Understanding Questions that Arise When Working with Business Documents

by

Farnaz Jahanbakhsh

MSc., University of Illinois at Urbana-Champaign (2017)
BSc., Sharif University of Technology (2015)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2021

Author .....
Department of Electrical Engineering and Computer Science
January 27, 2021

Certified by .....
David R. Karger
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by .....
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Understanding Questions that Arise When Working with Business Documents

by

Farnaz Jahanbakhsh

Submitted to the Department of Electrical Engineering and Computer Science
on January 27, 2021, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

While digital assistants are increasingly used to help with various productivity tasks, less attention has been given to employing them in the domain of business documents. To build an agent that can handle users' information needs in this domain, we must first understand the types of assistance that users desire when working on their documents. In this work, we present results from two user studies that characterize the information needs and queries of authors, reviewers, and readers of business documents. In the first study, we used experience sampling to collect users' questions in-situ as they were working with their documents, and in the second, we built a human-in-the-loop document Q&A system which rendered assistance with a variety of users' questions. Our results have implications for the design of document assistants that complement AI with human intelligence including what types of human respondents are needed and the challenges around such systems.

Thesis Supervisor: David R. Karger
Title: Professor of Electrical Engineering and Computer Science

Acknowledgments

This research was conducted while I was an intern at Microsoft Research. I would like to thank my wonderful mentors Adam Fourney, Ryen White, Robert Sim, and Elnaz Nouri for their help and guidance in all the stages of this work.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Systems and software are increasingly incorporating intelligent digital assistants in order to help people be ever more productive even as systems, documents, and information spaces become more complex. For example, at home, people can rely on their voice assistants to manage processes such as shopping, cooking, and to quickly retrieve information from the web [33, 37, 18]. At work, digital assistants help with scheduling meetings, triaging emails, and managing task lists [8, 9, 34].

One workplace area of interest that has not seen significant progress is document-centered assistance. In this scenario, people engage with an intelligent digital assistant to consume and operate over written documents to perform complex tasks more quickly. Such assistance can also be useful in contexts that have been previously reported as challenging for example, when using a mobile phone [21].

To achieve this vision and build an agent that is prepared to handle a wide variety of requests from the user, we must first understand and characterize the types of assistance that people would want as they are working on their documents. While prior work has studied the types of queries that people would generate given one or a collection of public documents [10, 28, 39, 47], it is conceivable that these queries may not generalize to private or business documents that are in various stages of preparation. The information needs in a document-centric context could be especially different given that users may have prior interactions with a document and may not want to only read, but also review, or add information to the document. Therefore,

to understand users' actual needs, it is important that we involve authors, reviewers, and readers of such documents in the process of generating queries. In addition, for these questions to have ecological validity, we must collect them in-situ as users are working on their documents and their information needs arise.

In this work, we conducted two user studies to gain insight into the distribution of questions with which users need support in a document-centric scenario. Our focus in these studies was on document consumption rather than activity-centric assistance. In the first study, we collected questions via an experience-sampling method by having users install a Microsoft Word add-in which, at random points in time while a user was working on a document, would prompt them and ask about their current information need. In the second study, users submitted their questions via an add-in and received answers from a human-in-the-loop document Q&A system. Given a document and a question, an AI extracted a passage out of the document as a candidate answer [43]. A human worker would then decide between transmitting the candidate answer to the user, or composing and transmitting their own answer in cases where the AI-provided answer was wrong or insufficient.

The settings of both of these studies, the first simply asking about information needs and the second augmenting the AI with human intelligence, allowed users to be liberal in the questions they posed and not limited to the capabilities of current digital assistants, or the in-application search capabilities integrated in some productivity software (e.g., [5]). Therefore, these approaches allow us to investigate what capabilities to incorporate in document assistants prior to making deep investments in their development. We found for instance, that while factual questions, i.e., questions the answers to which could be extracted as a passage out of a document, are the focus of state-of-the-art Q&A datasets and models [35, 39], these types of questions are in fact not asked often in the domain of business documents. In addition, the study results give us an understanding that, for the foreseeable future, effective assistance will likely involve at least some human-in-the loop or hybrid intelligence—important classes of questions are best answered by document authors, collaborators, domain experts, and other types of human respondents. Even in these cases, an AI may assist

in question triage, and in routing questions to the appropriate parties.

The main contributions of this work are the characterization of the types of assistance that users desire when working on business documents and implications for the design of digital assistants that can provide such assistance.

Chapter 2

Related Work

We situate our work in the context of prior work on document question answering and automated and hybrid digital assistants.

2.1 Document Question Answering

Recent years have seen significant progress in research on finding answers to questions given one or a collection of documents. This body of work includes factoid Q&A, document understanding, summarization, and comparison [38, 22, 6, 49, 46, 15, 13, 14].

Most datasets for document question answering contain factoid questions generated by crowd workers or search queries on documents found on the web [10, 24, 28, 35, 48, 39, 47]. Models trained on these datasets therefore may fail to generalize to personal and business documents with which people have richer context and possibly prior interactions. Indeed, previous work in the context of email and Web search has shown that people's information needs are different depending on whether they are a co-owner of a document [3].

Closer to our scenario, Ter Hoeve et al. investigated the conversational assistance that people would want in a document consumption scenario by recruiting crowd workers to generate questions for a set of public documents retrieved from the Web. The workers were trained to imagine having some familiarity with the document

subject presented to them and to produce questions about a document based on its summary [43]. However, all documents were presented in the form of a final published draft, and participants often assumed the role of "reader", with only limited knowledge of document content or provenance, when forming their questions.

We extend the body of work on question answering to better understand people's information needs while working with business documents. To gain insight into the types of queries that people would ask about the business documents they read or work on, it is imperative that we involve the authors, reviewers, and readers of such documents in the process of generating these queries to capture their actual needs which may be unique to their role and context. It is also desirable for these questions to be captured in-situ as the users' needs arise. Therefore, in this work, we use two methods, experience sampling and a human-in-the-loop document Q&A system, to capture users' information needs while they are working on their documents.

2.2 Digital and Hybrid Assistance

Digital agents are increasingly used in areas such as entertainment, e-commerce, healthcare, and to help with various productivity tasks [18, 12, 36]. As systems, documents, and information spaces become more complex, in the workplace, such assistants help with scheduling meetings, triaging emails, managing task lists, and even performing data science tasks [8, 34, 9, 11].

Although these automated systems can accomplish many tasks, they still have limited capabilities in understanding what people want or in performing the requested tasks. To augment their abilities, a body of work has attempted to incorporate human intelligence into the workflow of such systems [25]. For instance, Lasecki et al. developed Chorus, a conversational agent that allows users to interact with a group of crowd workers as if they are a single conversational partner [31]. Commercial services such as Facebook M, Clara, and X.ai also employ a hybrid of machine and human knowledge to run errands for consumers [16, 19, 32].

One workplace area that is relatively under-explored is document-centered assis-

tance. In this scenario, people engage with an intelligent digital assistant to consume written documents and to perform complex tasks more efficiently. Closest to this scenario and to our work is Soylent, a word processor add-in that employs crowdworkers to help users edit Word documents [4]. Indeed, Soylent serves as an inspirational model for our work, but is limited in that it considers only a few specific editing scenarios, and largely overlooks opportunities to facilitate document consumption. In consumption scenarios, it is conceivable that different questions should be routed to different types of respondents, similar to social Q&A systems such as *IM-an-Expert* [41], *Aardvark* [17], and *Zephyr* [2] which routed questions to individuals with expertise or interest in the subject matter of the question.

More generally, document-centric assistance can also be useful in contexts that have been reported as challenging, e.g., when using a mobile phone [21]. Likewise, while AI advancements have made assistance with pre-defined tasks within documents such as PowerPoint theme suggestions or intelligent placeholders possible [1, 7], less is known about the types of assistance that people desire with their documents.

Our work aims to bridge this gap by examining the types of document-centric questions with which people need support. It is conceivable that for the foreseeable future, accommodating some requests in this domain requires some degree of human assistance. Understanding users' needs, including what types of human intelligence we should incorporate into such a system is the first step towards designing systems that can accommodate these needs.

Chapter 3

Research Questions

Our work investigates the following research questions:

- What types of document-related questions do people desire support with from a digital assistant?
- How are the distribution of questions different depending on whether the user is an author, reviewer, or reader of the document?
- What types of human respondents are needed to accommodate requests in a human-in-the-loop document Q&A assistant?

Chapter 4

Method

To investigate these research questions, we designed a two-phase study. In the first Phase, we collected document-oriented questions with which people reported needing support via an experience-sampling method. In the second Phase, we designed and developed a human-in-the-loop system that collected users' questions about their documents and provided responses to them. This system consisted of an AI component and human operators who supervised the AI and answered questions if the AI failed to produce a satisfactory answer.

The tool we built in Phase 2 was a prototype of the Q&A system that we envisioned and which motivated our study. The experience sampling Phase was a step along the path which informed our choices for the second Phase of the study. The technology probe enabled by the Q&A system allowed us to gain insight into the feasibility of generating answers in a document Q&A context, view the documents on which questions were asked, and obtain feedback about how useful such a tool would be and the quality of the answers. In addition, after Phase 1, there was still an open question on whether the nature of the questions would vary depending on whether participants expected to receive answers.

For the questions to have ecological validity, it was important that we collect them in-situ as users were authoring, reviewing, or reading a document. Therefore, we opted for using an add-in that would open on the side of a document, so that users could ask questions when they were working on a document and without having to

leave it. For each Phase of the study, we designed and developed an add-in that would work in Microsoft Word documents.

The scope of document assistance in both studies was document consumption rather than activity-centric assistance which we conveyed through the wording of the prompts as well as in the recruitment emails and consent forms where we outlined that the study purpose is to understand what kinds of questions users have about documents they visit. In both Phases, after consenting to the study, participants filled out a survey asking about demographics, their Word document consumption, and the readership of the Word documents they authored. Then they received instructions on how to install the add-in. Our study was approved by our institutional Review Board.

4.1 Phase 1 - Experience Sampling

At random points in time, the experience-sampler add-in would prompt the user asking what question they had about the document at that moment, in addition to a few follow-up questions to better understand the context (Figure 4-1).

To communicate the type of system we were envisioning, the first question was *"Imagine Word could connect with a skilled assistant that could answer your questions about a document. Is there any question about this document you would like to ask the assistant right now to help with your work?"*. To understand the perceived complexity of the question, the add-in then asked *"How long do you think it would take you to answer this question?"*. To gain insight into what kinds of questions can be routed to what kinds of respondents in a human-in-the-loop Q&A system, the add-in also asked *"Who do you think could answer this question?"* The answer choices to this question included: *"The author"*, *"A domain expert"*, *"A Microsoft Word expert"*, *"Someone familiar with the doc"*, *"Someone with enough time to read the doc"*, *"Other"* along with a free-form text for elaboration, and *"N/A"*. Because we expected the role of the user in the document to impact the questions that they ask, we collected information about their contribution to the document: *"What best describes your primary role in*

this document?" (Reader, Reviewer, Author/Co-author), *"Have you contributed to this document?"* (I have edited the document, I have commented on the document, I have not contributed to the document), *"How long ago did you contribute to this document?"* (In the past 24 hrs, In the past week, Longer ago, Never). In addition to capturing user responses, at the time of submitting the add-in questionnaire, the add-in collected contextual information about the document including the document's number of words, images, tables, lists, comments, timestamps (e.g., creation date, last modified date, date comments were posted, etc.), number of authors, filename, file size, URL, the location of the user's cursor when asking the question, and whether the user submitting the questionnaire was the creator, author, or the most recent contributor to the document. While we did not directly collect document content, we note that filenames often resembled document titles[1].

To minimize fatigue, we limited the number of prompts to 6 per day. Participants could answer the experience sampling questions whenever they wanted without waiting for the prompt. At the conclusion of the one-week period of the study, we had collected a total of 101 questions. Then we asked participants to open Word or their cloud documents homepage, select one or a few Word documents from their Recommended or Most Recently Opened lists of documents, and submit up to five questions for these documents as if they were working with the documents then. The completion of this step was voluntary and participants were not compensated additionally for submitting these questions. This step resulted in 38 more questions. We inspected these questions and found them to be extremely similar to those input earlier in the week, and thus we include them in our analyses. In total, questions from Phase 1 were asked about 103 distinct documents.

The add-in was deployed as a static app written in Vue.js which connected to a Node.js server. The server in turn interfaced with a SQL server and a Redis server.

[1]Microsoft Word uses the document title as the default filename when saving.

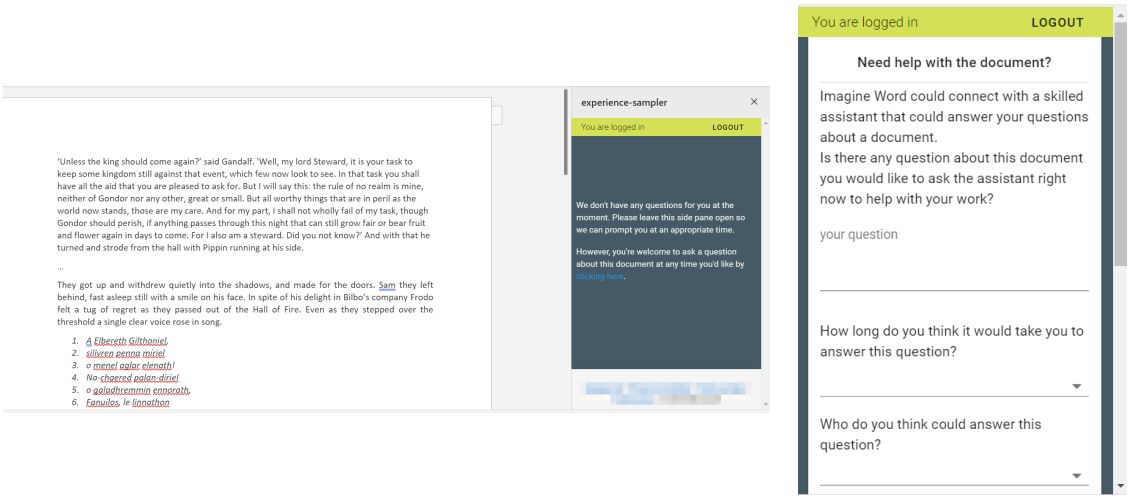

Figure 4-1: The figure on the left shows the Experience Sampler add-in opened in a side pane in a Word document. The screenshot is captured at a time before the user is prompted about their information need. The figure on the Right shows the questionnaire that the add-in presents to the user when the user is prompted.

4.2 Phase 2 - Human-in-the-loop Q&A System

We built a human-in-the-loop Q&A prototype for connecting users with knowledge workers who would supervise and complement an AI trained to answer document-centered questions [43]. The AI was a BERT Large model fine-tuned on the SQuAD2.0 [38] and DQA datasets [43]. Given a question and a document, the model would extract a passage out of the document if it detected an answer for the question. Participants submitted and received answers to their questions through a Microsoft Word add-in. With each question, participants submitted a share link to the document if they consented to the digital assistant or a human knowledge worker accessing the document to answer their question. We informed the participants that for the purpose of the study, the knowledge workers were limited to the researchers involved in the project, and noted that the researchers were employed at the same technology company as the participants, thus minimizing concerns about confidentiality—indeed the document share links required corporate credentials to access. If a user did not submit a share link to their document, we did not have access to the document's content. When submitting a question, the add-in would collect the same contextual information about the document as in Phase 1. For each document, the add-in would

display the user's previously asked questions in the form of expandable tiles. Because of the hybrid nature of the system, and the high level of human supervision, questions were not instantaneously answered. Participants were informed that their questions would be answered on a best effort basis and there may be a delay in the responses that they would receive from the system. Once a question had been answered or deemed unanswerable, the answer or the explanation for why the question could not be answered would appear below the question on the question tile. An icon on the question tile would signal whether the question was answered yet or marked as unanswerable and whether the user had seen the answer or the explanation yet. In addition, users would receive email notifications about received answers. Figure 4-2 displays different states of the Q&A add-in.

The knowledge workers' view contained all the questions submitted by users (see Figure 4-3). As the questions arrived, workers assigned tags to them and iterated over tags already assigned to prior questions. This approach helped with the organization of the questions and routing them to the different components of the system. In addition, the tags served as a basis for the taxonomy that we developed from the question pool. The workflow for answering questions was based on the question type. Workers used the AI model to answers questions related to the content. If the AI-provided answer was unsatisfactory upon an initial inspection (e.g., if the selected passage did not seem to answer the question), workers would answer the question by reading through its corresponding document. Workers answered the questions about metadata either by investigating the metadata captured by the add-in at the time of question submission or by accessing the document through the share link if the requested metadata was not among the contextual information captured by the add-in. Questions about style were similarly answered by accessing the document. Some questions sought external information that was available on public resources. Workers answered these questions by retrieving relevant information using a search engine. If a question requested external information available only within the company, workers used the internal repository of documents shared with employees to find the requested content. The median response time in this Phase was 17 minutes, with an average of

1.5 hours. Even outside work hours, we tried to keep answering questions, but those questions at times experienced longer delays.

At the conclusion of the two week period of the study, users had submitted a total of 133 questions asked about 61 distinct documents. We then distributed an end-of-study survey to participants asking about their experience with using the system. The survey questions asked for examples of good and poor answers that they received from the system. For each example, the survey asked about the type of question they had asked and the delay in receiving the response. The survey also asked about the overall satisfaction with the system (5 point Likert), how the system could be improved, and if they would recommend using the system to colleagues (5 point Likert). A total of 31 participants from this phase completed the questionnaire.

Similar to the add-in in phase 1, this add-in was deployed as a static app written in Vue.js. The worker view was also a separate Vue.js static app. Both clients connected to an OAuth provider to retrieve authentication tokens which they then used to send requests to the backend, a Node.js server interfacing with a SQL server. The add-in also opened and maintained a WebSocket connection to the server, so that it would be notified upon receiving answers to user's already asked questions.

4.3 Taxonomy Development

To develop a taxonomy of the types of questions with which people need support when working on business documents, we combined the questions we collected from both phases of the study. The questions from Phase 2 already had preliminary labels assigned to them as the workers had organized the questions to decide the course of action for answering them. We divided those submissions from Phase 1 that contained more than one question into idea units. With this division, the total number of questions from Phase 1 and Phase 2 was 272. A member of the research team then used open-coding to inductively develop codes that encompassed thematically related idea units. The categories assigned to each question were not a property of only the question but also of the document about which the question had been submitted.

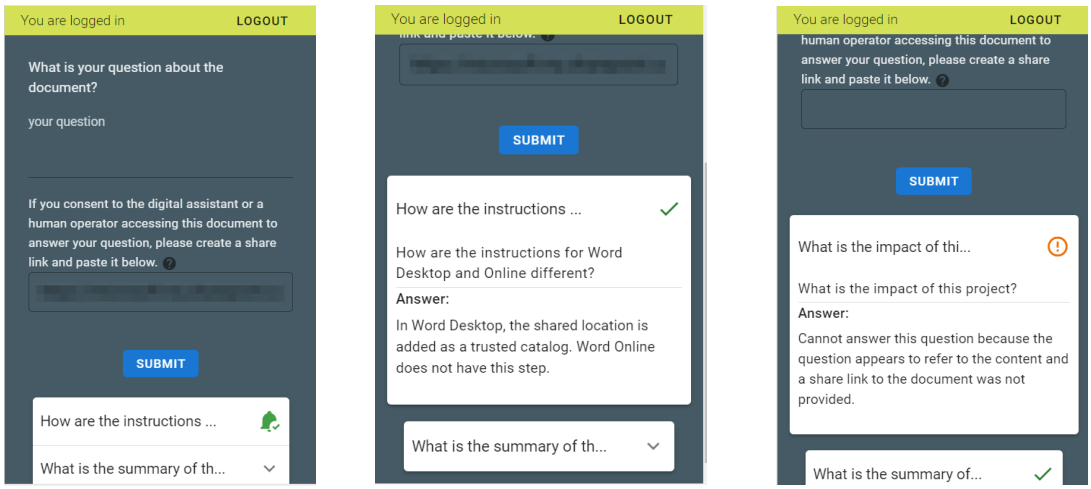

Figure 4-2: The Q&A add-in from Phase 2. The image on the left shows some questions that the user has asked about the document on the collapsed tiles. The green notification icon indicates that the question has been answered but the user has not yet viewed the answer. The image in the center displays the full question and the answer. Once an answer is viewed, the notification icon is changed to a checkmark. The image on the right shows another question from another document which could not be answered. The status is shown with a warning icon.

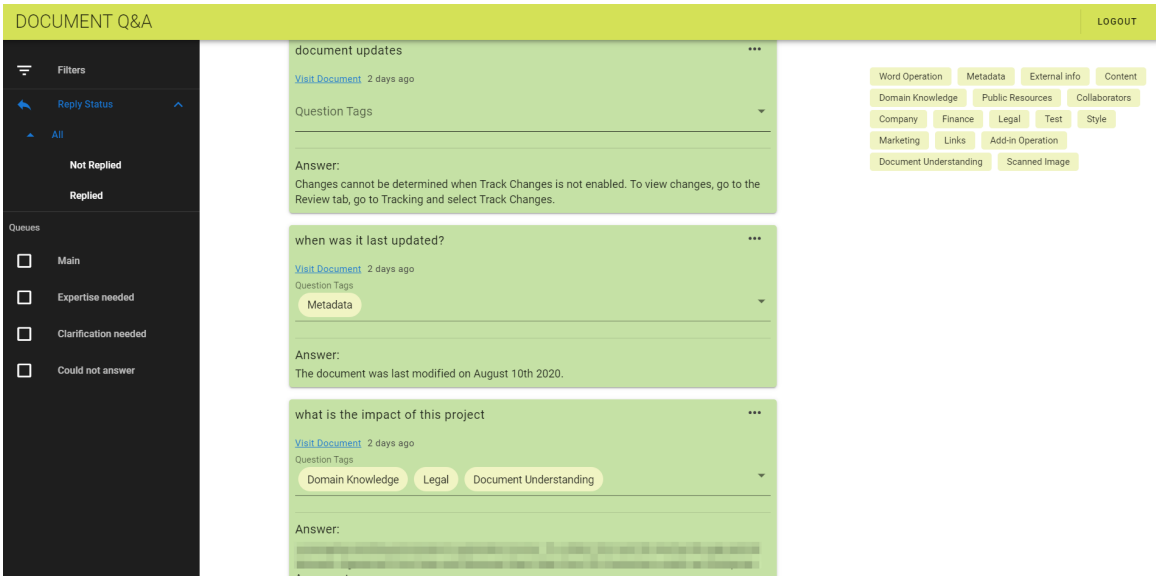

Figure 4-3: The document Q&A worker view. Each question that was submitted via the Q&A addin would be visible in this system and workers could visit the document to answer the question. Additionally, workers could move the questions to different queues (filters seen on the left) or assign tags to each question (seen on the right).

For instance, the question *"What is client sdk?"* can be a content-related question or one seeking external information depending on whether "client sdk" is defined in

the document or simply mentioned in the document in some context.

Through subsequent passes, the labels with too much overlap were consolidated and the ones showing distinct ideas were further split into separate categories. Another member of the research team was then trained on the categories and used them to label a randomly sampled set of 80 idea units (29% of the total). Cohen's Kappa for the high-level categories was 0.86. To determine the inter-rater reliability for the subcategories, we assigned each category that did not have nested subcategories a subcategory with a value equal to the value of the category. Cohen's Kappa for the subcategories was 0.75. Both Kappa values exceeded the recommended threshold for accepting the results [29].

4.4 Participants

We recruited participants in both Phases by randomly sampling email addresses from and sending invitations to employees of a large technology company. A total of 59 and 41 users participated in the Phase 1 and Phase 2 studies respectively. Across all study participants, 37% were female. The medians of reported age and highest education achieved were 35-44 and Bachelor's degree. Participants came from a diverse set of roles at the company including software and hardware engineering, sales, content writing, program management, finance, communications, etc. The medians for how frequently the participants read or edited Word documents were both a few times per week. Participants copy-edited Word documents less often (a few times per month).

We compensated participants from the Phase 1 study with a base amount of $20 in the form of an e-gift card. To encourage more involved participation, for each question that a participant submitted in each day, they would secure an entry into a raffle for 5 gift cards, each with a value of $50. We limited the number of entries per day for each participant to 25. The compensation for Phase 2 of the study was $20 e-gift cards. Although the duration of Phase 2 was longer than Phase 1, receiving answers to questions about documents was another form of value that participants received from the study.

Chapter 5

Results

5.1 Taxonomy of Question Types

We categorized the questions of our study based on what type of information they sought and where that information was available. Table 5.1 shows the full taxonomy as well as examples for each category. Throughout the paper, where we present participants' questions, we identify them with a string of the form 'p-' + phase number + participant number.

Participants used different languages for articulating their information needs. Some queries were rather verbose—*"I pasted in a table from a PPT slide and the table dimensions are larger than the page which makes it cut off. Is there a way to have it auto-scale to the page, without having to manually resize the table?" (p-2-9)*, and others, short—*"app service blog" (p-2-26)*. Both types of queries can cause challenges for AI systems to interpret user intents [27].

Among the questions that we categorized as *Workflow & Operation Help*, some were posed in the format of help queries and others were commands or direct requests from the assistant. The help queries were either about general procedures—*"How to create a table?" (p-1-39)* or specific to the user's context—*"Why is there the warning "Upload blocked" as I already have a copy in my local storage?" (p-1-11)*. The command type submissions were more common in the Phase 1 study in which participants were free to imagine an assistant that could help with any of their document-related

tasks. In the Phase 2 study however, because the system afforded help with document consumption and not manipulation, participants rarely submitted commands. The majority of the commands that the users submitted requested features that were not implemented in Word—*"Find all the pink words and review for modifications" (p-1-8).*

A number of the questions that we categorized as seeking *External Information* were tied to the content of the document—*"Love for the tool to find market research data regarding the topic. Both in our own data and from known trusted research companies online." (p-1-52).* The reason for this categorization however, was that the information that they sought resided outside the document.

Table 5.1: Taxonomy of the types of questions with which users need support when working with business documents.

Category	**Definition**	**Subcategories**	**Examples**
Metadata (N=32)	Questions about one or more of the following: actions (e.g., edits or comments), time of actions, actors, and document's properties (e.g., word count)	—	*"Was this file's location moved in the last 6 months?" (p-2-6)* *"When was the last time this info was updated?" (p-2-25)*
Workflow & Operation Help (N=61)	Questions about how to or commands to perform a specific task in Microsoft Word or the Q&A add-in	**Help Query** (N=54)	*"How can I tell if an embedded Excel table is current and could a reviewer view the data source?" (p-1-42)*
		Command or Request (N=7)	*"Show me my unresolved comments." (p-1-47)*
Content (N=56)	Questions that could be answered from the content of the document	**Factual**: A question the answer to which can be retrieved as a passage from the document (N=19)	*"How much morale money did everyone get in June?" (p-2-32)*

Continued on next page

Table 5.1 – continued from previous page

Category	Definition	Subcategories	Examples
		Reasoning: A question answering which requires complex reasoning and/or use of external information (N=25)	*"Are there any action items in this document?" (p-2-14)* *"How many different databases are mentioned"? (p-1-4)*
		Overview: Refers to the document as a whole—including document type, topic, impact of the document, etc. (N=9)	*"What is the focus of this document?" (p-2-35)*
		Summary: A special case of overview questions, seeks the summary of the document or a section of the document (N=3)	*"Can you summarize the main points of each section of this document?" (p-1-13)*
Written Style (N=10)	Questions about the organization, syntax, or semantics of the text or the language of the document	—	*"Are there any statements in this document that could be unclear to the reader, or are not inclusive in nature?" (p-1-8)* *"Does this document contain unnecessarily wordy paragraphs?" (p-1-50)*
Visual Style (N=9)	Questions about the document's current formatting, layout, typography, etc.	—	*"Why are there weird spaces in my document?" (p-1-1)*

Continued on next page

Table 5.1 – continued from previous page

Category	Definition	Subcategories	Examples
External Information (N=72)	Questions seeking information that is external to the document. This information could be available:	**On public resources**, e.g., public documentations (N=21)	*"What is executive bias?" (p-2-32)* [Asked in the context of a document that mentioned executive bias but did not define it.]
		Within the company, e.g., usage data of specific apps (N=12)	*"Where can I find a template of a communication planning document?" (p-1-42)*
		To the author or collaborators of the document (N=22)	*"What is our target date to publish this document?" (p-1-37)*
		To experts (N=15)	*"What are some general topics that are covered when building a training agenda for a specific program?" (p-1-54)*
		Across applications (N=1)	*"Can word check a date and time against my outlook calendar for conflicts?" (p-1-18)*
		Other (N=1)	*"How old is my son?" (p-2-26)*
No Question (N=28)	Users specifying they did not have a question at the time of submitting the Experience Sampler or the Q&A form. These submissions were more common in the Phase 1 study where we asked users about their questions at random points in time.	—	*"I do not have any questions at this time." (p-1-10)*

Continued on next page

Table 5.1 – continued from previous page

Category	Definition	Subcategories	Examples
Unknown (N=4)	Questions that may be either content-related or seek external info but that we did not have access to the content of the document to tease them apart	—	*"What activities are needed?" (p-2-34)*

5.2 Difference in the Distribution of Questions by User's Role

To gain insight into whether the distribution of questions differs by the user's role in the document, we investigated the contextual information that the add-in had collected to understand whether the user was an author, reviewer, or a reader of the document about which they were asking a question. This information included whether the user was the creator of the document and whether the user had either edited or commented on the document. If the user was the creator, we assigned them the role *author*. In all the documents of our study, if a user was the creator, they had also edited the document. If the user had not created the document but had edited or commented on it, we assigned them the *reviewer* role. If the user had done neither, we labeled them a *reader*.

While these signals are reasonable heuristics for understanding users' role, they may not be exact. A user who has not created a document but has contributed significantly to it may in fact be a co-author rather than a reviewer. Although we directly asked about the user's primary role in the document in the Experience Sampler questionnaire of Phase 1, to keep the Q&A experience seamless, we did not ask a similar question in Phase 2. In addition, the perception of what amount of contribution makes one a co-author rather than a reviewer may vary across users. Therefore, for the whole dataset, we used the contextual signals collected via the add-in to infer

users' roles.

We then performed a Chi-squared test of independence on the contingency table of question categories and users' roles. We excluded queries of the type *No Question* or *Unknown* from the data. The test revealed that the distribution of questions is in fact not independent of the user's role ($\chi^2(10) = 37.78$, $p < 0.001$).

Figure 5-1 displays how the distributions of questions across categories vary by the user's role. Because the numbers of questions asked by authors, reviewers, and readers of documents are different ($N_{Author} = 118$, $N_{Reviewer} = 66$, $N_{Reader} = 56$), the bar for each question category and each role is normalized by the number of all questions asked by users with a similar role. The figure suggests that the questions that authors ask are more often concerned with performing specific operations in Word— *"How to save as pdf without showing the comments?" (p-1-11)* or finding information external to the document— *"How would one approach a point of view paper?" (p-1-7)*, both of which help with authoring tasks. Interestingly, readers also ask more questions about external information rather than the content of the document, although it is potentially the content of the document that gives rise to such questions— *"Who's the competition?" (p-1-41)*. Reviewers however, are more concerned with the document's content— *"...What action [is] needed from me?" (p-1-35)* or its metadata— *"Where is the last change?" (p-2-10)*, with both types of questions helping the reviewer find the information that is relevant to them. As expected, readers are not concerned with the visual style of the document as often as authors or reviewers, although the datapoints in this question category may not be enough for generalization.

5.3 Types of Human Respondents

Many of the submitted queries required a level of understanding, analysis, or need of a knowledge base that would be challenging for even the state-of-the-art ML systems. To answer these queries therefore, a document Q&A assistant would need to employ human intelligence. In our study, the type of human respondents that were needed to answer participants' questions included the author or document collaborators—

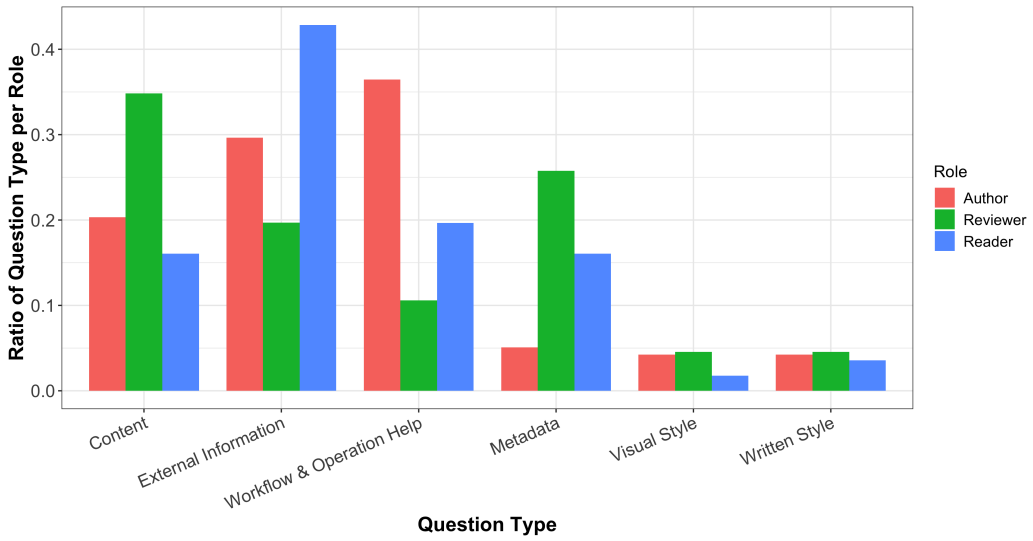

Figure 5-1: Distributions of question types by the role of the user in the document. Each bar shows the ratio of the question type relative to all questions asked by users with similar role.

"When is this slated for release?" (p-2-16), experts in various domains— *"Does [cloud service product] use my [OAuth provider] account?" (p-1-6)*, and human workers with a more general skill set— *"Can the assistant make recommendations to other ways to phrase my questions?" (p-1-12).*

The types of questions that needed human intelligence were not limited to those that asked for external information or an opinion, but also included those that referred to the content of a document understanding which required expertise, for instance, in the domain of finance, legal, marketing, or software engineering. The desire for connecting with colleagues familiar with the document or domain experts within the company explicitly surfaced in Phase 1, where participants had also indicated who could answer their question. We present a sampling of questions, and suggested human respondents below:

"What information is missing? Do the readers understand the content and what questions do they have?"—suggested respondent: "People who have written similar documents or documents on this topic" (p-1-27)

"Why aren't images in my clipboard being pasted/transferred to the word document?"— suggested respondent: "Someone from the [TEAM NAME] team that can troubleshoot

this." (p-1-46)

The end-of-study survey responses also indicated the need for domain experts in such Q&A systems where the most cited question type for which participants stated they had received an unhelpful answer was one needing domain expertise. When the knowledge workers of our study received a question answering which required expertise they did not possess, they marked the question as unanswerable, explaining the reason.

5.4 Users' Experience with the Q&A System

Overall, users were fairly satisfied with the answers they received from the Q&A system ($\mu = 3.45$, $\sigma = 0.98$). Only 4 out of the 31 respondents rated the answers unsatisfactory (rating of 1 or 2). We explored participants' free-text responses to understand what they liked or disliked about the system. Participants saw value in the tool helping them be more efficient at their task— *"It helped me by providing answers quickly, I could have probably checked myself some of them by going through previous versions of the documents etc, but this was more efficient." (p-2-29)* or find solutions or workflows appropriate for their particular context— *"I was able to ask questions about a few different capabilities that I was unfamiliar with before, which was nice." (p-2-9).* A number of participants stated that they would have liked to know about possible use cases of the tool— *"I think it would be helpful if the system could present some sample questions so the user could know what type of answers the system could provide." (p-2-14).* We had deliberately withheld examples of usage from users because we did not want to prime users to ask questions of a particular type. Some participants asked for a faster response time— *"More expedient answers" (p-2-5).* Others had concerns about privacy— *"be able to search some properties of a document without full access" (p-2-38).*

5.5 Additional Analyses

To further examine what types of questions a document Q&A system could be most valuable for, we explored how the type of questions affects the time it takes users to answer them. To do so, we probed the data from the experience-sampling phase in which participants submitted their estimate of the time it would take them to answer their query along with the query. We excluded the queries of the type *No Question* or *Unknown* from the data partition, leaving 112 queries in the partition. Figure 5-2 shows the distribution of self-reported response times across question categories. The bar for each question type and response time is normalized by the number of all questions that belong to that question type. The Figure shows that most content-related questions can be answered by the participant in a relatively short time (less than 20 minutes) and questions about style and metadata can take longer. Interestingly, the questions that participants indicated would take them hours to answer or that they could not answer at all were related to External Information or Workflow & Operation Help.

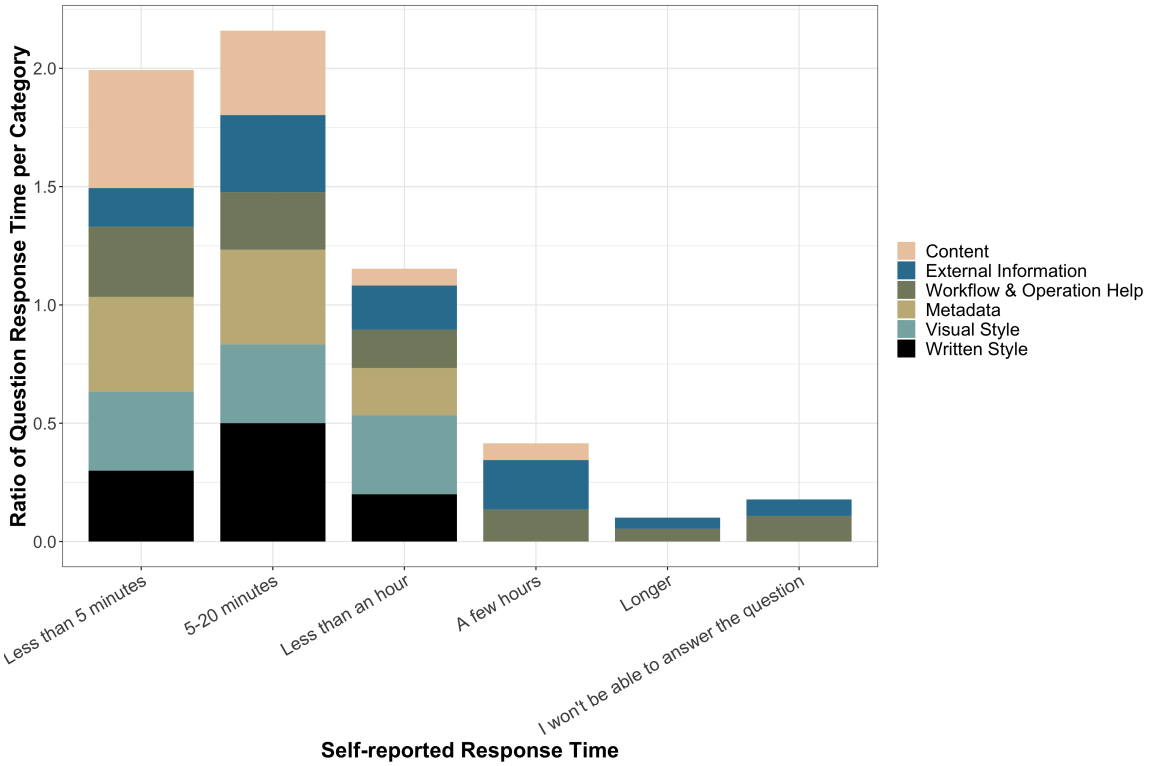

Figure 5-2: Distributions of self-reported response time by the type of questions. Each bar shows the ratio of the questions in the specified category that could be answered within the indicated response time relative to all questions that belong to the question category.

Chapter 6

Discussion

The results of our studies contribute empirical understanding of users' information needs when working with their documents. Understanding the types of assistance that users require is a stepping stone for designing systems that can provide that assistance.

The human-in-the-loop document Q&A prototype in our study was a first attempt at the kind of document digital assistants that we envision. It served as a technology probe to understand the needs of users in a real-world setting, examine the feasibility of generating answers with the aid of AI and other humans, and understand the design requirements of such a technology [20]. Using a document Q&A prototype and collecting questions in-situ as users were working on their documents naturally and when they had questions enhanced the ecological validity of our study with respect to both the nature of questions asked and the frequency with which such a tool would be used. While this data collection method did not result in a large number of questions per participant, it yielded a sample that represented actual questions in the wild. Indeed, some participants told us that they anticipated accessing only a few Word documents during the period of the study.

In the context of our study, because human operators had access to the entirety of the documents through the share links that users provided, in asking questions, users were not bound to the capabilities of current digital assistants. This versatility allowed users to receive help with and us to collect a variety of questions before

attempting to develop a document Q&A system which may not align with users' actual needs.

6.1 Content-Related Questions

One surprising finding of our study was that despite the significant investment in research on extractive question answering (used for factual questions) [39, 35], the majority of the content-related questions that the participants asked could not be answered by returning an excerpt from the document. The shortcomings of existing question answering models could be partly attributed to the datasets they use for training which consist of questions and answers generated by crowd workers or search queries on documents on the web. These queries have been on finalized public documents authored by someone other than the query generator and on which query generators have had little prior knowledge. Therefore, most queries obtained in this corpus lack the diversity of attributes that can influence the type of content-related questions.

One such attribute is the role of the user asking the question, which can potentially affect the type of content-related questions they ask, similar to how user's role affects their overall types of questions. For instance, in the cases where the user is an author of the document, they already have enough familiarity with the document and the context of the information they seek to look for it using navigation or the "find" tool instead of submitting their query to our system which would take longer to respond. As one participant said: *"[M]ost of the questions I had could either be found directly in the doc with a quick Ctrl+F or, for metadata, just looking in file Properties dialog box (right click on file)..." (p-2-25).* In the cases where the user is reading a document they have not authored, their content-related questions may more often involve an overview or finding an answer to a specific question that requires some degree of reasoning. In cases where factual questions were asked, the end-of-study survey responses suggest that the participant may have been posing questions to familiarize themselves with the system: *"Even before asking questions that'd be*

actually useful, I began by asking questions that test its ability." (p-2-37).

Another attribute that can conceivably affect the type of content-related questions a user asks about a document is the stage of development at which the document is. For instance, questions about style or metadata are not often asked on an already published document.

6.2 Opportunities and Challenges of Incorporating Humans in the Loop

Another insight from the range of the questions we received was that a number of them required human assistance from the document's authors, domains experts, or workers with more general skill sets. A document assistant would therefore need to classify and route questions to the appropriate respondents. This component of the system would be akin to IM-an-Expert, a social Q&A system which located and contacted potential respondents with expertise or interest in a subject matter [41, 45]. While questions submitted to a human-in-the-loop document Q&A tool may experience delays, the delay could be justified if the tool enables users to be more efficient or help them accomplish tasks that they could not otherwise perform. Indeed, in Section 5.5, we observed that there are certain types of questions such as those seeking information external to the document or workflow and operation help which would take participants hours to answer or that they would not be able to answer at all. These questions are of the type that could be routed to and handled by other human respondents.

Many questions were beyond the capabilities of the document Q&A model used in this study. Although human respondents are necessary for some of these questions, others have great potential for automation, if only the model could consult other resources. For instance, for answering the questions seeking external information available on public resources, and for some questions about the operation of Word, a Q&A system could make use of the research in information retrieval and search.

Questions seeking information within the company's documents could use the same approach while also leveraging enterprise knowledge bases, together with rich contextual information such as the role or expertise of candidate documents' authors, their organizational distance from the user asking the question, or the authors' ownership of similar documents in addition to the content of the documents. Finally, questions on metadata could be answered by mapping the questions to the correct application interface (API) calls to retrieve the necessary information [42].

While many document-centric requests can be automated and therefore accomplished without the user having to share their content with someone who does not otherwise have access to it, there still exist information needs that cannot be satisfied by pure automation. Because the knowledge workers in this study were the researchers on the project and employees of the same company as the users, most users agreed to sharing the content of their document with the workers. However, privacy may be a challenge in cases where users work with confidential documents or if workers are recruited from outside the company. This issue was addressed in a similar hybrid system, Calendar.help [8], by having human workers signing a non-disclosure agreement and by designing microtasks that include only the information needed to complete the scheduling task. Future work can investigate other approaches to maintaining confidentiality of documents and the user's privacy for instance, by chunking the document into different segments and assigning each to a different knowledge worker in a manner similar to the context-free microtaks explored in [21]. In fact, research has examined how to withhold personally identifiable information in images from crowdworkers by showing only small segments and iteratively zooming out to identify visual information and by leveraging workers' prediction of adjacent segments that are not displayed [30, 26]. In the context of documents, some tasks such as stating the impact of a document or critiquing its overall language style may require a holistic view. However, it is conceivable that teams of knowledge workers could dynamically be organized similar to flash teams in [44, 40], where outputs or summaries provided by some teams are given as inputs to others, eliminating the need for any one worker to have access to the specifics of the document.

Chapter 7

Limitations and Future Work

In this work we focused only on Word documents as a common document type where users can author, copy-edit, or read content. Indeed, the organization where we deployed our studies primarily uses Word for business documents. The focus on Word rather than e.g., PDFs or web pages also allowed us to obtain questions about documents at various stages of development, not just finalized manuscripts. Future work can investigate the types of questions that arise when users interact with other file types such as PDF, Excel, and Powerpoint. Because each file type is used for different purposes (e.g., Excel documents for long-term book-keeping [23]) and possibly containing content at different levels of abstraction, the extent to which question answering in these documents can be automated and the kinds of expertise knowledge workers need may be different from the Word documents in our study.

Because all questions had to first pass through the knowledge workers' system, the majority of the responses that participants received from the Q&A system had a delay. Therefore, it is possible that participants would have asked different questions if the responses had been provided instantaneously. However, because in the Experience Sampling Phase participants could imagine an assistant with any or no delay, the set of questions in our study could generalize to settings where not all questions necessarily experience a delay. In addition, although we had specified in the consent form that there may be a delay in the responses that participants would receive from the Q&A system, some end-of-study survey responses indicated that a number of

participants had in fact not noticed this point— *"I was initially confused by the delay of asking the question in the document and then waiting for an email that told me to go back to the document. It seemed a bit redundant to get an email about it vs. just a notification in the Word doc itself and telling me it was working on it or something. For a plug-in, however, I would expect less of a delay." (p-2-18).* Therefore, the first few questions from these participants could further help with the generalizability of our results.

The prompts that we presented to the participants in both phases slightly nudged them to ask questions that would help them with consuming the document or accomplishing a particular task, rather than delegating a task to the assistant. While we received some questions in the form of commands directed to the assistant in the Phase 1 study, the majority of questions were not task-oriented commands. Future work can investigate the types of tasks that users desire to delegate to a document assistant.

It is conceivable that the types of questions about a document may vary with the document type. In Phase 2 where we had access to the documents, we observed that the document distribution was in fact very varied and included project proposals and timelines, value propositions, design specifications, service instructions, management training, protocols, FAQs, whitepaper reports, strategy planning, customer feedback, research findings, etc. from various domains. With this diverse set of document types, investigating the relationship between types of questions and types of documents would require collecting far more questions by running the study for a long time.

Chapter 8

Conclusion

We studied users' information needs when working with their business documents as a first step towards building document assistants that can handle a variety of user requests. To understand users' actual needs, it was important to collect their document-centeric questions in-situ. Therefore, we conducted two user studies. In the first study, we performed experience sampling of users' questions via a Microsoft Word add-in as users where working with their documents. In the second, users submitted their questions via an add-in and received answers from a human-in-the-loop document Q&A system that complemented a question-answering AI with human intelligence. We characterized the distributions of questions and observed that the types of questions do indeed vary by whether the user is an author, a reviewer, or a reader of the document. In addition, the questions gave us insight into what types of request can be automated and what type of human respondents are needed in a document digital assistant that is co-powered by artificial and human intelligence.

Bibliography

[1] Create professional slide layouts with powerpoint designer. *https://support.microsoft.com/en-us/office/create-professional-slide-layouts-with-powerpoint-designer-53c77d7b-dc40-45c2-b684-81415eac0617*.

[2] Mark S Ackerman and Leysia Palen. The zephyr help instance: promoting ongoing activity in a cscw system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 268–275, 1996.

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

[4] Michael S Bernstein, Greg Little, Robert C Miller, Björn Hartmann, Mark S Ackerman, David R Karger, David Crowell, and Katrina Panovich. Soylent: a word processor with a crowd inside. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 313–322, 2010.

[5] Horatiu Bota, Adam Fourney, Susan T Dumais, Tomasz L Religa, and Robert Rounthwaite. Characterizing search behavior in productivity software. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, pages 160–169, 2018.

[6] Danqi Chen, Adam Fisch, Jason Weston, and Antoine Bordes. Reading wikipedia to answer open-domain questions. *arXiv preprint arXiv:1704.00051*, 2017.

[7] Jared Spataro Corporate Vice President for Microsoft 365. Collaborate with others and keep track of to-dos with new ai features in word. *https://www.microsoft.com/en-us/microsoft-365/blog/2018/11/07/collaborate-with-others-and-keep-track-of-to-dos-with-new-ai-features-in-word/*, November 2018.

[8] Justin Cranshaw, Emad Elwany, Todd Newman, Rafal Kocielnik, Bowen Yu, Sandeep Soni, Jaime Teevan, and Andrés Monroy-Hernández. Calendar. help: Designing a workflow-based scheduling agent with humans in the loop. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2382–2393, 2017.

[9] Mark Dredze. Intelligent email: Aiding users with ai. *University of Pennsylvania, Philadelphia, PA*, 2009.

[10] Matthew Dunn, Levent Sagun, Mike Higgins, V Ugur Guney, Volkan Cirik, and Kyunghyun Cho. Searchqa: A new q&a dataset augmented with context from a search engine. *arXiv preprint arXiv:1704.05179*, 2017.

[11] Ethan Fast, Binbin Chen, Julia Mendelsohn, Jonathan Bassen, and Michael S Bernstein. Iris: A conversational agent for complex tasks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.

[12] George Ferguson, James Allen, Lucian Galescu, Jill Quinn, and Mary Swift. Cardiac: An intelligent conversational assistant for chronic heart failure patient heath monitoring. In *2009 AAAI Fall Symposium Series*, 2009.

[13] Ning Gao and Silviu Cucerzan. Entity linking to one thousand knowledge bases. In *European Conference on Information Retrieval*, pages 1–14. Springer, 2017.

[14] Alexander Gelbukh, Grigori Sidorov, and Adolfo Guzman-Arenas. Document comparison with a weighted topic hierarchy. In *Proceedings. Tenth International Workshop on Database and Expert Systems Applications. DEXA 99*, pages 566–570. IEEE, 1999.

[15] Shuguang Han, Xuanhui Wang, Mike Bendersky, and Marc Najork. Learning-to-rank with bert in tf-ranking. *arXiv preprint arXiv:2004.08476*, 2020.

[16] Jessi Hempel. Facebook launches m, its bold answer to siri and cortana. *Wired. Retrieved January*, 1:2017, 2015.

[17] Damon Horowitz and Sepandar D Kamvar. The anatomy of a large-scale social search engine. In *Proceedings of the 19th international conference on World wide web*, pages 431–440, 2010.

[18] Matthew B Hoy. Alexa, siri, cortana, and more: an introduction to voice assistants. *Medical reference services quarterly*, 37(1):81–88, 2018.

[19] Ellen Huet. The humans hiding behind the chatbots. *Bloomberg. com (April, 18, 2016) https://www. cnet. com/news/facebook-is-killing-m-its-personal-chatbot-assistant*, 2016.

[20] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, et al. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 17–24, 2003.

[21] Shamsi T Iqbal, Jaime Teevan, Dan Liebling, and Anne Loomis Thompson. Multitasking with play write, a mobile microproductivity writing tool. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, pages 411–422, 2018.

[22] Mohit Iyyer, Jordan Boyd-Graber, Leonardo Claudino, Richard Socher, and Hal Daumé III. A neural network for factoid question answering over paragraphs. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 633–644, 2014.

[23] Farnaz Jahanbakhsh, Ahmed Hassan Awadallah, Susan T Dumais, and Xuhai Xu. Effects of past interactions on user experience with recommended documents. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, pages 153–162, 2020.

[24] Mandar Joshi, Eunsol Choi, Daniel S Weld, and Luke Zettlemoyer. Triviaqa: A large scale distantly supervised challenge dataset for reading comprehension. *arXiv preprint arXiv:1705.03551*, 2017.

[25] Ece Kamar and WA Redmond. Hybrid intelligence and the future of work. In *Productivity Decomposed: Getting Big Things Done with Little Microtasks Workshop (CHI 2016). http://research. microsoft. com/en-us/um/people/eckamar/papers/HybridIntelligence. pdf*, 2016.

[26] Harmanpreet Kaur, Mitchell L Gordon, Yi Wei Yang, Jeffrey P Bigham, Jaime Teevan, Ece Kamar, and Walter S Lasecki. Crowdmask: Using crowds to preserve privacy in crowd-powered systems via progressive filtering. In *HCOMP*, pages 89–97, 2017.

[27] Giridhar Kumaran and Vitor R Carvalho. Reducing long queries using query quality predictors. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 564–571, 2009.

[28] Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, Kenton Lee, et al. Natural questions: a benchmark for question answering research. *Transactions of the Association for Computational Linguistics*, 7:453–466, 2019.

[29] J Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.

[30] Walter S Lasecki, Mitchell Gordon, Jaime Teevan, Ece Kamar, and Jeffrey P Bigham. Preserving privacy in crowd-powered systems. In *Proceedings of AAMAS 2015 Workshop on Human-Agent Interaction Design and Models*, 2015.

[31] Walter S Lasecki, Rachel Wesley, Jeffrey Nichols, Anand Kulkarni, James F Allen, and Jeffrey P Bigham. Chorus: a crowd-powered conversational assistant. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pages 151–162, 2013.

[32] Cade Metz. Ai helps humans best when humans help the ai. *Wired.com (2015) https://www.wired.com/2015/09/ai-helps-humans-best-humans-help-ai*, 1, 2015.

[33] Kevin Murnane. Ifttt survey provides insight into what people do with amazon's echo and google's home. *(Forbes 2017) https://www.forbes.com/sites/kevinmurnane/2017/07/12/ifttt-survey-provides-insight-into-what-people-do-with-voice-controlled-assistants/2d5966b643e6*, July 2017.

[34] Karen Myers, Pauline Berry, Jim Blythe, Ken Conley, Melinda Gervasio, Deborah L McGuinness, David Morley, Avi Pfeffer, Martha Pollack, and Milind Tambe. An intelligent personal assistant for task and time management. *AI Magazine*, 28(2):47–47, 2007.

[35] Tri Nguyen, Mir Rosenberg, Xia Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder, and Li Deng. Ms marco: A human-generated machine reading comprehension dataset. 2016.

[36] Christi Olson and Kelli Kemery. Voice report: From answers to action: customer adoption of voice technology and digital assistants. *Microsoft Search and Market Intelligence, Tech. Rep*, 2019.

[37] Emma Persky. Now we're cooking – the assistant on google home is your secret ingredient. *https://www.blog.google/products/assistant/cooking-with-the-assistant-google-home-your-secret-ingredient/*, April 2017.

[38] Pranav Rajpurkar, Robin Jia, and Percy Liang. Know what you don't know: Unanswerable questions for squad. *arXiv preprint arXiv:1806.03822*, 2018.

[39] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250*, 2016.

[40] Daniela Retelny, Sébastien Robaszkiewicz, Alexandra To, Walter S Lasecki, Jay Patel, Negar Rahmati, Tulsee Doshi, Melissa Valentine, and Michael S Bernstein. Expert crowdsourcing with flash teams. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 75–85, 2014.

[41] Matthew Richardson and Ryen W White. Supporting synchronous social q&a throughout the question lifecycle. In *Proceedings of the 20th international conference on World wide web*, pages 755–764, 2011.

[42] Yu Su, Ahmed Hassan Awadallah, Madian Khabsa, Patrick Pantel, Michael Gamon, and Mark Encarnacion. Building natural language interfaces to web apis. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 177–186, 2017.

[43] Maartje ter Hoeve, Robert Sim, Elnaz Nouri, Adam Fourney, Maarten de Rijke, and Ryen W White. Conversations with documents: An exploration of document-centered assistance. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, pages 43–52, 2020.

[44] Melissa A Valentine, Daniela Retelny, Alexandra To, Negar Rahmati, Tulsee Doshi, and Michael S Bernstein. Flash organizations: Crowdsourcing complex work by structuring crowds as organizations. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 3523–3537, 2017.

[45] Ryen W White, Matthew Richardson, and Yandong Liu. Effects of community size and contact rate in synchronous social q&a. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2837–2846, 2011.

[46] Han Xiao, Feng Wang, Jianfeng Yan, and Jingyao Zheng. Dual ask-answer network for machine reading comprehension. *arXiv preprint arXiv:1809.01997*, 2018.

[47] Yi Yang, Wen-tau Yih, and Christopher Meek. Wikiqa: A challenge dataset for open-domain question answering. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 2013–2018, 2015.

[48] Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William W Cohen, Ruslan Salakhutdinov, and Christopher D Manning. Hotpotqa: A dataset for diverse, explainable multi-hop question answering. *arXiv preprint arXiv:1809.09600*, 2018.

[49] Jin-ge Yao, Xiaojun Wan, and Jianguo Xiao. Recent advances in document summarization. *Knowledge and Information Systems*, 53(2):297–336, 2017.