Taylor & Francis
Taylor & Francis Group

International Journal of Human–Computer Interaction

ISSN: 1044-7318 (Print) 1532-7590 (Online) Journal homepage: www.tandfonline.com/journals/hihc20

Exploring the Mental Model of Web Page Scrap: Design Suggestion of an AI-Powered In-Browser Scrap Tool and Its Usability Evaluation

Jeongmin Jo, Sangyeon Kim & Sangwon Lee

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2024.2402119

Taylor & Francis
Taylor & Francis Group

Exploring the Mental Model of Web Page Scrap: Design Suggestion of an AI-Powered In-Browser Scrap Tool and Its Usability Evaluation

Jeongmin Jo[a], Sangyeon Kim[b], and Sangwon Lee[a]

[a]Department of Industrial and Management Engineering, Korea University, Seoul, Republic of Korea; [b]Division of Artificial Intelligence Engineering, Sookmyung Women's University, Seoul, Republic of Korea

ABSTRACT
People use diverse read-it-later tools to save webpages for future uses; however, difficulties in scrap retrieval remain. Our study aims to understand users' mental models of read-it-later tools and suggests a new tool, Read-It-Now, for web scraping. We conducted interviews to identify web scraping habits and developed four main design recommendations for read-it-later tools: tool hierarchy, contextual aids, proactive triggers, and multiple navigation strategies. With a focus on the proactive trigger that facilitates scrap retrieval, Read-It-Now is designed to recommend relevant web scraps based on a comparison of the title of the current browser tab. We tested the design prototype against traditional bookmarks and found that users preferred Read-It-Now because of its simplicity and effectiveness; however, explanations for scrap recommendations could be improved.

KEYWORDS
Personal information management; information scrap; user interface; recommender system

1. Introduction

Owing the deluge of information available on the Internet, users are easily overwhelmed by attempts to store and retrieve information on the Internet (Shen & Prior, 2013). Personal information management (PIM) systems have become increasingly important in this context. The first PIM tool on the Internet was Bookmark—in the earliest web browser NETSCAPE—which a file-system-like feature used for scraping web pages (Bruce et al., 2004; Kaasten & Greenberg, 2001). Modern web browsers have also implemented their own in-browser web scraping features, such as reading lists (Chrome), collections (Microsoft Edge), and favorites (Safari).

As the use of these bookmarks has become more prevalent, researchers have identified the necessity of a user-oriented PIM system that reflects users' thoughts and activities in terms of information scrapping (Bernstein et al., 2008). These studies investigated how people scrap information in everyday life and proposed information scrap life-cycles (Bernstein et al., 2008; D. K. Barreau, 1995, Lin et al., 2004). However, previous studies have been merely focused on the use of in-browser bookmarks, while neglecting different types of PIM tools (Boardman & Sasse, 2004; Jones et al., 2001; Kleek et al., 2011; Lee et al., 2012; Shen & Prior, 2013). People often use out-of-browser web scraping tools in addition to in-browser features. For example, users send e-mails or messages to themselves or use Google Keep, Pocket, and other note-taking tools (e.g., EverNote, OneNote, and Notion). These out-browser uses highlight the need for a comprehensive investigation of both in- and out-browser PIM tool usage.

Regardless of in- or out-of-browser tools, all existing PIM tools aim to support users in saving scraps. These tools adopt the same functional characteristics to store information first and retrieve scraps later, and are referred to as *read-it-later* tools. However, these tools have some limitations in terms of information retrieval and organization (Abrams et al., 1998; Shen & Prior, 2013; Swearngin et al., 2021). For instance, Abrams et al. (1998) found that only 6% of bookmarks were revisited within a month, 37% of users never organized their bookmarks, and 73% experienced difficulties in revisiting them (Shen & Prior, 2013). This implies that existing read-it-later tools impose a heavy burden on users to organize scraps well and consciously memorize where they can find them later (Swearngin et al., 2021).

To overcome the drawbacks of these storage-centric approaches, recent studies have proposed different solutions that require users to actively perform particular tasks when using PIMs such as Scraps (Swearngin et al., 2021), Tabs.do (Chang et al., 2021), Crystalline (Liu et al., 2022), and ForSense (Rachatasumrit et al., 2021). These tools provide contextual information for recalling stored data, suggesting task-centric tab bundles, and helping to organize information for future use. These solutions commonly facilitate user actions for addressing the problems in existing storage-centric approaches. However, previous studies have been focused on only a few problems without considering the user requirements and have neglected various issues in web

CONTACT Sangwon Lee upcircle@korea.ac.kr Department of Industrial and Management Engineering, Korea University, Seoul, Republic of Korea

scraping. This research gap shows that a task-centric tool is required to be designed in accordance with the user's thoughts and activities related to web scraping.

This study proposes a novel task-centric web-scrapping tool, *Read-It-Now*, that fully reflects the user requirements. Therefore, Read-It-Now technically adopts natural language processing to offer proactive scrap recommendations that facilitate user information retrieval. With the objective of designing a user-oriented tool, we conducted a full-fledged interaction design using the following steps: (1) determination of user requirements: we studied the users' web scrapping patterns and constructed a mental model based on exploratory interviews; (2) definition of design strategies: we suggested design concepts for resolving difficulties in information retrieval based on the mental model; (3) design of a prototype: we developed a concrete design prototype based on design recommendations; and (4) prototype evaluation: we evaluated the efficacy and usability of our prototype. Using this design process, we suggested a final version of our design proposal for Read-It-Now. In terms of mental-model development, (1) and (2) were based on the content of our previous semi-archival research (Jo et al., 2023). The overall research process is illustrated in Figure 1.

Our study contributes to the field of PIM by investigating users' mental models for web scraping and broadening our understanding of user interactions with conventional read-it-later tools. Furthermore, we propose a novel web scraping tool, Read-It-Now, that satisfies user requirements using artificial intelligence (AI) techniques such as semantic embedding and similarity search. It successfully improved the overall usability and user satisfaction, thus demonstrating the effectiveness of task-centric, context-aware recommendations over traditional bookmarking methods. These findings demonstrate the potential of AI-powered PIM tools for enhancing information retrieval, thus offering design insights for the development of future user-centric digital organizational tools.

2. Discovering user requirements

To gain realistic user requirements for managing web page scraps across the entire lifecycle, we first conducted formative user research with semi-structured interviews to investigate user thoughts and activities related to web scraping. The interview results were then organized according to the user's mental model. To construct the mental model, we followed Young's (2008) conceptual framework, which consisted of mental space, task towers, tasks, atomic tasks, and supporting quotes (Young, 2008). Using this mental model, we systemically discovered the user requirements for establishing design strategies.

Figure 1. Overall research process.

2.1. Semi-structured interview

2.1.1. Participants

A total of ten individuals (average age = 28.3 years; standard deviation = 2.54; number of females = 9; and number of males = 1) were selected to engage in a semi-structured interview aimed at examining the usage patterns of web scrap tools across different devices and tools. Each interview session lasted approximately 90 min. In-person interviews were conducted with two participants, and remote interviews were conducted with the remaining eight participants. The participants were instructed in advance to bring their personal electronic devices such as personal computers (PCs), tablet PCs, and smartphones. All the participants reached a consensus for capturing and documenting their collaborative screens. The participants were rewarded with a sum of KRW 15,000 in exchange for their involvement.

2.1.2. Procedure

The pre-interview session began with inquiries regarding the participants' general knowledge of the Internet, browsers, devices, and scraping tools. In the first interview phase, participants were instructed to organize the order of importance of the scrap tools in accordance with eight queries (see Table 1). Real-time classification of the tools was performed using Google Slides and screen sharing. The following inquiries were focused on determining the frequency of scrap creation and visits of each tool on the desktop, tablet, and mobile devices. The participants were instructed to articulate their thoughts verbally and justify their decisions while setting the orders.

Table 1. Detailed interview queries in phase 1.

No.	Instruction
1	Please add any missing web page scraping tools that you use, and delete the tools you do not use.
2	Please pick up the PC-related tools from the first response. Kindly let us know which tools are your favorites.
3	Please pick up the mobile/tablet tools from the first response. Kindly let us know which tools are your favorites.
4	Please organize the PC-related tools in the order that corresponds to how often you generate scraps.
5	Please organize the PC-related tools in the order that corresponds to how often you use the tool.
6	Please organize the mobile/tablet tools in the order that corresponds to how often you generate scraps.
7	Please organize the mobile/tablet tools in the order that corresponds to how often you use the tool.
8	Please organize all scrap tools based on the typical proportion of scraps that are revisited.

During the subsequent phase, a semi-structured interview was conducted in accordance with a predetermined set of inquiries and accompanying open-ended discussions. Questions were asked to determine the basic life cycle of information scrap (Bernstein et al., 2008; Lin et al., 2004): collecting, translating, organizing, reusing, and deleting. Moreover, as the CRUD (Create, Read, Update, and Delete) life cycle of online content is widely recognized (Kilov, 1990), we incorporated the deletion stage into our interviews.

2.2. Interview results

2.2.1. Device and tools

2.2.1.1. Internet usage. Participants used computers daily for an average duration of 7.8 h (standard deviation = 3.12), while they accessed the Internet for an average of 6.4 h per day (standard deviation = 2.50). The Internet was predominantly used for work or education, with an average proportion of 75% (standard deviation = 0.16).

2.2.1.2. Device ecosystem. The five individuals possessed a uniform ecosystem throughout their electronic devices, which comprised iOS/MacOS or Windows/Android. The remaining half of the group had disparate ecosystems, including both iOS and Windows/Android platforms.

2.2.1.3. Browser usage. Ten people used Chrome, five Naver Mobile, five Safari, four Naver Whale, one Samsung Internet, and one Microsoft Edge. The participants had an average of 2.75 internet browsers, with a standard deviation of 1.05.

2.2.1.4. Scrap tool usage. The average number of scrap tools used by the participants was 7.6 (standard deviation = 1.42). In particular, all the 10 participants employed both the "Kakao Talk Chat with Myself" and "Notion" applications simultaneously. Table 2 lists the frequency with which the participants used each scrap tool. KakaoTalk is a popular messaging program in South Korea. It is interesting to note that, throughout the conversation, tool-free scraping methods such as "add to home" and "keep internet tabs open" were also brought up.

2.2.2. User mental model

Following the conceptual framework (Young, 2008), the mental model was developed as follows. Initially, the transcripts

Table 2. Numbers of users for each scrap tool, retrieved from our prior research (Jo et al., 2023).

Location	Tool	Respondents	Location	Tool	Respondents
In-Browser	Chrome–Bookmark	9	In-Browser	Safari–Bookmark	1
In-Browser	Safari–Favorites	4	In-Browser	Whale–Bookmark	1
In-Browser	Safari–Reading List	3	In-Browser	Whale–Scrapbook	1
In-Browser	Chrome–Reading List	2	In-Browser	Samsung–Bookmark	1
Out-Browser	KakaoTalk–Chat with Myself	10	Out-Browser	PC Memo Application(iOS)	2
Out-Browser	Notion	10	Out-Browser	Naver Keep	2
Out-Browser	Mobile Memo Application	6	Out-Browser	GoodNotes	1
Out-Browser	PC Sticky Notes	5	Out-Browser	LiquidText	1
Out-Browser	OneNote	1	Out-Browser	Google Drive	1
Tool-Free	Screenshot	7	Tool-Free	Add to home in mobile	2
Tool-Free	Keep tabs opened	3	Tool-Free	Download as html	2

were partitioned into 594 speech segments. We categorized each segment according to the tool names and life stages and identified the atomic tasks. Subsequently, tasks exhibiting similar patterns were divided into groups to form task towers. Mental spaces, which are abstract constructs representing the underlying purpose of users, are ultimately formed through a bottom-up classification. We thus constructed a mental model of scrap in a series of smaller spaces: encounter, create, revisit, reuse, and organize. Because users in the interview viewed "Translate" and "Delete" in terms of organization, these functions were included in the organizing stage. We extracted these key findings from each mental space to obtain adequate design recommendations for read-it-later tools. The complete mental model is presented in Table 3.

3. Defining design strategies

The mental model of read-it-later provides significant insights into what users need for web scraping. Based on these findings, we defined design strategies for creating a conceptual design that satisfies the user requirements. In particular, we first discuss the user requirements from the mental model and then propose four main design recommendations for designing a prototype.

3.1. [D1] considering mental hierarchy of users

3.1.1. Rationale

The key findings are (a), (b), (e), (i), (j), and (l) in Table 3. According to the interview results, users initially build a mental hierarchy regarding the tool to use for creating scraps. Initially, "a temporary warehouse" or "add to cart" is used to depict the area in which the majority of scraps are piled. As the terminology suggests, their primary function is to temporarily store webpages for short-term usage. This first location for a scrap is referred to as "primary storage." Applications such as memos, messengers, and screenshots belong to this category because they are portable, quick, and easily accessible. The user reclassifies the "stacked up" scraps over time and transfers them to an alternative scrap tool that is suited to the role of each scrap and easy to reuse.

This location to which a scrap is moved is referred to as "secondary storage." The purpose of secondary storage is archiving and long-term use. They serve two distinct functions and roles, namely, shortcuts and archives. Favorites and bookmarks are considered "shortcuts" that hold scraps that are frequently accessed. The "Archive" section allows users to organize relevant scraps on a single page and enhance them with contextual and visual information. Individuals may also implement their own systemic structures, such as contextual information, folders, or toggles. “Notion” is an exemplar of this archival secondary storage. In primary storages and secondary "shortcuts," both work-related and personal information is combined. However, work-related information is often transferred to a secondary “archive” because users need to analyze, summarize, and organize the content.

Scrap tools can thus be categorized within the users’ mental hierarchy. Owing to these differences, different tools have different uses and mental models. Therefore, the following design recommendation is made: *(D1) In accordance with its mental hierarchy, the tool design strategy should be distinct.*

3.1.2. Design recommendation

Users have distinct mental models for various read-it-later tools. The design technique that should be emphasized depends on whether the scrap tool is stored in primary or secondary storage. Portability, speed, compatibility, number of access points, and the ability to share with oneself are all desirable benefits of primary and secondary storage (Jones et al., 2001). Secondary storage as an archive should have the following advantageous features: rich contextualization, organization, visualization, representation, and the ability to be shared with others (Swearngin et al., 2021).

3.2. [D2] Encouraging users to create more scraps

3.2.1. Rationale

The relevant key findings are (c), (d), (j), and (k) in Table 3. In the scrap production stage, judgments are made in a moment while taking into consideration a variety of contextual factors. A user should consider not only the tool context, but also the external context of potential revisit scenarios when determining where to store a scrap. The channel factor (Bernstein et al., 2008) states that even seemingly insignificant tasks, such as turning on a computer or looking up a scrap application on the phone, can have a significant positive or negative impact. Consequently, context-supportive features that reduce cognitive costs and serve as facilitators for making a scrap should be included in scrap tools. According to Jones et al. (2001), the context, portability, reminders, and number of access points are important considerations when selecting a scrap tool (Bruce et al., 2004). During the interviews, we discovered that shareability is another important consideration. Users frequently transfer scraps among tools and devices, and the mental model demonstrates how the reuse process frequently involves sharing requirements. Thus, the following design recommendation is made: *(D2) To encourage users to create more scraps, scrap tools should reduce the cognitive load associated with the context-specific judgments of users.*

3.2.2. Design recommendation

Participants shared examples of helpful features that enhanced contextual convenience during their scrap lifetime. Firstly, to enable users to maintain the original workflow as much as possible, a straightforward and quick creation interface without back-and-forth behavior is necessary. At the time of creation, a lightweight organization function is also effective. Secondly, users appreciated the features of automatic preview loading. The application automatically crawls the image, title, and first part of the body paragraph to display them as a preview in the scrap list. Large thumbnails and text previews were particularly popular because they assist memory recall and prevent users from having to create descriptions of the content. Finally, a convenient transfer

Table 3. Mental model and key findings of web page scraping behavior, retrieved from our previous research (Jo et al., 2023).

Mental space	Description	Task tower	Task – *Atomic Task*
Encounter	User finds a new web page	1. Personal recommendation	1.1 In conversation
			1.2 In lectures
			1.3 In online chatting
		2. Web surfing	2.1 Search for Shopping
			2.2 Search for Hobby/Leisure
		3. SNS Following	3.1 Online Community
			3.2 Subscribe to paid contents
		4. Work (Study)	4.1 Research for work
			4.2 Search for arising questions
Key Findings	(a) Contents associated with the workplace were considered to be more labor-intensive, suitable for archival purposes, and compatible with personal computers.		
	(b) Contents associated with the individually captivating material were considered as more transient, unimportant, and appropriate for a mobile device.		
Create	User Creates a Scrap	1. Determine whether to scrap the web page or not	1.1 Read the page content
			– *1.1.1 Decide if the situation is suitable for reading*
			– *1.1.2 Read title only*
			– *1.1.3 Skimming*
			– *1.1.4 Read all*
			1.2 Evaluate its perceived usefulness
			– *1.2.1 Potential future use*
			– *1.2.2 Personal Interest*
			– *1.2.3 General usefulness*
			1.3 Evaluate its importance
			– *1.3.1 Estimated storage duration*
			– *1.3.2 Certainty of return*
			– *1.3.3 Anticipated frequency of use*
		2. Determine where to keep	2.1 Available time remaining for creating
			– *2.1.1 Leave a memo while having a conversation*
			2.2 Ease of use
			– *2.2.1 Prefers familiar tools*
			– *2.2.2 Need for additional information*
			– *2.2.3 Shareability with other devices*
			2.3 Future need of sharing with other people
			– *2.3.1 Need for online publication*
			– *2.3.2 Need for collaboration with others*
			2.4 Context of the web page
			– *2.4.1 Consider not only the page itself, but also the platform to which it belongs*
			– *2.4.2 Consider a device suitable for the web page*
			– *2.4.3 Consider the relationship with other scarps*
Key Findings	(c) Decisions must be made within a restricted timeframe, while considering a multitude of contextual factors.		
	(d) Users prefer not to interrupt the flow of work by generating an information scrap.		
	(e) The majority of scraps are initially stored in easily accessible tools before being reexamined.		
Revisit	Revisit a Scrap	1. Intention to find a scrap arises	1.1 At the start and end of the day
			1.2 New task starts offline
			1.3 Share information in communications
			1.4 Needed for work
		2. Search for the desired scrap	2.1 Keyword search
			2.2 Locational navigation
			– *2.2.1 Read thumbnail, title, and paragraph preview*
			– *2.2.2 Scroll down and check one-by-one*
			2.3 My own organizational hierarchy
			2.4 Guess the purpose of a scrap with a date
			2.5 Lost in scraps while surfing
		3. Explore a tool aimlessly	3.1 Aimlessly open a scrap tool
			– *3.1.1 At the start and end of the day*
			– *3.1.2 On the road*
			– *3.1.3 When bored*
			– *3.1.4 While working (studying)*
			3.2 Evaluate the importance of a scrap again
			3.3 Rediscover scraps that were not known to exist
Key Findings	(f) The majority of retrieval triggers occur when people communicate with one another outside, which makes shareability necessary.		
	(g) Internal triggers rely heavily on the memory of the user.		
	(h) Users' cognitive load from a long list of scraps makes it difficult for them to recall, which encourages locational search.		
Reuse	Reuse the page content	1. Summarize/Organize	1.1 Add contextual and additional memo
			1.2 Beautify relevant arrangements
			1.3 Archive the URL after the arrangement
		2. Share the summarization	2.1 Unspecified disclosure for viewing
			2.2 Share directly with specific people
Key Findings	(i) When the scrap is intended for archival purposes, users want to organize related scraps into a page structure, decorate it, and add contextual information to it.		
	(j) Individuals often want to easily share links with others.		
	(k) Owing to the transmission between tools or devices, compatibility and shareability are crucial.		

(*continued*)

Table 3. Continued.

Mental space	Description	Task tower	Task – *Atomic Task*
Organize	Organize scrap lists	1. Transfer to a more suitable tool	1.1 Move from a temporary tool to a permanent tool 1.2 The scrap is discarded from the temporary tool
		2. Organize for findability	2.1 When creating, classify according to its theme 2.2 Sort hierarchically by topic – *2.2.1 Folder* – *2.2.2 Toggle* – *2.2.3 Database list* 2.3. Change the folder structure when the interests change significantly
		3. Realize effective influence	3.1. Proud, satisfied 3.2. Indebtedness, shame, burden
		4. Delete a scrap	4.1. Discard non-critical scraps 4..2. Leave them till they no longer irritate eyes
Key Findings	(l) Reexamined scrap is transferred to an appropriate tool—such as an archive or shortcut—based on its function and goal. (m) Users may encounter difficulty while trying to classify scraps into a single folder or category. (n) The user frequently overlooks scraps in visual blind zones (the bottom of the list, a folder inside another folder, etc.).		

option is required for further shareability. Sharing a single link is sufficient for primary storage. However, secondary storage as an archive requires the ability to export multiple webpages simultaneously.

3.3. [D3] Encouraging users to revisit more scraps

3.3.1. Rationale

The key findings are (f) and (g) in Table 3. The study of the mental model revealed two problems with scrap retrieval. During the revisit stage, real-world events serve as the primary source of the external recall triggers frequently encountered in interpersonal communication. These triggers rely on coincidences and are uncontrollable. Furthermore, internal stimuli (such as requirements for studying) are extremely memory-dependent. This suggests that scrap is lost and never found again if a user forgets about it. The following design recommendation is made: *(D3) To encourage users to revisit more scraps, it is important to provide proactive reminder triggers that assist with user recall.*

3.3.2. Design recommendation

We assert that proactive triggers are insufficient for contemporary scrap tools. This problem arises from the storage-oriented characteristics of scrap tools. Users are required to go between several windows, remember the position of the scrap, and retrieve it, particularly if the storage site differs from the usage location (Chang et al., 2021; Swearngin et al., 2021). Prior research has suggested task-centric design solutions, such as enhancing the visual prominence of the currently active task or providing context-aware automatic suggestions (Chang et al., 2021). Therefore, it is conceivable that implementing task-centric triggers can reduce the costs associated with information retrieval for users.

3.4. [D4] Increasing the ease of search

3.4.1. Rationale

The key findings are (h), (m), and (n) in Table 3. We discovered challenges in navigating scraps. Users use direct keyword searches if they can recall the exact keywords from scrap titles. In addition, they used their thematic organization system (i.e., folders) or navigated through each item individually. We discovered that a single scrap frequently contains numerous properties in the user's mind, such as "recipe," "to-do," and "Christmas," making it impossible to classify it into a single folder. This is distressing to users and lengthens the time required to organize scraps. In addition to making scraps less accessible, a disorganized list of web page scraps provides visual blind spots (e.g., bottom of the list and folders in folders). Therefore, the following design recommendation is made: *(D4) To increase ease of search, it is advisable to provide a variety of organizing strategies.*

3.4.2. Design recommendation

The implementation of multiple organizational techniques effectively lowers the organizational burden and facilitates seamless navigation. As previously described, while keyword searching and foldering are the most common organizational techniques, they are insufficient. The simultaneous incorporation of multi-class classification and location-based browsing techniques (i.e., visual filtering) into scrap tools enhances the ease of search.

4. Designing a prototype

4.1. Visioning

Based on the above design strategies, we present the main visions with a concrete design for our task-centric webpage scrap tool, Read-It-Now. The primary vision of Read-It-Now involves making it easier for users to revisit as many scraps as possible, which aligns with the concept of primary storage in D1. During the webpage scrap life cycle, our design specifically attempts to realize the following visions:

- [V1] Develop appropriate design techniques for primary storage (as suggested in D1)
- [V2] Enable a task-centric reminder for triggering scrap retrieval effectively (as suggested in D3)

- [V3] Enable simple creation, automatic visualization, and convenient transfer. (as suggested in D2)
- [V4] Enable the use of various organizational methods (as suggested in D4)

Furthermore, we formulated functionalities that aligned with our predetermined design objectives. The functions within the interface are briefly described in Figure 2 and elaborated on in the subsequent section.

4.2. Prototype design and development

4.2.1. [V1] Develop appropriate design techniques for primary storage

The primary design objective of Read-It-Now is V1, and all the other design choices were made with this in consideration. We constructed three basic functions: storing webpage scraps, enabling scrap classification, and providing semantic scrap recommendations. This is because primary storage prioritizes speed over the quality of contextualization. Read-It-Now does not have any beautifying or additive editing features. The most important feature is the proactive task-centric trigger, which is provided using semantic recommendations between titles. Other user-interface elements were also developed to align with implications from the mental model.

4.2.2. [V2] Enable task-centric reminder for effectively triggering scrap retrieval

Read-It-Now is a web extension that functions as an in-browser application that utilizes extension application programming interfaces (APIs) to seamlessly interact with web browsers. The primary purpose is to identify the user's active tab and offer relevant recommendations for scrap retrieval. The extension operates in the browser's side panel (see Figure 2-a), conveniently situated within the same window as the user's workspace. This arrangement eliminates the need for users to move back and forth between windows to obtain the relevant information. In addition, users can effortlessly recall fragments with the aid of semantic recommendations. Owing to these capabilities, Read-It-Now is a task-centric scrap tool that overcomes the drawbacks of storage-centric tools.

The initial step for developing the prototype involved gathering a set of sample scraps. To include a variety of user interests, we randomly selected 200 websites from 10 media blogs that cover different topics such as Hobbies, Style, News, Technology, and Trivia (Geikhman, 2023). The provided sample scrap file underwent a conversion process to a .CSV file, which involved organizing the data into two distinct columns: one for the title and the other for the corresponding URL. To develop a semantic recommender system, it is crucial to perform preprocessing and embed all the

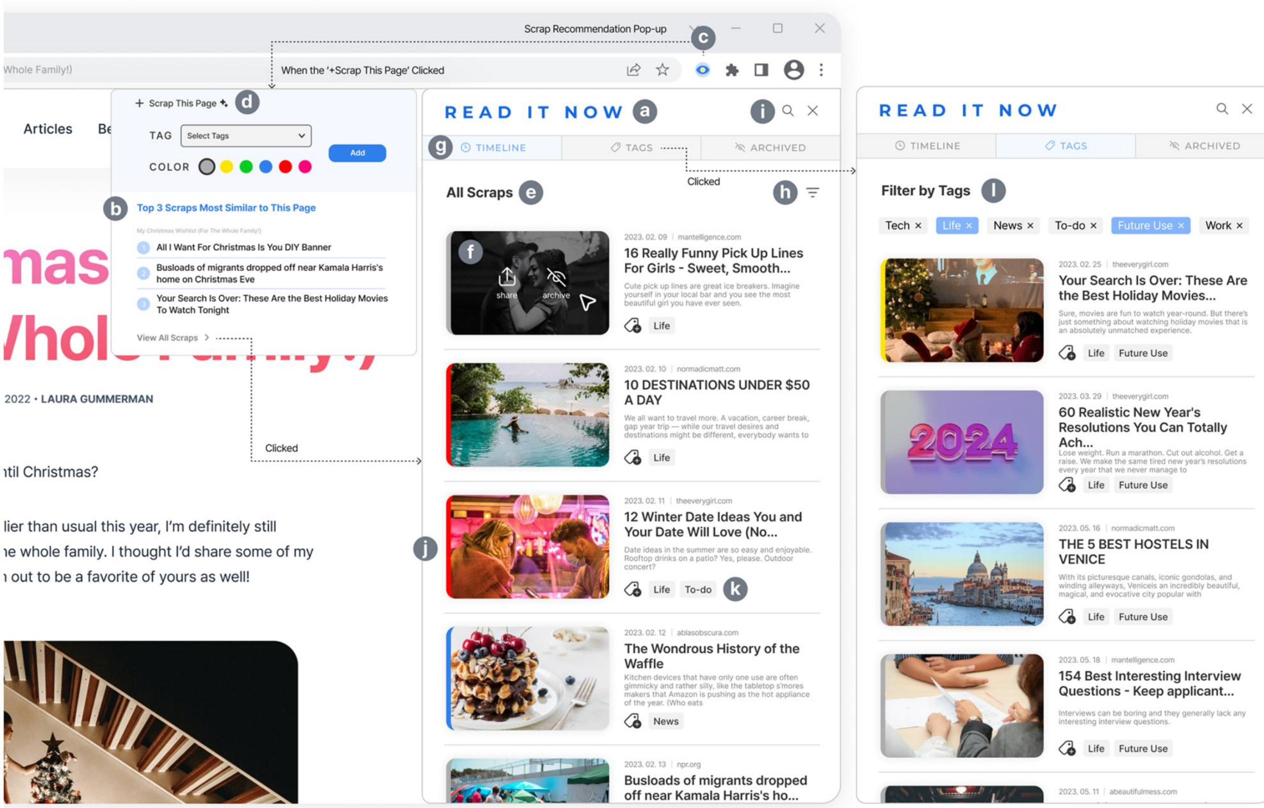

Figure 2. Overview of the Read-It-Now interface design. The functions for realizing V2 are the use of a side panel in the browser (a) and integrating semantic scrap recommendations as a task-centric retrieval trigger (b). The functions for realizing V3 are simple and speedy creation (c) and (d), automatic preview loading (e), and convenient transfer (f). The functions for realizing V4 are multiple navigation strategies (g)–(k), especially including the tags view (l). All the thumbnail images are derived from Unsplash.

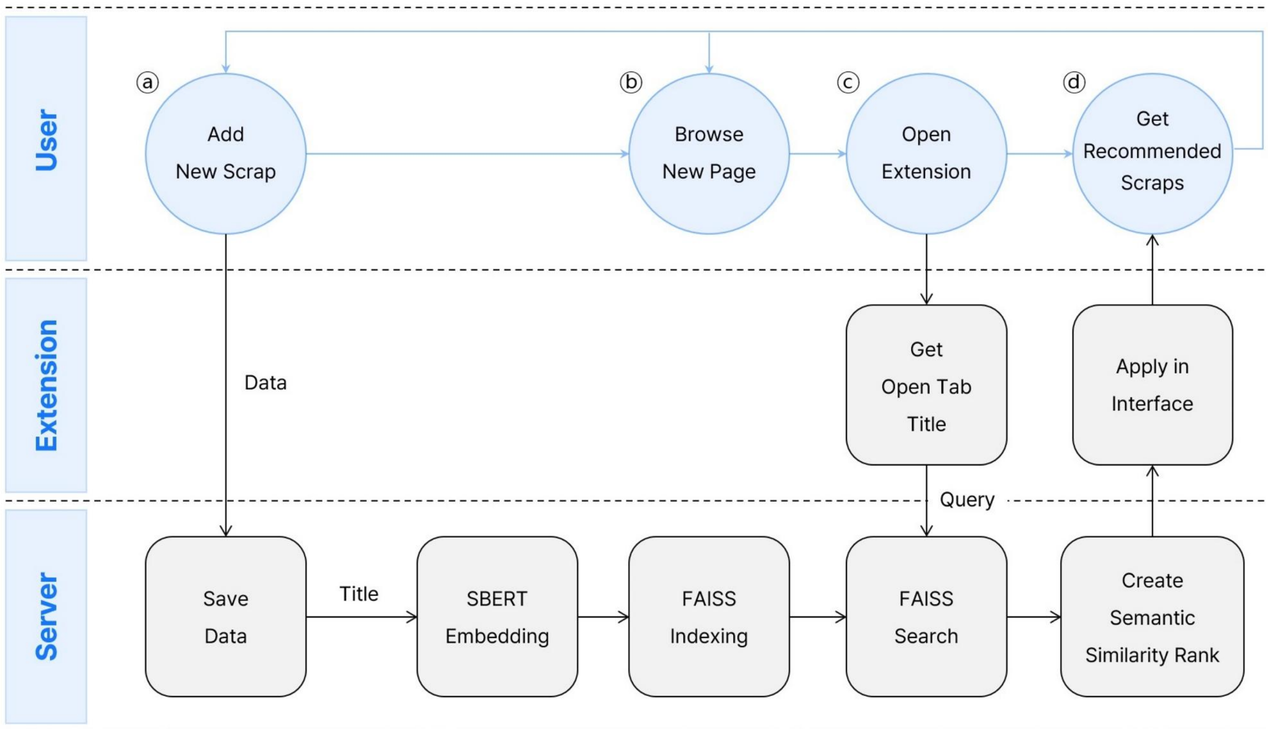

Figure 3. Workflow of read-it-now.

scrap titles into semantic vectors, which allows each sentence to be positioned inside the vector space. The proximity of two sentences in the semantic vector space reflects their degree of similarity. We adopted SBERT (Sentence-BERT), a contemporary pre-trained transformer network designed for generating semantic sentence embeddings (Reimers & Gurevych, 2019). An embedding process was employed, followed by the use of the FAISS library developed by Facebook AI Research for conducting semantic searches (Facebookresearch, 2017).

Figure 3 provides a basic overview of the system workflow. When a user adds a new webpage to Read-It-Now (Figure 3-a), SBERT converts the title of the saved scrap to a semantic sentence vector, and FAISS indexes it for ease of finding it in the future. The user continues to browse the Internet (Figure 3-b) and discovers a page that piques their interest and makes them want to learn more. Next, the user can launch the Read-It-Now extension (Figure 3-c) and enable it to identify the tab that is currently open in the browser. Figure 3-d shows how the title queries FAISS semantic search and retrieves similarity rankings for relevant scraps.

For example, when we entered "15 Pumpkin Desserts To Bake This Fall" as the title query, the following were the top-ranked scrap titles: The best match went to "8 Hosting Hacks That Will Take Your Halloween Party to the Next Level," the second best match was “40 Easy 30-Minute Dinner Ideas,” and “Busloads of migrants dropped off near Kamala Harris’s home on Christmas Eve” was the third best match.

Using another prototype test result, we produced an example of the interface design (see Figure 2). When the icon is clicked, a pop-up window is displayed with the section “Top 3 Scraps Most Similar to This Page” highlighted (Figure 2-b). The interface then analyzes the semantic similarity between the open-tab title and the list of scrap titles to provide adequate scrap recommendations for the context.

4.2.3. [V3] Enable simple creation, automatic visualization, and convenient transfer

Read-It-Now can be accessed quickly through the fav icon of the browser toolbar. The app icon (Figure 2-c) is a single-click button that adds a page as a new scrap entry. The color scheme and scrap category can be easily selected by users in a pop-up box (Figure 2-d). The default settings made it possible to bypass this process. When the scrap is created, an addition is made to the "All scraps" list (Figure 2-e). The scrap list is presented chronologically by default. When adding a new scrap to the list, Read-It-Now automatically scans the page’s HTML document to extract the image thumbnail, title, URL, and first paragraph to produce a preview. Because thumbnails are the most popular recall suggestion, they are arranged to occupy relatively large spaces. Hovering over the thumbnail causes the "Share" and "Archive" buttons to appear (Figure 2-f). Clicking the "Share" button copies the scrap’s URL link to the clipboard instantly, making transfer easy.

4.2.4. [V4] Enable use of various organizational methods

The tab bar, as depicted in Figure 2-g, exhibits three discrete categories for organizing scraps. At first, "Timeline" accumulates scraps to make it easy for visitors to search for what they are looking for by scrolling down the list by date.

Location-based navigation is a widely adopted design approach for primary storage tools. For convenience, Read-It-Now supports several other navigational techniques. Several sorting criteria can be used, including color priorities and the oldest and latest criteria (Figure 2-h). A keyword search is also available (see Figure 2-i). Each scrap may exhibit a single color scheme (see Figure 2-j) and contain different categories representing multiple thematic hierarchies (see Figure 2-k). Users can filter content by tags using the "Tags" menu located in the tab bar, as shown in Figure 2-l. Finally, the "Archive" menu is designated for the storage of discarded scraps. Sometimes, users save scraps that they might come across again "just in case" and find it difficult to delete them. By enabling an archiving option, Read-It-Now helps users avoid visual blind spots and stress by relieving them of the burden of permanent deletions.

5. Prototype evaluation

Based on the suggested webpage scrap interface, we conducted a user evaluation study to examine the effectiveness and usability of the interface. In particular, we aimed to investigate how the Read-It-Now interface enhances user behavior in scrap retrieval compared to conventional webpage scrap interfaces. Accordingly, we designed an experiment using a short interview to validate the effectiveness and usability of our design recommendations. In the evaluation study, we endeavored to elucidate how Read-It-Now compares to traditional scrap interfaces, how users engage with the retrieval trigger, and how the scrap recommendation of Read-It-Now can be improved.

5.1. Method

5.1.1. Participants

The experiment included tasks that involved reading and writing in English, thus requiring the participants to have a certain level of language proficiency. To ensure this, potential participants were invited to describe their experience in English through a screening survey. We specifically sought individuals with the Test of English for International Communication (TOEIC) score of at least 700 or an equivalent competency in English. We finally recruited 32 participants (average age = 28 years, standard deviation = 3.10; number of females =19; number of males = 14) and considered their English proficiency to ensure a balanced distribution of the experimental conditions. The interviewees were compensated with KRW 12,000 for their participation.

5.1.2. Experimental setup

We created the following test environment to evaluate the Read-It-Now interface. All experimental materials were designed using the software Figma. For this study, we constructed a fictitious website named "Everything on Earth" and populated it with 80 articles. The articles were selected from ten famous websites that publish online articles on lifestyle (A Beautiful Mess, The Every girl, Life Hacker, Mantelligence, and Be More with Less) and technology (Tech Crunch, The Verge, Techradar, CNET, and Digital Trends). We randomly selected eight articles from each website. Next, the Read-It-Now tool was preloaded using 106 diverse web page scraps. In particular, 106 scraps were randomly selected from the existing list used to develop the prototype, and six additional articles were selected for the experimental task. In summary, we ensured that scrap recommendations were available on all 186 web pages.

Two distinct experimental conditions were established for comparison. The Read-It-Now interface we proposed previously was termed as the recommended design group, and for the baseline, the comparative condition was termed as the comparison group. Figure 4 presents the difference between the recommended design and comparison groups.

The comparison group was deliberately designed to resemble the most conventional scrap tool, a bookmark. No scrap-recommendation features were provided for the comparison group. When users clicked the extension icon, they could see a "Scrap This Page" button, excluding recommendations, and were able to create a scrap by assigning it to a single folder instead of multiple tags. In the recommended design group, the scraps contained one or two tags relevant to their topic. The folders were configured to match the first tag of each scrap sample in the experimental group. The scrap list was also listed in the folder in the order in which it was created, which is similar to the conventional bookmark interface. Sixteen participants were allocated to the recommended design and comparison groups each.

5.1.3. Task and procedure

The experiment was conducted in the following sequence: a tutorial session, experimental session, post-hoc survey, and short interview. Initially, in the tutorial session, the participants were introduced to the experimental environment and engaged in two simple activities to familiarize themselves with the scrap tool. First, they were asked to create a scrap from the "Everything on Earth" website. Owing to the limitations of the Figma tool, it was impossible to add new scrap data; therefore, participants were informed that no new scraps would be created during the experiment. The second tutorial activity involved quickly browsing through the list of scraps in the Read-It-Now tool, selecting the three web pages they liked, clicking on them, and then investigating which articles were recommended by the scrap recommendation feature. The comparison group was instructed to check the three web pages they liked while browsing the scrap list. This skimming activity was intended to implant a partial memory of the scrap list in the participants, thereby simulating the perception of "scraps that were previously saved but forgotten."

Subsequently, the participants were allocated to either the recommended Design or comparison groups. In the experimental session, they were instructed to use the Read-It-Now scrap tool on the "Everything on Earth" website to perform two distinct tasks. The reason for implementing the two tasks was to allow for a rich experience of the potential uses of Read-It-Now through tasks of varying contexts and levels

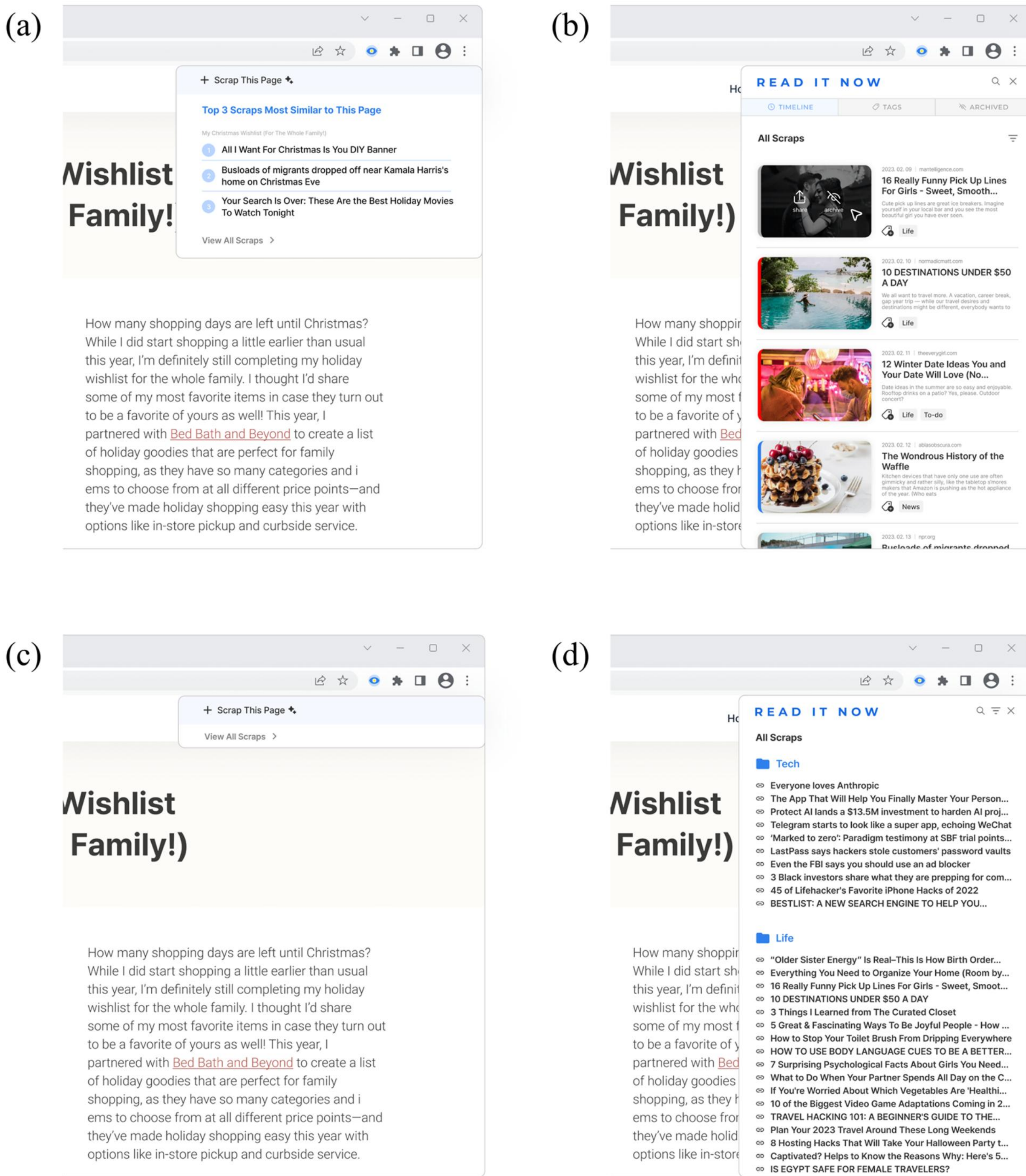

Figure 4. Difference between the Read-It-Now and baseline. (a) Recommended design group: pop-up, (b) recommended design group: all scraps (timeline view), (c) comparison group: pop-up, (d) comparison group: all scraps (timeline view with folders). All the thumbnail images are obtained from Unsplash.

of difficulty. Task A is as follows: "You are a parent with two young children and a popular daily blogger. Please write a blog post about how you plan to create a wonderful home party with your family this Christmas." Task B was as follows: "You work at a venture capital firm. Please write an introduction about the American IT start-up 'Instacart' to report to your boss." The participants were required to write a minimum of 450 characters within a maximum time limit of 20 min. For each task, a total of six highly relevant scraps were pre-included to provide necessary information for task completion: three scraps on "Everything on Earth" and three in the scrap tool. The order of the two tasks was counterbalanced. This means that participants were allocated to one of the following four segments: A-1 (recommended design group: Task A-Task B), A-2 (recommended design group: Task B-Task A), B-1 (Comparison Group: Task A-Task B), or B-2 (comparison group: Task B-Task A).

5.1.4. Measurements

After completing both tasks, the participants participated in a post-hoc survey session. In this session, participants were asked to respond to two different questionnaires to evaluate their experiences using the Read-It-Now scrap tool. The first questionnaire used was the system usability scale (SUS), which is a convenient but reliable tool for quantitatively measuring a product's usability (Brooke, 2013). The SUS consists of 10 question items rated on a five-point Likert scale that ranges from "strongly disagree" (0) to "strongly agree" (4). With these responses to the 10 questions, the SUS score for each participant was calculated as follows. To normalize the scores in terms of percentile ranking, the responses were first added and then multiplied by 2.5 for normalizing the scores in terms of percentile ranking. Thus, we finally obtained SUS scores of the participants ranging from 0 to 100.

Furthermore, the recommended design group responded to the Recommender Systems' Quality of User Experience (ResQue) questionnaire to assess the effectiveness and usability of the scrap recommendation feature (Pu et al., 2011). We modified the original ResQue survey items to suit the objectives of our study. For example, items for which our tool did not have a corresponding feature were excluded. The selected items are listed in Table 4.

It has been suggested that the ResQue questionnaire items are sufficient, with only one question per construct, and that the number of questions can be adjusted according to the context and objectives of the researcher (Pu et al., 2011). Therefore, to reduce participant fatigue, we used fewer sub-questions for constructs that were less relevant to the context of our evaluation, reducing the original list of thirty-two questions to 19. All items in the ResQue survey were rated using a seven-point Likert scale, ranging from "strongly disagree" (1) to "strongly agree" (7). This approach was employed to provide a wider range of response options, considering that users may be unfamiliar with scrap recommendation features. The expanded scale facilitated a more accurate reflection of the subtle differences in user responses. After the completion of all the surveys, a brief interview lasting approximately 10 min was conducted with the experiment conductor, wherein participants discussed their method of using the scrap tool and the subjective pros and cons they experienced.

5.2. Results

5.2.1. Perceived usability–SUS scores

All responses were carefully checked for quality, and one outlier in the recommended design group was excluded due to inconsistent responses obtained for the positive and negative items in the SUS. Figure 5 illustrates the SUS scores for Read-It-Now and the baseline. The recommended design group, consisting of 15 participants, had a mean score of 81.5, with a median score of 82.5. The standard error of the mean for this group was calculated as 2.94, and the standard deviation was 11.4. In contrast, the comparison group, with 16 participants, had a lower mean score of 75.0 and a median of 73.8. The standard error of the mean for this group was slightly higher at 3.39, with a standard deviation of 13.6. The average SUS score in the experimental group was 6.5 points higher. According to the normality test, the SUS scores in the cases of both Read-It-Now and the baseline did not significantly deviate from a normal distribution ($p > 0.05$).

A statistical comparison of the SUS scores between the recommended design and comparison groups was conducted. A student's t-test was conducted, yielding a t-value of 1.44 with 29 degrees of freedom. The significance level (p-value) was found to be 0.08. The mean difference

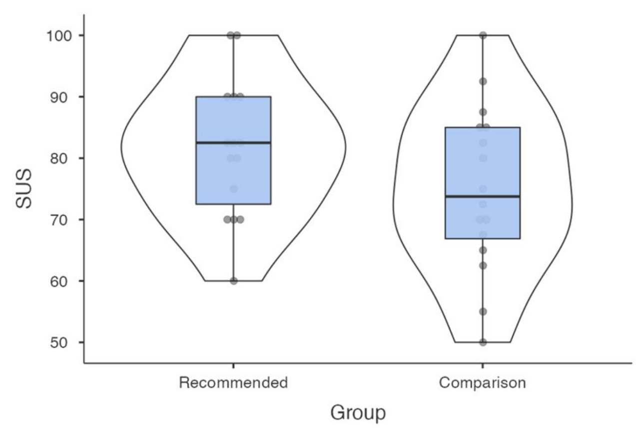

Figure 5. Box and violin plot depicting the SUS results for each group.

Table 4. Shortened list of ResQue questions derived from a previous study (Pu et al., 2011).

Question number	Constructs	Question
Q1	Accuracy	The items recommended to me matched my interests.
Q2	Novelty	The recommender system helped me discover new items.
Q3	Diversity	The items recommended to me are diverse.
Q4	Interface Adequacy	The layout of the recommender interface is adequate.
Q5	Explanation	The recommender explains why the items are recommended to me.
Q6	Information Sufficiency	The information provided for the recommended items is sufficient for me to make a selection.
Q7	Perceived Ease of Use	I became familiar with the recommender system very quickly.
Q8	Control	I feel in control of modifying my taste profile.
Q9	Transparency	I understood why the items were recommended to me.
Q10	Perceived Usefulness	The recommender helped me find the ideal item.
Q11	Perceived Usefulness	Using the recommender to find what I like is easy.
Q12	Perceived Usefulness	The recommender gave me good suggestions.
Q13	Overall Satisfaction	Overall, I am satisfied with the recommender.
Q14	Confidence & Trust	I am convinced of the items recommended to me.
Q15	Confidence & Trust	I am confident I will like the items recommended to me.
Q16	Confidence & Trust	The recommender made me more confident about my decision.
Q17	Confidence & Trust	The recommender can be trusted.
Q18	Use Intentions	I will use this recommender again.
Q19	Purchase Intention	I will tell my friends about this recommender.

between the groups was 6.50, with a standard error of the difference calculated at 4.51. Furthermore, the effect size, measured using Cohen's d, was determined as 0.518, indicating a moderate effect (Figure 6). The result indicates that the average SUS score of the recommended design group was statistically significantly higher than that of the comparison group at a significance level of 0.1.

5.2.2. Recommender system UX–ResQue scores

The responses to each construct of the ResQue survey are listed in Table 5. The Cronbach's alpha value obtained in this study was 0.914, indicating a high level of internal consistency among the survey items and thus confirming the reliability of the results.

Given that a score of 4 represents a neutral response on a seven-point Likert scale, an analysis of constructs that exhibited notably higher or lower average scores facilitates an understanding of user perceptions towards the Read-It-Now scrap recommendation features. The interface adequacy and perceived ease of use received the highest average scores of 5.38. The novelty and diversity were also appreciated, with average scores of 5.25. The perceived usefulness was also positively rated with a score of 4.96. However, the lowest score was for explanation, at 2.81. The control and information sufficiency had scores of 3.69 and 3.88, respectively.

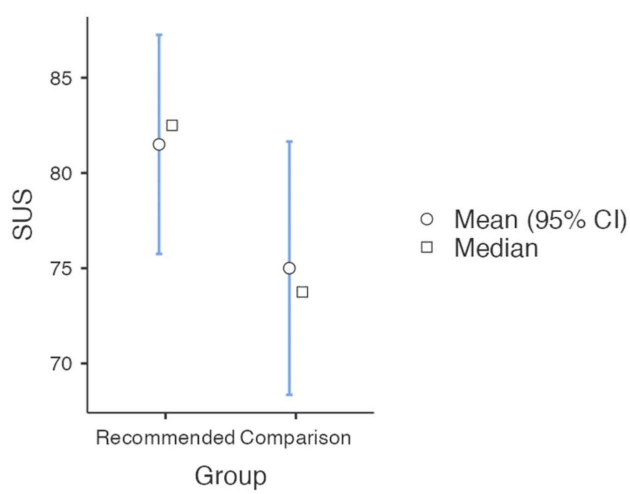

Figure 6. T-test result for SUS.

Table 5. Mean and standard deviation of each construct.

Construct	Mean	Standard deviation
Accuracy	4.38	1.31
Novelty	5.25	0.856
Diversity	5.25	1.44
Interface Adequacy	5.38	1.31
Explanation	2.81	1.68
Information Sufficiency	3.88	1.31
Perceived Ease of Use	5.38	0.957
Control	3.69	1.54
Transparency	3.88	1.45
Perceived Usefulness	4.96	1.27
Overall Satisfaction	4.50	1.15
Confidence & Trust	4.39	1.08
Use Intentions	4.75	1.61
Purchase Intention	4.94	1.44
Total	4.53	0.756

6. Discussion

In this study, we developed a mental model of users' webpage scraping behavior and derived design implications for read-it-later tools based on this. Subsequently, we proposed a scrap tool, Read-It-Now, designed to assist users in revisiting their scraps. Read-It-Now has a task-centric characteristic with a scrap recommendation feature that proactively suggests the most relevant scraps by detecting the user's current context. The evaluation of the proposed design from a user perspective in terms of usability and satisfaction allowed us to obtain insights into how users interact with the AI-powered personal information management tool. In this section, we intend to convey interesting findings from our experiments and interviews, categorized into the following three main themes: How does Read-It-Now compare to the traditional scrap interface? How do users interact with the retrieval trigger of Read-It-Now? How can scrap recommendations be improved?

6.1. Comparison of Read-It-Now to traditional scrap interface

6.1.1. Overall usability

The comparison group received a SUS score of 75.0, corresponding to a B + grade on the Sauro–Lewis curved grading scale (Lewis & Sauro, 2018). The recommended design score was 81.5, corresponding to an A grade. Considering that an A-grade (approximate score of 80) is generally a reasonable SUS benchmark for most applications (Lewis & Sauro, 2018), this may indicate the superiority of the Read-It-Now interface. However, the results of the statistical comparison indicated that the differences between the recommended and the comparison groups were marginally significant. This could be derived from variations in individual scraping usage patterns and preferences due to small sample sizes. Furthermore, because the comparison group's interface was conventional, this could have resulted in higher SUS scores owing to familiarity. Regarding this issue, one participant responded that "I can't see how this is different from typical bookmarks. [B-2-7]," and another "I'm not digitally savvy, so I'm not sure if I have used it properly and effectively. [A-1-4]." In contrast, although some participants found Read-It-Now unfamiliar, they reported a higher SUS score than the comparison group, suggesting that the proposed interface is more user-friendly.

6.1.2. Retrieval support

We observed that the users' navigation strategies for scrap retrieval differed between the two interfaces. First, in the comparison group, participants relied heavily on folders, exhibiting folder bias, which means that they did not read folders that seemed irrelevant. They predominantly adopted location-based browsing (Barreau & Nardi, 1995), and skimmed scrap titles individually from the top of the list. Interviews revealed that many found the folders well organized yet expressed a desire for more flexibility in categorization, e.g., assigning multiple categories to a single scrap using subfolders or tags. Considering that some interviewees responded that they usually do not organize scraps at all as well as the general difficulties people face in folder organization (Boardman & Sasse, 2004; Swearngin et al., 2021), the efficiency of revisiting traditional interfaces could be significantly dependent on how users organize scraps manually.

In the recommended design group, participants primarily used the scrap recommendation feature because they expected it to be relatively efficient than location-based retrieval with tags serving as secondary tools. For example, one participant responded that "to trigger the recommendation algorithm, I initially searched for a Christmas-related article, but going back and scanning through everything seemed inefficient. So I tried to make the most of the recommendations I got [A-2-1]." Several participants highlighted that tags were not sufficiently prominent, and some even relied solely on scrap recommendations and location-based browsing without using tags. Considering that tags correspond to folders in the comparison group, this negligence of the tags leaves room for improvement. The spatial separation of the default scrap list and tag-based categorization into different tabs may have caused the participants to forget about the tag feature, which indicates a need to enhance its salience.

Other feedback regarding the features designed for scrap navigation included the following aspects: for location-based browsing, it was preferred that the scrap titles be fully visible and not truncated. The preview text was rarely used by the participants. Thumbnails were found to be effective in attracting interest. There were mixed opinions on color coding, with some finding it intuitive, and others finding it difficult to perceive. There was a misunderstanding among the three participants, who believed that the color assigned during creation was directly linked to the tags. Overall, the participants appreciated the presentation of a large amount of contextual information, but this also resulted in longer scrolling times, which made it challenging to grasp everything at a glance. Some participants expressed a desire for a more compact layout, suggesting the possibility of reducing the size of the thumbnails to achieve this.

6.2. How users interact with retrieval trigger of "Read-It-Now"

6.2.1. User engagement

A consistent behavioral pattern emerged in the recommended design group, wherein participants used the scrap recommendation feature to obtain the necessary information. Initially, users accessed articles that, although not exactly what they were looking for, were predicted by the recommendation algorithm to obtain the desired content. They would then surf forward through the recommended articles, continuously moving from one article to another. If the flow diverged too far from the original topic, they would stop and revert to the first step, repeating the process. This strategy differed significantly from that of the comparison group users, which primarily involved location-based skimming through specific folders. The majority of participants believed this approach to be more time efficient than directly browsing through a scrap list. This point might be a prime reason for the superiority of the recommended group in the usability testing. Participants also found enjoyment in the surfing process, which encouraged continuous viewing. For example, one participant responded that "if it really targets my interests and preferences, then I think I would just mindlessly browse through it, much like how I get carried away by the YouTube algorithm [A-1-8]."

Furthermore, the behavior of checking scraps unrelated to a given task was observed exclusively in the recommended design group. This occurred for reasons such as "I checked why it was recommended despite seeming unrelated or a new interest was evoked while reading the title [A-2-4]" when viewing the scrap recommendation interface or because of the "eye-catching thumbnails [A-1-6]." Such entertaining scrap surfing and unpredicted revisitation suggest that, while scrap recommendation features were initially designed to function as task-centric triggers, they can also inadvertently serve as triggers for serendipitous scrap retrieval, providing utility beyond the task at hand.

6.2.2. Differences by task

Although the perceived effectiveness was positively rated in the quantitative evaluation, interestingly, the interview results demonstrated that it varied depending on the type of task. Users predominantly found the recommendation more effective in the Christmas task as a task-centric trigger. This difference can be attributed to the nature of the tasks. In the case of highly specific topics, such as company research, users tended to prefer direct keyword searches, thus diminishing the perceived need for scrap recommendations. In contrast, the Christmas task allowed for more freedom in content organization, causing participants to explore a wider range of topics aligned with their own interests. Consequently, the diversity offered by scrap recommendations in the Christmas task facilitated more efficient exploration than direct search because the users' memories were incomplete. Participants believed that using the recommendation system would reduce "the cognitive effort of thinking about search terms [A-1-6]" and help in exploring a wider variety of content as opposed to arriving at "limited scrap search keywords on their own [A-2-5]." In summary, the scrap recommendation feature could be preferred in the case of topics wherein relevance in various aspects is more important than accuracy.

6.3. Improving scrap recommendations

6.3.1. More diverse recommendation rationales

As observed earlier, the participants noted that the Christmas task yielded a superior recommendation quality than the Instacart task. This is inferred to be due to the nature of the natural language embedding process, where more commonly used terms tend to reflect relationships with other words more accurately. Participants felt that the recommendation performance was poor when "the results differed from my predictions [A-1-2]," which mainly occurred when the type (e.g., news or essay) or role of the scrap varied. They expressed concern that repeated recommendations for seemingly unrelated items could diminish their trust in the system.

Therefore, diversifying scrap recommendation methods beyond semantic similarity, which are limited by the embedding of constraints in their nature, is necessary to prevent the lowering of trust. The specific suggestions the participants mentioned include diverse data sources (e.g., search history, reading time, content creator, word frequency, keywords, usage pattern, and social proof) and interactive user interfaces (e.g., allowing users to select from a few recommended keywords).

6.3.2. Enhanced explanations

While the ResQue score indicated the "control" to be low, the recommendation result was designed to be modifiable based on the scraps users capture. Therefore, these response results may have been obtained owing to the participants' lack of understanding of the actual working mechanism of the recommendation algorithm. The quantitative results indicated that participants gave the lowest score in terms of explanation among the ResQue items. Users just speculated that the recommendations were based on various criteria such as recent history, article keywords, and tags set by the user.

This misunderstanding could stem from a lack of explainability. In the interviews, two types of information needs emerged that could enhance explainability. One is an explanation of the black box in the recommendation algorithm. On this issue, a participant mentioned "there was a guidance message stating that the more pages you scrap, the more accurate the recommendations might become, which was somewhat helpful in encouraging more use of this service [A-2-5],"

The second need pertains to the capacity to articulate the underlying rationale and reliability of each recommended item. Another participant responded that "it would be beneficial to know the reason behind the recommendations. I was annoyed by a seemingly irrelevant article - about Venice - constantly showing up, but if it was explained as 'The city most desired by people for this Christmas,' focusing on the relevance to Christmas, it would be more convincing. However, in its current state, it seems too disconnected, making it hard to accept [A-2-4]."

6.3.3. Information sufficiency

Regarding information representation in the pop-up window for scrap recommendations, the following comments were made. Firstly, the majority opinion was that the current number of three recommended items was the most appropriate. The scrap list remains whenever the user returns, but in the case of the recommendation interface, one might not return after selecting the next article and engaging in a surfing pattern. This seems to create a cognitive burden when deciding what to select. This burden may increase if there are more choices, thus making decision-making difficult. For example, one participant said that "The number of recommendations should ideally be between three to five. Even if the number increases from five, I can still only select one item to view. If it keeps adding more, it would be no different from the 'All scraps' list. So, I think providing a moderate number of options, rather than too many, would help reduce the time spent on making a choice [A-2-6]."

However, some participants expressed a desire for an option such as a "See More" button to selectively display more items when they were satisfied with the scrap recommendations, but they agreed that the maximum number of items should be five.

The majority of the participants responded that the title alone was sufficient for selecting the most likely article among the recommendations. Additional information such as thumbnails, preview text, and the date of scrap creation seemed unnecessary and potentially distracting. However, there were a few who desired thumbnails for the visual information they were provided (e.g., website favicon, company logo) that could not be obtained from the text information alone.

6.4. Updated design proposal

In this research, we first derived design recommendations for read-it-later tools based on a mental model and accordingly proposed a new in-browser web page scrap tool, Read-It-Now. Finally, we propose an updated version of Read-It-Now based on insights derived from the evaluation results. The current recommendation algorithm was unchanged to reflect the identified improvements as much as possible solely through changes in the interface design. These changes can be summarized in terms of three fundamental aspects: recommendations, representations, and organizations. Figure 7 presents an overview of the updated interface.

- ***Scrap Recommendation:*** To enhance the transparency of the recommendation algorithm, the ambiguous label "Top 3 scraps most similar to this page" was changed to a more concrete explanation, "Top 3 scrap titles most similar to the current page: (title information)." Furthermore, we attempted to display a confidence level calculated from the similarity scores such that it could assist users in calibrating their trust and inferring the recommendation rationale by themselves (Kim et al., 2023).
- ***Scrap Representation:*** To increase the number of scraps visible at once, the size of the thumbnails was reduced, the preview texts were eliminated, and the font size of the scrap title was decreased. The color schema for each scrap was modified to be more noticeable, thus facilitating more intuitive location-based browsing.
- ***Organization:*** The separation between the "Timeline" and "Tags" tabs was eliminated, and the "Tags" view was set as the default. To prevent confusion owing to the numerous color options, the number of color schemas was reduced to four common colors that intuitively represented the order of priority.

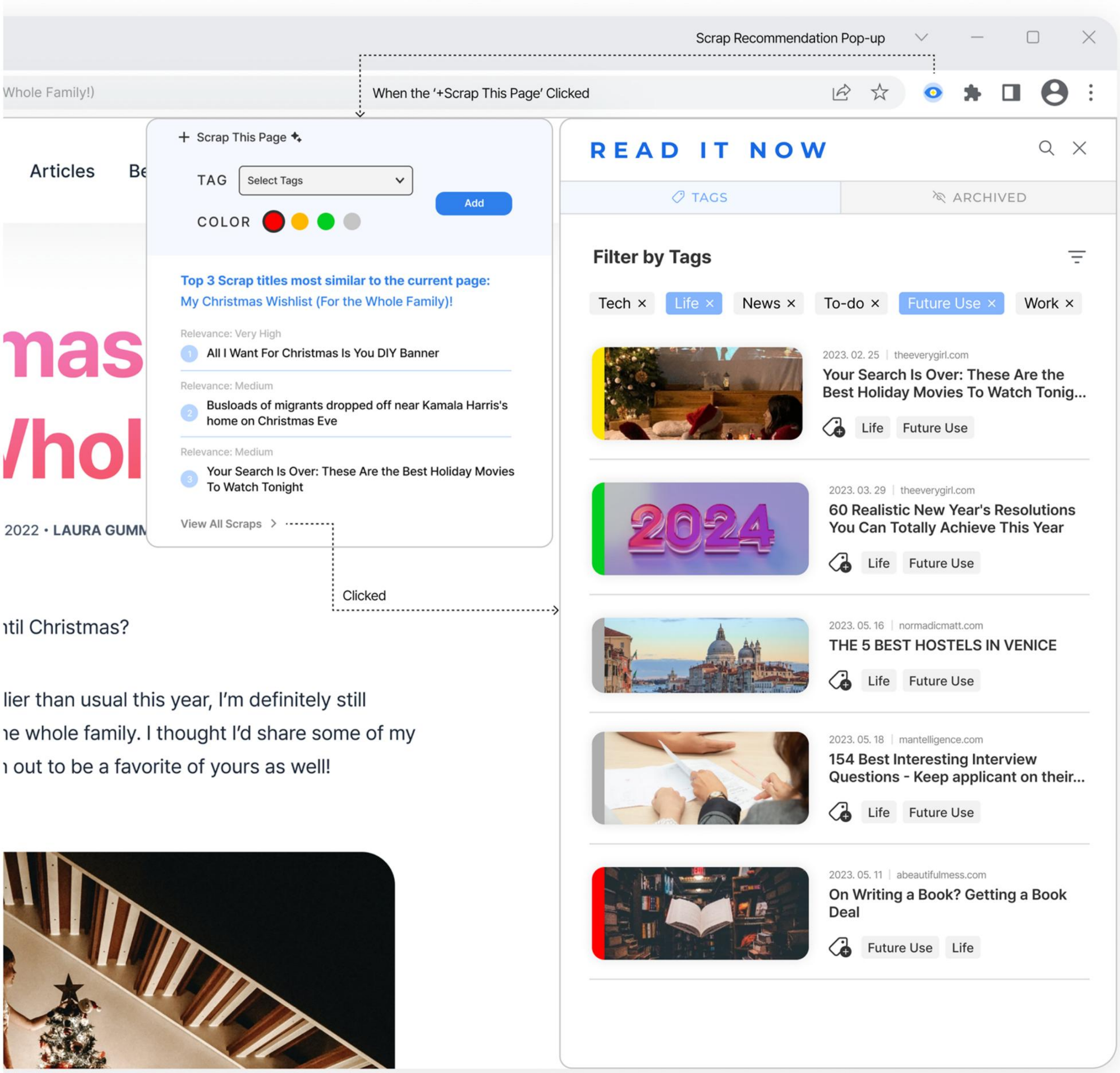

Figure 7. Modified interface overview of Read-It-Now.

6.5. Limitations

Our study has several limitations in terms of sample size and development. Firstly, we sample sizes used for the interview for discovering user requirements and the usability tests for evaluating the prototype were limited. We interviewed 10 interviewees and recruited 32 participants for usability testing, which was a small sample size. However, these sample sizes were acceptable in accordance with previous studies. For example, in some studies, investigations were conducted with only three users for designing web products such as bookmarks, emails, and filings (Boardman & Sasse, 2004; Lee et al., 2012). In addition, according to one study that investigated the appropriate sample size for usability test, even 5–10 participants were sufficient for formative usability testing, and the minimum sample size was 30 for summative usability testing with the majority of parametric statistical methods (Lewis, 2014). Moreover, the usability test included simple comparisons based on descriptive statistics and t-tests. Furthermore, we conducted a post-hoc interview to obtain more detailed information from the users. The results of the quantitative analyses were mostly aligned with the results of the interviews. Hence, despite the small sample size, we believe that the results could guarantee the reliability of our research.

Secondly, our experimental setting had a limitation in that it did not allow the manual selection and creation of scraps by the participants. To mitigate this, the participants were allowed to preview the scrap list to foster partial recall during the tutorial phase and encouraged to regard these as their own "forgotten" scraps. As this study aimed to propose an interface that enhances the rediscovery of forgotten scraps, we intentionally aligned this setup with our objective.

However, a notable drawback was the underemphasis on scrap creation dates and colors within the task execution context, as scraps were not personally generated by the participants. To overcome this, our future work will include the development of a real browser extension application that facilitates real-world experimentation with user-generated scrap data, thereby providing a more realistic environment for exploring the diverse potential of Read-It-Now.

7. Conclusion

This study proposes a new task-centric web page scrap tool that triggers scrap retrieval via semantic recommendation. Nowadays, people store the majority of their information in the digital form via various read-it-later tools, and the difficulty of information retrieval is increasing. After analyzing how users interact with diverse read-it-later tools throughout the scrap lifecycle, we developed a mental model with four mental spaces: encounter, create, revisit, reuse, and organize. Design recommendations for addressing the difficulties in each stage and facilitating active tool usage were suggested. Accordingly, our research attempted to develop a task-centric in-browser web page scrap tool interface, called Read-It-Now. In the evaluation experiment, we were able to confirm that our prototype is easy to use and has higher perceived usability than the conventional bookmark interface. However, the interview responses indicated that the information and explanations provided were insufficient for understanding the internal principles of the scrap recommendation feature, which was implemented as a retrieval trigger. Therefore, in light of these user opinions, we proposed an improved interface design as the final version.

The theoretical contribution of study is in filling the research gap in the bookmark-centered research model in PIM. Our user-centered mental model sheds light on the design implications for reducing user cognitive load and facilitating scrap retrieval. This study is expected to provide insights for practitioners designing task-centric information scrap tools.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No.2022-0-00078, Explainable Logical Reasoning for Medical Knowledge Generation). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00343599). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2023-00210250).

ORCID

Sangyeon Kim http://orcid.org/0000-0001-9316-7982
Sangwon Lee http://orcid.org/0000-0002-9825-0854

References

Abrams, D., Baecker, R., & Chignell, M. (1998). *Information archiving with bookmarks: Personal web space construction and organization* [Paper presentation]. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 41–48. https://doi.org/10.1145/274644.274651

Barreau, D. K. (1995). Context as a factor in personal information management systems. *Journal of the American Society for Information Science*, *46*(5), 327–339. https://doi.org/10.1002/(SICI)1097-4571(199506)46:5

Barreau, D., & Nardi, B. A. (1995). Finding and reminding. *ACM SIGCHI Bulletin*, *27*(3), 39–43. https://doi.org/10.1145/221296.221307

Bernstein, M., Kleek, M. V., Karger, D., Bernstein, M., Van Kleek, M., Karger, D., & Schraefel, M. C. (2008). Information scraps: How and why information eludes our personal information management tools. *ACM Transactions on Information Systems*, *26*(4), 1–46. https://doi.org/10.1145/1402256.1402263

Boardman, R., & Sasse, A. (2004). Stuff goes into the computer and doesn't come out" a cross-tool study of personal information management. In *Proceedings of the 2004 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/985692

Brooke, J. (2013). SUS: a retrospective. *Journal of Usability Studies, 8*(2), 29–40.

Bruce, H., Jones, W., & Dumais, S. (2004). Keeping and re-finding information on the web: What do people do and what do they need? *Proceedings of the American Society for Information Science and Technology*, *41*(1), 129–137. https://doi.org/10.1002/meet.1450410115

Chang, J. C., Kim, Y., Miller, V., Liu, M. X., Myers, B. A., & Kittur, A. (2021). Tabs.do: Task-centric browser tab management. *UIST 2021 - Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology*, *14*(21), 663–676. https://doi.org/10.1145/3472749.3474777

Facebookresearch. (2017). *GitHub - facebookresearch/faiss: A library for efficient similarity search and clustering of dense vectors*. Retrieved November 27, 2023, from https://github.com/facebookresearch/faiss/

Geikhman, Y. (2023). *12 English reading websites: Made for native speakers and great for learners| FluentU* English. January 15. https://www.fluentu.com/blog/english/english-reading-practice/

Jo, J., Kim, S., & Lee, S. (2023). *Read it now, not later! integrating semantic recommendation into in-browser web page scrap tool* [Paper presentation]. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, April 19. https://doi.org/10.1145/3544549.3585858

Jones, W., Bruce, H., & Dumais, S. (2001, October). *Keeping found things found on the web* [Paper presentation]. Proceedings of the Tenth International Conference on Information and Knowledge Management, 119–126. https://doi.org/10.1145/502585.502607

Kaasten, S., & Greenberg, S. (2001). *Integrating back, history and bookmarks in web browsers* [Paper presentation]. CHI '01 Extended Abstracts on Human Factors in Computing Systems, 379–380. https://doi.org/10.1145/634067.634291

Kilov, H. (1990). *From semantic to object-oriented data modeling* [Paper presentation]. Proceedings of the First International Conference on Systems Integration, 385–393. https://doi.org/10.1109/ICSI.1990.138704

Kim, S. S. Y., Watkins, E. A., Russakovsky, O., Fong, R., & Monroy-Hernández, A. (2023). *"Help Me Help the AI": Understanding how explainability can support human-AI interaction* [Paper presentation]. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 17. https://doi.org/10.1145/3544548.3581001

Kleek, M. V., Styke, W., & Karger, D. (2011). *Finders/keepers: A longitudinal study of people managing information scraps in a micro-note tool* [Paper presentation]. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,. https://doi.org/10.1145/1978942.1979374

Lee, S. C., O'Brien-Strain, E., Liu, J., & Lin, Q. (2012). *A survey on web use: How people access, consume, keep, and organize web content* [Paper presentation]. Proceedings of the 2012 CHI Conference on

Human Factors in Computing Systems, 619–628. https://doi.org/10.1145/2212776.2212834

Lewis, J. R. (2014). Usability: Lessons learned … and yet to be learned. *International Journal of Human-Computer Interaction*, *30*(9), 663–684. https://doi.org/10.1080/10447318.2014.930311

Lewis, J. R., & Sauro, J. (2018). Item benchmarks for the system usability scale. *Journal of Usability Studies*, *13*(3), 158–167. https://doi.org/10.5555/3294033.3294037

Lin, M., Lutters, W. G., & Kim, T. S. (2004). *Understanding the micronote lifecycle* [Paper presentation]. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 687–694. https://doi.org/10.1145/985692.985779

Liu, M. X., Kittur, A., & Myers, B. A. (2022). *Crystalline: Lowering the cost for developers to collect and organize information for decision making* [Paper presentation]. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, April 29. https://doi.org/10.1145/3491102.3501968

Pu, P., Chen, L., & Hu, R. (2011). *A user-centric evaluation framework for recommender systems* [Paper presentation]. RecSys'11 - Proceedings of the 5th ACM Conference on Recommender Systems, 157–164. https://doi.org/10.1145/2043932.2043962

Rachatasumrit, N., Ramos, G., Suh, J., Ng, R., & Meek, C. (2021). *ForSense: Accelerating online research through sensemaking integration and machine research support* [Paper presentation]. International Conference on Intelligent User Interfaces, Proceedings IUI, 608–618. https://doi.org/10.1145/3397481.3450649

Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence embeddings using siamese BERT-networks*. ArXiv Preprint, ArXiv 190810084; Association for Computational Linguistics, August 27. https://doi.org/10.18653/v1/d19-1410

Shen, S. T., & Prior, S. D. (2013). *My favorites (bookmarks) schema - One solution to online information storage and retrieval* [Paper presentation]. ACM International Conference Proceeding Series, 33–40. https://doi.org/10.1145/2503859.2503865

Swearngin, A., Iqbal, S., Poznanski, V., Encarnacion, M., Bennett, P. N., & Teevan, J. (2021). *Scraps: Enabling mobile capture, contextualization, and use of document resources* [Paper presentation]. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, May 6. https://doi.org/10.1145/3411764.3445185

Young, I. (2008). *Mental models: Aligning design strategy with human behavior*. Rosenfeld Media.

About the authors

Jeongmin Jo is a graduate from the Department of Industrial and Management Engineering. She received her Master degree from Korea University in 2023. Her academic interests lie in human-centered AI and user experience design strategy.

Sangyeon Kim is an Assistant Professor in the Division of Artificial Intelligence Engineering at Sookmyung Women's University. He obtained his PhD degree from Sungkyunkwan University in 2022. His academic interests include human-centered AI, intelligent user interface, responsible AI, and accessible computing.

Sangwon Lee is a Professor in School of Industrial and Management Engineering at Korea University. He obtained his PhD and Master degrees from the Pennsylvania State University in 2010 and 2006. He graduated as BS from Korea University in 2004. His academic interests lie in HCI, UX, and XAI.