Information Retrieval Behavior on Social Applications

Anyi Dai
RIT
Rochester, NY
ad2599@rit.edu

Joseph Hoeltke
RIT
Rochester, NY
jghisd@rit.edu

Shuishi Fang
RIT
Rochester, NY
Sf8859@rit.edu

Lingfei Li
RIT
Rochester, NY
Ll6067@rit.edu

ABSTRACT
As everyone embraces modern applications and technologies in their daily lives it is important to see how people retrieve the vast amount of digital information available to us. To address this, we conducted user interviews on digital media retrieval with the intent of sharing information with another person. We evaluated our interview findings against other research and came up with some initial evaluations about the problem that gives some potential insight. There are many conditions and possibilities to consider when searching for information.

Author Keywords
Digital information, retrieval, media, IM, instant messenger, UGC, app, application.

INTRODUCTION
Nowadays modern instant message applications allow users to share multimedia content in conversations [Grinter et al 2006]. There are certain moments that a user might wish to send content he or she encountered before during a conversation. Losing digital information has been a problem as long as the internet has existed; it started off when people found it hard to retrieve a single email or an attachment [Denning 1982]. Since then a lot of work has been done to address such problem [Gina et al 2001] yet instant messaging brings new flavor into this issue. In an instant message chatting scenario, the clock is ticking to respond to someone; So, users need to find their content quick. Although there has been formal work done about how to manage personal generated photos [Kerry, Kenneth, 2003], few studies have been done regarding content that people randomly encounter online but want to retrieve later. This research focuses on such moments.

METHODS
For this study we conducted brief interviews with a small group of participants on their current digital media retrieval methods. By doing this we hope to understand more about the users so that we can properly identify current issues. Once these issues have been identified we can consider new potential alternatives or methods for users to retrieve digital media easier.

2.1. Research Questions
We created seven research questions:

1. What types of digital media do they search for?
2. About retrieving digital information in life, can people relate to what are we talking about? Is there a difference between personal and professional use?
3. Do people have an expectation of how much time it will take to retrieve a certain piece of information? If they have:
    a. Will they give up if they think it might take too long to find it? How long is too long?
    b. Are these expectations usually accurate?
4. Do people have certain strategy to find information quicker?
5. Do people tend to archive interesting information more?
6. How long do they typically keep the media?
7. Overall, how do they feel about the current (technology, software, applications) for information retrieval?

2.2. Participants
To ensure the final results of interview would be comprehensive enough considering the wide variety of users; We selected fourteen interviewees from different majors, who were undergraduate students and graduate students from Rochester Institute of Technology. In this group of users, 6 students were international students, and 8 were from the United States.

2.3. Data collection
Selected participants were interviewed individually in relatively quiet environments. In the interview, we asked seven prepared general questions to help participants share their feelings and behaviors of information retrieval which also included some context of the situation. After the interview, to conduct a possible subsequent study, we noted contact information of the participants who would like to join in follow-up research.

2.4. Data analysis
After we interviewed these fourteen interviewees, we created sections for each question with common answers and ratings of how many people answered that question. This included specific keywords, phrases, or mentions which were easier to group the data from different interviewers. According to the organization of data, we generated an affinity diagram which would help researchers classify and conclude behaviors of the information retrieval.

RESULTS

Roughly one third of the participants didn't get what we are talking about. After being explained further and given examples they all said they had similar experiences but 2 participants gave quite off topic information.

The majority of participants (8 out of 14) (see Table 1 for results) said they only retrieved images and gifs, video and music, and web pages that can be represented by a single url, such as literature and Github/Stackflow pages. Four participants also mentioned email and documentation. Very few people (2 out of 14) said they retrieve "all kinds of information", which extend the scope to recipes, forum responds, hand taken notes, scraps of rich content document not as a single file, .dll files, and other unusual things.

Nine participants rely on Google to find information, 7 of them only use Google, which highly overlap with computer science backgrounds. Five participants mentioned using application provided cloud archiving. And one of them mentioned using WeChat history. People who use Google always mention "choose the term you search" or "change to another term and search again." When searching for something they the pick the term that they think will give them the best results; most consider it a skill. One participant said "I hone my (searching) skill over time". People using the cloud archives didn't mention similar things.

The majority of participants (7 out of 14) don't archive information and another 4 participants said they don't archive things unless it is important or noteworthy. When asked to elaborate, they said archiving has changed over the years but nowadays mainstream apps and services always provide functions which include history, notification, timeline, favorites and these allow people to easily reach old piece content. Very few (3 out of 14) users mentioned bookmarking information. Five participants mentioned they archive things locally, using photo album function on their device. One said he keep things in flash drives and another said he will send it via text or email. No matter which way they archive things, all but one user said they don't manage archiving manually and rely on chronological organization provided by default.

The time expectation of finding media can be clearly grouped into 3 groups: less than 5 minutes (57.2%), 15~20 minutes (28.6%), one hour(s) or more (14.2%). However, the people who say they give up sometimes are averagely spread into these 3 groups. If they feel they are spending too much time on retrieving one certain content, they will think maybe give up on it.

Table 1
The prominent findings from the interviews

	Total
Type of retrieved information	
Images and Videos	8
Professional Research	4
All kinds of information	2
Total	14
Time of retrieval	
Less than 5 minutes	8
5 - 20 minutes	4
Above 20 minutes	2
Total	14
Achieve things for retrieving	
Yes	7
-When thing is important	4
-Always achieve	3
No	7
Total	14

In addition to gathering information with these general research questions we planned to have the participants do a walk-through demonstration of them retrieving a piece of information that we asked for. After a few participants tried this unfortunately we decided to cancel this section as it was confusing for the users and they didn't see the benefit to it.

Discussion

There is clearly a gap between the needs of retrieving something and really doing it. It might take seconds for someone to think of something they can share with other people, less than 20 seconds for them to describe the multimedia content to us roughly during the interview, but it take minutes or sometimes even longer to retrieve it share it; and they said it was easy. So what would be hard?

By reading the interview notes and interpreting them inside our group it seems to us that at least one third of the participants don't share multimedia content on a regular basis. They do have experience sharing content when having conversations with other people but it isn't something that normally happens

As mentioned before, it just takes less than half a minute for someone to roughly describe the multimedia content to an audience. During the interview, one of the participants said to us that she will retrieve a picture that shows "a cute dog that is given a birthday cake and it is so excited and it break that cake! It's so cute!" (https://gifrific.com/dog-knocks-down-cupcake/). In this example, even most details are lost, the core idea of the picture is shown clearly and we can recognize from a bunch of similar pictures that this is the one she is talking about.

We also identified a few interesting things. There was a Chinese Instant Messenger app called WeChat, and they provided a special function for multimedia retrieve by providing an album of all the multimedia content inside each conversation (see Figure 1 for example) so later you can go through them very quickly. WeChat is the IM app champion in the Chinese market.

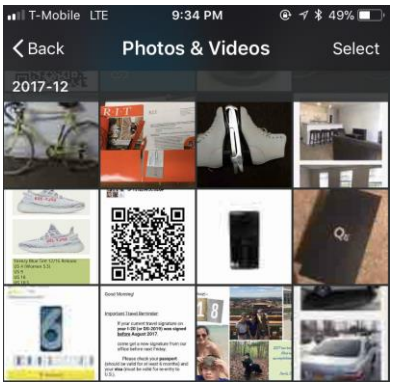

Figure 1: WeChat Images Album inside a Conversation

People who don't retrieve multimedia content from Google will use the application provided cloud archive function, and all those apps are picture/meme/video, which is the main field of what we call "UGC". Movies and music aren't the main focus on the app.

About one third of the participants said "I always encounter situations that I want to retrieve old content", but only one said the chance of retrieving old media is rare; So we don't know what happens to the rest of the participants. Maybe they are not sure about it because they can't identify such scenario so they can't give a comment upon that?

So what kinds of information can be shared by description so you don't need to really find and then share it? What kinds of information must shared to revoke feelings or deliver knowledge that can't be done by words? Thus, there is a possibility that only when people feel retrieving something is necessary during a conversation will they really do it. On the other hand, we know that somethings are harder to find on the internet than others, for example some memes that are less popular are harder to find than others [Xuetao et al 2012][Raphaël et al 2013]. It is surprising that participants have such a high confidence in finding things. Are they really so good at this or is it just because what they set out to find is popular/trending and is easy to locate.

We think it is possible that where you get your daily information is important and subcultures might be related to this. A subculture example could be anime fans, video game fans, photographers, and illustration artists might share and exchange ideas within their community. It will then be harder for them, compared to other people, to find different multimedia content because they aren't related to their interests or isn't popular/trending. The fact people use note taking tools in their professional life to help organize that information they are dealing with from a day to day basis suggests such a possibility exists.

CONCLUSIONS AND FUTURE WORK

When searching for digital media there are multiple things to consider. The user might be basing their search on unique terms or keywords to best find the content or they might simply look until they find what they are looking for. With the introduction of many different social media, instant messaging, and communication applications/tools there are many ways for users to search for their information.

With what we have learned from this initial study there is potential for future work. If we or someone else took this as an initial step and worked around our potential concerns and shortcomings they would be able to narrow down the field of scope into a more concise study.

REFERENCES

1. Grinter, R., Palen, L. and Eldridge, M. (2006). Chatting with teenagers: Considering the Place of Chat Technologies in Teen Life. *ACM Transactions on Computer-Human Interaction*, 13(4), pp.423-447.
2. Kerry Rodden , Kenneth R. Wood, How do people manage their digital photographs?, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, April 05-10, 2003, Ft. Lauderdale, Florida, USA [doi>10.1145/642611.642682]

3. Denning, P. (1982). Electronic Junk. Communications of the ACM, 25(3), March 1982.
4. Venolia, G., Dabbish, L., Cadiz, J., & Gupta, A. (2001, September 01). Supporting Email Workflow. Retrieved February 19, 2018, from https://www.microsoft.com/en-us/research/publication/supporting-email-workflow/
5. Wei, X., Valler, N., Prakash, B. A., Neamtiu, I., Faloutsos, M., & Faloutsos, C. (october 2012). Competing memes propagation on networks: a case study of composite networks. *ACM SIGCOMM Computer Communication Review*, *42*(5), 5-12. doi:10.1145/2378956.2378958
6. Troncy, R., Milicic, V., Rizzo, G., & Redondo-Garcia, J. L. (may 13 - 17, 2013). MediaFinder: collect, enrich and visualize media memes shared by the crowd. *WWW '13 Companion Proceedings of the 22nd International Conference on World Wide Web*, 789-790. doi:10.1145/2487788.2488048