Bento Browser: Complex Mobile Search Without Tabs

Nathan Hahn, Joseph Chee Chang, Aniket Kittur
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213
{nhahn, josephcc, nkittur}@cs.cmu.edu

ABSTRACT
People engaged in complex searches such as planning a vacation or understanding their medical symptoms are often overwhelmed by opening and managing many tabs. These challenges are exacerbated as search moves to smartphones and mobile devices where screen real-estate is limited and tasks are frequently suspended, resumed, and interleaved. Rather than continue to utilize tab-based browsing for complex search, we introduce a new way of browsing through a scaffolded interface. The list of search results serves as a mutable workspace, where a user can track progress on a specific information query. The search query serves as a gateway into this workspace, accessed through a task-subtask hierarchy. We instantiate this in the Bento mobile search system and investigate its effectiveness in three studies. We find converging evidence that users were able to make progress on their complex searching tasks with this structure, and find it more organized and easier to revisit.

Author Keywords
Exploratory Search; Sensemaking; Mobile

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
People spend an enormous amount of time making sense of the world online, whether patients trying to determine the causes of their conditions, scientists trying to understand an emerging field, or citizens trying to understand the effects of a proposed health care bill. They may read through hundreds of web pages, discussion forums, blog posts, tweets, reviews, and scientific, news, and magazine articles trying to collect the evidence they need to build a mental representation of the information space that is useful for their goals [29]. Although fewer in absolute number, complex searches consume a disproportionate amount of attention: a recent Forrester poll puts the amount of time an U.S. household spends online at approximately 12 hours per week [3], with studies showing that up to 33% of this time is related to complex exploratory search tasks [29, 23, 38]. Helping people make sense of complex information, diverse viewpoints and evidence could significantly improve learning and decision making across a variety of domains and user populations.

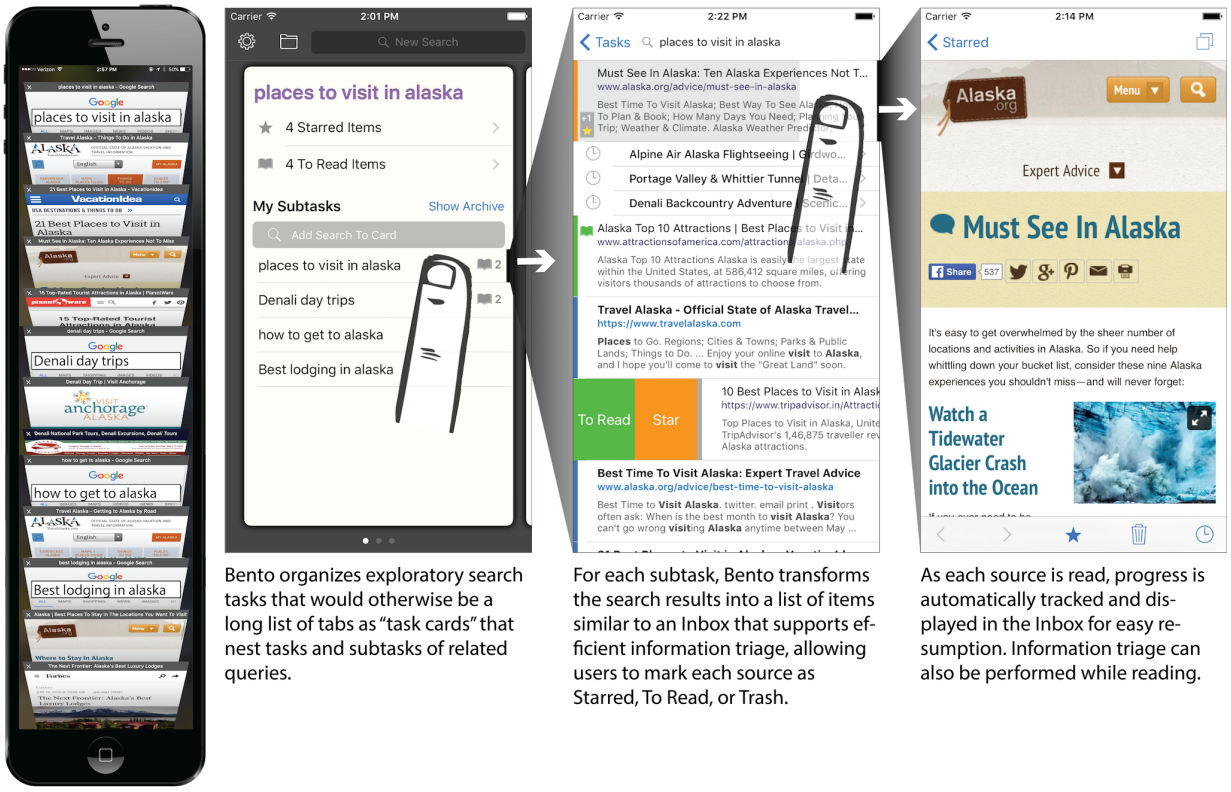
As search increasingly shifts to mobile environments – in 2015 the number of searches from mobile devices overtook desktop searching [14] – complex sensemaking becomes even more difficult. Consider a person planning a trip to Alaska: on a desktop they may create multiple tabs for each location or point of interest, which quickly multiplies as the user drills down into restaurants, hotels, and activities for each of those locations – potentially resulting in dozens of tabs open at once. Many of these tasks may be going on in parallel (e.g., investigating alternate destinations such as Anchorage vs. Homer), may be suspended and resumed in various states of progress over time, and may be interleaved with other tasks (e.g., finding a place to eat tonight). On a mobile device, limited screen real estate, short bursts of use, frequent interruptions, difficulties in saving and organizing information, and loss of context pose even more serious challenges. Addressing sensemaking in a mobile device context thus is not only timely and important, but provides additional generative constraints for new approaches.

Our core contribution in this paper is introducing an alternative approach to tabbed browsing that also addresses the additional constraints involved in a mobile context. The key insight we build on is that tabs are often performing two distinct functions: 1) they serve as a way to organize and juggle multiple tasks that may be going on at once; and 2) they serve as a workspace to triage and build a mental model for a given task, for example queuing sources for later consumption, performing comparisons between sources and saving information of uncertain value for further review. Because tabs are overloaded in such a manner, we argue that they accomplish neither task very well, especially when used in a constrained mobile environment.

To overcome the limitations of tabs, we introduce a scaffolded process that separates the task management and workspace functions of tabs into two distinct interfaces. Instead of having many open tabs we transform the search results page into a mutable workspace that allows users to triage and keep track of their progress on any given search, with those searches collected into tasks and subtasks. We instantiate this approach in a novel mobile web browser, Bento Browser, and evaluate

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

ISBN 978-1-4503-5620-6/18/04…$15.00
DOI: https://doi.org/10.1145/3173574.3173825



Bento organizes exploratory search tasks that would otherwise be a long list of tabs as "task cards" that nest tasks and subtasks of related queries.

For each subtask, Bento transforms the search results into a list of items similar to an Inbox that supports efficient information triage, allowing users to mark each source as Starred, To Read, or Trash.

As each source is read, progress is automatically tracked and displayed in the Inbox for easy resumption. Information triage can also be performed while reading.

Figure 1. Comparing a typical list of tabs (left) with Bento's search centered navigation from the same exploratory search task.

its effectiveness through three user studies. Our results suggest opportunities for the development of novel systems of online information seeking for both mobile and desktop platforms that both better suit the nature of complex searching as well as constrained mobile environments.

DESIGNING FOR MOBILE EXPLORATORY SEARCH

Search is a critical activity in daily life – from looking up a phone number, to finding the right apartment. Some of these information needs are simple and straightforward – we can find the answer within the first search result or even in a search result description [7]. However, for more complicated tasks, like learning about linear algebra or trying to figure out what places to visit on a vacation to Alaska, we often depend on multiple competing information sources and have to interpret and synthesize what we find. These "exploratory" searches require multiple iterations of searching to understand and form a mental model of the information space [29]. This is one instance of the more general concept of sensemaking – an iterative process where a user is building up an understanding of an information space in order to achieve a goal [39]. Theories and related empirical work point out that unlike simple factual information finding (e.g., what is the weather, when was someone born), for complex sensemaking tasks like shopping or making health decisions finding relevant information sources is only the first step in the search process [39, 45]. Users must also perform additional synthesizing to produce an actual understanding. A number of models of sensemaking have been proposed, including Russell et al.'s cost structure view [39], Dervin's sensemaking methodology [13], Klein et al.'s data-frame model [25], organizational process views [17], organizational adaptation views [12, 31], and the notional model by Pirolli and Card [36].

Based on these models, there have been a number of attempts to better assist desktop users performing exploratory search. Researchers have identified and improved upon areas such as triage [19, 40, 30], comparison and synthesis [24, 11], and task-based organizations for searches [33, 44]. However, despite the importance of exploratory search, it is still poorly supported by existing browsers and search engines [46, 29].

Mobile Sensemaking

More searches are now happening on mobile devices than on desktop devices [14] and large populations of mobile-only device users that now exceed desktop-only internet users [28]. Existing research and systems relevant to mobile search have primarily focused on supporting simple informational queries, with recent advances including factual question answering, cognitive and task assistants, information snippets, location aware ranking, and chatbots [5, 41, 8, 6, 37, 27, 47].

However, considering that smartphones are many users' primary computing devices, we posit that users may be engaged in exploratory searching as well. To explore this we performed a short survey, first partially published in [9], with 164 smartphone users (98 Male, 66 Female, Age: M= 32.29, SD= 8.72) on Amazon's Mechanical Turk platform. We asked a series of questions about a user's exploratory searches, how often they

perform them, what were some past searches, as well as the interface tools they use. Surprisingly, we found that people reported frequently conducting complex exploratory searches either partly (70%) or completely (45%) on their phones, ranging from planning a vacation to researching woodworking projects. However, 47% of the users also agreed with the statement that "It would be frustrating to do a complex search on a smartphone".

We asked participants a number of questions about their current habits, based on a 5 point Likert scale (Rarely - A Great Deal). When queried about which exploratory search activities they currently perform on their phones, the most common activity was simply "Reading web pages" (M = 4.03, SD = 0.89). Text entry during search (M = 3.59, SD = 1.14) and keeping track of multiple pages (M = 3.24, SD = 1.12) were the next to most common activities, with saving web pages (M = 2.40, SD = 1.17) and collaborating (M = 2.21, SD = 1.14) being the two most uncommon activities.

We then asked about future support. 80% of participants agreed with the statement "I would find it valuable if smartphones had better interfaces for doing complex searches". Delving deeper into this question, at least 1/3 of the respondents reported extreme difficulty (highest Likert rating) with "saving web pages", "keeping track and switching between pages" and "sharing findings with others". These suggest that the current browser interfaces on smartphones do not well support the constant context switching and task suspension present in exploratory search. Conversely, participants cited the advantage of being able to do searches "on the go" and the general "convenience" of smartphones. These results suggest users think there are significant problems with managing exploratory searches on smartphones, even though they currently do them, and would like to continue to do them. This suggests addressing complex searching in the mobile context may have both real world practical value as well as being a source of potentially generative design constraints that could also translate to less constrained device footprints such as the desktop.

Understanding Tabs

Tabs are a ubiquitous feature in every major web browser today, where they serve multiple functions ranging from organization to triage to reminding [21, 15]. In particular, they serve two primary functions in the exploratory search process. First, tabs provide task management functions – by separating out tasks [44], acting as a reminder to resume a task, and allowing for quick efficient switching between tasks [21]. Second, they provide a workspace for triaging sources [19], performing comparisons between sources [24], and saving good resources for further review.

We note three specific problems with the ways that tabs try to support these two function simultaneously. First, tabs are only loosely coupled to their generating activity. As a result, tabs during exploratory search become disconnected from their search results page, potentially causing negative effects such as users losing track of why they opened a tab, where they were in their task progression, and which pages belong to which tasks [15]. This is particularly problematic early in the exploratory search process as users are uncertain about the future value of the information contained within them [24]. Second, tabs are ordered based on the sequence in which they were opened and which tab spawned them, in order to keep them co-located to the other tabs in their task. This can become inconsistent as an organizational model as tabs are closed or opened in the middle of other tabs, and also misses an opportunity to provide more meaningful organizational structure, either for separating tasks or as a workspace. Lastly, tabs have limited context (e.g., a favicon and partial title) which can make it difficult to find a tab, know the state of progress on using it, or understand which tabs belong to which tasks. All three of these challenges are exacerbated in the mobile context, where there is little space to show multiple tabs at once or to provide context for them.

SYSTEM DESIGN

We introduce Bento, a novel interface for scaffolding complex search which obviates the use of tabs while supporting their underlying functions of task management and serving as a workspace. This can be seen in see Figure 1, which shows how a user might perform planning a trip to Alaska with the current paradigm of tabbed browsing on a mobile device versus Bento. The fundamental component enabling the approach is transforming the search results page in place into a mutable workspace that allows the user to queue page to read (analogous to the common practice of opening a search result link in a new tab), star pages they found useful, trash unhelpful pages, and, critically, to see the progress they have already made in reading each page they opened (See Figure 1). Unlike previous approaches (e.g., collaborative search [34], history management [33, 2], or activity workspaces [20, 43, 22]) which require a separate interface for managing and surfacing individual tabs, Bento's approach provides a natural centralized workspace in the search results page that is already a fundamental and familiar element of navigating complex searches and obviates the need for tabs altogether.

Of course, for complex searches a single search is often not enough; for example for planning a trip to Alaska one might have additional searches for day trip destinations, how to get there, and where to stay. For managing tasks and subtasks Bento bundles search result pages together into task cards, drawing inspiration from previous search-based task management tools such as SearchBar [33]. However, one difference from tools which focus on surfacing the past history of searches is the prospective nature of Bento's scaffolding, in which users can (and did) create searches as placeholders and reminders of subtasks they would need to work on (like finding a place to stay) before actually doing any of the work. Together, these elements suggest a radically different way for people to manage complex searching than traditional tabbed browsing. Below we describe the design rationale for developing Bento and details about its various interface elements.

A Sensemaking Workspace

When creating this workspace, we initially considered leveraging approaches utilized by previous information triage systems (e.g., [40, 19]). However, these tended to rely on spatial organization which were not a good fit for the limited real estate

of smartphones. After a number of design iterations, we settled on a representation evoking the affordances of an email inbox. Email inboxes are designed for quick and efficient triage by users, providing information to users about what information is important, has been dealt with, and what still needs to be read / triaged. They accomplish this in a simple list format – not requiring the larger spatial requirements of other information triage systems. Email inboxes provide users with organization strategies ranging from flagging or starring items (which can pin them to the top of the list), archiving undesired items, and marking items to be read later (e.g., through marking as unread). We found these strategies useful for organizing searches, allowing search results to be flagged as important, archived if irrelevant or not needed in the future, marking items as potentially relevant and of interest to come back to, and supporting an awareness of where search results are in the list (in this case, ranked by relevance to the query if acted on, or in their original search result order if not). The items in the inbox are ordered with starred items at the top, followed by to read items, and finally any other search results in their original relevancy order. Trashed items are placed at the very bottom of the search results list to enable undo if needed. The search results also have a natural progression of states: unviewed search results show up with bold text and a blue dot similar to an "unread" email message. Users can then manually mark a page as being in an intermediate state, with a 'to read later' annotation, or as a particularly useful reference source, with a 'star'.

Figure 2. The different manipulations that can be applied to a search result

We not only considered triage to be important in this interface, but also resumption and information provenance (See Figure 2). Resumption is managed through a couple of factors: the search results persisting in appearance, read / unread indicators, and progress bars. Initially, we combined the progress indicator and read/unread indicator into the same space, however this caused a number of misinterpretations or disregard of the progress indicator all together. To increase visibility, we separated the progress indicator from the priority indicator (see Figure 4). We instead represented the progress indicator as the background fill of each search result. As the user scrolled further down the page for a result and spent more time on that result, a gray bar would fill up the background of the row. In early piloting users found this to be much more intuitive and easily parsed. The read / unread indicator was adjusted to be a colored bar on the far left of the row. Additionally, when a user reopens a page, they are automatically scrolled to their last position on that page, letting them quickly resume what they were reading. We believe this novel approach of showing progression information directly on an information source gives users a way to understand progress without having to visit the source.

Lastly, to maintain information provenance, all subsequent pages visited from a search result are associated with that result. In a normal, tabbed environment, there is no obvious connection from a new tab to a previous page. Even the tacit relationship of being next to each other can be broken if tabs are opened in between. In Bento, there is a fundamental connection between each page and the search result it was opened from. If a user is reading a page that is a deep link from the search result, and they return to the list of results, that page is surfaced under the result with a small clock icon, representing it was a page the user was reading and paused. Similarly, if a user stars, or marks a subsequent page as to read later, it appears in the search results list under its parent search result (see Figure 4). This provides an easy way for a user to resume their progress even from a page nested deeply within a search result.

Managing Sensemaking Tasks

Bento not only assist with managing the information from one searching task, but also gives users a way to juggle multiple information seeking tasks at once. Bento features a separate, second interface for managing the higher level tasks users are working on, and their multiple sub-components. Noting how users utilize tabs, this management interface needed to allow for quick switching between tasks, act as a reminder to resume a task, and create separate workspaces for each task. Inspired by the previous work [20, 43, 22], we decided to utilize an activity-centric management interface based on a task-subtask hierarchy, with the search queries acting as the task unit [33]. However, instead of using this hierarchy as a post-hoc way of revisiting and organizing tasks, we have users actively build their tasks in this structure.

The tasks and subtasks are organized into cards, designed to give the user a quick overview of the current status of their sensemaking activities (See Figure 3). In order to make the structure as lightweight as possible, when a user creates a

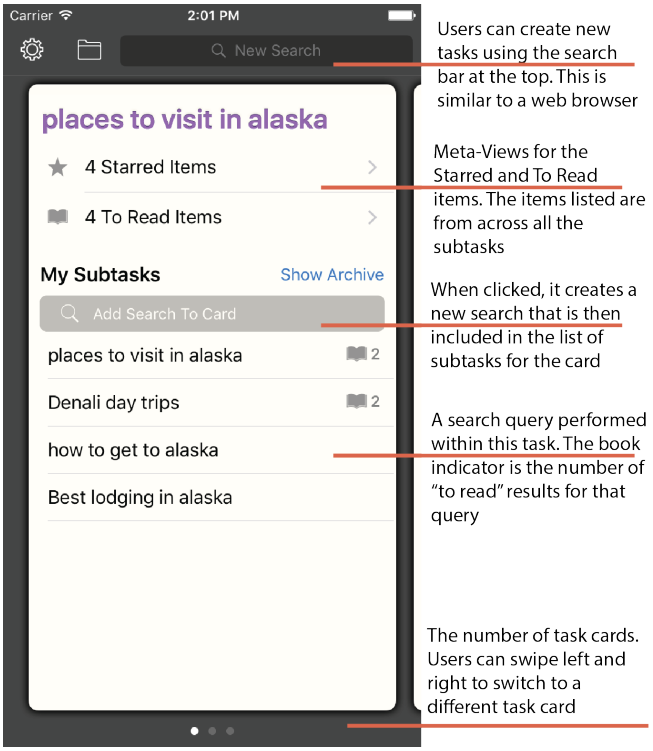


Figure 3. The task screen for the task "Places to go to Alaska"

search based on a new topic of research, we create a new task for them, naming it the title of their search. If a user adds a search to an existing task, a subtask appears under the task, with the query as the subtask's name. These task cards allow users to actively switch between tasks using a simple swipe, and their isolated "card" presentation provides the user a clear and present workspace for organizing, and quickly resuming their tasks activities. This is dissimilar from a tab environment, where everything is presented in a flat, undifferentiated structure.

To assist with the tedium of reorganizing tasks, the cards are automatically reordered based on recency – similar to adaptive human memory [4]. When an activity is performed in a task, it is pushed to the top of the list, and older ones never resumed slowly fall to the bottom. The subtasks within a task card follow a similar ordering. More recent queries, shown towards the top, serve as both a reminder for users about subtasks they need to complete, and allows users to orient their work chronologically as their understanding of the information changes. The temporal organization also allows users to scroll down their list of searches so that they can gain a retrospective understanding of how their mental model has evolved over time and restore the context they had when they stopped the search. This structure removes the ambiguity from tab ordering, creating a consistent interaction for later resumption [35].

To create a new task, a user types in a new information need (query) into the top search bar. Initially, we required users to create a task card, and then used it's naming as the information query. However, a normal browser provides a one touch experience in order to create a new search, with a large target for initiating the search. Recognizing that our initial approach broke the user's mental model of searching, we modified the workflow to use a more traditional search bar. The search provides results / suggestions for existing cards, subtasks, as well as normal auto completion results.

We utilized a similar approach for creating subtasks (or sub queries). Each task card features a separate search button titled 'Add Search To Card' modeled in the appearance of a normal search bar. This reduced the cost for creating a new search to a single tap on a large salient target, in comparison to needing to tap a small button to create a card then subsequently search. On the task card, we provide rich information about the status of the individual subtasks. Beside each subtask, we note the number of starred and to read items from the search results. Besides the ordering of the subtasks, these provide the user with information on the completion of each subtask, as well as the usefulness of that particular information query.

Initially, we required users open up their individual subtasks to view important saved information, or to make progress. After the first study, one participant noted that "it would be nice ... to see all of my starred items in one place for easy reference." As a result, we added summarization lists on the to-do list card view (see Figure 4). These summarization lists allowed a user to immediately look at all of their to read results and starred results across all the searches in a task. The "to read" summarization list became a reading list for the task, while the starred summarization list as a collection of the most important information an individual had collected for the task. These views serve as a way for users to get an understanding of either the important information they have collector for a task quickly, or what information they need to process next in a task.

Implementation

The Bento Browser application was built for the iOS platform and was available to participants running iOS 8.0 and above. The application utilizes Google's Firebase real-time database and analytics platform to collect telementry data from users. We utilize the Bing API to fetch search results for queries made by users.

EVALUATION

We completed three studies to provide converging evidence on the value of our approach: a lab evaluation, a qualitative real world deployment, and an expanded quantitative deployment. Between studies we used the feedback to iterate and refine the design of the system. We explore whether our dual interfaces of task management and an information triage workspace are able to more effectively accomplish what tabs try to. We specifically look at the pressure points caused by a tabbed interface: organization of tasks, a workspace for information processing and sufficient context for quick resumption of activities. In each of these studies, we optimized the design to promote maximum motivation to actually work on the complex searching tasks by having users work on their own tasks. Rather than trying to evaluate an outcome from a fixed search

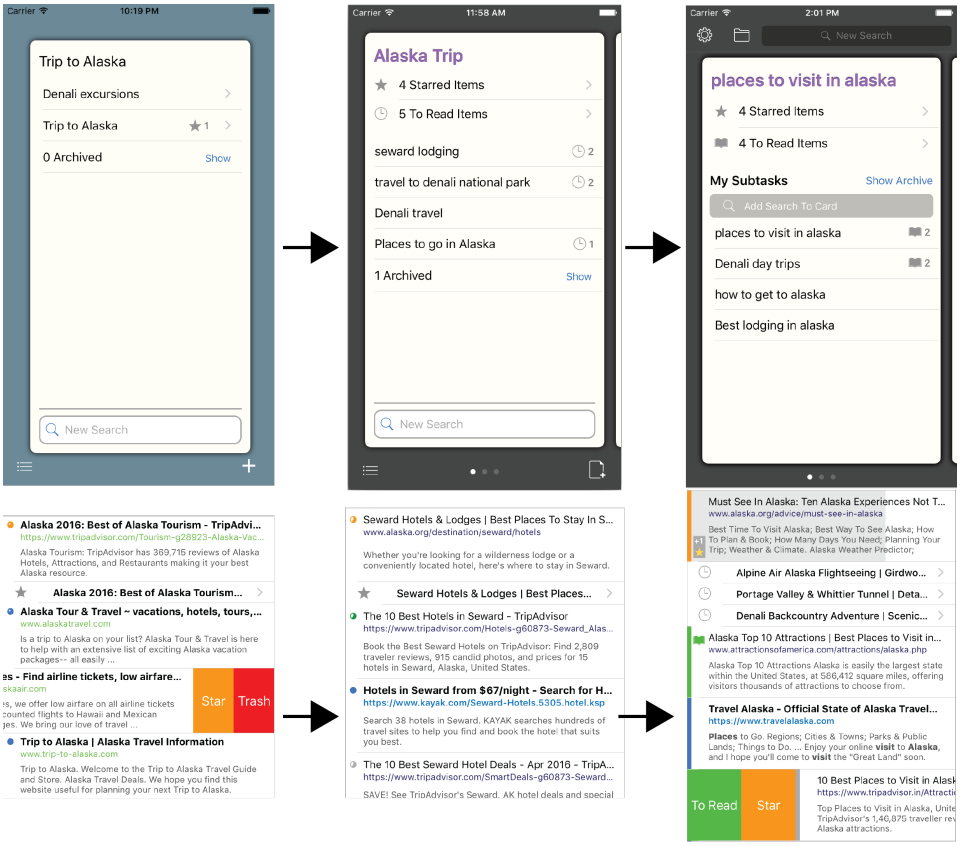


Figure 4. The progression of the Bento Brower design. From left to right: Study 1, Study 2, Study 3

task, we wanted to capture how individuals found Bento to be useful for a variety of complex searching tasks.

Study 1 - Understanding Triage

The goal of the lab study was to evaluate our approach while controlling for differences in the complexity and nature of the tasks that users engaged in, which would otherwise vary in a field trial. We focused on the triage interface in this study, having users only work on one task while in the lab. It also provided an opportunity to get feedback and iterate on the system's features that might otherwise cause critical issues in a lengthier deployment.

In order to make the lab study task realistic and providing internal motivation, we collected multiple real search tasks from each participant and randomly assigned one to the Bento condition and another to an environmentally valid control, for which we chose the Safari mobile web browser (the default tabbed browser on the iPhone). The study was a within-subjects design in which participants used both browsers (counterbalanced across participant) and provided feedback on them.

Procedure

We recruited 22 participants through a local behavioral research participant pool. Participants ranged in age from 20 to 59, with the majority of participants being local undergraduate and graduate students. 10 participants identified as male and 12 as female. We required that study participants own and use an iPhone to ensure they would be familiar with the existing iPhone operating system and Safari browser. All participants were provided with an iPhone 6s with Bento preloaded onto for use during the study.

Since a single fixed search task might not provide internal motivation to every participant [26], we instead elicited four search tasks of potential interest from the participants themselves during a prescreen. We selected two of these searches based on how participants rated the topics across 4 scales: their knowledge of the topic, the importance of the topic, the expected research time to learn about the topic, and the estimated number of web pages they would have to visit to fulfill their information need. To select one of their proposed searches, we required participants mark it as at least moderately important, have less than a moderate amount of knowledge about the topic, the search would take at least several hours, and the search would require least 8 different web pages. For each participant, we selected two searches that met the criteria (if more than 2 met the criteria, we chose those with the highest values) and randomly assigned them to either the Safari condition or the Bento Browser condition. Some example searches include "How to create an Android application" and "Advice on how to enjoy being a tourist in Japan".

Participants were asked to search for 20 minutes using either Bento or Safari, then to switch to the other search with the other tool (with order of browser counterbalanced across participants). For the Bento condition, before they began they

were provided with a brief tutorial that walked them through the interface and features; all participants were already highly familiar with Safari from their own phone use. After completing both searching tasks participants completed a post-survey about their experiences. The survey asks participants to directly compare their experience with the Bento tool to the Safari mobile browser, as well as review some of the features of the Bento tool.

Results

Overall, we found that participants appreciated the Bento interface, finding that compared to Safari, it helped keep them significantly more organized ($M = 4.25, 95\%\ CI[3.91, 4.59]$), and would be more useful for helping them restart where they left off ($M = 4.15, 95\%\ CI[3.66, 4.64]$). Despite participants' high familiarity with Safari, we did not find significant differences in the ease of search creation using our tool, nor did individuals feel less effective using it. Participant did note that Safari was much easier to learn (85% of the participants stated this), however 70% thought that Bento was more helpful in finding pages. This was especially notable considering the prototype status of the system during Study 1 and the additional steps individuals had to go through in order to make and organize their searches.

The comparison was also well supported by feedback on the individual features of the search. On average, participants reported in our post-survey (using 7-point Likert scales) that they enjoyed the software and the features provided by it ($M = 4.95, 95\%\ CI[4.18, 5.71]$). They thought that Bento Browser amplified their search effectiveness on a mobile phone ($M = 5.25, 95\%\ CI[4.49, 6.01]$), they felt confident searching using the tool ($M = 5.05, 95\%\ CI[4.30, 5.80]$), and they reported wanting a tool like Bento Browser for searching on their mobile phones ($M = 5.15, 95\%\ CI[4.37, 5.93]$).

Of all the features, participants found starring pages to be the most useful tool (over 90% reported starring being moderately useful). When asked more about this, they noted that starring pages made it "incredibly easy to save pages" and in general "it was easy to collect a large amount of relevant webpages to read and delete the irrelevant ones." This suggests that the triage interface made it easy to quickly sort through the search results, and persist the important information for later use. A participant directly agreed with this, stating "... I could refind my pages for future viewing. This is very useful for searches that I am more likely to come back to."

Study 2 - Task Management

We iterated on the initial version of Bento based on the feedback from the previous study. Several participants noted that the interface was âĂIJclunky compared to the web browserâĂİ and it needed to be more attractive. A few others were confused by some of the interactions, such as what happens when they trashed a search result. From the qualitative feedback participants provided in the lab post-study questionnaire, we worked on three areas for improvement: visual attractiveness, better feed-forward and feed-back cues, and the summarization views.

In order to better evaluate the utility of the iterated version of Bento in a more ecologically valid setting we conducted an exploratory field study, in which participants used Bento daily for a period of 4 to 6 days.

Procedure

We recruited 8 participants from the same local participation pool as in Study 1. We required that participants own an iPhone with iOS 8.0 or above installed and had not participated in the previous study. Participants ranged in age from 18 to 24, with four identifying as male and four as female.

Individuals installed the Bento Browser application on their personal mobile device in the lab, completed a short tutorial, then spent 15 minutes working on a search of their choice in the lab so that they could ask questions and get used to the tool. They were then were instructed to use it for at least 10 minutes each day. The application provided a reminder three times a day if the individual had not yet used it for 10 minutes that day. Aside from the time requirement, we did not instruct the individuals to utilize the application in any particular way. We were interested in knowing how individuals used the different features of the application, and which features were the most useful to each individual.

After a 4 to 6 day period, participants returned to the lab for an interview and to complete a post-survey. During the interviews, we asked participants to walk through their usage of the application, showing off any of the concrete tasks that they did, as well as exploring their individual queries. This probe was designed to help users ground their experience of using the app in the specific tasks that they were performing. After the interview, participants completed the same post-study questionnaire as in Study 1.

Results

Post-survey results were very similar to the results from the lab study, with participants significantly preferring Bento over Safari for the questions "If you wanted to keep searching later, which tool would be better for picking up where you left off?" ($M = 4.25, 95\%\ CI[3.38, 5.11]$) and "Which tool makes your information more organized?" ($M = 4.125, 95\% CI[3.43, 4.82]$). Additionally, participants also preferred Bento for the question, "It was easier to refind information with (Bento Browser)" ($M = 4.125, 95\%\ CI[3.30, 4.95]$) in favor of Bento. No other questions showed significant differences. Feedback about Bento was similar to the lab study survey, except more individuals cited a desire for a desktop companion ($M = 5.25, 95\%\ CI[3.72, 6.78]$), suggesting that additional usage in different contexts incurred the desire to switch between devices with different characteristics.

The interviews provided further insight into how individuals used the application for their own needs. Participants used Bento for a variety of exploratory tasks, ranging from learning about gardening techniques to product comparison to learning about political candidates. Several of the participants brought up Bento's value in capturing their mental model during the search process, which helped them get an overview of their search and suspend and resume it more easily. P7 specifically noted that you could "see everything that you Googled ... in

Question	Study 1 Mean	Study 1 CI	Study 2 Mean	Study 2 CI	Study 3 Mean	Study 3 CI
Which tool did you like better	3.15	[2.45, 3.85]	3.125	[2.18, 4.06]	3.01	[1.94, 3.89]
Which one was easier to create new searches in?	3.4	[2.82, 3.98]	3.126	[1.99, 4.26]	3.38	[2.76, 3.99]
If you wanted to keep searching later, which tool would be better for picking up where you left off?	4.15*	[3.66, 4.64]	4.25*	[3.38, 5.12]	4.44*	[4.05, 4.83]
Which tool makes you feel more at peace?	2.9	[2.16, 3.64]	2.63	[2.01, 3.25]	2.69	[2.05, 3.32]
Which tool makes your information more organized?	4.25*	[3.91, 4.59]	4.13*	[3.43, 4.82]	4.25*	[3.89, 4.61]
I felt more effective using:	3.2	[2.56, 3.84]	3.125	[2.18, 4.06]	3.01	[1.94, 3.89]
It was easier to refind information with:	3.47	[2.96, 3.99]	4.13*	[3.30, 4.95]	3.31	[2.65, 3.98]
I felt more confident that I didn't miss any important sources of information with:	3.0	[2.39, 3.61]	3.38	[1.96, 4.78]	2.53	[1.89, 3.31]
* Significantly different based on 95% Confidence Interval						

Table 1. The direct comparison questions were asked on a 5-point likert scale. A higher score indicates preference for Bento Browser, while a lower score indicates preference for the Safari browser. A score of 3 indicated no preference for one over the other. This table covers Studies 1, 2, and 3.

a straight sequence." and the different triage lists let you "archive what you were thinking about in a single moment ... it was like a screen shot of what you were thinking about." P5 mentioned that he enjoyed "just being able to quickly look at the task list and know what to do next".

Perhaps the most important value perceived by participants was in how Bento structured searches into organized workspaces in which they could make progress. Organizing searches into tasks and recording searches as subtasks were rated highly in the survey (5.5 and 5.9, respectively). This led to some unexpected benefits, such as one participant noting "how easy it is to compare prices this way rather than with a traditional browser". Participants seemed to actively want to keep their subtasks organized, with 6 of 8 mentioning that they enjoyed utilizing the "trash" feature to throw out irrelevant results. We found this interesting because eliciting explicit feedback from users about search results is traditionally challenging, as users could just skip over the search result without having to put in the extra effort to trash it. One interpretation of this is that when users perceive the search results screen as a workspace rather than simply a launching pad they are more willing to invest effort into personalizing it.

Consistent with this, all participants utilized either the starring feature or the to-read functionality. Some of them (P2, P3) indicated that they weren't sure what the point of the to-read functionality was, since they would just immediately read a web page and star it if it was good. In contrast, P1 thought that the "to-read" functionality was one of the most useful features of the application. P1 cited that the feature allowed her to "cue up what she wanted to do next", effectively creating a future list of information to absorb. In a similar vein, P4 created several subtasks at once, and didn't visit them immediately. This allowed her to "just record all of the things she might need to think about for her trip ahead of time, and then just come back to them later". This prospective task encoding was a unique and unexpected benefit of the ability to structure sets of searches together.

Finally, transforming the search results into a workspace made some participants feel a sense of stability and organization; P5 specifically noted that he "like[d] that the results froze from when you went to them" unlike when you traditionally requery a search. These results suggest that a key benefit of the approach was being able to organize and evolve their mental model through a relatively stable workspace.

Participants were also queried about how the mobile form factor of the application either enhanced or detracted from their experience. All of the users (6) who noted something positive about Bento cited the convenience and portability of the application. For example, P3 noted "The ability to search whenever I wanted to. ie waiting in line for something". Two of the users didn't cite anything positive, saying that they preferred larger screens and physical keyboards.

We also noted a number of challenges that users faced with Bento. Some were relatively straightforward, such as confusion around the progress indicator, leading to redesigns for Study 3. However, some were more substantive issues for the general approach. Some participants found Bento useful for complex searches, but overly high overhead for simple informational searches, suggesting that they would like "being able to toggle on/off the organizing part" or "not hav[ing] to create a new task for simple searches which would not require detail planning and organization". Another common complaint about mobile phone searching more generally was the difficulty of typing, e.g., "typing on a phone screen can be arduous." Exploring the tension between low overhead for simple searches and supporting complex searches – especially when the former can transform into the latter – may be a fruitful area for further research.

Study 3 - Behavioral Traces

Studies 1 and 2 provide converging evidence about the value of Bento's two interfaces of task management and a triage workspace. However, although the field trial in Study 2 provides suggestive evidence and scenarios of how participants used Bento, one weakness is that it relies on self-report data

which could be biased or incomplete. In order to collect richer quantitative data about Bento's usage in the field we conducted a third study in which we instrumented the browser with data collection capabilities and analyzed participants' actual usage data. This also gave us an opportunity to iterate on the design to address the issues discovered in Study 2, e.g., confusion around the progress indicator and lack of support for quick, simple searches.

Again, with the previous study, we performed some modifications to Bento's appearance based on feedback. We focused on improving the readability of the search results and improving the learnability and "first use" experience. We introduced a more coherent first use scenario, adjusted the progress indicators on the search results screen to their final form, and modified task and subtask creation.

Procedure

Study 3 followed the same procedure as Study 2 but with the updated Bento application, with more participants, and for a longer period of time (10-13 days). Utilizing the local participation pool, we recruited 16 participants with ages ranging from 18 to 25. Participants who participated in the previous field study or lab study were not eligible to participate. Five participants identified as male, and 11 as female. Twelve of the participants were undergraduate students, while 4 of them were graduate students. Afterwards participants completed the same interview as the first field study, and completed a slightly extended version of the questionnaire.

Results

Survey results and feedback were overall similar to the first field and the lab study, with significant preference towards Bento for the two questions: "If you wanted to keep searching later, which tool would be better for picking up where you left off?" ($M = 4.44, 95\%\ CI[4.05, 4.83]$) and "Which tool makes your information more organized?" ($M = 4.25, 95\%\ CI[3.89, 4.61]$).

The key research question for Study 3 was quantitatively investigating whether participants were in fact managing complex searches and utilizing the different features of Bento. Each individual created on average 13 tasks (M= 13.06, SD= 8.433) with on average 3 subtasks (M= 3.13, SD= 1.48), suggesting that participants were indeed engaged in complex searches with multiple subtasks. There was high variability between users, with some users having as many as 14 subtasks within one task. Drilling down further, for each subtask participants opened an average of 7.7 pages (M= 7.7, SD=5.41), suggesting that they were engaging in tasks that involved significant exploration. To check this against participants' own perceptions we asked them to classify their searches as either complex or simple when they came back into the lab at the end of the study. Participants classified 35% of their tasks as complex searches, suggesting that they were engaged in complex searches but also using the system for simple searches, addressing an issue raised in Study 2. Participants found value in Bento's organization and resumption capabilities for complex searches including researching "fandom", bus routes, and radiation oncology internships. For example, one participant explicitly mentioned "I learned the sort of tasks that bento is good for – [it] requires several (subtask) searches. for e.g. transferring money internationally there's wire transfers, exchange rates, foreign check processing fees, different charges for diff banks."

The mobile form factor in this longer study offered some surprising and unexpected benefits for the sensemaking process. One user mentioned that the mobile versions of web pages were actually easier to parse: "Many result pages are mobile optimised, such that the content delivered may be more condensed and the design of the webpage more minimalistic." Another user cited a scenario where it is actually impossible to have a laptop – cooking in the kitchen. In this case, her mobile phone is the only tool she can use to perform sensemaking: "When I'm cooking and I have a recipe loaded, I will prop up my phone on the counter. My laptop would take up too much space."

Participants consistently used many of the features of Bento. Individuals reopened subtasks on average about 2.2 times, starred 7.05% of pages visited, marked 5.84% as to read, and trashed 4.24% of results (note only 20 results were loaded at the time of the search). Each individual had approximately 23 sessions over the study period, so about 2 application sessions per day. When asked what feature they liked most, participants mentioned the organization of tasks and subtasks (9); being able to come back to searches (3); marking pages to come back to later (2); the gray background progress bar (2); starring pages (1); and the overall design (1). When asked what they would like to change most there were a large variety of suggestions, most having to do with not having the features of a full search engine like Google they were familiar with (e.g., access to google scholar or images, answering questions directly after a search, having better search results). Two participants mentioned "quick search" as desired, suggesting there may still be a need to support simple searches more easily than in the current approach.

There was high variability around the use of features and types of searches participants engaged in, with some focusing on simpler searches and some more complex ones. To examine whether the type of search affected perceptions of the tool, we correlated the ratio of complex:simple searches with perceptions of the tool from survey responses. Those with a higher number of complex searches:

- Liked Bento better than a mobile web browser ($r(14) = 0.638, p < 0.01$)
- Felt more at peace using Bento ($r(14) = 0.71, p < 0.01$)
- Felt more organized using Bento ($r(14) = 0.55, p < 0.05$)
- Felt more effective with Bento ($r(14) = 0.834, p < 0.01$)
- Wanted to keep Bento on their phones ($r(14) = 0.644, p < 0.01$)
- Felt Bento improved their effectiveness on mobile phones ($r(14) = 0.65, p < 0.01$)

Summary

We performed three studies to evaluate the use and usefulness of Bento in web search. Across a controlled lab study, qualitatively-focused deployment, and quantitatively-focused

deployment, we found evidence that users appreciated both the task based organization interface, as well as the search results workspace interface. Together, these consistently made users feel more organized and feel like they could resume their activities more easily. Based on these findings, we also have a couple of additional takeaways.

Users appeared to use a few strategies, aligned with many goal activation theory approaches [1]. For example, in study 2 P4 noted that she queued up searches for later exploration, largely a prospective planning task. Conversely, we had individuals like P2 and P3 who didn't really understand the to-read feature, another planning tool we had incorporated into our design. This suggests that different user populations might practice different planning techniques for their exploratory searches, and while the structure appeared to be amenable to most of them, having tools for both planning and retrospective recounting could be essential to the design of these systems. This information was similar to what was found in Study 3 with user preference for different features of the system. Most users liked the overall organization for easy re-finding, however some users liked the specific planning features, such as the to-read feature and the grey progress bars. Supporting both of these resumption use cases will be key proceeding forward.

We had several complaints from users about the overhead of Bento for simpler searches. In both studies 2 and 3, individuals noted that they wish they didn't have to make an entire task card for just simple searches. However, in some of our interviews, individuals noted that their simple searches, such as looking up an actor in a movie, often blossomed into more complex searches, such as looking at what else that actor was in, what roles they typically play, etc. Having a low overhead, while also supporting this transition of simple search into complex search was an issue in Bento that was not entirely resolved.

DISCUSSION

We introduced Bento, a novel way to manage complex searching tasks on mobile devices. Bento creates a scaffolded process that separates the task management and workspace functions of tabs into two distinct interfaces. Its two focused interfaces provide users with the necessary affordances to make progress on their sensemaking tasks even within the constraints of mobile devices. This structure is able to meet the complex demands of sensemaking and mobile work, and can be used for later transfer, hand-off, and resumption.

This structure present in Bento provides clear, functional units that can be leveraged in future work. For example, consider the goal of making sensemaking independent of person. The tasks present in Bento could serve as the key unit of collaboration. Since they represent a specific, independent information goal, a task could be shared and worked on by a team of individuals. Subtasks could be delegated out to certain individuals to explore, and because all of the information is tracked, mechanisms such as the star feature could be expended into a voting feature. Similarly, the task could be handed-off to another person. The details about which pages were found important and which queries were used could provide a valuable starting point to another individual researching the same topic [16].

Yet another possible approach could involve crowdsourcing parts of the information collection. The natural task composition and segmentation of searches in Bento could support work in short bursts by crowds. Systems such as the Knowledge Accelerator and Alloy [18, 10] could be hooked into specific subtasks, and provide assistance in performing research on a topic. Additionally, individuals subasks and pages could be "selfsourced" to give users ways to complete tasks in even smaller increments of time [42]. For example having a feed of Bento subtasks that need to be triaged, or pages that need to be parsed could be beneficial to extremely time and context constrained environments.

Along another line, the structure could be used to allow for easy transfer and resumption from other computing systems. Indeed, one participant (P6) noted that he wished there was a web based version. He utilized a number of different devices, many of them not his own (such as those provided by his university). Having this tool available on any platform would let him pickup his searching or find some information that he needed. For example, a similar approach could be instantiated as a virtually identical desktop browser that syncs with the mobile version. However, moving to the desktop may provide other design opportunities given the additional screen real estate; for example, the three level hierarchy (task $>$ subtask $>$ page) of Bento on a smartphone might be flattened to two levels (e.g., task cards and a subtask pane of search results with a reading list, similar to an email application) or even a single level (by incorporating the task card into the view). These changes would keep the integrated tab management and exploratory search support of Bento while being a more efficient way of reading and exploring. Better support for text entry and annotation on desktops could also benefit a future version of Bento on the desktop.

In Bento's current iteration, it primarily only provides support for the "foraging" portion of the sensemaking process [39]. However, there is a large opportunity for also addressing the rest of the sensemaking process, including synthesizing collected information into useful structures. This could include a way to capture information from pages easily on mobile devices, coupled with a system for generating a structure or comparing pages to each other.

Finally, it is possible that using an approach such as Bento may change the strategies that people use in search [32]. For example, although we expected users to add subtasks as they encountered new information, one participant found it useful to do the opposite: "my searches were more focused because i tended to brainstorm at the start of a task and added subtasks at that time". This is an interesting change in search strategy that may have been evoked by the task management interface. On the other hand, the organization of pages into subtasks might make it more difficult to flip back and forth doing comparisons between individual pages, as noted by one participant. Further investigation in a large scale deployment with a diverse user base could help shed light on the advantages and drawbacks of the approach, as well as possibly creating a valuable dataset of exploratory search behavior.

The most recent version of the application is available for download on the Apple App Store: https://itunes.apple.com/us/app/bento-browser/id1101530325?mt=8. Please feel free to download it and try it out for yourself.

REFERENCES

1. Altmann, E. M., and Trafton, J. G. Memory for goals: An activation-based model. *Cognitive science 26*, 1 (2002), 39–83.
2. Anderson, C. R., and Horvitz, E. Web montage: A dynamic personalized start page. In *Proceedings of the 11th international conference on World Wide Web*, ACM (2002), 704–712.
3. Anderson, J. Forrester Market Report: Consumer Behavior Online: A 2009 Deep Dive. http://www.forrester.com/go?docid=54327, 2009. Accessed: 2017-09-10.
4. Anderson, J. R., and Milson, R. Human memory: An adaptive perspective. *Psychological Review 96*, 4 (1989), 703.
5. Berant, J., Chou, A., Frostig, R., and Liang, P. Semantic parsing on freebase from question-answer pairs. In *EMNLP*, vol. 2 (2013), 6.
6. Blanco, R., Cambazoglu, B. B., Mika, P., and Torzec, N. Entity recommendations in web search. In *International Semantic Web Conference*, Springer (2013), 33–48.
7. Bota, H., Zhou, K., and Jose, J. M. Playing your cards right: The effect of entity cards on search behaviour and workload. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, CHIIR '16, ACM (New York, NY, USA, 2016), 131–140.
8. Bota, H., Zhou, K., Jose, J. M., and Lalmas, M. Composite retrieval of heterogeneous web search. In *Proceedings of the 23rd international conference on World wide web*, ACM (2014), 119–130.
9. Chang, J. C., Hahn, N., and Kittur, A. Supporting mobile sensemaking through intentionally uncertain highlighting. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, ACM (New York, NY, USA, 2016), 61–68.
10. Chang, J. C., Kittur, A., and Hahn, N. Alloy: Clustering with crowds and computation. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM (2016), 3180–3191.
11. Cutting, D. R., Karger, D. R., Pedersen, J. O., and Tukey, J. W. Scatter/gather: A cluster-based approach to browsing large document collections. In *Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM (1992), 318–329.
12. Daft, R. L., and Weick, K. E. Toward a model of organizations as interpretation systems. *Academy of management review 9*, 2 (1984), 284–295.
13. Dervin, B. *An overview of sense-making research: Concepts, methods, and results to date*. The Author, 1983.
14. Dischler, J. Building for the next moment. http://adwords.blogspot.com/2015/05/building-for-next-moment.html, 2015.
15. Dubroy, P., and Balakrishnan, R. A study of tabbed browsing among mozilla firefox users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2010), 673–682.
16. Fisher, K., Counts, S., and Kittur, A. Distributed sensemaking: improving sensemaking by leveraging the efforts of previous users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2012), 247–256.
17. Gioia, D. A., and Chittipeddi, K. Sensemaking and sensegiving in strategic change initiation. *Strategic management journal 12*, 6 (1991), 433–448.
18. Hahn, N., Chang, J., Kim, J. E., and Kittur, A. The knowledge accelerator: Big picture thinking in small pieces. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM (2016), 2258–2270.
19. Hearst, M. A., and Degler, D. Sewing the seams of sensemaking: A practical interface for tagging and organizing saved search results. In *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval*, ACM (2013), 4.
20. Henderson Jr, D. A., and Card, S. Rooms: the use of multiple virtual workspaces to reduce space contention in a window-based graphical user interface. *ACM Transactions on Graphics (TOG) 5*, 3 (1986), 211–243.
21. Huang, J., and White, R. W. Parallel browsing behavior on the web. In *Proceedings of the 21st ACM conference on Hypertext and hypermedia*, ACM (2010), 13–18.
22. Kaptelinin, V., and Czerwinski, M. *Beyond the desktop metaphor: designing integrated digital work environments*, vol. 1. The MIT Press, 2007.
23. Kellar, M., Watters, C., and Shepherd, M. A field study characterizing web-based information-seeking tasks. *Journal of the American Society for Information Science and Technology 58*, 7 (2007), 999–1018.
24. Kittur, A., Peters, A. M., Diriye, A., Telang, T., and Bove, M. R. Costs and benefits of structured information foraging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2013), 2989–2998.
25. Klein, G., Moon, B., and Hoffman, R. R. Making sense of sensemaking 1: Alternative perspectives. *IEEE intelligent systems*, 4 (2006), 70–73.
26. Kules, B., and Capra, R. Creating exploratory tasks for a faceted search interface. In *Second Workshop on Human-Computer Interaction (HCIR 2008)* (2008).

27. Lasecki, W. S., Wesley, R., Nichols, J., Kulkarni, A., Allen, J. F., and Bigham, J. P. Chorus: A crowd-powered conversational assistant. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, ACM (New York, NY, USA, 2013), 151–162.
28. Lella, A. Number of mobile-only internet users now exceeds desktop-only in the u.s. https://www.comscore.com/Insights/Blog/Number-of-Mobile-Only-Internet-Users-Now-Exceeds-Desktop-Only-in-the-U.S, 2015.
29. Marchionini, G. Exploratory search: from finding to understanding. *Communications of the ACM 49*, 4 (2006), 41–46.
30. Marshall, C. C., and Shipman III, F. M. Spatial hypertext and the practice of information triage. In *Proceedings of the eighth ACM conference on Hypertext*, ACM (1997), 124–133.
31. Milliken, F. J. Perceiving and interpreting environmental change: An examination of college administrators' interpretation of changing demographics. *Academy of management Journal 33*, 1 (1990), 42–63.
32. Moraveji, N., Russell, D., Bien, J., and Mease, D. Measuring improvement in user search performance resulting from optimal search tips. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, ACM (2011), 355–364.
33. Morris, D., Ringel Morris, M., and Venolia, G. Searchbar: a search-centric web history for task resumption and information re-finding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2008), 1207–1216.
34. Morris, M. R., and Horvitz, E. Searchtogether: an interface for collaborative web search. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*, ACM (2007), 3–12.
35. Oulasvirta, A., and Saariluoma, P. Surviving task interruptions: Investigating the implications of long-term working memory theory. *International Journal of Human-Computer Studies 64*, 10 (2006), 941–961.
36. Pirolli, P., and Card, S. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, vol. 5 (2005), 2–4.
37. Quarteroni, S., and Manandhar, S. Designing an interactive open-domain question answering system. *Natural Language Engineering 15*, 01 (2009), 73–95.
38. Rose, D. E., and Levinson, D. Understanding user goals in web search. In *Proceedings of the 13th international conference on World Wide Web*, ACM (2004), 13–19.
39. Russell, D. M., Stefik, M. J., Pirolli, P., and Card, S. K. The cost structure of sensemaking. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, ACM (1993), 269–276.
40. Shipman III, F. M., Hsieh, H., Maloor, P., and Moore, J. M. The visual knowledge builder: a second generation spatial hypertext. In *Proceedings of the 12th ACM conference on Hypertext and Hypermedia*, ACM (2001), 113–122.
41. Teevan, J., Karlson, A., Amini, S., Brush, A. J. B., and Krumm, J. Understanding the importance of location, time, and people in mobile local search behavior. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, ACM (New York, NY, USA, 2011), 77–80.
42. Teevan, J., Liebling, D. J., and Lasecki, W. S. Selfsourcing personal tasks. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, ACM (2014), 2527–2532.
43. Voida, S., Mynatt, E. D., and Edwards, W. K. Re-framing the desktop interface around the activities of knowledge work. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, ACM (2008), 211–220.
44. Wang, Q., and Chang, H. Multitasking bar: prototype and evaluation of introducing the task concept into a browser. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2010), 103–112.
45. White, R. W., Kules, B., Drucker, S. M., et al. Supporting exploratory search, introduction, special issue, communications of the acm. *Communications of the ACM 49*, 4 (2006), 36–39.
46. White, R. W., and Roth, R. A. Exploratory search: Beyond the query-response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services 1*, 1 (2009), 1–98.
47. Williams, K., Kiseleva, J., Crook, A. C., Zitouni, I., Awadallah, A. H., and Khabsa, M. Detecting good abandonment in mobile search. In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, International World Wide Web Conferences Steering Committee (Republic and Canton of Geneva, Switzerland, 2016), 495–505.