Searching in Professional Instant Messaging Applications: User Behaviour, Intent, and Pain-points

Ismail Sabei
Jazan University
Gizan, Saudi Arabia
The University of Queensland
Brisbane, Australia
i.sabei@uq.edu.au

Mahmoud Galal
Independent Statistician
Brisbane, Australia
ph.galal@gmail.com

Bevan Koopman
CSIRO
Brisbane, Australia
bevan.koopman@csiro.au

Guido Zuccon
The University of Queensland
Brisbane, Australia
g.zuccon@uq.edu.au

Abstract

This study provides a thorough investigation of search within instant messaging applications (IMA) used in workplace settings (e.g., Slack and MS Teams), investigating search intents, user behaviour, points of friction, and missing functionalities. While IMAs are extensively used to accelerate communication between workers, user search behaviour in IMAs is still poorly understood, and search functionalities currently appear primitive and possibly lacking features that support the specific nature of search in this context.

We designed a mix-methods analysis based on three core studies that help us unveil search within IMAs. First, we created an in-depth diary study to capture user interactions when searching in professional IMAs, involving 17 participants spanning diverse geographies and yielding a total of 298 diary entries. The study comprised a pre-study interview with participants, followed by a structured diary form that captured essential metrics along with qualitative insights. The diary study was followed by a post-study interview focused on identifying and understanding failures, struggles and missing functionality. Subsequently, the insights learnt from the thematic coding of post-study interviews were utilized to develop a large-scale survey involving 222 participants out of 400 recruited through the Prolific platform.

Our findings suggest that while users find search functionalities within IMAs useful, there remains significant scope for improvement. The study sheds light on common search intents driving users to explore their message histories and the types of content that fulfil these intents. We outline potential features and required enhancements to improve searching within professional IMAs.

CCS Concepts

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

Keywords

Search in Instant Messaging Applications; User Understanding

1 INTRODUCTION

Instant Messaging Applications (IMAs)[1] such as MS Teams and Slack, have become ubiquitous in digital communication, especially in professional/enterprise settings. Serving as complementary channels to traditional email, these platforms facilitate both one-to-one and many-to-many communication mechanisms. Slack is projected to reach 79 million active users by 2025, while MS Teams reported 300 million users in 2023 [36]. With the increasing volume of information exchanged and archived on these platforms, the necessity for effective search functionalities is evident. Despite advancements in search technology in other domains, IMAs have been slower to adapt, often retaining basic and unrefined search capabilities. Furthermore, there is a limited understanding of how people search on these platforms and whether it differs from other areas (such as email or web search).

This study aims to thoroughly investigate how the search functionality of professional IMAs can be enhanced. We seek to clarify and understand three major dimensions that impact domain-specific search engines: (1) search intents and their relation to corresponding chat objects, (2) user and search behaviour, and (3) user satisfaction, including issues, challenges, and potential improvements. Our specific research questions are:

RQ1: What are the primary **search intents** behind users' searches within chat conversations, and what types of information and chat objects (messages, threads, documents) are they seeking?

RQ2: What **strategies** do users employ to navigate and search through chat conversations, and how do they cope with unsuccessful search attempts?

---

[1]We use 'Chat Conversation Applications' and 'Instant Messaging Applications' interchangeably.

	# of Paricipants	Sampling Method	Retained/Filtered Participants
Diary Study	25	Snowballing	17
Post Interview	10	Random Sampling	8
Large Scale Survey	400	Prolific	222

Table 1: Overview of sampling methods and participant continuation across study phases.

RQ3: How **satisfied** are users with the current search functionality of chat applications, including specific features that affect satisfaction, common challenges encountered, and possible areas for improvement?

To address these questions, we conducted a mix-method investigation comprising a series of user studies including a 14-day diary study, focused interviews, and a large-scale survey as showed in Table1 . The diary study and focused interviews revealed the issues and challenges faced by participants, as well as collected suggestions for potential improvements. The large-scale survey enhanced the accuracy of our findings, drawing on themes identified from the interview content to validate their results. The study materials, including the interview, diary, and large-scale survey questions, are publicly available at https://github.com/ielab/ima-user-study.git.

Our results reveal various user behaviours related to search in IMAs, indicating that users often struggle to search effectively. We outline the different factors that impede users and, based on the analysis of behaviours and challenges, propose a set of recommendations for improving search within chat applications.

2 RELATED WORK

Investigations like ours have taken place for search tasks in applications similar to IMAs. For instance, email search behaviour and email ranking have been extensively characterized and studied [1, 8]; likewise for conversational search [2, 39, 43, 44], where users iteratively and interactively retrieve information. However searching in IMAs is currently understudied [34], and it is unclear if findings from similar tasks apply to this task.

Our work specifically focuses on investigating IMAs within professional and enterprise settings. IR systems are widely utilized by professionals across various domains [33, 38]. These systems serve two primary purposes: gathering evidence from unstructured data, such as medical and legal documents, and retrieving archived correspondence, e.g., within a corpus of organizational emails.

This section provides an account of research across two related search tasks: (1) research focused on searching within conversational archives, such as email search, and (2) studies aimed at understanding professional searches.

Email is the most common communication method for both individual and organizations [18], and search is a key capability to enable access to information contained in large email archives [17]. Previous research has aimed to better understand user behaviour in email search. Ai et al. [1] adopted a mix-method approach to analyse email search behaviour consisting of a large-scale log analysis and a user survey. They found that the nature of email search is often that of a known item retrieval task, characterised by short queries and only a few search results considered. Our study, which also uses a mixed-method approach, reveals a broader set of intents underlying search in IMAs, with similarly short queries but much more conversations browsing.

While in other domains like web search, there are highly effective ranking methods to order content by relevance, chronological

ordering is still the main ranking method offered by email search functionalities. This ranking method has been reported to make it difficult and frustrating to search older messages [8]. Carmel et al. [8] challenged the prevalent chronological ranking approach, discussing its major limitations and proposing alternative, more effective ranking strategies based on relevancy and rich content features. We note that currently chronological ordering is the key ranking method offered in IMAs search functions. These lessons from email search underpin the motivation for our comprehensive user study to answer similar questions in IMAs search.

Professional search entails the need to locate unstructured information for performing professional services; both descriptive and behavioral definitions of professional search have been reported and examined in the literature [22, 27, 28, 38, 41]. Professional search often takes place within domain specific settings, e.g. in legal and healthcare domains. Numerous studies have focused on how search takes place in these domains. For example, Russell-Rose et al. found that lawyers typically search legal storage systems using keywords combined with logical operators to construct complex boolean queries; while others have reported that also natural language queries are nowadays used in this context [29, 30]. Similar querying behaviour are found also across healthcare settings: for example complex boolean queries are often used for systematic review creation, while natural language queries are typically used for searching through health records [21]. Similarly, studies have reported that users in these contexts often struggle with selecting appropriate search keywords and recall features associated with the resources to be searched [42].

Chat applications have become repositories of enterprise and professional knowledge. Searching through chat conversations however differs significantly from traditional professional search tasks. The unique challenges include contextual complexity, as chat conversations involve dynamic and contextually rich interactions; unit of retrieval, due to the fragmented and informal nature of chat data; and temporal aspects, since chats are time-sensitive and the relevance of information can change rapidly.

3 METHODOLOGY

To gain a comprehensive understanding of how users navigate IMA conversations, we developed a methodology encompassing a sequence of user studies: a diary study, focused interviews, and a large-scale survey. Each study is designed to build upon the observations made in the previous, creating a cohesive and cumulative research process, depicted in Figure 1. Ethics was granted by The University of Queensland. A detailed explanation of each component of our methodology is provided next.

3.1 Diary Study

Recall bias has been identified as a significant issue in in-lab based methods for data collection such as interviews, surveys, and focus groups [20]. To overcome this, methods of data collection have been designed to capture authentic user-system interactions within natural settings, minimizing observer bias and providing more genuine insights into user behavior [4, 26]. These methods include Diary Studies [4, 9, 16, 19, 25, 26, 35] and the Experience Sampling Method (ESM) [13–15, 40]. Both methods use questionnaires or pre-built surveys to gather data from participants, but there is a
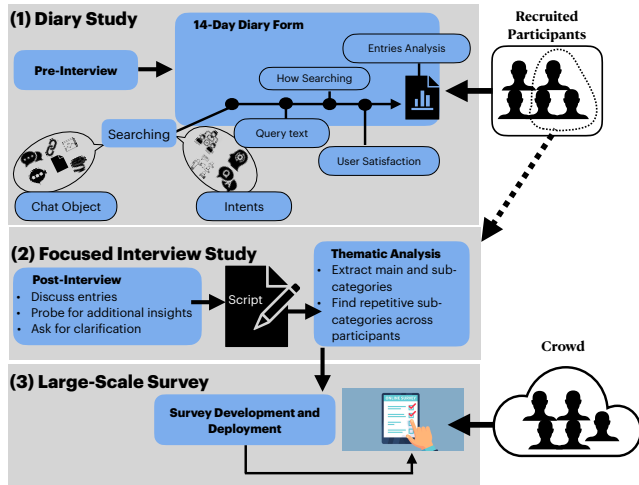
Figure 1: Overview of our methodology, which comprises of a diary study (1), a focused interviews study (2), and a large scale survey (3).

notable difference in how and when data is captured. ESM relies on notifications to prompt participants when data is to be collected, whereas Diary Studies are more appropriate when participants need to self-report events or entries in situ [32].

Due to these considerations, we have designed our user study to rely primarily on the diary study method, where participants record their diary entries whenever they search their chat conversations. However, diary study methods have limitations. Notably, participant engagement may wane over time, leading to inconsistent or incomplete diary entries [4]. To address this challenge, we designed a user-friendly, bespoke diary entry form aimed at capturing essential data with minimal cognitive load. Over a 14-day period, participants were prompted to respond to a series of succinct, multiple-choice questions following each chat search event. Additionally, at the end of each day, they were asked to review their professional messaging application's search logs to extract and document their search queries in the diary form. This structured approach was intended to reduce the daily cognitive burden and sustain participant engagement throughout the study.

As depicted in Figure 1, our diary study unfolds in three distinct phases: a pre-study interview, the diary entry phase, and the quantitative analysis of the diary entries.

3.1.1 Participant Recruitment. Our recruitment focused specifically on professional environments where chat applications serve as primary communication tools alongside email. We limited our choice to users of Slack and MS Teams, leveraging the in-built search log features of these platforms to facilitate easy recording of daily search queries. This approach not only enhanced data reliability but also reduced the cognitive load for participants, thereby maintaining their ongoing engagement. Our recruitment strategy employed 'snowball sampling' [31], where respondents were encouraged to share a self-recruitment link with colleagues who might be willing to participate, thus expanding our volunteer pool. We recruited a total of 22 volunteers (14 male, 8 female) from a broad geographical spectrum, including Australia, Egypt, New Zealand, Qatar, Saudi Arabia, and the United Kingdom. Ages ranged from

25 to 55, and participants held educational qualifications from associate degrees to postgraduate qualifications. Their professional roles varied: engineers, academics, administrative officers, and data analysts/scientists. All participants were either native English speakers or operated in English-speaking environments, eliminating the need for translating our diary entry forms from English into other languages. The participants were informed that their participation was entirely voluntary, and they could withdraw at any time without any conditions. Five participants discontinued their participation, leaving 17 who completed the 14-day diary study. We provided compensation after the fnal post interview, or upon the end of the diary study, in the form of an electronic voucher.

3.1.2 Pre-Study Interview . In the pre-study interview, we collected demographic information and ensured that our participants had adequate skills to meet their information needs generally. We asked participants about their purpose in using chat applications (e.g., personal or work) and specifically inquired how often they revisit their archived conversations using the search functionality of these chat applications. This question helped us confirm that participants would provide reliable information and evaluation of the search functionality of the current chat application, as the study primarily aims to measure user satisfaction. Finally, we clarified the nature of our study by presenting a dry run example to ensure participants fully understood their role.

3.1.3 Diary Entry Form. The diary entry form comprises questions designed to delve into the intricacies of participants' search activities, strategies, outcomes, and satisfaction levels. It probes the professional instant messaging applications used, search intent, precise queries employed, search methods adopted, the presentation of search results, and satisfaction with both the search function and the results obtained.

As the diary entry form is central to our data collection process, it includes a carefully designed set of questions aimed at capturing a comprehensive understanding of participants' search activities. These questions cover various aspects, such as the specific professional messaging applications used, search intent, exact queries entered, search methods employed, and the format of search results. Additionally, the form evaluates participants' satisfaction with both the search functionality and the results obtained. Specifically, we asked participants to record answers to the following questions:

(1) Can you share the query text you typed in your chat application Search Textbox? This question is Optional due to ethical considerations.
(2) What was your search intent when issuing this query to search your historical chat?
(3) What chat objects satisfied your search intents when you used the search function?
(4) How did you search your previous chat conversations?
(5) How were the results displayed when you performed this search?
(6) How satisfied were you with the quality of the search results?

3.1.4 Quantitative Entries Analysis. The diary form, a key component of our study, is primarily quantitative in nature. Consequently, the analysis of entries will be conducted quantitatively to explore IMA search intents and the content searched. A significant aspect

of this analysis involves mining entries to discern relationships between IMA content (e.g., threads, documents, etc.) and search intents, aiming to understand their interconnections. Moreover, we will assess user satisfaction, as participants are required to rate their search experience following each diary entry. A total of 298 diary entries were collected. Out of these, 162 entries included shared queries that participants used to search in their IMAs. The remaining entries did not share their queries with us, primarily due to privacy concerns.

3.2 Post Interviews

Three primary interviewing approaches are widely utilized in research: structured, semi-structured, and unstructured open-ended interviews [6]. Structured interviews consist of a predetermined list of questions that interviewees must answer in a specific order. Unstructured interviews are ideal for eliciting in-depth, open-ended responses that aim to gain a comprehensive understanding of the interviewee's perspective [3]. Given that our research aims to explore the search functionality of IMAs in terms of searchers' behaviors and to enhance current functionalities, neither structured nor unstructured interviews would fully serve our objectives. Therefore, semi-structured interviews were deemed the most appropriate method for this study, allowing for both prepared questions and the flexibility for participants to expand on their responses, reflecting on their diary entries—a key advantage of this approach.

3.2.1 Semi-Structured Interview. After closely reviewing our participants' diary entries, we contacted 10 participants who expressed interest in further elaborating on their experiences by participating in our post-study semi-structured interviews. Due to scheduling conflicts, however, two were unable to attend. Consequently, we conducted interviews with eight participants. They were provided with semi-structured interview questions that related to their recorded diary entries from the 14-day study period. Initially, participants responded to these questions briefly, which was followed by more focused interviews to gain deeper insights into their diary entries and responses. The primary aim of these interviews was to seek additional clarifications, identify challenges faced, and gather suggestions for improvements. The following questions were sent to our interview participants:

(1) Looking back at your entries, can you identify any recurring challenges you faced during the search activities? If yes, could you please describe them?

(2) Were there any instances where you were particularly satisfied or dissatisfied with the search results? Could you please elaborate on those instances?

(3) Did your search behavior or strategy change over the course of the diary study to improve your search results? If so, how and why?

(4) Did you discover any new features or functionalities of your chat application during this study that you were unaware of before?

(5) Are there any features you wished the chat application had to improve the search experience? You could look back at the diary questions as well as your entries.

3.2.2 Qualitative Thematic Analysis. The objective of our research is to develop a comprehensive understanding of the search process
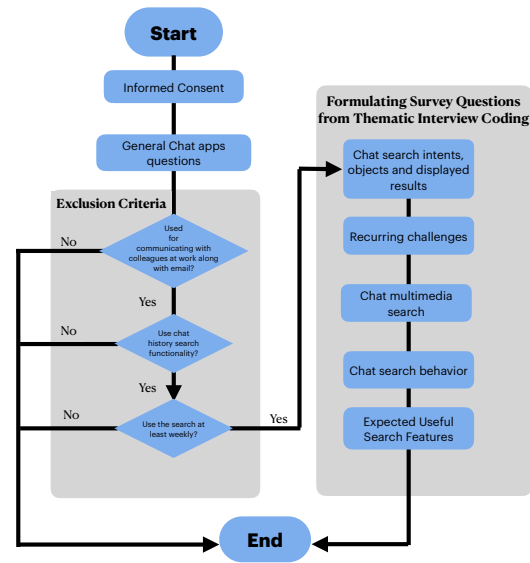

Figure 2: Flowchart detailing the survey structure.

within IMAs. To achieve this, we adopt an inductive qualitative analysis approach, specifically a thematic analysis [5], to identify and integrate themes from participants' narratives in focused interviews into pre-existing categories derived from semi-structured interview questions. These themes are reviewed and coded by the authors to ensure accuracy and consistency in thematic interpretation.

Our study employs a qualitative methodological design for several reasons. First, as the investigation of archived chat conversations is a relatively new area in Information Retrieval, our research aims to explore this field by collecting and analyzing data that generates a theoretical explanation through an inductive approach. This is in contrast to our deductive approach, which has been designed to evaluate our diary entries. Second, Brinkmann's characterization of qualitative research interviews as capturing "the world from the subjects' point of view" [7, p.3] supports our research questions. These questions seek to understand searchers' behaviors and evaluate the current search functionalities of IMAs, focusing particularly on the "HOW" aspect of these processes.

3.3 Large-Scale Survey

3.3.1 Participants. We recruited survey participants through Prolific[2] and offered compensation for their participation. As our questionnaire was in English, we restricted our participants to those from English-speaking countries. The main target group was IMA users. More strictly we only included those users who use IMA search functionality either daily or weekly at most. We explicitly explained this criteria before participants commenced the survey. For the sake of gathering accurate information, we implemented a criteria in the survey so that if a participants reported to not using IMA search functionalities daily or weekly, the survey would be terminated automatically. In addition, we excluded those participants who did not use IMA within professional work places.

3.3.2 Survey Questionnaire. Our survey was divided into two parts: exclusion criteria and thematic coding-dependent questions as depicted in Figure 2. The exclusion criteria section aimed to filter
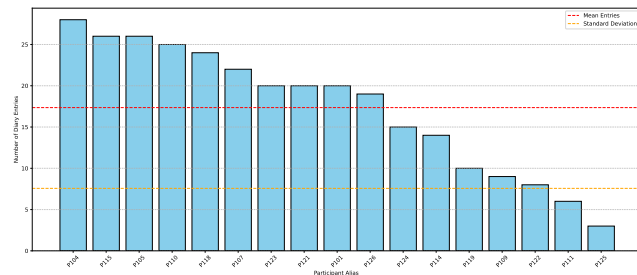
---

[2]https://www.prolific.com/

Figure 3: Number of diary entries per participant.

out participants who did not use IMAs for work, did not utilize the search functionality of IMAs, or used it infrequently. The thematic coding-based survey was designed to validate our primary data collected from the diary study and the focused interview. The survey, constructed using Qualtrics[3], enabled us to analyze our questionnaire results quantitatively.

4 RESULTS

4.1 Analysis

In theis section, we describe key basic information analysis derived from our user studies.

Table 2 summarizes the demographic and search functionality usage data, as well as response counts of our study participants. Participants were aged between 18-24 and 55-64 ($M = 35.9$), and of the 17 participants, 41% ($n = 7$) identified as female, with the remaining 59% ($n = 10$) male. Our participants held a range of occupations, including engineering, academia, information technology, and management. Noteworthy, more than a third of the participants have completed a postgraduate degree, and the remaining are graduates. Of our 298 valid diary entries, there was a mean of 17.1 entries per participant. Given the prevalence of numerous chat apps, we selected MS Teams and Slack as they have been popular in professional workplaces. In the pre-interview questionnaire, participants were asked how frequently they search their archived chat conversations. The majority of our participants reported having searched for chat conversations on a daily basis (Table 2). This initial finding underscores the importance of this work, highlighting its relevance and justifying a detailed investigation. The pre-interview also collected self-assessments of prior search knowledge, specifically their ability to fulfill their information needs. On a scale from 1 (poor ability) to 5 (excellent ability), the average rating was 4.12 as calculated in Table 2, suggesting a high level of self-reported search proficiency among participants.

After completing the diary study, which yielded 298 valid entries, 8 participants continued with the structured post-interview. The post-interview was structured around three categories: (1) recurring challenges, (2) search behaviors, and (3) suggested search features. Subsequently, we coded the post-interview contents, leading to 13 themes, listed in Table 3. We then distributed our large-scale survey, collecting data from 222 online participants.

4.2 RQ1 – Chat Search Intents and Objects

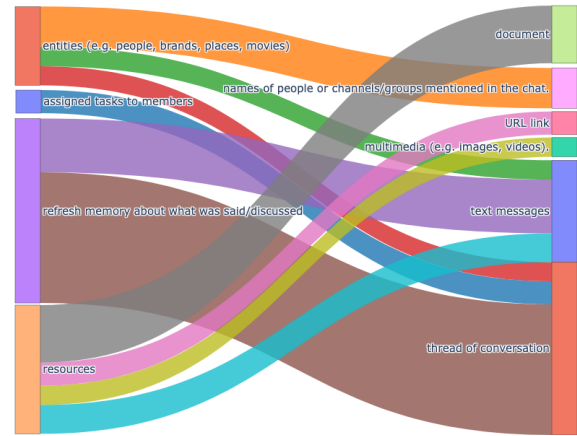In this section, we introduce qualitative and quantitative analyses to understand people's intents in searching IMA.

Figure 4: The interaction between search intents and chat objects as identified from the diary study data.

We quantitatively analyze the diary study entries by categorizing them according to chat search intents. This analysis is aimed at answering RQ1, which focuses on understanding why people search chat conversations. Results are reported in Table 4. We found that over a third (36%) of the diary entries were associated with the goal of recollecting information from past conversations (**D**) – this intent was reported by every respondent. Another popular entry was to find resources shared in chat (**E**).

In terms of commonly searched chat objects, reported in Table 5, approximately a third of the entries (34.01%) related to retrieving a conversation thread (**O1**), while the retrieval of a single message (**O2**) was the goal for approximately a fourth of the diary entries.

Figure 4 illustrates the interaction between search intents and chat objects identified from the diary entries. Notably, the discrepancy between Figure 4 and Tables 5, 4 was intentional since the figure was designed to focus on the most prominent intents-to-objects for clarity and readability, restricting the analysis to only those search intents with more than 10 entries, while the tables provide an exhaustive list, including additional intents/objects.

The strong interaction between the search for threads or individual messages and the goal of refreshing memory about a conversation is evident. This association is further validated by the results of a ranked ANOVA on our large-scale survey data, which indicated a statistically significant relationship between this intent and the retrieval of single messages (p-value = 0.001, Cohen's f = 0.34) and threads (p-value = 0.01, Cohen's f = 0.27). The moderate effect size of these relationships implies practical significance in the context of search behaviours. Nevertheless, the nuances of the relationship between chat objects and search intents warrant further exploration. These quantitative findings are supported by qualitative insights from one of our focused interview participants who emphasized the value of retrieving entire conversations for complete context, thus enhancing the understanding of user needs in IMA search functionalities (Table 13, *Theme: Conversation Retrieval, quote # 7*: "*I would love to retrieve the whole conversation, which gives me the full context.*"). We note that current chat tools often do not provide full conversations as the unit of retrieval in the search result page, with access to the conversation occurring only as a second step after clicking on an individual message.

Alias	Age	Gender	Occupation	Country	Chat App	# of Diary Entries	Search Functionality Usage	Self-assessment of Search Skills	Participated in Post-Interview?
P101	45-54	M	Engineer	AU	MS Teams	20	Daily	4	X
P104	25-34	M	Transportation Consultant	US	MS Teams	28	Daily	4	✓
P105	25-34	M	Logistic process assistant	SA	MS Teams	26	Multiple times per day	4	✓
P107	45-54	M	Mining Supervisor	AU	MS Teams	22	Daily	4	✓
P109	25-34	M	Project engineer	AU	MS Teams	9	Daily	5	X
P110	55-64	F	Senior business analysis	NZ	MS Teams	25	Daily	4	✓
P111	25-34	M	Procurement process specialist	SA	MS Teams	6	Daily	4	X
P114	25-34	M	Researcher	AU	Slack	14	Multiple times per day	4	X
P115	35-44	M	Senior officer	AU	MS Teams	26	Multiple times per day	5	X
P118	35-44	M	Principle engineer	Egypt	MS Teams	24	A few times a week	4	X
P119	35-44	F	Data analyst	AU	MS Teams	10	Multiple times per day	4	✓
P121	25-34	M	Lecturer	UK	Slack	20	A few times a month	5	✓
P122	35-44	F	Senior project manager	Egypt	Slack	8	Daily	4	X
P123	18-24	F	Research assistant	Qatar	Slack	20	Daily	4	✓
P124	35-44	F	Accessible communication advisor	AU	MS Teams	15	Multiple times per day	3	✓
P125	45-54	F	Researcher	AU	MS Teams	3	A few times a month	4	X
P126	25-34	F	Researcher	AU	Slack	19	A few times a month	4	X
Average	**35.9**					**17.1**		**4.2**	**Total 8**

Table 2: Demographic information of diary study participants.

Category	Theme	# of qoutes
Recurring Challenges	Manual Browsing(Scrolling up and down) as a Fallback/Long queries	2
	Keyword-Based Search Limitations	6
	Precision-based Search Limitation	2
	Content-Type Specific Search Challenges	2
	Order and Organization of Search Results	2
Search behaviours	Rigid Search Behaviour	3
	Persistent and Adaptive Query Refinement	5
Suggested Search Features	Semantic Search	1
	Labelling Messages/Conversations	1
	Advanced Search	3
	Search with Regular Expressions	1
	Multimodal Search	
	Conversation Retrieval	1

Table 3: Themes emerging from the focused interviews associated with challenges, search behaviours, and suggested search features.

	Search Intent	% of Diary Entries	# of Participants
A	Assigned tasks to members	8.75	11
B	Find entities (e.g., People, Brands, Places, Movies)	16.11	14
C	Question-answering	3.70	5
D	Refresh memory about what was said/discussed	36.24	17
E	Resources	28.28	15
F	Work / Social announcements	7.1	10

Table 4: Search intents and their prevalence in diary entries.

	Chat Object	% of Diary Entries	# of Participants
O1	Thread of conversation	34.01	15
O2	Single text messages	24.24	16
O3	Document	12.12	11
O4	Specific keywords or phrases mentioned in the chat	7.4	10
O5	Names of people or channels/groups mentioned in the chat	10.44	9
O6	Multimedia (e.g. images, videos)	4.71	7
O7	URL link	6.73	9

Table 5: Searched chat objects and prevalence in diary entries.

4.3 RQ2 – Search Behavior

Our research has examined the navigation strategies and search behaviors of IMA users. Generally, it was observed from our diary study that IMA's search functionalities impose constrained search methods, as shown in Table 6, where in the majority of cases keyword-based search or manual up and down scrolling were utilized. This consequently resulted in the behaviors reported in Table 7, which were qualitatively analyzed and coded from the focused post-interview.

4.3.1 Query Refinement. In previous studies focused on web search [11, 24, 37], strategies for query refinement have been extensively investigated. In this section, we report the query refinement behavior of IMA searchers based on qualitative analysis, validated by our large-scale survey. After asking participants **Q3** in Section 3.2.1 and analyzing the interview transcripts, we identified five quotes

Search Methods	% of Total Entries	# of Diary Entries	# of Participants
Keywords within a textbox	42.76	127	16
Keywords with operators (AND, OR, quotes, temporal ranges, user (from/to), has_attachment)	9.43	28	7
Manually (e.g., scrolling up and down)	42.42	126	17
Natural language phrases/questions	5.39	16	5

Table 6: Search methods utilization in the diary entries.

Theme	#	Quote
Rigid Search Behaviour	1	"My behaviour probably hasn't changed as it seems that using keyword search is the only available option plus scrolling up and down when there is a long list of results."
	2	"I was simply looking through my past search queries and my search strategy on the platform appears to be consistent over time (simple keywords, often filtered by channel)."
	3	"If I'm searching for words within chat, it is more difficult and tricky. It takes me back to the problem of the keyword thing that I'm searching with. I need to change it. I need to keep remembering what else to use until I get the most relevant results."
Persistent and Adaptive Query Refinement	4	"I did very much strategy at different times during the study depending on the results that I was looking for. For example, sometimes I use quotation if I know the exact phrase I am looking for."
	5	"Yes, starting to add '+' between words (like in Google!). Also after using this for a while, you get to understand the algorithms that produce the search results so you understand to input better search criteria. I don't know whether they have wildcard. I tried to do wildcard search and it didn't really work."
	6	"Searched as standard subject matter singular word but did understand the importance of correct wording to ensure search results satisfied intent."
	7	"I tried to keep a variety of searches. The outcome was pretty much the same. It is a game of changing words until one of them wins."
	8	"For one search, I did try multiple queries to find a match. Then, no success. So, I have to look by scrolling up and down."

Table 7: Search behaviours coded from interview data.

illustrating their struggle with refining queries using different techniques. We named this theme *Persistent and Adaptive Query Refinement*, as shown in Table 7. Examples of these techniques are mentioned in quotes #4 and #5, where participants used advanced operators (e.g., "+" between keywords and quotation marks) to improve search results.

4.3.2 Rigid Search Behavior. A recurrent theme identified in the interviews was the rigid search behavior users had to follow. This highlights the frustration expressed by all participants who regularly had to alternate between keyword-based searches and manual browsing due to the limitations of keyword-based search functionalities. This sentiment is captured by one participant's quote in Table 7 (quote #1): *"My behavior probably hasn't changed, as it seems that using keyword search is the only available option, plus scrolling up and down when failing."*. This pattern was validated by our survey, which showed a strong, significant relationship between the frequency of manually browsing the chat history due to search functionality failures and the challenge of manual scrolling, indicated by a Chi-Squared Test (p-value $< 1 \cdot 10^{-5}$, Cramer's V = 0.294). Other quotes in Table 7 confirm that IMA's search functionalities obstruct users from meeting their information needs.

4.3.3 Quantitative Analysis. To further investigate these behaviors, we quantitatively analyzed our survey to understand the causes.

	Query Refinement	Rigid Search
Exact Keyword Usage	✓	✓
Weak Precision in Search Results	✓	✗
Multimedia Search Difficulty	✗	✓
Advance Search	✓	✗
Search Failure	✗	✓

Table 8: Relationship between search behaviours and search usability aspects.

	Usability Statement	Disagree (%)	Agree (%)
US1	Since the search is keyword-based, I often need help to recall the exact keywords required to locate specific information.	26.2	73.8
US2	When I search for information, the search results give me various search results related to my topic. However, I still need to look through them to find exactly what I want.	11.0	89.0
US3	A lot of the time, the amount of information returned is too much to search through, and advanced search options are not available	41.0	59.0
US4	I often have a poor experience searching for multimedia content such as images or audio because I can't find what I'm looking for if I don't use the exact word I used when sharing the content.	45.1	54.9
US5	The optimal implementation would allow me to control when I wish to conduct a search based on the overall meaning and relevance of the conversation (contextual search) and when I prefer a straightforward, word-for-word search (exact keyword search).	13.9	86.1
US6	The search functionality of my chat application usually fails, leading me to browse my chat history manually.	88	12

Table 9: Agreement on IMA search feature usability.

We cross-tabulated questions on the need to refine searches and the agreement on the necessity of exact keywords (Table 9, Statement #US1). Table 8[4] revealed a strong, significant relationship between the necessity of using exact keywords and the frequency of search refinement. It is evident that the more users agree on the necessity of exact keywords, the more they tend to refine their searches. Our analysis also identified that the frequency of search refinement is strongly associated with the excessive amount of retrieved information (US2, Table 9), compelling users to refine their search behaviour. Based on this correlation, it is recommended that the anticipated themes in Table 13 be incorporated into the IMA search functionality to improve precision with minimal query refinement.

4.4 RQ3 –User Satisfaction

To measure user satisfaction during search activities within chat archives, our diary entry form asked participants to rate their experience on a seven-point scale immediately after each search activity was completed. The average satisfaction level reported by participants was moderate ($\bar{x} = 4.49$ out of 7) in terms of the quality of search results. To enable a more in-depth analysis, we aggregated user satisfaction rates by the chat objects they searched for, applying a reweighting formula to the participants' ratings:

$$W(O_i) = \frac{|O_i|}{|O_i| + m} \cdot R + \frac{m}{m + |O_i|} \cdot C \qquad (1)$$

where:

- O_i: The chat object for which we are recalculating the weighted rating mean.
- $|O_i|$: The number of ratings for O_i.
- m: The quantile threshold, representing the minimum number of ratings required for significance.
- R: The unweighted average rating for O_i.

---

[4]The significant relationships, shown in Table 8, were validated using a Chi-Squared Test (p-value $< 1 \cdot 10^{-5}$, Cramer's V > 0.5).

Chat Object	Reweighted Satisfaction Rate W(O_i)
URL link	4.46
Document	4.45
Multimedia (e.g., images, videos)	4.44
Names of people or channel groups	4.70
Specific keywords/phrases mentioned chat	4.51
Text messages	4.47
Thread of conversation	4.30

Table 10: Reweighted user satisfaction ratings across chat objects (7-point scale).

- C: The mean rating value across all chat objects.

This reweighting ensures that chat objects with only a small number of ratings do not skew the overall average. By combining the actual ratings R with the global mean C, weighted by the number of ratings $|O_i|$ and the threshold m, we achieve a more reliable and robust estimation of user satisfaction for each chat object.

The overall user satisfaction rates across chat objects are presented in Table 10. The results highlight potential for improvement across all types of searches. This suggests that refining search algorithms and enhancing user interfaces could lead to better user experiences and satisfaction.

4.4.1 Challenges with Current Search Functionality of Chat Apps. Related to user satisfaction was the categorization of challenges identified in the analysis of the semi-structured focused interviews and the subsequent theme-building process (see Table 3).

With this respect, the interviewees highlighted significant challenges that hindered their ability to retrieve relevant content from archived chat conversations during the diary study. Table 11 analyses 14 responses from the interviewees, coding them into five different themes. Quote #1: "*Documents/chats were not returned although I typed exact match queries. Then I had to [...] scrolling up and down. [...] it happens when my query is [...] 3 words or more.*" prompted us to perform a quantitative analysis of the query lengths collected from our diary study. The analysis of the 162 queries revealed an interesting trend: the average length of these queries was 1.7 words. This suggests that current IMA search functionalities may inadvertently lead users to adopt such concise search behaviour because of the issue mentioned in the quote.

The theme Keyword-Based Search Limitations emerged as a significant challenge across most participants, being the most frequently mentioned challenging theme (with six quotes all concerning keyword-based retrieval). For further validation, questions related to these challenges were included in our large-scale survey to be analyzed quantitatively. A Chi-Squared Test confirmed a statistically significant relationship between Keyword-Based Search Limitations and the refinement and retrying of search queries ($p = 1.06 \cdot 10^{-4}$). Another prominent statistically significant relationship ($p = 0$ and large Cohen's f effect size of 0.63) suggests difficulties when users search for multimedia content (theme Content-Type Specific Search Challenges in Table 11).

To further investigate our findings, we asked the large-scale survey participants to rate the challenging themes (1-5 where 1 is not challenging at all and 5 is extremely challenging), which were extracted from the interviews. Figure 5 compares the participants' responses, highlighting that all current search aspects were highly challenging to the participants. To identify which search challenges
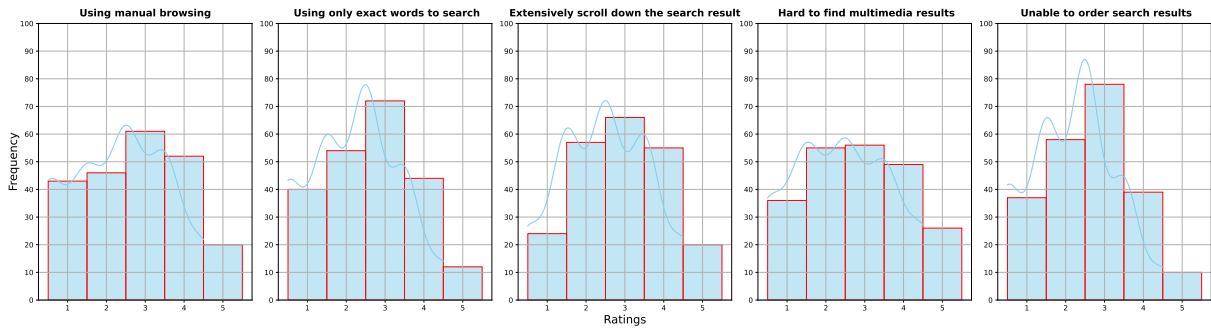
Figure 5: Analysis of current search functionality challenges for IMAs; ratings: 1 =Not Useful at all; 5 = Extremely Useful.

Theme	#	Quote
Manual Browsing as a Fallback/ Long queries	1	"Documents/chats were not returned although I type exact match queries. Then have to search manually scrolling up and down. I realized that it happens when my query is lengthy like 3 words or so."
	2	"Having to manually search results to find desired content, search results are condensed with varying search subject responses but still require review to locate desired result."
Keyword-Based Search Limitations	3	"I do not always remember the exact match message so I write something else close to what is in my mind but the search in slack does not retrieve what I am looking for."
	4	"Since the search appears to be keyword-based, I sometimes found it challenging to remember the right keywords to find what I was looking for."
	5	"Very rigid search func. that only depends on keywords used in chat, otherwise can't retrieve any results."
	6	"Exact keywords are required to retrieve my information needs. Otherwise, no results are returned."
	7	"When I forget the words we discussed in chat, I have only the option to use the same keyword to refresh the memory."
	8	"I think it's key to enter the correct information to start with. If you're unsure of what a document's called or a resource's called or if you misspell it, you could get lost in here. It's essentially getting the right search criteria. Because I hadn't done this a lot, it took me a bit to figure out the way it works. I spent a lot of time redoing my search because I didn't word it the way that it was going to pick it up. If you don't know 100% what the exact name of the document is, you cannot use synonyms. You can't search like sounds like or you know that kind of thing. It won't search like that."
Precision-based Search Limitation	9	"Trying to find specific information if you only have a select amount of data to search on. A lot of the time the amount of information returned was too much to search through and the search conditions had to be more succinct."
	10	"Sometimes the desired results still had to be searched to find what you were after; that might be because the wording wasn't correct for the initial search. It was a little bit more results."
Content-Type Specific Search Challenges	11	"Teams has the option to record but to search recordings for content is challenging due to having to listen to the entire content. If these recordings were auto-converted to text then text would be available to search and use as reference material."
	12	"I had a very poor experience searching for media/images. For instance, in the query 'raclette' I was looking for an image that I shared with my family. Since I didn't use the word raclette when sharing the image, I wasn't able to find the item I was looking for."
Order and Organization of Search Results	13	"The problem that I was facing was that in some cases what I really wanted was chronological, but the default is by relevance and the relevance ranking doesn't seem to work very well."
	14	"Sometimes the search results are random, they were not displayed in a timely order!"

Table 11: Thematic analysis of recurring challenges.

	Odds Ratio	p	[95% Conf. Interval]
Using manual browsing	1.005	0.960	0.821 - 1.2305
Using only exact words to search	0.842	0.029	0.721 - 0.982
Extensively scroll down the search result	1.217	0.005	1.061 - 1.396
Hard to find multimedia results	1.085	0.373	0.904 - 1.3029
Unable to order search results	0.805	0.004	0.821 - 1.2305

Table 12: Ordered logistic regression results.

most hinder the search process within chat conversations, we statistically analyzed the responses to produce a reliable significance analysis aimed at ranking the current search challenges.

To achieve this, we first eliminated intrapersonal variations by using the individual median response as an offset. By doing so, each participant was assigned a weight factor. We then utilized Ordered Logistic Regression to compute the Odds Ratio and p-value, as shown in Table 12. To conclude our analysis, 'Using only exact words to search' (Odds Ratio: 0.842, p-value: 0.029) and 'Extensively scroll down the search result' (Odds Ratio: 1.217, p-value: 0.005) are identified as significantly challenging. In other words, our findings indicate that the current search functionality

in chat conversations indeed suffers from low semantic matching capabilities and precision when trying to find relevant information.

4.4.2 Expectations from Future Search Functionality of Chat Apps. A significant portion of our qualitative analysis focused on identifying new features participants wished having when searching IMAs (Suggested Search Features in Table 3).

For example, one participant stated: "*I would love to have the option to run semantic searches in MS Teams*" (Theme: Semantic Search), emphasizing the necessity of semantic understanding to address the limitations of keyword-based searches. Notably, 3 out of 8 quotes support the theme *Advanced Search*, indicating a user demand for more sophisticated search capabilities, such as filtering by sender, channel, or specific time frames. Our large-scale survey further confirmed the importance of these proposed features. As illustrated in Figure 6, a majority of large-scale survey participants expressed a preference for implementing these features within the current IMA search functionality. Our quantitative analysis reinforced that "Advanced Search" is a highly desired feature for future implementation. The significance analysis, conducted using Ordered Logistic Regression, yielded a p-value of 0.002 and an Odds Ratio of 1.43. Additionally, our analysis reported that Context-based Search/Semantic Search and Conversation Retrieval also achieved relatively high Odds Ratios of 1.06 and 1.01, respectively.

To further investigate this, we cross-tabulated[5] what functionalities IMA users would prefer to have in future to help them overcome the current IMA challenges. The results of this analysis are reported in Table 14, showing a statistically significant relationship tested by a Chi-Squared Test (p-value = $3.51 \cdot 10^{-4}$ and Cramér's V effect size = 0.22). This is corroborated by Table 9 (Statement 1), in which 73.8% of our large-scale survey participants agreed that IMA keyword-based search is often a pain point. At the same time, from the same table, the majority accepted Statement 5 that suggested enhancing the current IMA search functionality by implementing semantic search along with keyword-based search.

5 STUDY LIMITATIONS

While significant effort was made to ensure the validity of our user studies and results, there are fundamental limitations inherent in our study approach, particularly the self-reported nature of diary studies. The information provided by end-users may contain

---

[5]For this, we considered the intersection of participants groups defined across two variables – e.g., the intersection between the respondents that found a feature useful with those respondents that found the challenge significant.
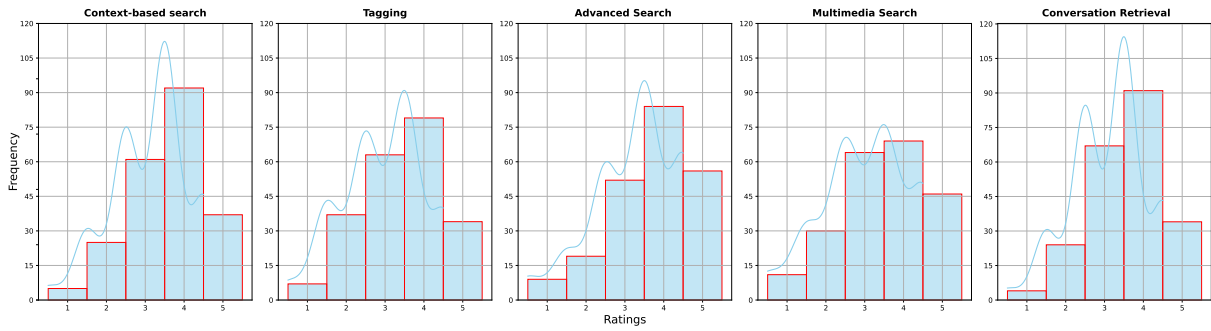
Figure 6: Analysis of future search functionalities for IMAs; ratings: 1 =Not Useful at all; 5 = Extremely Useful.

Theme	#	Quote
Semantic Search	1	"I would love to have the option to run semantic search in MS Teams. The best possible implementation of it would be for MS Teams to ask me or let me control when I want to do semantic search and when I want to do classic literal search."
Labelling Messages / Conversations	2	"Tagging messages or conversations by specific keywords."
Advanced Search	3	"Potential improvement is to have search option (i.e., drop down box) for search content results to be listed as selected rather than pre-decided (i.e., chronological, document only, chat only, etc.). Ability to refine searches to chosen content results."
	4	"The oldest and most recent results options. Search by the most relevant context."
	5	"I am old enough and have enough of every search background that, to my mind, any search should come within an advanced search option that users can directly control. Because I would really like to do an advanced search, you know, be able to set that chronological window and they only return search results from Y to X period."
Regular Expression Search	6	"Maybe a wildcard search option. There may be one (i.e., % for MS Teams) but when I tried to use '%,' it didn't really help."
Conversation Retrieval	7	"I would love to retrieve the whole conversation, which gives me the whole context."
Global Search	8	"Since I'm in over 20 Slack workspaces and there is substantial overlap between the people in each one, sometimes it was challenging to remember which ones I had DMs to specific people. A cross-workspace search might be helpful."

Table 13: Suggested search features.

	Not at all challenging	Slightly challenging	Moderately challenging	Very challenging	Extremely challenging
Not at all useful	2.6	0.0	1.4	6.8	0.0
Slightly useful	21.1	13.0	8.3	9.1	0.0
Moderately useful	31.6	25.9	34.7	20.5	8.3
Very useful	36.8	50.0	44.4	36.4	25.0
Extremely useful	7.9	11.1	11.1	27.3	66.7

Table 14: Cross-tabulation of the portion of participants respondents for the variables future useful feature "Context-based search" and current challenging search functionality "Searching with exact words".

inaccuracies due to the type of platform, whether mobile or desktop environments. Our analysis of query length revealed that the queries were generally short. The cause of this remains unclear; it could be related to search habits developed in web search [23], or mobile search behavior[10, 12] since participants primarily search their chat conversations on mobile devices. It also might be due to search functionalities not supporting verbose queries. For instance, one participant reported that Slack returns no results when a three-term query is used. However, we are still uncertain about the exact cause, as we do not have sufficient information.

Another potential limitation is related to our focused interviews' thematic analysis. The coding was performed by a single author and reviewed by another. Ideally, the coding process should involve multiple rounds and multiple coders to extract well-refined themes.

An important limitation of our diary study was the participant recruitment method, which is often a significant challenge in diary-based research. To address this, we opted for snowball sampling. While we were aware of the risk that participants from the same

organization might have similar behaviours, we initiated snowball sampling across diverse locations and countries as a mitigation strategy; to further enrich the diary data we also conducted post-study interviews. Although the sample size of the study may appear small, we collected in total 298 diary entries, 162 of which contained real queries that participants shared with us. This comprehensive dataset enabled us to draw meaningful conclusions and compensate for a smaller participant pool.

6 CONCLUSION AND FUTURE WORK

While widely used, there is a lack of research on search-related functionalities within instant messaging applications. We conducted a series of user studies to address the following questions: What search functionalities do people use? How satisfied are they with these functionalities? What resources do they look for? What challenges do they face?

Our mixed-methods approach helped compile both qualitative and quantitative insights related to these research questions. Participants reported using the IMA search functionality regularly (daily to weekly), but their satisfaction level was only 4.4 out of 7, highlighting the need for better search technology in this context. Key challenges identified include the limitations of exact match, keyword-based searching, poor performance on long queries and specific content types (e.g., images), and the lack of control over ranking (chronological vs. relevance).

Through this study, we aim to stimulate further research in this domain and to aid practitioners in improving the search functionality of instant messaging applications used by millions every day.

To conclude, two core recommendations have been derived from our extensive theoretical methodology. While most of the current search functionality of instant messaging applications (IMAs) have not kept pace with recent trends in representation learning, our comprehensive study strongly suggests the inclusion of semantic search technologies like dense vector retrieval within the IMA domain. Representation learning has proven to be an effective alternative to exact term matching. The second recommendation is to enhance current IMA search functionalities to allow users to retrieve entire conversations as a unit of retrieval, rather than just individual messages. This approach to the retrieval task has been experimentally investigated by Sabei et al. [34], demonstrating its significant performance in retrieving entire conversations.

References

[1] Qingyao Ai, Susan T Dumais, Nick Craswell, and Dan Liebling. 2017. Characterizing email search using large-scale behavioral logs and surveys. In *Proceedings of the 26th International Conference on World Wide Web*. 1511–1520.

[2] Mohammad Aliannejadi, Hamed Zamani, Fabio Crestani, and W Bruce Croft. 2019. Asking clarifying questions in open-domain information-seeking conversations. In *Proceedings of the 42nd international acm sigir conference on research and development in information retrieval*. 475–484.

[3] Rita SY Berry. 1999. Collecting data by in-depth interviewing. (1999).

[4] Niall Bolger, Angelina Davis, and Eshkol Rafaeli. 2003. Diary methods: Capturing life as it is lived. *Annual review of psychology* 54, 1 (2003), 579–616.

[5] Virginia Braun and Victoria Clarke. 2022. Conceptual and design thinking for thematic analysis. *Qualitative Psychology* 9, 1 (2022), 3.

[6] Bonnie S Brennen. 2021. *Qualitative research methods for media studies*. routledge.

[7] SK Brinkmann. 2015. S.. Interviews: Learning the craft of qualitative research interviewing Thousand Oaks.

[8] David Carmel, Guy Halawi, Liane Lewin-Eytan, Yoelle Maarek, and Ariel Raviv. 2015. Rank by time or by relevance? Revisiting email search. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. 283–292.

[9] Scott Carter and Jennifer Mankoff. 2005. When participants do the capturing: the role of media in diary studies. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 899–908.

[10] Karen Church, Antony Cousin, and Nuria Oliver. 2012. I Wanted to settle a bet! - Understanding why and how people use mobile search in social settings. *MobileHCI'12 - Proceedings of the 14th International Conference on Human Computer Interaction with Mobile Devices and Services* December 2011 (2012), 393–402. https://doi.org/10.1145/2371574.2371635

[11] Karen Church, Barry Smyth, Keith Bradley, and Paul Cotter. 2008. A large scale study of European mobile search behaviour. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*. 13–22.

[12] Karen Church, Barry Smyth, Paul Cotter, and Keith Bradley. 2007. Mobile information access: A study of emerging search behavior on the mobile Internet. *ACM Transactions on the Web* 1, 1 (2007). https://doi.org/10.1145/1232722.1232726

[13] Sunny Consolvo and Miriam Walker. 2003. Using the experience sampling method to evaluate ubicomp applications. *IEEE pervasive computing* 2, 2 (2003), 24–31.

[14] Mihaly Csikszentmihalyi and Reed Larson. 1987. Validity and reliability of the experience-sampling method. *The Journal of nervous and mental disease* 175, 9 (1987), 526–536.

[15] Mihaly Csikszentmihalyi, Reed Larson, et al. 2014. *Flow and the foundations of positive psychology*. Vol. 10. Springer.

[16] David Dearman, Melanie Kellar, and Khai N Truong. 2008. An examination of daily information needs and sharing opportunities. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. 679–688.

[17] Susan Dumais, Edward Cutrell, JJ Cadiz, Gavin Jancke, Raman Sarin, and Daniel C. Robbins. 2003. Stuff I've seen: a system for personal information retrieval and re-use. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval* (Toronto, Canada) *(SIGIR '03)*. Association for Computing Machinery, New York, NY, USA, 72–79. https://doi.org/10.1145/860435.860451

[18] Jacek Gwizdka. 2002. Reinventing the inbox: supporting the management of pending tasks in email. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems*. 550–551.

[19] Jonna Häkkilä, Farnaz Vahabpour, Ashley Colley, Jani Väyrynen, and Timo Koskela. 2015. Design probes study on user perceptions of a smart glasses concept. In *Proceedings of the 14th international conference on mobile and ubiquitous multimedia*. 223–233.

[20] Gabriella M Harari, Nicholas D Lane, Rui Wang, Benjamin S Crosier, Andrew T Campbell, and Samuel D Gosling. 2016. Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science* 11, 6 (2016), 838–854.

[21] William R Hersh. 2009. *Information retrieval: a health and biomedical perspective*. Vol. 3. Springer.

[22] Terry Ann Jankowski. 2016. *Expert Searching in the Google Age*. Rowman & Littlefield.

[23] Bernard J Jansen and Amanda Spink. 2006. How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information processing & management* 42, 1 (2006), 248–263.

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

[25] Tero Jokela, Jarno Ojala, and Thomas Olsson. 2015. A diary study on combining multiple information devices in everyday activities and tasks. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 3903–3912.

[26] KarenChurchBarrySmyth. 2009. Understanding the Intent Behind Mobile Information Needs. (2009). http://searchengineland.com/071228-150455.php

[27] Youngho Kim, Jangwon Seo, and W Bruce Croft. 2011. Automatic boolean query suggestion for professional search. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. 825–834.

[28] Cornelis HA Koster, Nelleke HJ Oostdijk, Suzan Verberne, and Eva KL D'hondt. 2009. Challenges in professional search with phasar. (2009).

[29] Daniel Locke and Guido Zuccon. 2018. A test collection for evaluating legal case law search. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 1261–1264.

[30] Daniel Locke and Guido Zuccon. 2022. Case law retrieval: problems, methods, challenges and evaluations in the last 20 years. *arXiv preprint arXiv:2202.07209* (2022).

[31] Charlie Parker, Sam Scott, and Alistair Geddes. 2019. Snowball sampling. *SAGE research methods foundations* (2019).

[32] Bradley Rey, Bongshin Lee, Eun Kyoung Choe, and Pourang Irani. 2023. Investigating in-situ personal health data queries on smartwatches. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–19.

[33] Tony Russell-Rose, Jon Chamberlain, and Leif Azzopardi. 2018. Information retrieval in the workplace: A comparison of professional search practices. *Information Processing & Management* 54, 6 (2018), 1042–1057.

[34] Ismail Sabei, Ahmed Mourad, and Guido Zuccon. 2022. SCC-A Test Collection for Search in Chat Conversations. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 4429–4433.

[35] Timothy Sohn, Kevin A Li, William G Griswold, and James D Hollan. 2008. A diary study of mobile information needs. In *Proceedings of the sigchi conference on human factors in computing systems*. 433–442.

[36] Jared Spataro. 2020. Remote work trend report: meetings. https://www.microsoft.com/en-us/microsoft-365/blog/2020/04/09/remote-work-trend-report-meetings/?wt.mc_id=AID2409697_QSG_SCL_424041&ocid=AID2409697_QSG_SCL_424041

[37] Amanda Spink, Dietmar Wolfram, Major BJ Jansen, and Tefko Saracevic. 2001. Searching the web: The public and their queries. *Journal of the American society for information science and technology* 52, 3 (2001), 226–234.

[38] John I Tait. 2014. An introduction to professional search. In *Professional Search in the Modern World: COST Action IC1002 on Multilingual and Multifaceted Interactive Information Access*. Springer, 1–5.

[39] Johanne R Trippas, Damiano Spina, Lawrence Cavedon, Hideo Joho, and Mark Sanderson. 2018. Informing the design of spoken conversational search: Perspective paper. In *Proceedings of the 2018 conference on human information interaction & retrieval*. 32–41.

[40] Niels Van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The experience sampling method on mobile devices. *ACM Computing Surveys (CSUR)* 50, 6 (2017), 1–40.

[41] Suzan Verberne, Maya Sappelli, Diana Ransgaard Sørensen, and Wessel Kraaij. 2013. Personalization in professional academic search. (2013).

[42] Alice J Vollaro and Donald T Hawkins. 1986. End-User Searching in a Large Library Network: A Case Study of Patent Attorneys. *Online* 10, 4 (1986), 67–72.

[43] Alexandra Vtyurina, Denis Savenkov, Eugene Agichtein, and Charles L. A. Clarke. 2017. Exploring Conversational Search With Humans, Assistants, and Wizards. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI EA '17)*. Association for Computing Machinery, 2187–2193. https://doi.org/10.1145/3027063.3053175

[44] Hamed Zamani, Johanne R Trippas, Jeff Dalton, Filip Radlinski, et al. 2023. Conversational information seeking. *Foundations and Trends® in Information Retrieval* 17, 3-4 (2023), 244–456.