree the choice of information sources is influenced by the information need and which role demographic factors play.

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation (NSF) grant III-1717488. We thank all of our participants and members of the InfoSeeking Lab for their helpful input on the survey instrument.

REFERENCES

[1] Naresh Kumar Agarwal. 2011. Verifying survey items for construct validity: A two-stage sorting procedure for questionnaire design in information behavior research. *Proceedings of the ASIST Annual Meeting* 48 (2011).
[2] Claire J Anderson, Myron Glassman, R Bruce McAfee, and Thomas Pinelli. 2001. An investigation of factors affecting how engineers and scientists seek information. *Journal of Engineering and Technology Management* 18, 2 (2001), 131–155.
[3] Susan J Ashford. 1986. Feedback-seeking in individual adaptation: A resource perspective. *Academy of Management journal* 29, 3 (1986), 465–487.
[4] Marcia J Bates. 1989. The design of browsing and berrypicking techniques for the online search interface. *Online review* (1989).
[5] Eric Bauer and Ron Kohavi. 1999. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine learning* 36, 1-2 (1999), 105–139.
[6] Nicholas J Belkin. 1980. Anomalous states of knowledge as a basis for information retrieval. *Canadian journal of information science* 5, 1 (1980), 133–143.
[7] Pia Borlund. 2003. The concept of relevance in IR. *Journal of the American Society for information Science and Technology* 54, 10 (2003), 913–925.
[8] Bertram C Brookes. 1977. The developing cognitive viewpoint in information science. *De Mey, M. and al Editors* (1977), 195–203.
[9] Michael K Buckland. 1991. Information as thing. *Journal of the American Society for Information Science (1986-1998)* 42, 5 (1991), 351.
[10] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. 2016. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? (2016).
[11] Katriina Byström. 2002. Information and information sources in tasks of varying complexity. *Journal of the American Society for information Science and Technology* 53, 7 (2002), 581–591.
[12] Katriina Byström and Kalervo Järvelin. 1995. Task complexity affects information seeking and use. *Information processing & management* 31, 2 (1995), 191–213.
[13] Rafael Capurro and Birger Hjørland. 2003. The concept of information. *Annual review of information science and technology* 37, 1 (2003), 343–411.
[14] Donald O Case and Lisa M Given. 2012. *Looking for information.* Emerald Group Publishing Limited.
[15] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 785–794.
[16] Erik Choi and Chirag Shah. 2016. User motivations for asking questions in online Q&A services. *Journal of the Association for Information Science and Technology* 67, 5 (2016), 1182–1197.
[17] Edward W Christensen and James R Bailey. 1997. A source accessibility effect on media selection. *Management Communication Quarterly* 10, 3 (1997), 373–387.
[18] Ian Cornelius. 2002. Theorizing information for information science. *Annual review of information science and technology* 36, 1 (2002), 392–425.
[19] Brenda Dervin. 1992. From the mind's eye of the user: The sense-making qualitative-quantitative methodology. *Qualitative research in information management* 9 (1992), 61–84.
[20] Karen E Fisher and Charles M Naumer. 2006. Information grounds: Theoretical basis and empirical findings on information flow in social settings. *New directions in human information behavior* (2006), 93–111.
[21] Jerome H Friedman. 2002. Stochastic gradient boosting. *Computational statistics & data analysis* 38, 4 (2002), 367–378.
[22] Tim Gorichanaz, Kiersten F Latham, and Elizabeth Wood. 2018. Lifeworld as "unit of analysis". *Journal of Documentation* (2018).
[23] Roma M Harris, C Nadine Wathen, and Jana M Fear. 2006. Searching for health information in rural Canada. Where do residents look for health information and what do they do when they find it. *Information Research* 12, 1 (2006), 12–1.
[24] Trevor Hastie, Saharon Rosset, Ji Zhu, and Hui Zou. 2009. Multi-class adaboost. *Statistics and its Interface* 2, 3 (2009), 349–360.
[25] Isto Huvila and Farhan Ahmad. 2018. Holistic information behavior and the perceived success of work in organizations. *Library & Information Science Research* 40, 1 (2018), 18–29.
[26] Peter Ingwersen. 1996. Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of documentation* (1996).
[27] Elihu Katz, Hadassah Haas, and Michael Gurevitch. 1973. On the use of the mass media for important things. *American sociological review* (1973), 164–181.
[28] Carol Collier Kuhlthau. 1991. Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science* 42, 5 (1991), 361–371.
[29] Carol Collier Kuhlthau. 1995. The process of learning from information. *School Libraries Worldwide* 1, 1 (1995), 1–12.
[30] Carol Collier Kuhlthau. 2004. *Seeking meaning: A process approach to library and information services*. Vol. 2. Libraries Unlimited Westport, CT.
[31] Sanna Kumpulainen and Kalervo Järvelin. 2010. Information interaction in molecular medicine: integrated use of multiple channels. In *Proceedings of the third symposium on Information interaction in context*. 95–104.
[32] Andy Liaw, Matthew Wiener, et al. 2002. Classification and regression by randomForest. *R news* 2, 3 (2002), 18–22.
[33] Yang Lin, Charles Cole, and Kimiz Dalkir. 2014. The relationship between perceived value and information source use during KM strategic decision-making: A study of 17 Chinese business managers. *Information Processing and Management* 50, 1 (2014), 156–174.
[34] Robert M. Losee. 1997. A discipline independent definition of information. *Journal of the American Society for Information Science* 48, 3 (1997), 254–269.
[35] Matt Lovett, Saleh Bajaba, Myra Lovett, and Marcia J Simmering. 2018. Data quality from crowdsourced surveys: A mixed method inquiry into perceptions of Amazon's Mechanical Turk Masters. *Applied Psychology* 67, 2 (2018), 339–366.
[36] Andrew D Madden. 2004. Evolution and information. *Journal of Documentation* 60, 1 (2004), 9–23.
[37] Gary Marchionini. 2010. Information concepts: From books to cyberspace identities. *Synthesis Lectures on Information Concepts, Retrieval, and Services* 2, 1 (2010), 1–105.
[38] Maureen McCreadie and Ronald E Rice. 1999. Trends in analyzing access to information. Part I: cross-disciplinary conceptualizations of access. *Information processing & management* 35, 1 (1999), 45–76.
[39] Meredith Ringel Morris, Jaime Teevan, and Katrina Panovich. 2010. A Comparison of Information Seeking Using Search Engines and Social Networks. *ICWSM* 10 (2010), 23–26.
[40] Sanghee Oh. 2012. The characteristics and motivations of health answerers for sharing information, knowledge, and experiences in online environments. *Journal of the Association for Information Science and Technology* 63, 3 (2012), 543–557.
[41] Andrea Passerini, Massimiliano Pontil, and Paolo Frasconi. 2004. New results on error correcting output codes of kernel machines. *IEEE transactions on neural networks* 15, 1 (2004), 45–54.
[42] Allan D Pratt. 1977. The information of the image. *Libri* 27, 1 (1977), 204–220.
[43] Brent D Ruben. 1972. General system theory: An approach to human communication. *Approaches to human communication* (1972), 120–144.
[44] Diane Liang Rulke, Srilata Zaheer, and Marc H Anderson. 2000. Sources of managers' knowledge of organizational capabilities. *Organizational behavior and human decision processes* 82, 1 (2000), 134–149.
[45] Miamaria Saastamoinen and Kalervo Järvelin. 2017. Search task features in work tasks of varying types and complexity. *Journal of the Association for Information Science and Technology* 68, 5 (2017), 1111–1123.
[46] Miamaria Saastamoinen and Sanna Kumpulainen. 2014. Expected and materialised information source use by municipal officials: intertwining with task complexity. (2014).
[47] S Rasoul Safavian and David Landgrebe. 1991. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics* 21, 3 (1991), 660–674.
[48] Shawon Sarka, Yiwei Wang, and Chirag Shah. 2017. Investigating relations of information seeking outcomes to the selection and use of information sources. *Proceedings of the Association for Information Science and Technology* 54, 1 (2017), 347–356.
[49] Shawon Sarkar, Matthew Mitsui, Jiqun Liu, and Chirag Shah. 2020. Implicit information need as explicit problems, help, and behavioral signals. *Information Processing & Management* 57, 2 (2020), 102069.
[50] Reijo Savolainen. 2008. Source preferences in the context of seeking problem-specific information. *Information Processing & Management* 44, 1 (2008), 274–293.
[51] Chirag Shah and Vanessa Kitzie. 2012. Social Q&A and virtual reference - comparing apples and oranges with the help of experts and users. *Journal of the American Society for Information Science and Technology* 63, 10 (2012), 2020–2036.
[52] Chirag Shah, Vanessa Kitzie, and Erik Choi. 2014. Modalities, motivations, and materials–investigating traditional and social online Q&A services. *Journal of Information Science* 40, 5 (2014), 669–687.
[53] Chirag Shah and Jennifer Sonne. 2015. Seeking Information in Online Environments - Where, Who, and Why? *iConference 2015 Proceedings* (2015).
[54] Chirag Shah and Ryen W White. 2020. Tutorial on Task-Based Search and Assistance. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2436–2439.

[55] Chirag Shah and Ryen W White. 2021. Task Intelligence for Search and Recommendation. *Synthesis Lectures onSynthesis Lectures on Information Concepts, Retrieval, and Services* 13, 3 (2021), 1–160.
[56] Claude E Shannon and Warren Weaver. 1949. *The mathematical theory of information*. University of Illinois press.
[57] Aakanksha Sharaff and Harshil Gupta. 2019. Extra-tree classifier with metaheuristics approach for email classification. In *Advances in Computer Communication and Computational Sciences*. Springer, 189–197.
[58] Pertti Vakkari. 2000. Relevance and contributing information types of searched documents in task performance. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*. 2–9.
[59] Pertti Vakkari and Nanna Hakala. 2000. Changes in relevance criteria and problem stages in task performance. *Journal of documentation* (2000).
[60] Yiwei Wang, Shawon Sarkar, and Chirag Shah. 2017. Investigating Information Seekers' Selection of Interpersonal and Impersonal Sources. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. ACM, 353–356.
[61] Yiwei Wang, Shawon Sarkar, and Chirag Shah. 2018. Juggling with information sources, task type, and information quality. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*. 82–91.
[62] Thomas D. Wilson. 2010. Human Information Behavior. In *Informing Science*, Vol. 3. 49–56.
[63] Wan-Ching Wu, Diane Kelly, and Avneesh Sud. 2014. Using Information Scent and Need for Cognition to Understand Online Search Behavior. In *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval*, Vol. 37. 557–566.
[64] Yunjie Xu, Cheng Yian Tan, and Li Yang. 2006. Who will you ask? An empirical study of interpersonal task information seeking. *Journal of the American Society for Information Science and Technology* 57, 12 (2006), 1666–1677.
[65] Yan Zhang. 2010. Contextualizing consumer health information searching: an analysis of questions in a social Q&A community. In *Proceedings of the 1st ACM International Health Informatics Symposium*. ACM, 210–219.