Metadata-Enabled Personal Archive Search Tool

A tool that tracks metadata stored in digital files and allows users to utilize different types of information for file retrieval and management

Eunjung Paik

Carnegie Mellon University
School of Design

Metadata-Enabled Personal Archive Search Tool

A tool that tracks metadata stored in digital files and allows users to utilize different types of information for file retrieval and management

Eunjung Paik, Author

Andrew Twigg, Co-Advisor

Peter Scupelli, Co-Advisor

A thesis submitted to the School of Design, Carnegie Mellon University, for the degree of Master of Design in Design for Interactions

Abstract

In the age of information overload, personal digital file organization and retrieval has been a longstanding problem. This is because the amount of information digital devices can hold is increasing and various cloud storage platforms that offer us nearly limitless amounts of data are marketed strongly to us. In contrast, the human capacity for managing and remembering digital information is limited. It also requires huge commitment and time for keeping track of digital items across multiple devices and platforms.

Based on the insights obtained from literature reviews and user interviews, important design implications and key "How might" questions were established, which guided me toward the final design solution, "Metadata-Enabled Personal Digital Archive Search Tool".

Finally, this thesis project explores design opportunities of harnessing metadata to help non-expert digital item organizers actively modify file seeking and retrieval strategies. Metadata is information that describes the characteristics of other information, ranging from location, time, device, camera information, and so on. The opportunity of utilizing metadata is highlighted through some of the design features with the use of user scenarios.

With the final design suggestion, the study proposes several features and visual representations that would allow users to keep track of digital items in personal archives and modify their strategies in a more appropriate way for various situation. The final goal is creating user engagement into their file management and retrieval activities.

Chapter 1
Introduction

In This Chapter

This chapter briefly introduces how the thesis has been explored with various design processes. It covers brief introduction of the problem area this thesis has been targeting, why this problem is important to be solved, the literatures (academic papers and existing products) that had been investigated for the target problem area, types of research methods being used, key insights obtained from the research, and the final design suggestion.

In the age of information overload, retrieving content can be a tough task for users with a lot of digital items in their digital devices and archives. The amount of information digital devices can hold is increasing and various cloud storage platforms that offer us nearly limitless amounts of data are marketed strongly to us.

In contrast, the human capacity for managing and remembering digital information is limited. Beyond a maximum capacity, not all of target users have effective organizing skills and they often lack the time and commitment to organizing digital information, especially when they are overwhelmed by its amount. As the number and variety of our files grow, it becomes increasingly crucial to have effective organization methods and skills in order to succeed at digital item retrieval.

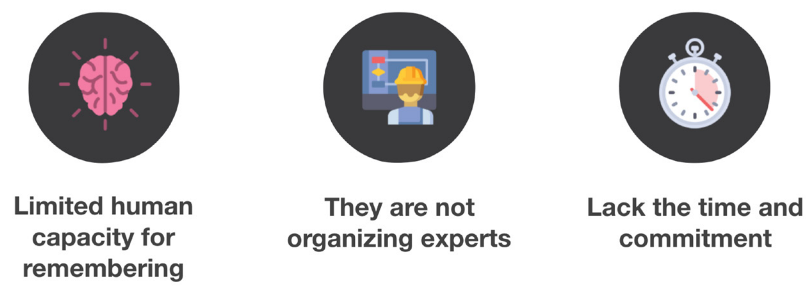


Fig. 1 Target user's key constraints.

From user interviews and literature reviews it became clear that people use different strategies in a different situation. It depends on the types of files users are trying to retrieve, the amount or the types of the information about the file that users remember, and the types of storage platforms they use.

The insights I got from my research then led me to ask these questions:

1. How might a user can see or control all digital items that are scattered over multiple devices and virtual locations?
2. How might a user is able to utilize one or multiple information in a proper manner for file retrieval in different challenging conditions?
3. How might a system can track the history or several pieces of information of digital files?

Based on the research insights and the key questions that are critical to answer in this problem area, this thesis explores various challenges that target users have been facing and opportunities for digital item management system that enables users to harness metadata, which is "any data that helps describe the content or characteristics of a file" [1], to help users retrieve and manage files easily.In the 1990s, Peter Pirolli and Stuart Card published Information Foraging Theory, which proposes human's adaptive Interaction with Information by drawing the analogy between animal's food-foraging activities in the real world and people's

1 Peter

And finally, the thesis suggests a new search tool with new features, interfaces, and interactions, which tracks metadata stored in digital items to provide users with various information across multiple devices and storages. With this tool, a user is able to get a holistic view of all scattered digital items that they possess, get specific information with visualization embedded in the files such as the relationships with other files (how their content is related to each) or with other people (with whom the user has shared the item with). By allowing users to take advantage of the metadata and modify their strategies effectively in different situations, I believe that they can be actively involved in file management and retrieval activities.

Chapter 2
Related Works

In This Chapter

This chapter illustrates the main literatures chosen for this thesis exploration, which guided and influenced my design directions, are covered. These literatures range from the term Information Overload to Information Foraging Theory that gave me fundamental insights of the problem area and supportive background, to previous research and implementation on Personal Information Management (PIM) tools, and to Metadata.

Information Overload

Information Overload nowadays is an everyday word, meaning receiving too much information. The burden of a heavy information load will lead to the individual's confusion, hinder people's attention to information, and make task/strategy priority setting difficult or inefficient. It has been a longstanding problem especially since personal computers have been widely used and World Wide Web has been publicly accessible. These factors have led to the increasing number of computer users all over the world as well as an increasing amount of information that users have been constantly creating and sharing.

Information Foraging Theory

Human Web Foraging Behavior

Information seeking activities in the digital environment. Information Foraging Theory has been adopted in this project as a fundamental theory for supporting the thesis. It's a scientific theory developed to understand human-information interaction.

The theory tackles the problem of increasing amount of information and following people's unique behavior of shaping themselves to the web environment in an attempt to maximize the chance of obtaining information: "understanding representative problems posed by the real-world information environment and adaptive cognitive solutions to those problems" [2]. The adaptive activities include assessing, seeking, and handling information sources.

It argues that in an information-rich world, the real problem of information foraging is not obtaining more information, but optimizing the user's strategies. It focuses on "user's adaptive task performances and variations in them in the information ecologies" [3].

The analogy between food foraging behaviors and human information foraging activities is that both of these activities have resulted from limitation in the ability or the environment. In the physical world, animals, in order to succeed in searching foods, has been evolving to have adaptive abilities and strategies based on the environmental situations.

The consideration for the animal food foraging is about the amount of food energy that can be extracted from the environment, compared to the amount of time they spend foraging. Whereas, in the digital environment, the consideration of how much useful or valuable information could be exploited with as less interaction time cost as possible is regarded as important.

So, the human has been evolving to improve the ability to process the information since their capacity to process information is inherently limited. For example, scientific models that Information Foraging Theory is proposing reveal several human information foraging methods: whether to continue pursuing something in this website or go somewhere else

2. Pirolli, Peter, and Stuart Card. "Information foraging." Psychological review 106, no. 4 (1999): 643.
3. Ibid

(Information Patch), a narrowing of the information (Information Diet Selection), following visual or textual cues to make decisions about where to go (Information Scent), etc.

Framing a new problem area

Focus on a different activity

Although the main focus of the theory is the web, it highlights the importance of the evolution of human adaptive behavior over the digital The variations of human adaptive behaviors, resulted from a unique digital environment where a wealth of information is widely distributed, lead to another problem. Combined with technology advancement and various product development, user's information foraging behaviors allow them to possess even more and actively share with others. For instance, the amount of information digital devices can hold is increasing and various cloud storage platforms, that users are encouraged to use in different collaborative environments, offer us nearly limitless amounts of data. That is, web foraging behavior leads to the increasing number of information we collect and finally the importance of management arises.

The information in the recent digital era is entangled in a complex way and distributed across various locations. That is, much of the information we possess now is more decentralized through various digital devices such as desktop, laptop, and mobile devices, and through virtual storage platforms. However, the more the information is distributed (or decentralized), the less powerful our ability become to keep track of every item.

Thus, more thorough management and control upon intertwined information within several locations and environment is necessary. This thesis explores the information management problems and opportunities

Fig. 2 An example illustrating a target user's personal digital archive circle.

Information Foraging v.s. Information Management Activities

that focus on post-foraging activities in various environment, including the web and the local devices.
As the amount and variety of our files grow, it becomes important to have effective and consistent organization methods in order to succeed at digital item retrieval.

However, target users have key constraints for organizing, for example, they lack of the time and commitment, it's impossible to remember every digital item or associated information because of given limited human capacity, and they are not expert organizers. So, I'm targeting on creating a new search tool that supports user's file retrieval and helps file management.

More specifically, the target area that this thesis explores is personal digital archives. It focuses on the user's activities of managing or organizing digital items which the user possesses, not of obtaining new items from the web. User's activities this thesis focuses on the activities after their information foraging behaviors.

Physical Spaces

Organization and Management Strategies in Physical Spaces

The biggest difference between file managing activity and information foraging activity is that user's memory is involved in managing activity. User's file creating or collecting activities will allow them to have a better memory of the target item they wish to retrieve, whereas, in web foraging activity, they confront the challenges of entirely new information and content.
First of all, I investigated academic research and papers on information organization systems in physical spaces. The reason why I explored this area of people's physical information repositories and strategies is because I wanted to see how people have been developing and utilizing various strategies for managing things in the real world, and eventually to learn how these strategies have been evolving within the digital environment either differently or similarly.

Thomas W. Malone, in "How Do People Organize Their Desks?", describes the function of desk organization is not only for finding but also reminding. Through conducted interviews with ten people, who often work at desks, eventually Thomas was able to get insights for the design of electronic office information systems, which is people do organize their desks in a specific but similar way: a) some people, based on the level of processing, designate areas and sort out documents, some others based on the types of contents.

The important implication from this study is that the function of desk organization is not only to help people find most wanted information but also to remind them of things to do. That is, "people organize their desks in part so that they can find things". The paper emphasizes that "By explicitly trying to facilitate reminding" [4], computer-based systems could become useful. This is because, even though there were individual differences, people often tend to keep top-priority things within their hand reach.

Finding and Reminding

Finding and Reminding
File organization from the desktop

In another study, Bonnie A. Nardi also describes that organizing activity is closely related to reminding. In the "reminding" section of the paper, it says "The location of information on the desktop also serves a critical reminding function" [5]. The other sections, "Location-based Search", "Three Information Types: Ephemeral, Working, Archived" also lead to "reminding" function since the user tend to put the items, which they are working on actively (="ephemeral" or "working" information), within easy reach areas (=location-based search) so that they can be reminded of things to work on.

Insights

The importance of understanding the corelation of user's activities and how organized thedifferent types of items accordingly is also described in Activity Theory.[6] In Personal Information Management, William Jones

4. Malone, Thomas W. "How do people organize their desks?: Implications for the design of office information systems." ACM Transactions on Information Systems (TOIS) 1, no. 1 (1983): 99-112.
5. Barreau, Deborah, and Bonnie A. Nardi. "Finding and reminding: file organization from the desktop." ACM SigChi Bulletin 27, no. 3 (1995): 39-43.
6. Kaptelinin, Victor. "Activity theory: Implications for human-computer interaction." Context

insists the importance by saying "The use, indeed the meaning, of objects in a person's world, including information items, paper-based and digital, should be understood primary with respect the roles they play in various activities the person is completing" [7].

From these 3 papers, I could find it critical that retrieving and reminding cannot be apart. And reminding function here does not mean that the system should provide users with an alert or notification of to-do lists or suggestion items, it means the user organizes items in a way they can remember or be reminded of which item to use for their specific tasks. And when the user is able to manipulate the locations of the items and the area, they feel like they are able to "command the space" (Nardi et al. 1995).

The potential design implication I was able to obtain from these studies of physical management activities is allowing the user to manipulate a space can be beneficial for people's management strategy and memory so that they can place top-priority items closely to them.

Personal Information Management (PIM)

In "Personal Information Management", William Jones draws conclusions with key implications for designing PIM tools: the importance of re-accessing is being emphasized again. Again, location, organizing folder, form, and associated devices and applications should be considered.[8]

3D User Interfaces and Environment

The Spatial Metaphor for User Interfaces: Experimental Tests of Reference by Location Versus Name

The area of new PIM tools with 3D digital environment and interfaces is also considered as worth-exploring domain based on previous literature reviews, which revealed potential opportunity of bridging the physical and digital organizing strategies and environments. Papers and researches in this area have had a lot of experimentations in an attempt to prove the "existing evidence from psychological studies of memory for the location of objects" [9].

One of the interesting experiments in this paper, which was conducted in order to assess the role of spatiality in an information retrieval system, shows research participants' better performance at a series of retrieval test, with "Name+Location-Combined" condition, in which they had four different types of conditions: "Name-Only", "Location-Only", "Name+Location-Combined", and "Name+Location-Separate".

Although, the study concludes with "the findings indicate that incidental

---

and consciousness: Activity theory and human-computer interaction 1 (1996): 103-116.
7. Jones, William. "Personal information management." Annual review of information science and technology 41, no. 1 (2007): 453-504.
8. Ibid
9. Jones, William P., and Susan T. Dumais. "The spatial metaphor for user interfaces: experimental tests of reference by location versus name." ACM Transactions on Information Systems (TOIS) 4, no. 1 (1986): 42-63.

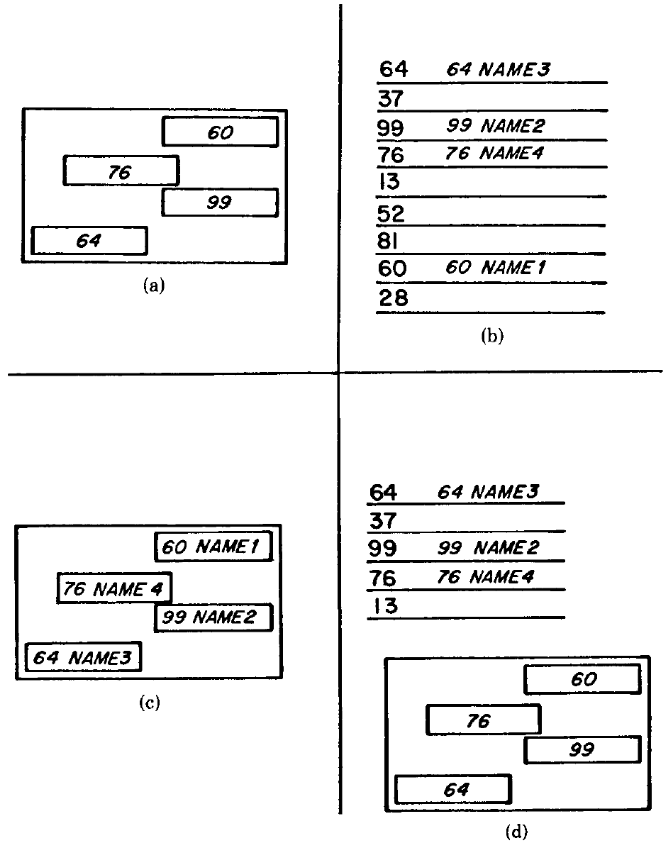


Fig. 3 Schematization of the four filing conditions of Experiment: (a) location only; (b) name only; (c) name+location combined; (d) name+location separate. Jones et al., "The spatial metaphor for user interfaces: experimental tests of reference by location versus name."

memory for location is better than chance, although the differences are usually modest". There are severe limitation and unreliability in "the efficacy of a spatially oriented approach to an object reference in a computing system" [10].Although, the study concludes with "the findings indicate that incidental memory for location is better than chance, although the differences are usually modest". There are severe limitation and unreliability in "the efficacy of a spatially oriented approach to an object reference in a computing system".

Data Mountain: Using Spatial Memory for Document Management

In 1998, Microsoft Research team developed a new document management tool called Data Mountain which breaks the common approach of using 2D spatial layouts in computer workspace environment, and leverages natural human capacities, which is spatial perception. The team implemented this new system which contains 3D virtual environments where users are allowed to "place documents at arbitrary positions on an inclined plane in a 3D desktop virtual environment using a simple 2D interaction technique" [11].

10. Ibid
11. Robertson, George, Mary Czerwinski, Kevin Larson, Daniel C. Robbins, David Thiel, and Maarten Van Dantzich. "Data mountain: using spatial memory for document management."

Fig. 4 Data Mountain with 100 web pages. Robertson et al., "Data mountain: using spatial memory for document management."

The Task Gallery

The Task Gallery, designed by Microsoft Research team later, is a 3D-based window manager with a 3D virtual environment that is more analogous to the real world. It says that during the experiment, participants did not have a problem perceiving front and back side of the environment or ordering their tasks, and they were rarely confused about ordering in their memory. The paper suggests that since this experiment is just a scratching the surface and needs further explorations.

Fig. 5 Robertson et al., "The Task Gallery: a 3D window manager."

In Proceedings of the 11th annual ACM symposium on User interface software and technology, pp. 153-162. ACM, 1998.

The Result

The experiment results from both Data Mountain and The Task Gallery show a surprising improvement in user's speed and accuracy of their retrieval.

Evaluating the Effectiveness of Spatial Memory in 2D and 3D Physical and Virtual Environments

With the several increasing researches and developments of 3D environment and tools such as Data Mountain, The Task Gallery, etc, Andy Cockburn and Bruce McKenzie, in an attempt to evaluate the effectiveness of 3D computer systems, conducted an experiment in which these 3D design concepts were thoroughly translated into the real world with physical prototypes.

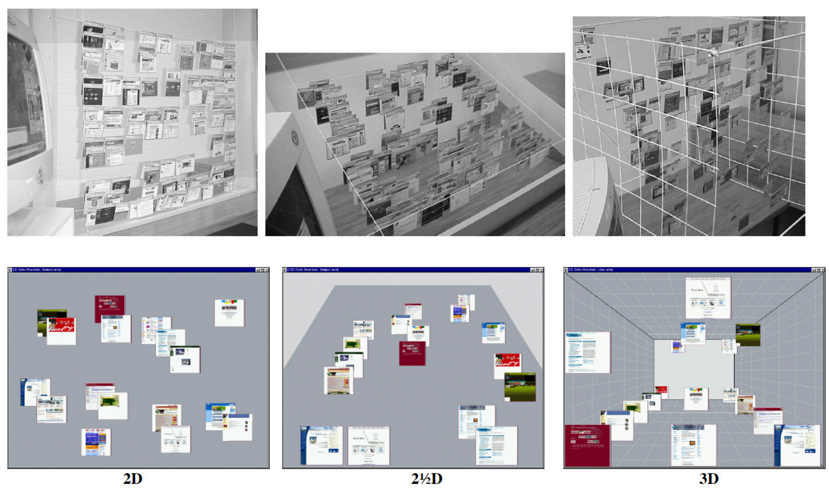


Fig. 5 The three physical (top row) and virtual (bottom row) interfaces. Cockburn et al. "Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments".

Cockburn et al. conducted an interesting experiment to examine They created 3 types of interfaces inspired by Data Mountain, which include 2D, 2½D, and 3D. The 2D interface allows "a vertical surface allowing the x and y coordinates" [12], the 2½D interfaces provide "a receding inclined plane on which pages can be located", and the 3D interfaces contain "the x, y, and z coordinates". For the experiment, 6 prototypes were built, 3 digital prototypes and 3 physical prototypes. The goal of this experiment was to identify the effectiveness of 3D displays: whether it would lead to user's outperformance of not.

However, during the retrieval test, participants interacting with 2½D and 3D were more active whereas those interacting with 2D were more static. The result is interesting since commonly we consider the 3D environment and interfaces would increase user's task improvement thus also decrease the retrieval time, yet participants' task performances reveal that both the 3D physical and virtual interface leads to poor performance.

12. Cockburn, Andy, and Bruce McKenzie. "Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments." In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 203-210. ACM, 2002.

Insights

I learned that there is a gap between the human-computer interactions in a virtual and physical environment because of the variety of movements that people make in the physical world when performing tasks. The 3D environment, even if it is designed to bridge the gap between virtual and physcal environment, since the way we interact with these 2 are different and there is a lack of sensory system people can leverage in the digital environment, it is difficult to expect user's high performance in 3D digital environment.

Thus, heavily relying upon utilizing location-oriented interface would create limited opportunity for a PIM tool design. The biggest reason is that there is still a huge gap between the spatiality in the virtual and physical world, the gap of how we can actually manipulate the two spaces (interactivity) and how we can perceive them (perception).

One of the suggestions that William Jones made is that by utilizing various types of retrieval cues, including shape, size, location, name, etc, users would remember items better and eventually manage document better. An important insight from these studies is that we should help users not just better recognize the visual representatives in the computer system that bridges users with the actual information, but should help them actually recall the information.[13] The difference between recognizing and recalling is the level of user's memory capacity. If users are able to recall the information by themselves, they would be better at proactively interacting with the information they are attempting to retrieve and commanding the environment.

13. Ibid

Metadata

An Introduction of Metadata

The decision of exploring what metadata is was made from my previous user and expert interviews. Metadata is a set of data that is embedded in digital items. Metadata provides information about other data. For instance, let's say one created an image file with my Android smartphone. As soon as one creates it, a variety of information of the image file, such as date & time, GPS, the dimension of an image, device information, camera information, etc, are created simultaneously.

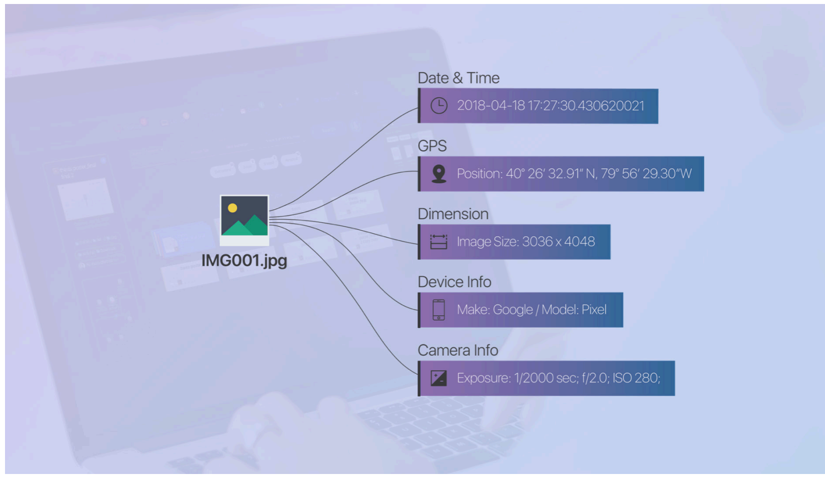


Fig. 6 Various types of Metadata embedded in a digital file.

Existing Product That Uses Metadata Enabled Technology: XMP, a File Labeling Technology

Currently, Adobe and a number of other companies are actively utilizing Metadata, with the technology called XMP. According to Adobe's description, "Metadata is any data that helps describe the content or characteristics of a file. You may already be accustomed to viewing and adding some basic metadata through the File Info or Document Properties box found in many software applications and some operating systems. Adobe's Extensible Metadata Platform (XMP) is a file labeling technology that lets you embed metadata into files themselves during the content creation process" [14]. Adobe emphasizes key benefits of using metadata technology:

1. Digital assets retain their context when traversing software, digital devices, and databases.
2. It enables powerful search and retrieval of rich content across varied file formats.
3. Manage relationships of assets.

The reason why I explore metadata is by using it, there is an opportunity to offer users the flexibility and variety in harnessing the pieces of memories they have, which can be implemented by involving metadata into the system. Since metadata is several sets of data that is automatically generated by the digital system, there is a huge design opportunity for using this data, which can be powerful information that the system and users can harness.

14. "Extensible Metadata Platform (XMP)." Adobe Captivate - Welcome to the World of Smart ELearning Authoring. Accessed May 13, 2018. https://www.adobe.com/products/xmp.html.

Fig. 7 Metadata of an image file enabled by XMP. Accessed through Adobe Photoshop CC 2017, "File Info > Camera Data"

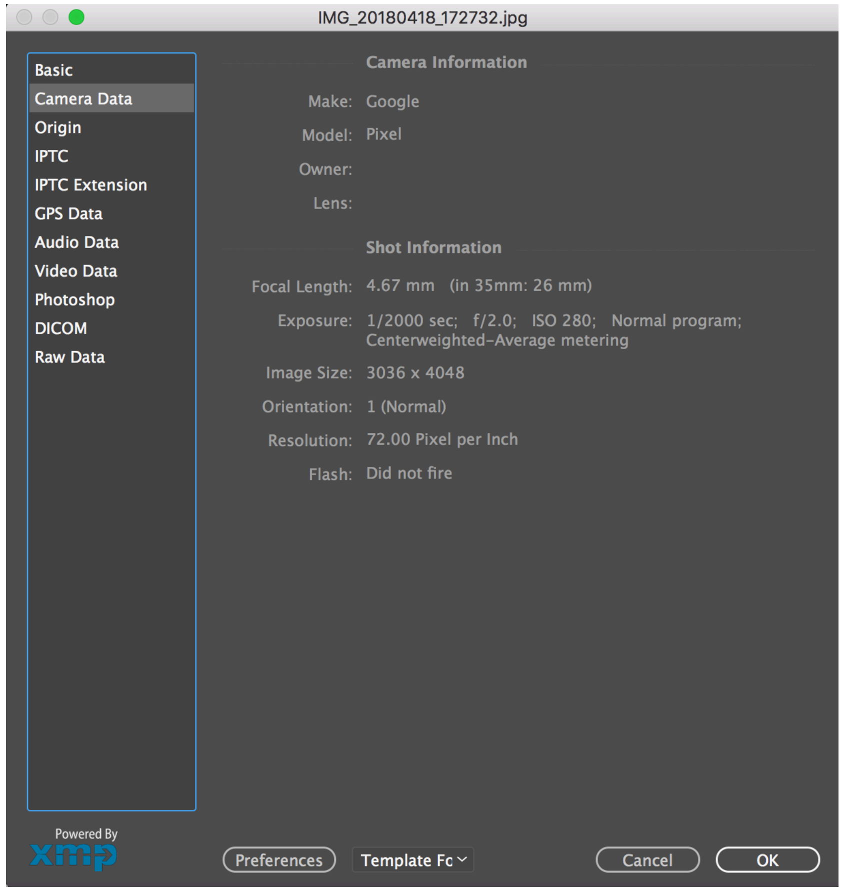
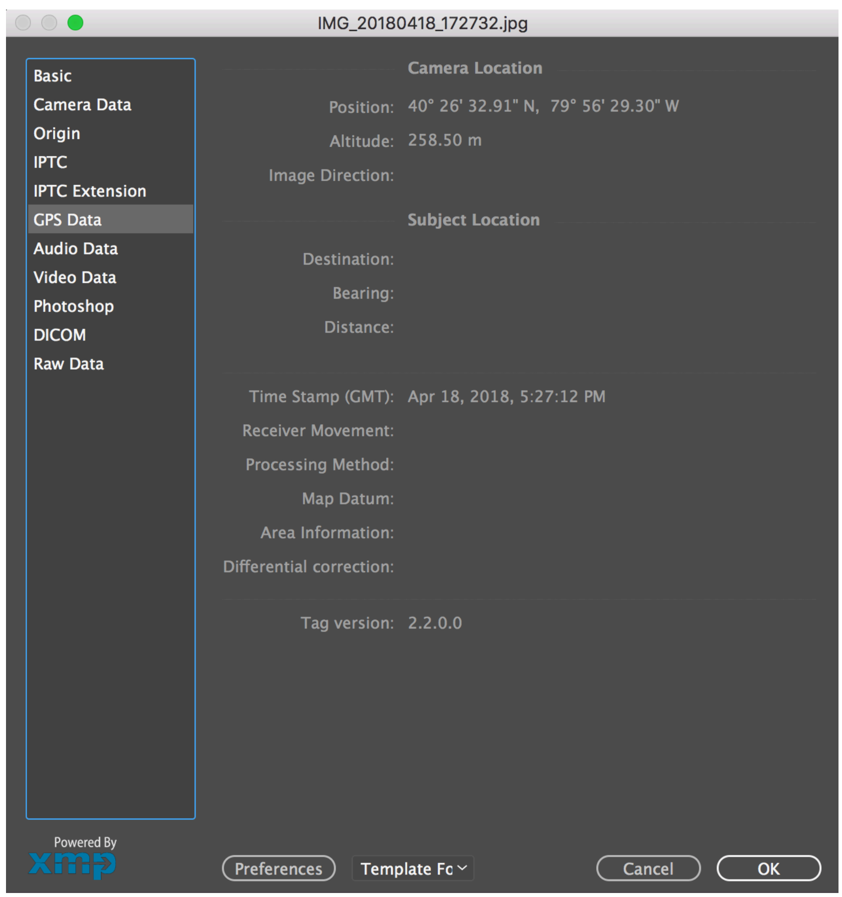


Fig. 8 Metadata of an image file enabled by XMP. Accessed through Adobe Photoshop CC 2017, "File Info > GPS Data"



Chapter 3
Design Process

Literature review
15+ papers and articles related to PIM tools, distributed cognition, and metadata.

Interviews
In-depth of interviews with 10 target users.

Lo-fi Prototyping
Low fidelity prototypes (paper + screen).

Expert Reviews
Expert interview with an expert organizer and review session with 2 advisors.

User Testing
User testing with 9 target users using lo-fi prototypes.

Design Refining
User testing with 9 target users using low-fidelity prototypes.

Hi-fi Prototyping
High fidelity prototypes based on users' feedback.

In This Chapter

This chapter covers the design process and methods used for the entire journey. Design process begins with exploratory research then generative, and finally evaluative research.

During exploratory research stage, more than 15 literatures were reviewed, that are related to Information Foraging Theory, PIM (Personal Information Management) tools, document organization strategies in the physical and virtual environment, and metadata. Also in-depth user interviews were conducted with 10 target users, who have been actively creating and collecting digital files and content. Questions given to the participants range from challenges and differences of item organization and retrieval in the digital and physical environments to their own organization strategies.

For generative process, I created 2 main design concepts with user scenarios and low fidelity wireframes. Lo-fi paper prototypes were used for user testing with 9 participants where I was able to obtain different perspectives and feedback on the concept, flow, and interfaces.

In evaluative stage, I refined the design concepts and graphical user interfaces based on the previous user tests and created high fidelity final design deliverable.

Explorative Process

1. Literature Reviews

In order to get insights ranging from the fundamental problem area to the previously experimented Personal Information Management (PIM) tools, I reviewed a wide range of academic research on several different areas: information foraging theory, activity theory, PIM tool designs with various approaches, document organization strategies in the physical and virtual spaces, and distributed cognition.

Methods

Literatures on different areas were chosen along with the guidance of the findings from the in-depth user interviews. Literature reviews and user interviews were not separately conducted as two independent steps, yet were simultaneously conducted and iterated by influencing each other.

2. In-Depth Interviews

I conducted in-depth interviews with 10 potential target users for this thesis. People were chosen to represent a wide range of people who create, collect, and share a wealth of digital items consistently. They included six design students, one design professor, two photographers, and one librarian. Design students and a professor tend to be involved in a variety of types of digital files, photographers are specifically dealing with image files, and a librarian mostly deals with both physical and digital document files, such as articles, papers, and emails.

Methods

Recruited interview participants were asked about their own experiences with organizing and retrieving both digital and physical items. The more specific questions included:

1. How they organize digital andphysical items
2. Factors that make managing digital/physical items andenvironments easy or difficult
3. Whether they have failed to retrieve digital and physical items
4. What types of an item is most difficult or easiest to retrieve

It was a verbal interview conducted at user's office spaces where they are involved in some item organization, management, and retrieval activities are occurring. I intentionally structured this interview to be taken in participant's office spaces so that I can naturally ask them to show their item management strategies and habits in both physical and digital environment.

Some of the main goals of the interviews were to learn various design conventions from devices and platforms that participants have been using but I'm not yet familiar with, learn different participants' strategies and habits which would be beneficial to apply to the final design idea, and the biggest challenges they have confronted when managing and retrieving items.

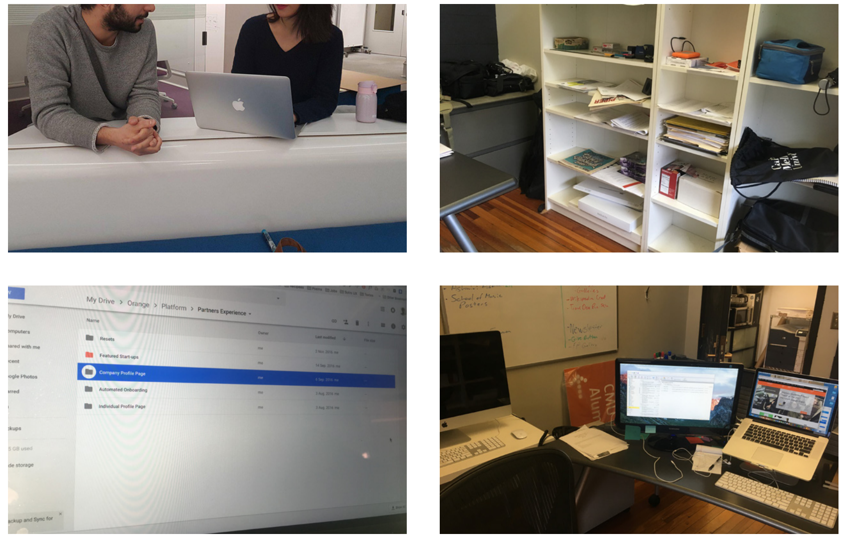


Fig. 9 10 in-depth interviews conducted at participants' work environments at Carnegie Mellon University, Pittsburh campus

3. Synthesis & Key Insights

By synthesizing the insights from the literature reviews and in-depth interviews, I was able to learn design opportunities. Especially, from the user interviews, I could see 3 repetitive patterns of target user's painpoints in the file management and retrieval. These 3 painpoints, which are the design opportunities at the same time, are:

1. **Different types of files are difficult to retrieve for different reasons**

   I found that users use different strategies according to specific types of files. (for retrieving different types of files.) Most of my interview participants said text files are easier to retrieve than any other types of files. But if they forget the content or the exact name, even text files can be difficult to retrieve sometimes.

   For instance, there is a case where a user remembers one keyword that is contained in the file name or content. The exact word is adjective yet the word the user remembers is a noun. In this case, file retrieval can be tricky because either the target file would not show up in the result list or the system would bring a list full of irrelevant files which will distract user's attention. Whereas, image files are difficult to retrieve for different reasons. Since they're often created with so many other images altogether, it's hard to find the exact one.

   > “” *A: Text based files are easy to retrieve than any other types of files.*
   > *B: Image files are easy to organize yet hard to retrieve if specific file name is not designated, which is my fault.”*

2. **Sometimes we remember only little pieces of information**

   There are moments where people don't have the main information of a target file they are trying to retrieve, yet have several little minor pieces of information. One of the interview participants depicted this example: they remember approximately where it was created (time information), what kind of content was in there (file format information), with whom they were at that time (content information), but they still could not find it.

“” *I know approximately when I downloaded it, from whom I got it, what kind of type it was, but I can’t find it. I don’t know why.*

3. The number of files is Increasing. We also upload and share our own files with others through diverse platforms.

Users cannot prevent the number of digital files from increasing. Thus they can’t track of or remember all these files because of the various products and services they are encouraged to use.

“” *I store too much but I don’t want to lose something.*

“” *My files are all scattered. They are everywhere.*

From the user interviews, I was able to come up with 3 key questions that can lead to powerful design opportunity. These questions are:

1. How might a system can track the history or several pieces of information of digital files?
2. How might a user is able to utilize one or multiple information in a proper manner for file retrieval in different challenging conditions?
3. How might a user can see or control all digital items that are scattered over multiple devices and virtual locations?

Also, from the literature review, 2 important design implications for designing personal archive search tool were acquired. These are:

1. Providing one retrieval cue leads to user’s poor task performance and memory. Various types of retrieval cues including location, color, size, and name should be provided to users.
2. The fundamental function of PIM tool should be focused on helping users actually recall the information or regain the memory associated with the information, not just recognize the visual representation of the information. That is, the new PIM tools should empower users so that they can be more actively engaged in retrieval activities.

Generative Process

1. Generative User Scenarios and Low Fidelity Prototypes (Screen + Paper)

With the insights and design implications obtained from previous research, I generated two main ideas with several variations thinking about potential interaction and interface designs. These ideas then were translated into low fidelity prototypes and user scenarios.

The possibilities of these ideas were considered as positive by my advisors, who are experts at design research, environment design, and interface design, so these ideas were continued to be developed further and to be validated through user testings later.

Fig. 10 Low fidelity paper prototypes which were later used for user tests.

Evaluative Process

1. User Testings with Lo-fi Prototypes

Screen-based prototypes were created first based on my preference and then printed out into the papers, with which I could easily test the ideas with research participants. I asked participants to look at paper prototypes with a user scenario at each step. The scenarios were created by the author in a way that no personal information is involved (personal information from either the author or the participants). The documents are fake designs and participants were asked to envision that these documents were created/possessed by them.

Methods

The basic structure of the user testing was verbal interviews, yet I also utilized think aloud method and low fidelity version of Wizard of Oz method. As I walked them through different steps, I pretended to be the system by placing the next piece of the paper prototype. I asked them to speak aloud whatever they think of when imagining they are interacting with the system for real. The main goal here was to gain diverse perspectives on the concepts and get general feedback on the flow, interface functionality, visual design, etc.

2. Design Refining & Generating High Fidelity Prototypes

The user testing with 9 participants allowed me to learn different perspectives. Some responses were insightful. For instance, participants pointed out the area and certain flow that caused confusion, opinions on the features (if the function will be helpful for them or not), and even gave me their knowledge on design conventions.

Based on what I learned, I began refining the design in order to not only reduce user's confusion with the previous interface designs but also make it high fidelity with precision.

Chapter 4
Early Iteration

In This Chapter

With three key design implications obtained from previous research in my mind, I generated several ideas and created low fidelity prototypes. Then I conducted quick usability tests with 9 participants, with printed low fidelity paper prototypes.

The first set of prototypes focused on a) the idea of integrating multiple devices and virtual storage platforms into one place, and b) the idea of allowing users to have a continuous series of conversations with the smart system, which would help them get to the answer by manipulating inputs according to the result.

Experiment Concept 1

Across multiple locations

From the 10 in-depth user interviews, one thing that repetitively arose was that target users are using multiple devices and virtual storage platforms (e.g., email, Google Drive, Dropbox, Box, etc) either to archive their own digital items or to share items with others. And for some of them, different interaction or interface conventions from software and digital platforms make it hard for them to get familiar with distinctive system conventions and guidelines, which often require their commitment and time into learning and familiarizing.

To address this problem, I propose the idea of integrating multiple locations into one centralized system with consistent interfaces and interactions. This would allow users to get a holistic view of what they possess, and better control digital possessions that are scattered across several devices and systems they are using.

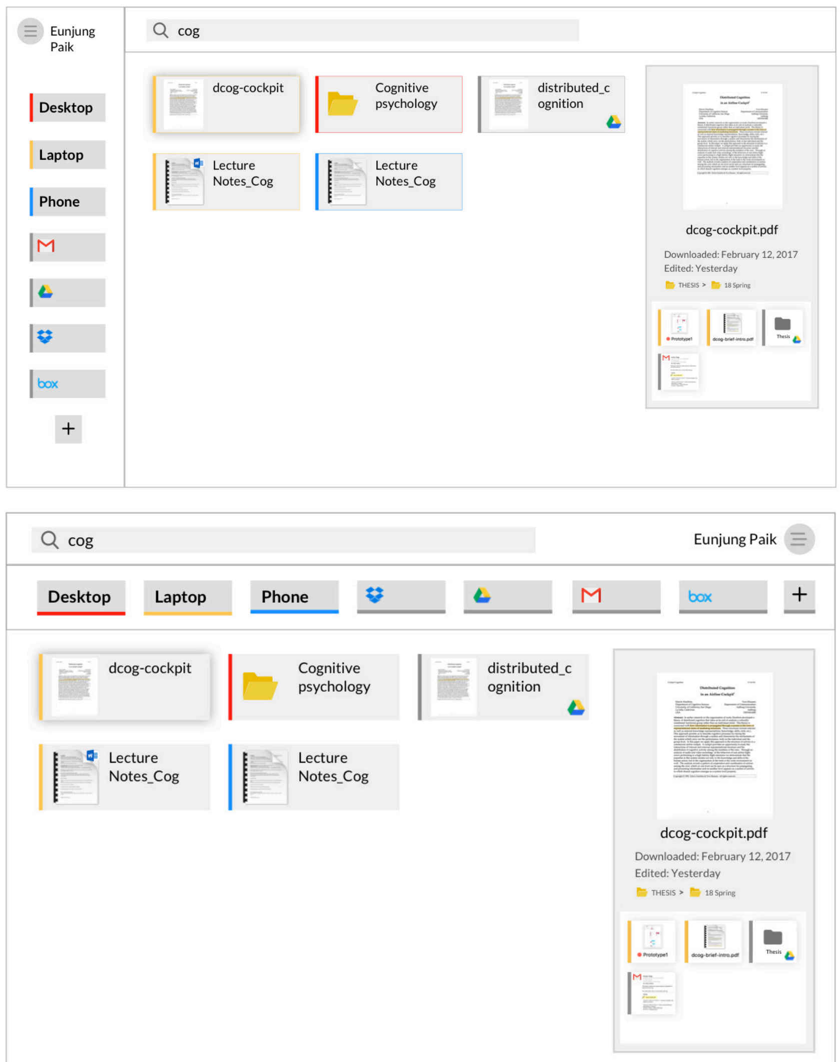
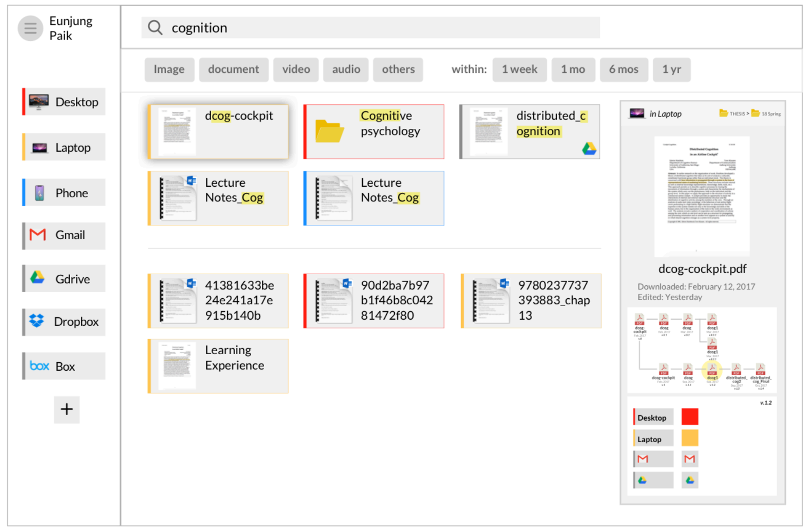


Fig. 10 Low fidelity wireframes. Top: The bar of the icons sits at the left side of the screen. Bottom: The bar of the icons sits at the top of the screen.

Revealing relationships between different files and their locations

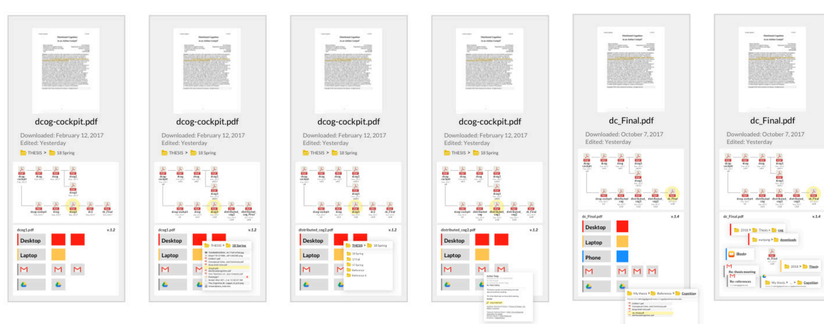
Fig. 12 Low fidelity wireframe with a search filter and information visualization for multiple versions of files with location information.



The information of the related files is added later. This is enabled by metadata which is stored in every digital item. Since the system is designed to keep track of not only the files in different locations but also the metadata stored in files, it is able to visually reveal the relationship between entangled files.

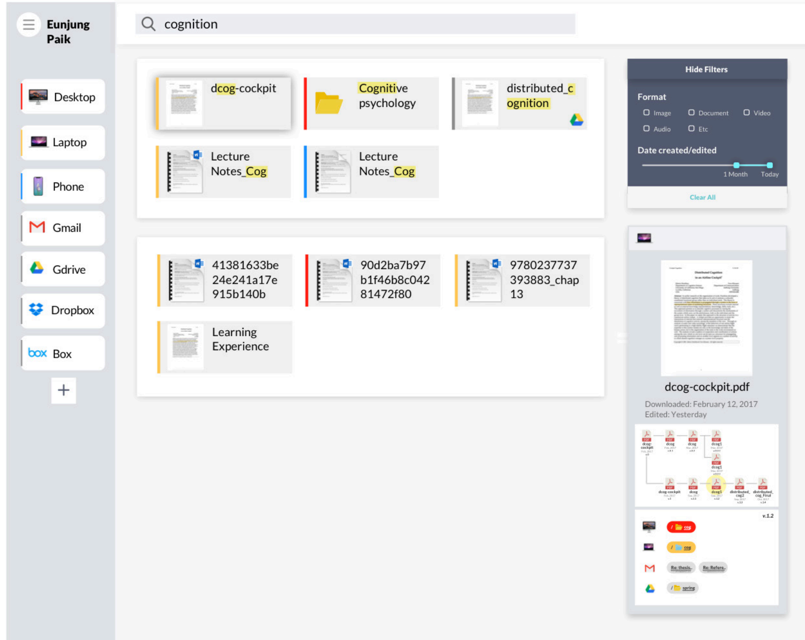
For example, if a user has been creating different variations for one file, the system keeps track of these variations every time a user duplicates an item. Eventually, it provides a visualization of the entire history of multiple files that are related to one file that the user retrieved.

Fig. 13 Iterations on the visualization for related files and locations.



A lot of iterations have been done based on different user scenarios. The user scenarios used here are based on previous in-depth user interviews. In refining this idea, several challenges and questions were considered, such as how the information for different locations can be displayed with designated colors and how much information detail will be displayed.

Fig. 14 Iterations on the visualization for related files and locations.



The inclusion of search filters was also considered and the first approach contains more conventional filter interfaces, with which users can toggle on and off the format of the files and can adjust the time duration when the file was created or edited.

The feature has been developed to the point where users don't need to select or deselect the filters. The system shows the user only relevant files based on what they type. For instance, the system analyzes the types of result files when there is a user's input, and if the analysis indicates that there is no audio file related to the input, the filter for "format > audio" is not given to users. This was inspired by some existing products such as Google image search and Pinterest search system.

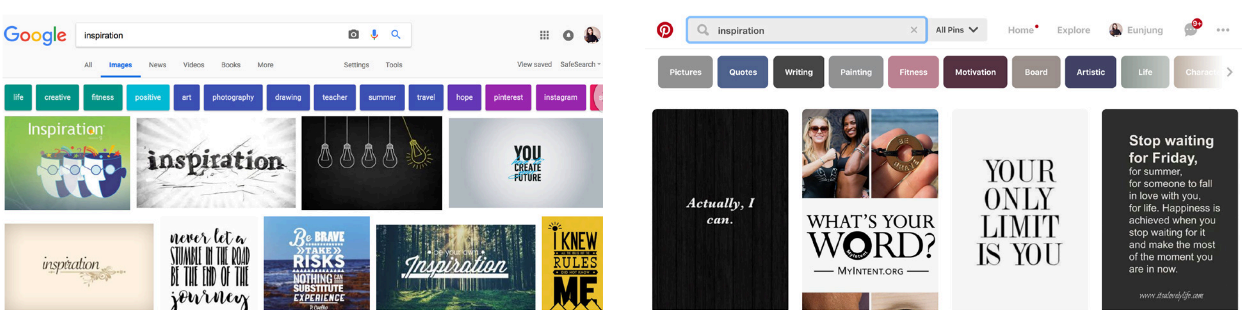


Fig. 15 Metadata enabled image search system. Left: Google image search. Right: Pinterest search. Both search system show "related topical category labels" as a user searches an image.

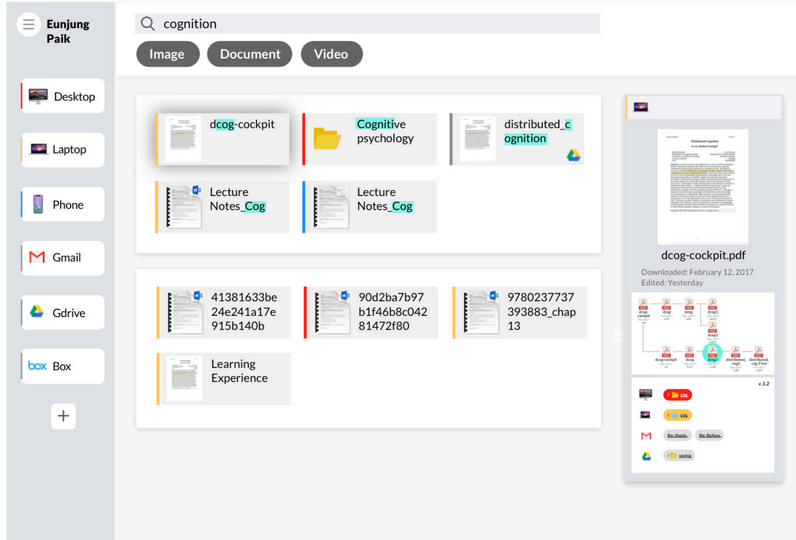


Fig. 16 Low fidelity idea wireframe on smart search filters. A user types "cognition" in the search bar and the system shows relevant filters.

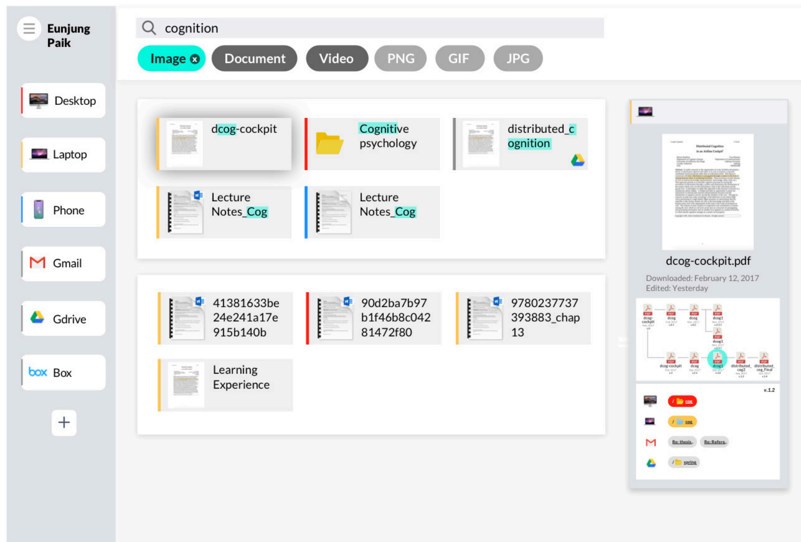


Fig. 17 A user selects an image filter among other filters. The system accordingly shows more detail options for image formats (like png, gif, jpg, etc)

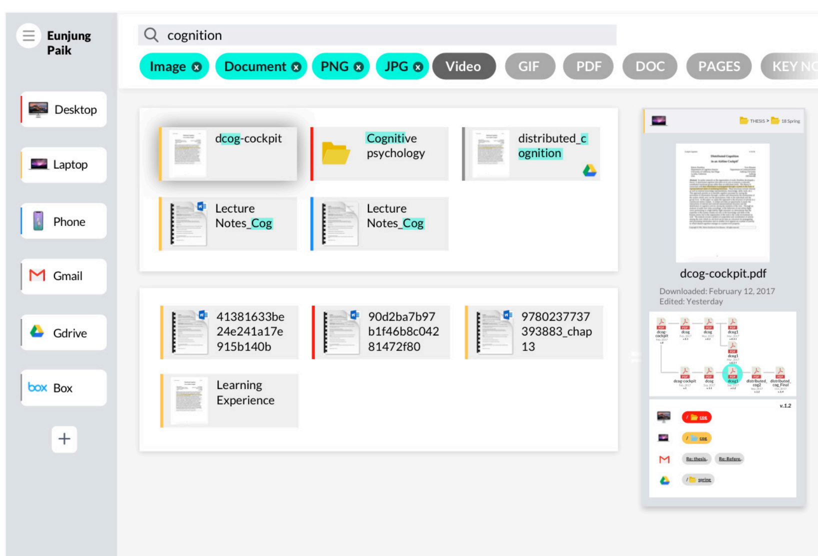


Fig. 18 A user selects an image filter, png, jpg, and document filters. The system also shows more options for document formats (like pdf, doc, pages, etc)

Experiment Concept 2

Deductive Conversation

I was inspired by one of the traditional games called Twenty Questions, which is a deductive reasoning game that focuses on the process of getting to the answer, the idea of a user having a conversation with the system was generated. Simply, this idea is to provide the result quickly based on user's saying and to allow them to adjust the conditions directly until they get to the target items.

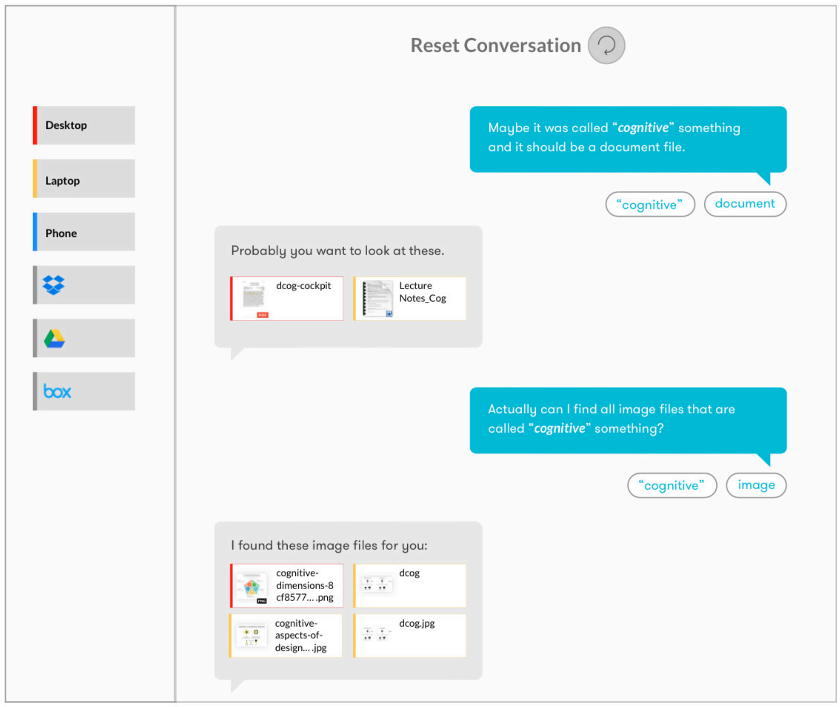


Fig. 19 First iteration on conversational UI and approach. The system picks up the keywords from user's question, which are used for search inputs.

The most relevant user scenario where this approach can be powerful and efficient is when a user has very blurry memories about the target item or content. Taylor et al, in his experiments on twenty question, proves the efficiency of using twenty question for improving learning skills and problem solving ability. The results from his experiments show that "there is rapid learning of the skill involved in the game". In the use of the traditional game, people's "motivation is easily sustained for a period of several days" [15]. It is largely because the entire session is focused on the continuous procedure of problem solving.

Just like the rule of traditional twenty questions where people start with broad questions at the beginning but ends with very specific question at the end, the conversation mode idea is designed to encourage people start file retrieval with vague questions and gradually "increase specificity" of the question until they find the desired item. In searching for the desired one, they are encouraged to ask a variety of questions based on the systems answer (result). It is a searching procedure between the user and the system and from this interaction, the system supports user's successful file retrieval and memory regain towards target file.

15. Taylor, Donald W., and William L. Faust. "Twenty questions: efficiency in problem solving as a function of size of group." Journal of experimental psychology 44, no. 5 (1952): 360.

After 1st Iteration User Testing

After rapidly generating low fidelity prototypes, I conducted user tests with nine participants to get their perspectives on the design concept and interaction/interface designs. They were only given the brief information about the purpose of this tool and the basic scenario for each stage, for instance, "This is where you typed "cog" for finding files" or "Now, you clicked this area". The information about how the interface works was not given to them.

They were asked to guess the functionalities of the features, functions, symbols, etc, yet they were allowed to ask any questions about what they should do at this stage or how they are supposed to interact with it, when they could not understand the design.

Throughout the user testing, most of the participants were thrilled about the fact that they are getting a holistic view over their items. They had no trouble with understanding the different colors for multiple locations.

However, some of them were confused about the order of location information, the visualization of the history of entangled information, and detail features such as bookmarking and time filter.

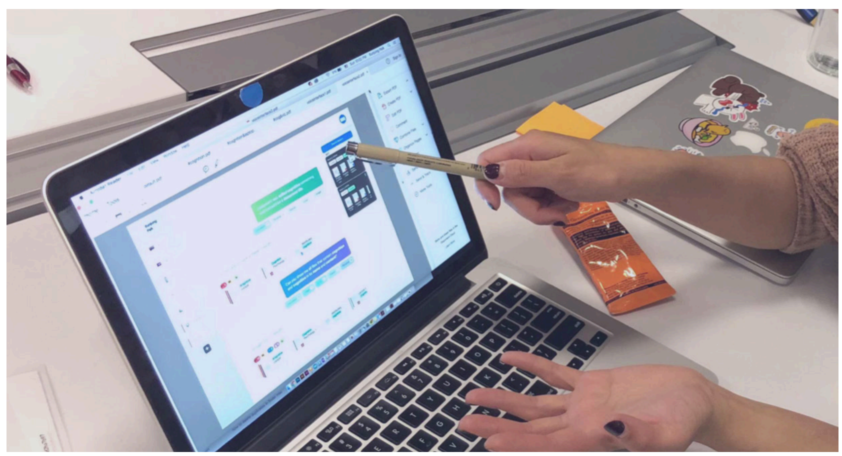


Fig. 20 Participants conducted user tests with low fidelity paper prototypes. The goal of this activity was to obtain various perspectives and feedback on the general concepts, flows, and graphical user interfaces.

Chapter 5
Final Design Suggestion

In This Chapter

This chapter illustrates the final design features and functions with multiple user scenarios that would help you understand how the ideal interaction between the system and a user looks like.

The final design solution is described in this chapter step by step, including its flow, user scenarios, and detail interface designs. This tool is designed to track metadata that is stored in digital files so that users can utilize different types of information for file retrieval and management. It provides the view of all the stored digital files, where they're located, with whom you're sharing, and all the duplications you have.


Desktop
Laptop
Phone
Gmail
Gdrive
Dropbox
Search
image file
last summer
from San Francisco
Document
Final
Shared
Personal
thesis poster_final final 2
Relevancy
Date Created
Date Edited
Type
thesis poster_1
thesis poster_1 copy
poster

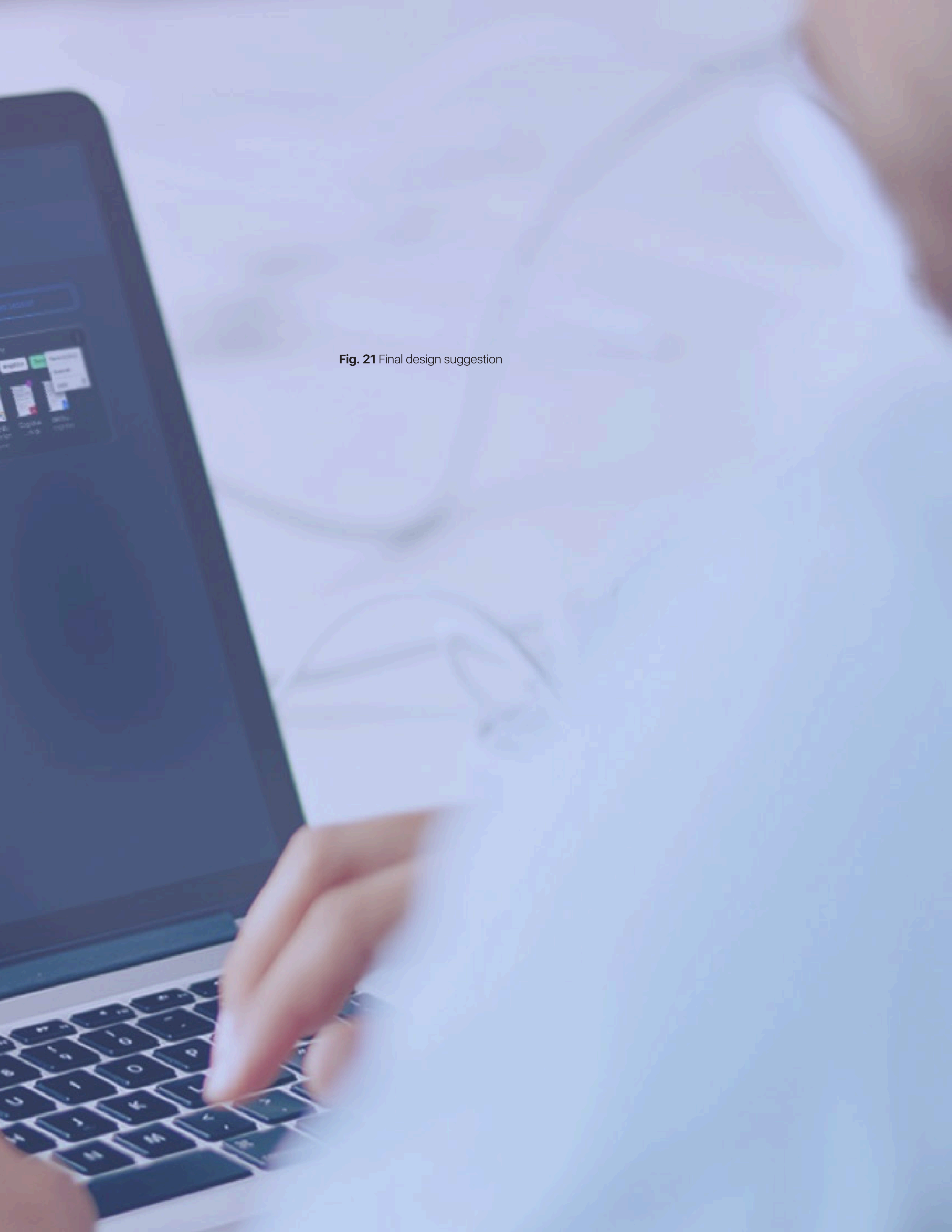
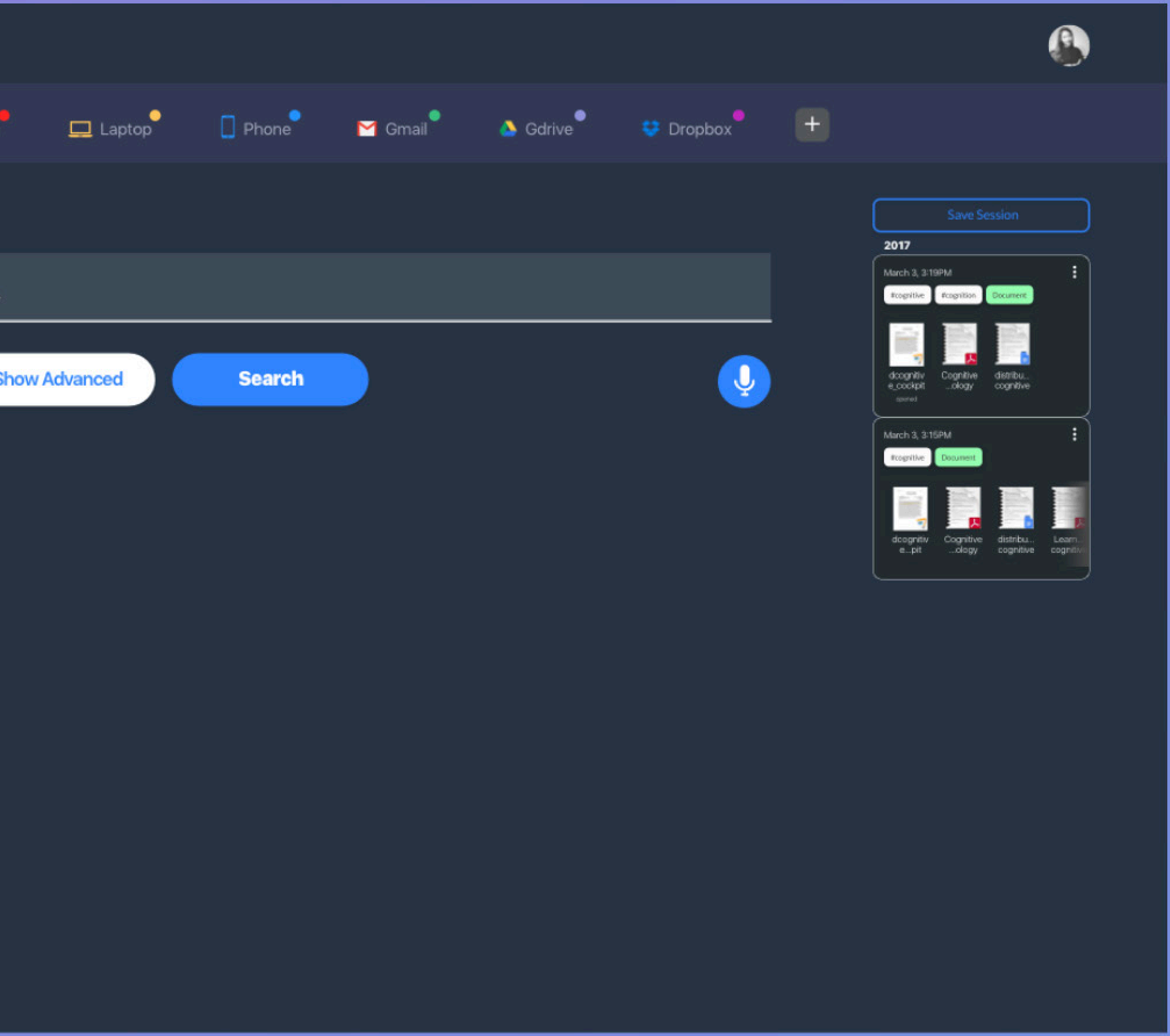
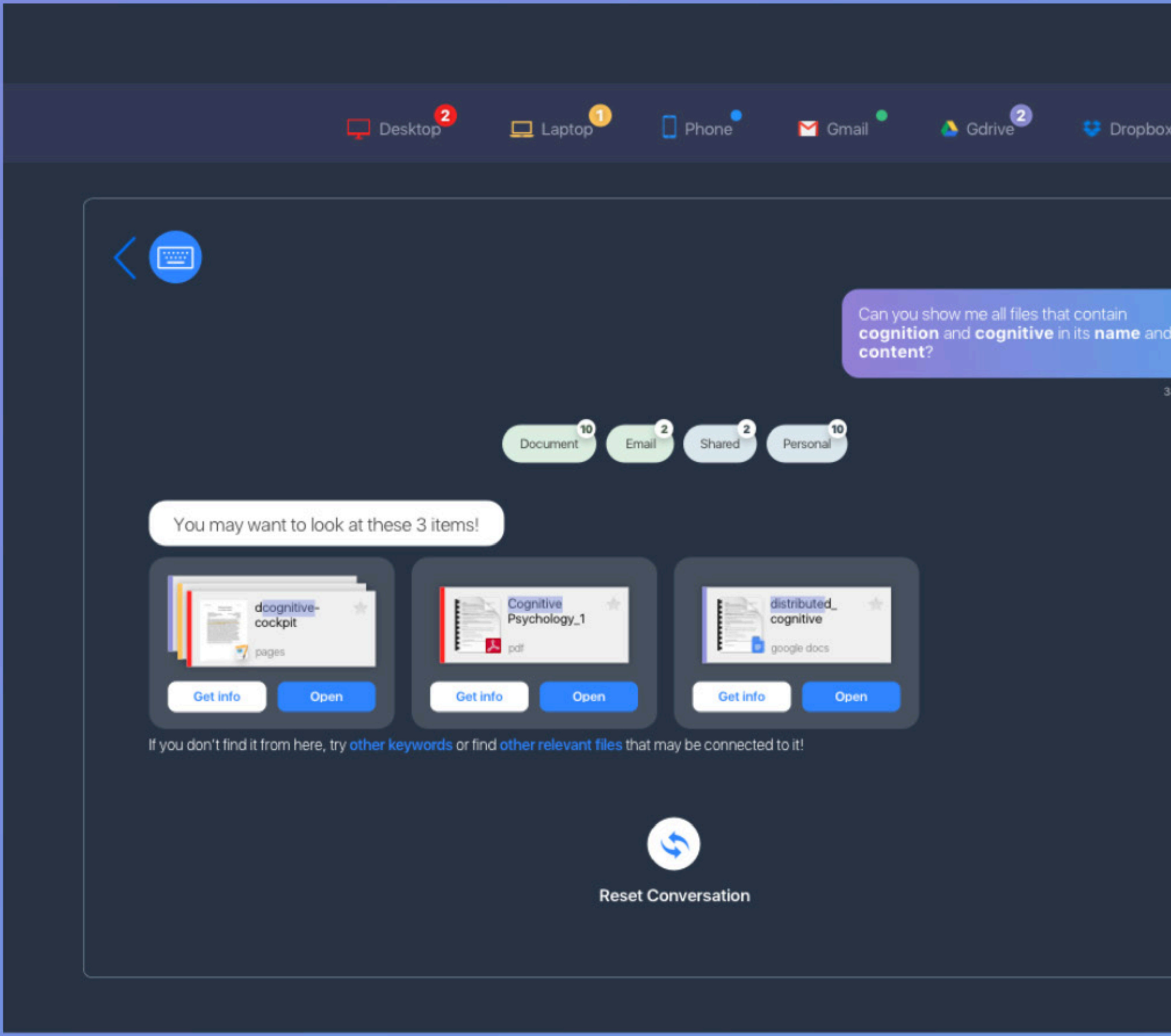


Fig. 21 Final design suggestion


Laptop
Phone
Gmail
Gdrive
Dropbox
Show Advanced
Search
Save Session
2017
March 3, 3:19PM
#cognitive
#cognition
Document
dcognitiv e_cockpit
Cognitive ...ology
distribu... cognitive
March 3, 3:15PM
#cognitive
Document
dcognitiv e...pit
Cognitive ...ology
distribu... cognitive

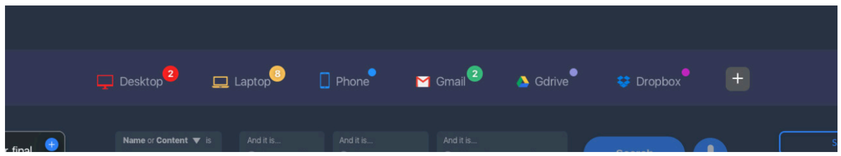

Desktop 2
Laptop 1
Phone
Gmail
Gdrive 2
Dropbox
Can you show me all files that contain **cognition** and **cognitive** in its **name** and **content**?
Document 10
Email 2
Shared 2
Personal 10
You may want to look at these 3 items!
dcognitive-cockpit
pages
Get info
Open
Cognitive Psychology_1
pdf
Get info
Open
distributed_cognitive
google docs
Get info
Open
If you don't find it from here, try other keywords or find other relevant files that may be connected to it!
Reset Conversation

Concept Overview

Metadata Enabled Personal Digital Archive Search Tool is a smart search tool with two different modes: conventional UI mode and conversational UI mode. It helps non-organizing experts, who are active digital item creators, collectors, or sharers, better manage and find their digital possessions that are scattered across multiple locations and are entangled with other possessions and other people. It is primarily designed to provide various flexible ways for users to retrieve content, but ultimately to help them manage items by having a comprehensive understanding of how many useful or non-useful digital items they have.

Main Feature 1

Location Bar



A bar of visual representations of user's personal digital archives sits at the top of the tool. It provides access to the multiple archives that the user uses the most (such as a laptop, a phone, Gmail, Dropbox, etc). Users can add their own devices and virtual platforms.

Designation colors to the registered devices and platforms plays a role in user's recognizing different locations. The bar provides a variety of information based on the search result.

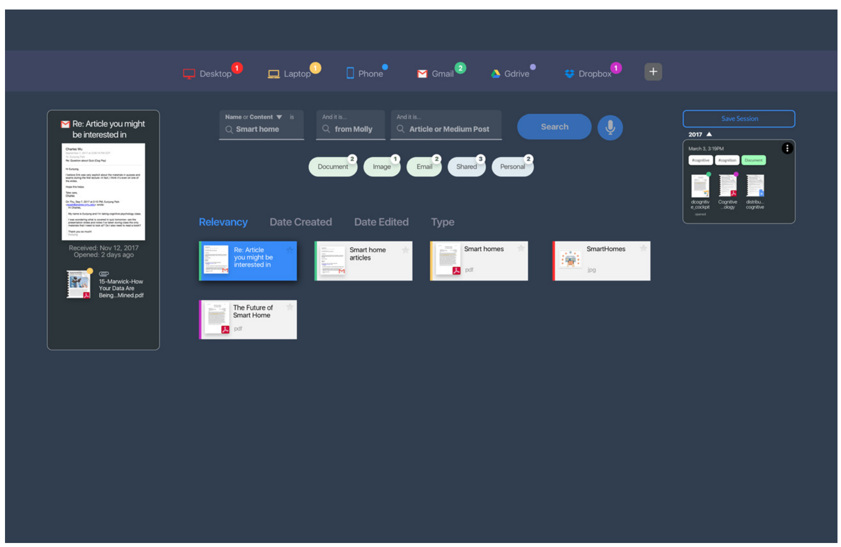


Fig. 22 File retrieval result 1: one file located in user's desktop, one in laptop, two in Gmail, and one in Dropbox.

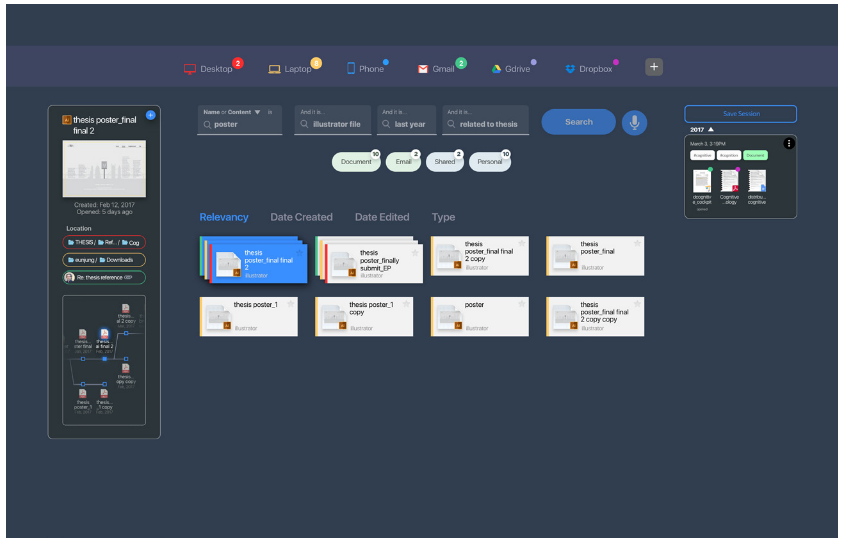


Fig. 23 File retrieval result 2: two files located in user's desktop, eight in laptop, and two in Gmail.

Main Feature 2

Advanced Search

Advanced search feature provides as many search inputs as possible if necessary, in which a user can type in any pieces of information that thye think is related to the target file. The system can sort out files with given pieces of information. The information that the user can provide ranges from time, location, related people (people who are shared with), format, content information, etc. These sets of information are equivalent to the metadata embedded in digital items, which is being tracked by the system.

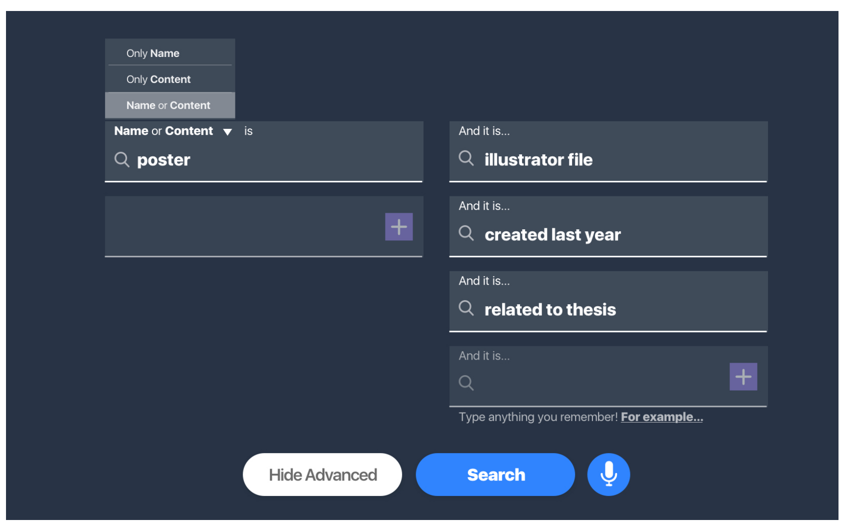


Fig. 24 Advanced search consists of main search bar for the name or contnet, and metadata search bar where diverse information can be typed in.

Main Feature 3

Shorcut for Ephemeral and Working Types of Information

The shortcut feature helps users certain files during the project quickly if they are working on a project for certain duration. Users can save a shortcut based on their save sessions. It allows them to bypass the complicated searching process.

The feature is helpful for dealing with ephemeral and working types of information (Nardi et al. 1995). Ephemeral and working types of information are equivalent to "hot" or "warm" life cycle (Jones et al. 2007). It is considered as important to have ephemeral and working types of information, that are in hot or warm cycle, within easy reach area.

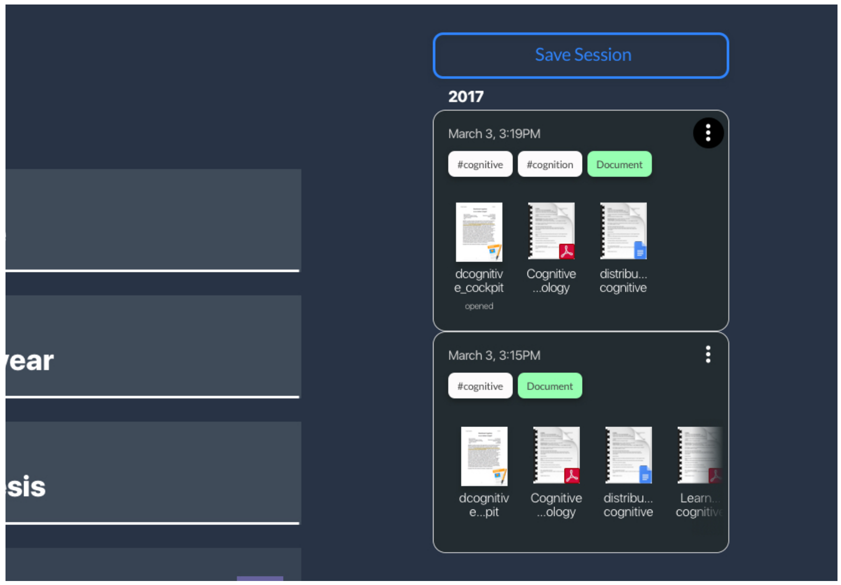


Fig. 25 A user, by saving search sessions, can keep certain files closely so that they can access them without any complicated search process. The function of this feature is analogous to our physical organization strategies of keeping important or frequently used objects near us.

Main Feature 4

History of Files that Reveals the Relationships Between Files

The feature provides information of related file that is entangled with the target file a user is looking at. The related fies can range from duplicated files containing different versions to image files that are embedded in a Adobe file. The feature can be powerful especially when a user failed to name different versions properly or to organize them properly since it reveals the visualization with which users can understand the entire relationship between files and the history of files chronologically.

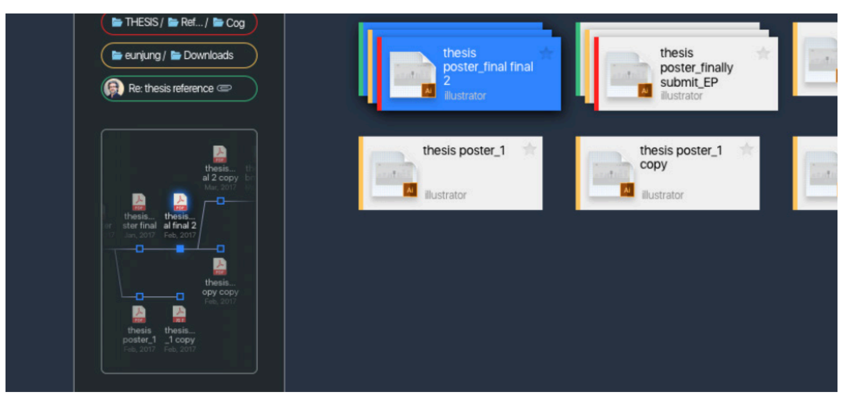
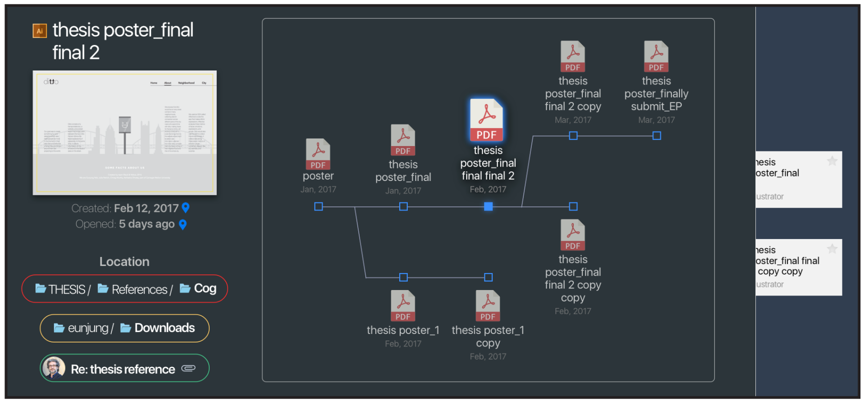


Fig. 26 A user scenario of when a user is confused with the vague file names created by them in the past.

Fig. 27 Expanded view of file relationship and history visualization. It provides interactivity, which users can obtain location information for other related files by clicking.



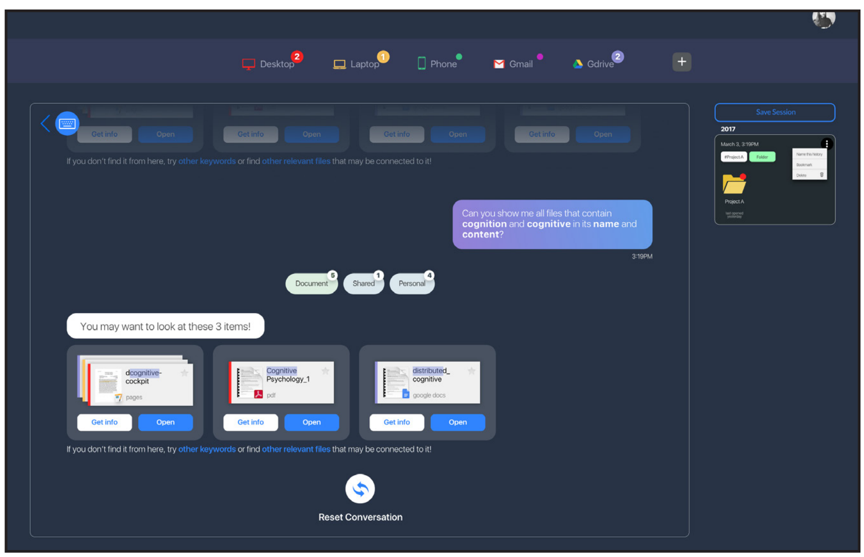
Main Feature 5

Conversational UI

The idea of having conversational retrieval system where users can search files by having a conversation with the system was inspired by Twenty Questions game. Based on the result that the system displays, users are encouraged to modify filters or keywords. The procedure helps them solving the problem, the file retrieval, with the system. As the conversation continues, users are able to be specific about their questions and eventually, they will regain the memory which is associated with the target item.

It focuses on the searching procedure between the user and the system and from this interaction, the system supports user's file retrieval and memory regain towards the target file.

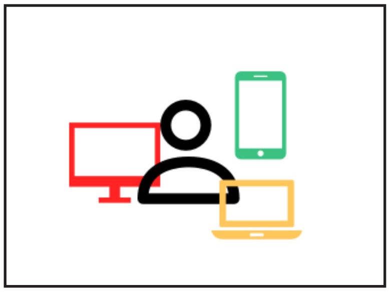
Fig. 28 Expanded view of file relationship and history visualization. It provides interactivity, which users can obtain location information for other related files by clicking.



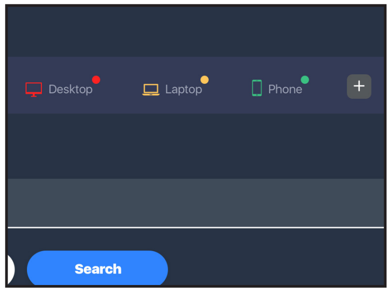
Scenarios

The following is user scenarios with different challenges and conditions. It illustrates how users can interact with the system and take advantage of it.

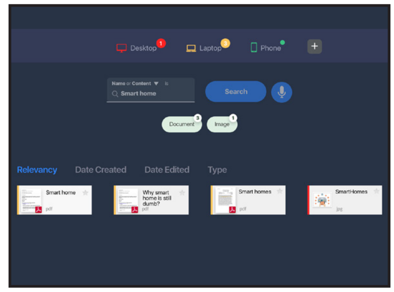
1. Register Multiple Locations



Jane has 3 digital devices, a smartphone, a laptop, and a desktop.

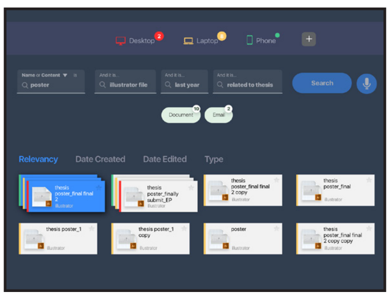


Jane registers her devices and the platforms she frequently uses into the system and designate different colors to them.The system synchronizes all digital files she has.

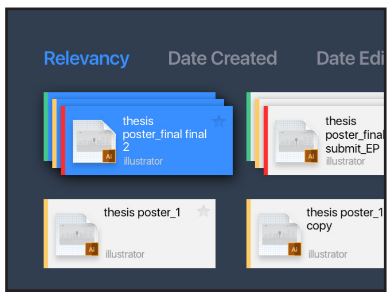


She searches "Smart home" and the system shows her the list of files with location information, the colors.

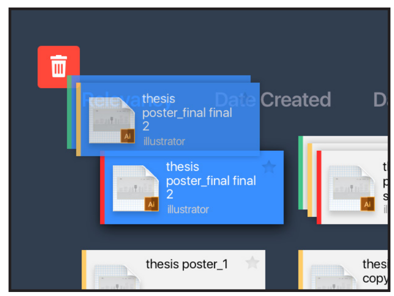
2. Managing Overlapped Content



Jane searches "thesis" and gets a list of files.

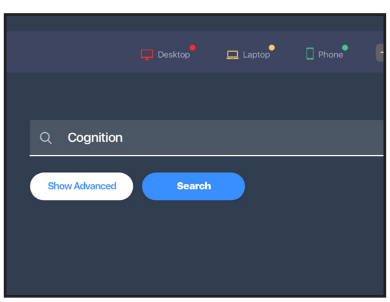


Jane recognizes there is some overlapped files that have same content yet are saved in multiple locations.

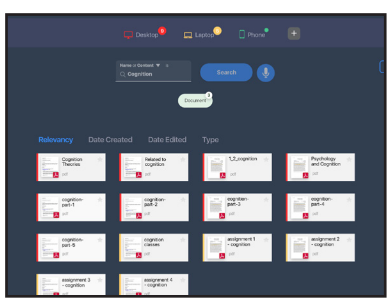


Jane deletes redundant files.

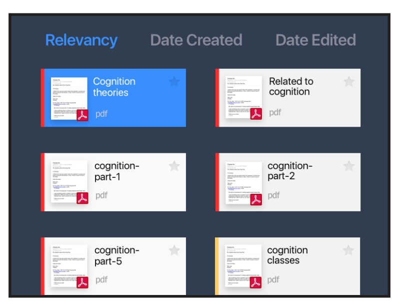
3. Searching with Smart Filters



Jane tries to search every file related to "Cognition", so she types it. She assumes there will be an image file.

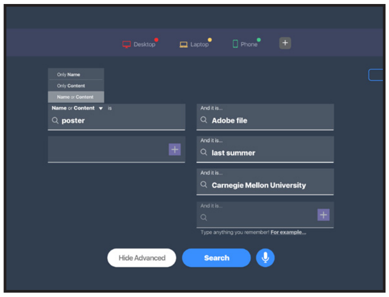


After Jane types, the smart filter system analyzes her files and offers relevant filters that reveal the information of the types and number of files that are related to "Cognition".

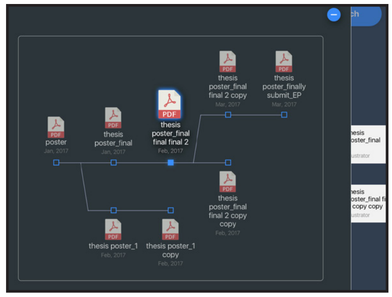


Jane realizes there is no image file associated with "Cognition" in its name and decides to look at document files instead.

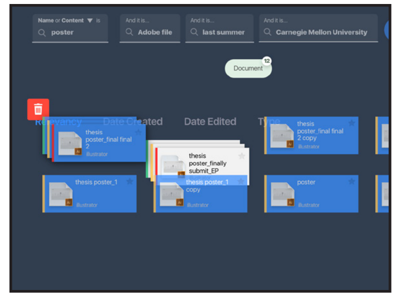
4. Managing Redundant Versions



Jane retrieves design files she was working on a year ago.

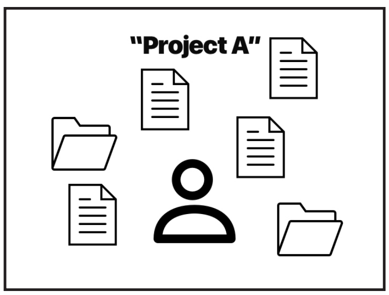


Jane selects one and gets information of related files by clicking. She realizes there are multiple versions of files.

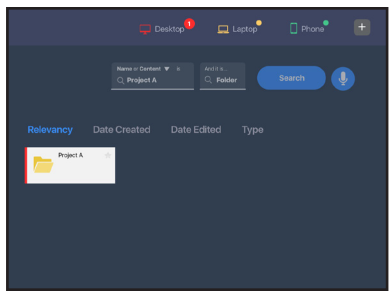


After reviewing, she decides to delete all other versions of files except the very final version of a file.

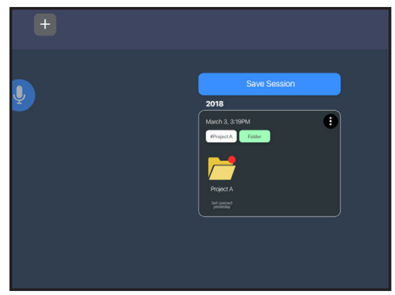
5. Dealing with “Working-on files” by Creating Shortcut



Jane is working on “A” project and until the project is done, she wants to retrieve certain files she has been using frequently for the project.

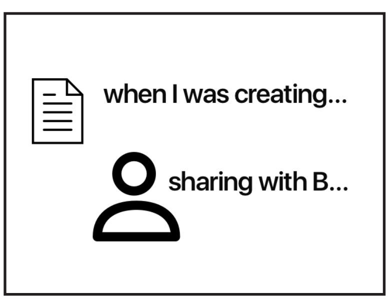


She retrieves a folder she needs for the project that has complicated folder structure. Then she saves the retrieval session for the folder.

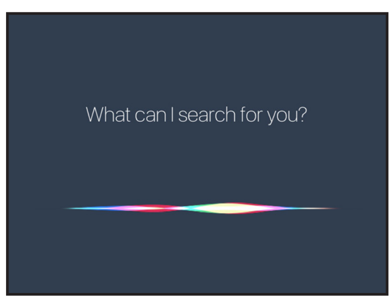


Now she can access the folder easily by bypassing the complicated process. Until she deletes, this folder stays at the left side of the screen.

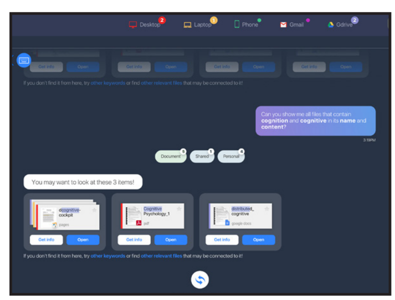
6. Having a Conversation with the System



Jane has only a few pieces of information, which is unclear, about a certain file she wants to retrieve.



Jane uses conversation mode of the system so that she can go through a process of getting to the goal.



By quickly modifying search inputs based on the result, she finally retrieves the desired file and obtains forgotten memory about the file.

Chapter 6
Discussion and Conclusion

In This Chapter

Throughout the process, some possible future challenges and considerations were identified. These need to be taken account into the problem area this thesis is targeting.

First, there is a lack of comprehensive integration across various relevant disciplines in solving the problem. Disciplines like information retrieval, database management, information science, human-computer interaction, cognitive psychology and artificial intelligence should be explored and considered for the issue of personal information management issue.

Second, defining the standard of metadata is critical since a lot of existing products, that are not originally designed for personal archive tools, but are used for archiving. Collaborative platforms such as Slack and Trello are widely used by a lot of global users.

The chapter concludes with the final summary of this journey, from where and how I started, to how I did, and to finally how I designed.

Discussion

Limitation of the Study

Personal Information Management is still a field of inquiry despite a wealth of research and experiments that have been conducted over decades of years. This is because a comprehensive integration across a number of relevant disciplines such as information retrieval, database management, information science, human-computer interaction, cognitive psychology and artificial intelligence has not been thoroughly made.[16]

This thesis focuses on addressing the problem of file management and retrieval in personal digital archives from mainly design perspective. For example, the approach to the problem heavily relies on qualitative design research methods, interaction and interface design, visual representations, and information visualization. For more thorough evaluation and analysis towards the efficiency of this design approach, accommodating diversified approaches and analysis, such as quantitative research, data, and cognitive and psychological experiments with technological implementation is critical. This is because Personal Information Management is a problem area where human long-term or short-term memory is involved. People retrieve personal information based on their memories.

This has not done in this project yet: even though the domains of artificial intelligence and cognitive psychology were explored briefly during the design process, it needs deeper exploration with the aid of experts in such areas. However, through several research such as literature reviews, user interviews, and expert interviews, some important design implications were able to be extracted, which will be covered in Conclusion part.

Standard of Metadata

How can we define the standard of metadata? Should we consider other platforms where people are actively sharing files with others and managing specific content? The expert I interviewed is a librarian at Carnegie Mellon University, Jill Chisnell, who uses the platform Mendeley a lot, which is a Reference Management Software & Researcher Network. This environment is where academic content is dealt with. Also, the librarian mentioned Slack, a cloud-based set of proprietary team collaboration tools and services, which increasing amount of people are using.

Today, there are so many software and digital platforms where people can upload, download, and share their digital content with others. Thus numerous types of metadata can be associated with the content. Also, emerging products originally designed as collaborative work platform and communication platform (like Slack, Facebook Messenger, Trelllo) are used as a space where a group of people can share ideas and digital items. The future consideration should include defining the standard of metadata according to the technology and product trends.

16. Ibid

Conclusion

The thesis began with the problem area exploration with Information Foraging Theory and framing the target area based on that, which is designing a search tool within the personal digital archive.

The findings of the qualitative research on user's item management and retrieval activities have revealed the important design opportunities and implications in which metadata can play a significant role for successful and flexible file retrieval.

1. How might a system can track the history or several pieces of information of digital files?
2. How might a user is able to utilize one or multiple information in a proper manner for file retrieval in different challenging conditions?
3. How might a user can see or control all digital items that are scattered over multiple devices and virtual locations?
4. Providing one retrieval cue leads to user's poor task performance and memory. Various types of retrieval cues including location, color, size, and name should be provided to users.
5. The fundamental function of PIM tool should be focused on helping users actually recall the information or regain the memory associated with the information, not just recognize the visual representation of the information. That is, the new PIM tools should empower users so that they can be more actively engaged in retrieval activities.

With the guidance of these questions and implications, I proposed the final design concept, Metadata-Enabled Personal Digital Archive Search Tool, that tracks metadata stored in digital items and provides various types of information across user's multiple devices and storages for non-expert organizers.

Chapter 7
Bibliography

Barreau, Deborah, and Bonnie A. Nardi. "Finding and reminding: file organization from the desktop." ACM SigChi Bulletin 27, no. 3 (1995): 39-43.

Cockburn, Andy, and Bruce McKenzie. "Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments." In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 203-210. ACM, 2002.

Jones, William P., and Susan T. Dumais. "The spatial metaphor for user interfaces: experimental tests of reference by location versus name." ACM Transactions on Information Systems (TOIS) 4, no. 1 (1986): 42-63.

Jones, William. "Personal information management." Annual review of information science and technology 41, no. 1 (2007): 453-504.

Kaptelinin, Victor. "Activity theory: Implications for human-computer interaction." Context and consciousness: Activity theory and human-computer interaction 1 (1996): 103-116.

Pirolli, Peter, and Stuart Card. "Information foraging." Psychological review 106, no. 4 (1999): 643.

Robertson, George, Mary Czerwinski, Kevin Larson, Daniel C. Robbins, David Thiel, and Maarten Van Dantzich. "Data mountain: using spatial memory for document management." In Proceedings of the 11th annual ACM symposium on User interface software and technology, pp. 153-162. ACM, 1998.

Robertson, George, Maarten Van Dantzich, Daniel Robbins, Mary Czerwinski, Ken Hinckley, Kirsten Risden, David Thiel, and Vadim Gorokhovsky. "The Task Gallery: a 3D window manager." In Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp. 494-501. ACM, 2000.

Taylor, Donald W., and William L. Faust. "Twenty questions: efficiency in problem solving as a function of size of group." Journal of experimental psychology 44, no. 5 (1952): 360.

Yee, Ka-Ping, Kirsten Swearingen, Kevin Li, and Marti Hearst. "Faceted metadata for image search and browsing." In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 401-408. ACM, 2003.

"Extensible Metadata Platform (XMP)." Adobe Captivate - Welcome to the World of Smart ELearning Authoring. Accessed May 13, 2018. https://www.adobe.com/products/xmp.html.