# Creating a Desktop Search Application That Utilizes RFID Ecosystem© and Google Desktop©

**Kyle Rector, Evan Welbourne, Magdalena Balazinska**
University of Washington
School of Computer Science and Engineering
Seattle, Washington 98195
{rectorky, evan, magda@cs.washington.edu}

## 1. INTRODUCTION

The amount of emails and computer files that we deal with daily are increasing. The amount of information stored on one's computer can be overwhelming, and parsing through this information can be very difficult. This is also true for the Internet, but search engines have made surfing the net much more feasible. On the computer, there are filing systems to help organize various files and emails. This becomes less helpful if someone has misfiled something or if they have forgotten the location of a file.

In the event that something is misfiled, one can use a desktop search in order to find the file. However, one has to remember the name or content of the file in order to do so. If a file was made a long time ago, this process could be very challenging. The goal of this project is to be able to use other methods of finding files including using the events involved with developing the file as search parameters.

The RFID Ecosystem provides a wealth of data dealing with locations of people, places, and things. This information can help people remember what they were doing at certain times. This is captured with Vannevar Bush's concept of memex [1]:

> *"…a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility."*

The ability to design tools with this idea can create a more user friendly way to organize emails and computer files.

The RFID Ecosystem Project is a part of the Computer Science and Engineering Department at the University of Washington. This project is seeking to find new applications with RFID, which stands for radio frequency identification. The data that is used with this project is a location associated with a person or object.

We built the RFID Ecosystem Desktop Search, which uses both Google Desktop Search© and the RFID Ecosystem© in attempt to make searching for emails, web history, and files more efficient. The RFID Ecosystem Desktop Search uses times where people were in the presence of others, or physical objects as filters to find a file. An example question one will be able to answer is "Find all the modified files when I met with Jane yesterday." Listed below are the new features that will be included in the search engine:

1. **Be able to use other people that one was with as a search parameter to find a file that was modified during that time.** This will be beneficial, as shown in Dumais et. al. [4]:

   > *"25% of the queries involved people's names suggesting that people are a powerful memory cue for personal content."*

   This can be helpful because other people can be used to filter the search query.

2. **Location of the person while touching or modifying a file or sending an email.** This contribution is beneficial because it creates another link in the human memory to events. It is observed in MyLifeBits that as more data is captured, more correlations can be made to find things [7]. An example of this includes that the user doesn't know when they sent an email, but they were in their office when Bob was visiting. Another positive aspect about this project is that the users do not have to put in any extra effort for this to work, with the exception of wearing an RFID tag.

The other part to the project is creating the necessary GUI for people to use for searching their computer. We have integrated what has worked for GUI's used in related work into the interface, while customizing it to include contextual information. Previous works have organized the results in frames, with a timeline on the left and the corresponding files on the right. According to Ringel et. al's work, they found that people generally liked this format of result [11]. Two other publications have used this as well [2, 8].

The rest of the paper is organized as follows. First, we will discuss the RFID Ecosystem and related work associated with this project, followed by the implementation of the system. Then, we will discuss the evaluation of the system and finish with the conclusion and future work.

### 1.1 RFID ECOSYSTEM

Antennas and RFID tags are deployed throughout the Computer Science and Engineering building at the

University of Washington [12]. The raw data coming from the RFID readers will be sent through a Particle filter which assigns probabilities of location in terms of room numbers, corridors, etc. One application that will be of benefit to the project is called Cascadia. It associates various events with the location data that comes from the Particle filter. For example, the user can develop their own events which include meeting with a particular person. The way that this meeting would be specified is saying that the user was with Joe in room 410. Any event involving the location of the user, other people, or objects can be created [15]. The RFID Ecosystem Desktop Search can use any of these created events.

## 2. BACKGROUND AND RELATED WORK

### Information Retrieval

Vannevar Bush believed that items could be linked together using time [1]. This means that computer activity can be linked to example work events such as a meeting, encounter, or break. Lifestreams provides a good example of indexing all of one's files from the past, present, and notifications for the future [6].

Three publications from EuroPARC have experimented with the logging of user activities. Their first system, PEPYS, tracks people's locations with the use of active badges. It creates a diary of what a person does throughout the day [10]. Instead of using Active Badges, we have implemented the same idea with RFID tags and Cascadia. Cascadia determines the events based on the raw data from the RFID Ecosystem. An improvement is that RFID tags don't need batteries, as opposed to the active badge system. In their research following PEPYS, they added the video taping of three subjects in their daily lives at work for an experiment to see how well they recall events. They seemed to remember more events using the video diary as opposed to their own memory or PEPYS, but the cues that were used in recall are useful. They said that the two major cues were objects and people. The physical location was also useful in aiding recall [5]. The RFID Ecosystem Desktop Search builds upon this by allowing for people, places, and objects as search parameters.

Lamming et. al. [9] reviewed PEPYS and the video diary. They found it would be best to implement the Active badge system in an unobtrusive environment, which is possible with RFID tags. They also thought that the ability to detect events would be beneficial, which is done with Cascadia. They have also concluded that "systems which aim to support human memory retrieval may require special attention to the user interface; otherwise the cognitive load imposed by interaction can outweigh the reduction in load on the user's memory". For this reason, we evaluate our system through a survey and an informal user study, which provide us preliminary results about the usability of our system. In addition, we would like to implement a long term user study in the future.

### Memory Landmarks

One way for people to refresh their memories about files or details about certain events is through memory landmarks. Memory landmarks are events that stick out in one's mind. Horvitz et. al. designed a Bayesian model to predict important memory landmarks based on a study they ran. Their data shows that some important variables in a landmark event include subject, location, attendees, and whether meeting is recurrent [8].

If the meeting is recurrent, it can be less memorable. Czerwinski et. al. conducted an experiment to see how well people remember events in their short term and long term memory, and one of their findings was that people tended to forget the repeatable events [3]. With the search engine, we are providing a way to distinguish the repeatable events by people who weren't at the meeting, or by objects that were brought that aren't normal at the meeting.

### Episodic Memory

Episodic Memory is the idea that memory can be organized into different episodes. Some aspects of an episode may include location of an event, and what happens before during and after [13]. In addition, the memory is also dependent on objects associated and the context [14].

Ringel et. al. made an application based on episodic memory that showed a timeline with events, and corresponding files alongside the events called their summary view. The people generally liked the look of the timeline results, with files next to an event indicator [11].

An event centered timeline can be used to refresh one's memory as to what computer tasks were occurring during specific times in the work day. The RFID Ecosystem Desktop Search uses this concept of organizing search results by the events which the user can specify. For example, if someone wanted to search for files modified during all of the meetings this week, they would receive results organized by the particular meeting, from most recent to least recent.

They ran the experiment by having all subjects browse their summary view with only time and with landmarks and time. There was a statistically significant time savings $p = 0.05$ from time only browsing to both landmark and time together. However, there were no queries made to their summary view [11]. This was also attempted by Horvitz et. al. with Memory Lens Browser [8]. They gathered location data by using their online calendar. There are no guarantees as to whether the calendar events happen or not, so a more reliable system is needed. Since the RFID Ecosystem gets accurate location data, the calendar is not necessary to get event information.

### Desktop Searching

Desktop Searching requires that files on a computer be indexed. One example of work that is done this is with

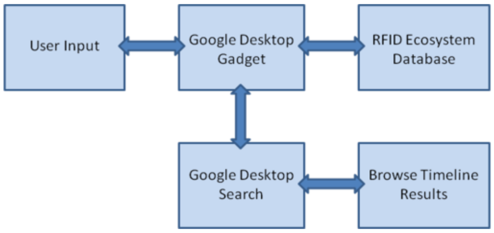


**Figure 1: System Architecture**

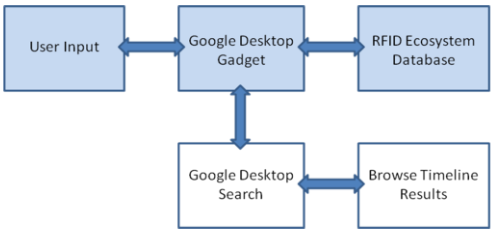


**Figure 2: Portion of system architecture used to configure the database**


**Figure 3: Appearance of gadget while configuring the database connection**

MyLifeBits [7]. However, Stuff I've Seen has created an index of email, files, web, and calendar. In addition, they have integrated the ability to search. They used contextual cues such as time, author, thumbnails, and previews. Some initial findings from their experiment are that time and people are important retrieval cues. 48% of queries involved a filter, most common being file type. 25% of queries involved people, suggesting that people are powerful memory cues. They also found that sorting by date is a good way for people to find items [4]. We are building on this because the RFID Ecosystem provides location data. This source also has a lot of good data that we can try to replicate in our user study.

Phlat was an update on Stuff I've Seen, where they conducted a long term user study. This system is implementing a desktop search using contextual cues. After allowing users to use their system over a long term of time, 47% of queries involved a filter, with the most common being people followed by file type. 17% of queries used only filters. Since they were using computer calendars, aliasing between people was an issue [2]. With the RFID Ecosystem, everyone has only one identity. In addition, we will have more accurate data, since computer calendars may not always be accurate.

## 3. HOW THE RFID ECOSYSTEM DESKTOP SEARCH WORKS

The Google Desktop Gadget© application is coded in JavaScript and xml, the connection to the database coded in Java, and the output page is coded in html.

At a high-level, the gadget enables a user to search through their files. In addition to specifying the file type and keywords as in standard desktop search tools, the user can also specify that the file was viewed, modified, or created within some time window preceding or following a physical event. Example events include meetings, encounters, and coffee breaks. Events are pre-defined and continuously detected by the Cascadia system. For example: Jill is searching for an email that she sent after she met with Joe.

The five required components, shown in Figure 1, for the RFID Ecosystem Desktop Search include the user, the gadget itself, RFID Ecosystem, Google Desktop Search, and the Browse Timeline feature. The following subsections list the steps that the system goes through to complete a successful search.

### 3.1 CONFIGURE THE DATABASE

The gadget connects with the RFID Ecosystem database to get a list of the possible events, and the user then selects any combination of those events. Figure 3, a screenshot, is shown during this process. After it is configured, the gadget sets up the interface accordingly with a drop down list of event choices. Figure 2 shows the interactions of the user, gadget, and database.

### 3.2 FILTER THE QUERY

There are many different options to filter a query. Nielsen has stated this in his article titled "Personalization is overrated". "The real way to get individualized interaction between a user and a website is to present the user with a variety of options and let the user choose what is of interest to that individual at that specific time".

Filters about the file itself include text filter, and file type. However, none of these filters are necessary, since one can filter by event. Event filters include people, places, and objects. An example screenshot at this step is located in Figure 4.

With regards to the example, Jill would specify that she was looking for an email during a meeting event. She would also specify Joe as a person. The gadget will communicate with the database to get all of the possible meeting events with Joe. After she got the available meeting times, she

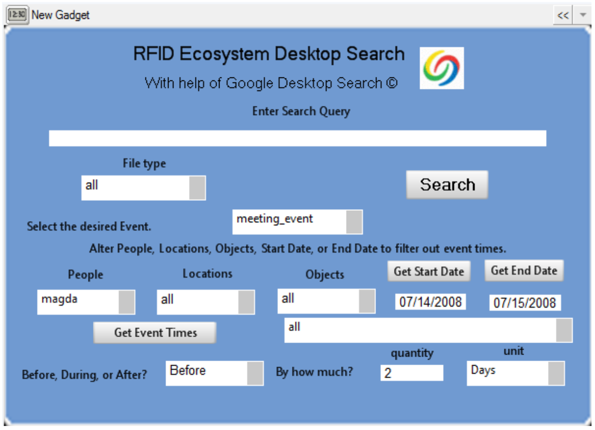


**Figure 4: Screenshot of gadget while filtering a search query**

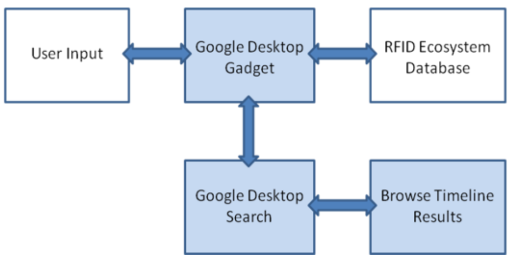


**Figure 5: Portion of system architecture used to execute the search query**

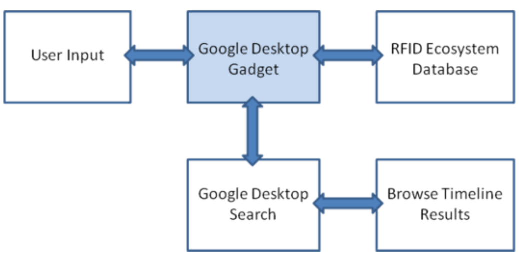


**Figure 6: Portion of system architecture used to display the results page**

would choose to search for emails created up to two hours after the meeting event with Joe. In addition, the choice can be made to filter of a unit of time during, or after the event.

### 3.3 SEARCH THE DESKTOP

The portion of the RFID Ecosystem Desktop Search being used is shown in Figure 5. The gadget then communicates with Google Desktop to continue to Browse Timeline, which has a chronological view of all the file edits over time. In the case of the example, it would grab all of the emails that fit within the specified times from the end of the meetings with Joe to two hours later.

### 3.4 THE RESULTS

The only portion of the system used is the Google Desktop Gadget which is shown in Figure 6. After the searching is complete, a results page pops up. Since many files may show up on a results page, there is the ability to expand and collapse the results based on the different event. In


**Figure 7: Portion of a results page**

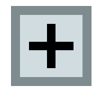

**Figure 8: Sample expand box**


**Figure 9: First prototype of RFID Ecosystem Desktop Search**

addition, the output states all of the query parameters at the top of the page. Figure 7, is an example results page, and a picture displaying the expand feature in Figure 8.

The Google Desktop Gadget was implemented in 1,818 lines of code across 4 files and 37 methods. There were 45 user interface elements involved with the main interface and the options window.

## 4. EVALUATION

A prototype system was created which is shown in Figure 9, and it was reviewed with our peers. Refinements were made to the system and we started testing it. First, we polled laptop users in the CSE building about their habits when using their laptop including location, proximate objects, and human interaction. There were seven questions in the survey. 33 people responded to this survey. Some important points we found and implemented were the importance of objects in their proximity and being able to look at files before, during, or after a work event.

### User Study

In order to prepare for the user study, one of the researchers made fictional event databases, files, emails, and web history. We then developed search tasks for people to try and answer that were both feasible with Google Desktop Search and the RFID Ecosystem Desktop Search. Initially, we wanted to compare how quickly users found the answers to particular questions between the two search engines.

Due to lack of time and participants, we were unable to make the comparisons that we originally wanted to. Two participants were able to go through the scenarios, but only one person was able to complete the experiment, answering five out of six questions. When the participant was using the RFID Ecosystem Desktop Search, they used only the event filters, and file type filters frequently. They frequently found the meeting time that led them to the correct solution of the problem. The unanswered question was from using Google Desktop Search.

Afterwards, an exit questionnaire was provided asking how they liked the system. Based on the two people who filled out this questionnaire, they generally liked all of the features of the system, as well as the overall system. Only 5 out of 56 possible Likert scores between the two participants were below a rating of 5 out of 7. Based on what was gathered, the system is useful, but it is very difficult to replicate the human memory and experiences needed to work with this application.

## 5. FUTURE WORK

### Integrating the System with Cascadia

When Cascadia is up and running real life data can be used to filter for files. Since real contextual events are what people really need to help their memory, this RFID Ecosystem Desktop Search will be much more useful. They users will be able to identify their own events, and the interface itself will be dynamic to fill the different events. This will also allow users to select as many people or objects as possible.

### Long Term User Evaluation

It is important to receive feedback from other people in order to determine how well the search engine has been developed. There will be periodic questionnaires that will have questions about whether it has improved their ability to search for files, and if it is intuitive enough to use. Some important questions that should be asked throughout the study are “did you find what you are looking for”, or “does this result make sense”. This will allow me to answer the question “does the system work?” so it can be used in the future. We will also test to see how well the RFID Ecosystem Desktop Search performs by seeing how long it takes to run a query. In addition, during the user study, the participants will have the option to opt in or out of the study.

### Security

There should be user consent to have their location data to be logged if they choose to participate in the long term study. Once Cascadia is working, a user login will be implemented so they can only use the search if they have successfully logged into the database. All users also have the option to lock their own Google Desktop Search so they can make it even more secure.

### Improve the usability with the interface and results

To make the RFID Ecosystem Desktop Search more intuitive, the user should be able to hit enter to start a search. Another goal is to integrate the output of the search within the gadget itself, so users don’t have to go back and forth. One comment from the informal study is that they would prefer to see files created before, during, and after an event all on the same results page. We are still thinking about how to implement this feature while keeping the results page functional.

### Write the research paper

The end goal is to submit a paper to one of numerous conferences including SIGIR, CHI, Pervasive, or LOCA. There are plans on continuing this into the next fall due to lack of time at CSE. More related work needs to be integrated into this research, but we are encouraged by the amount of support we have found.

## CONCLUSION

So far, we have been able to implement a Google Desktop Gadget which successfully captures files based on simple work event parameters. When testing this interface, we have found it is best to have people try the interface out in their real life work scenarios. It seemed challenging for people to search for files they never created based on parameters with people they didn’t know and events that never existed. If a longer deployment is issued, we want the results will be much more extensive and detailed than currently. However, related work and the survey have supported the different features that we have included in the RFID Ecosystem Desktop Search interface. In the future, we hope to have a successful interface which can be used over long term period of time.

## REFERENCES


1. Bush, V. As we may think *Atlantic Monthly 176*, 101-108 (1945).
2. Cutrell, E., Robbins, D., Dumais, S., Sarin, R. Fast, flexible filtering with Phlat – personal search and organization made easy, *Proceedings in CHI 2006, April 22-27, 2006, Montreal, Quebec, Canada* (2006).
3. Czerwinski, M., Horvitz, E. An investigation of memory for daily computing events. *Proceedings of HCI 2002 230 – 245* (2002).
4. Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., Robbins, C. Stuff I’ve seen: a system for personal information retrieval and re-use, *SIGIR’03, July 28 – August 1, 2003, Toronto, Canada. (2003*).
5. Eldridge, M., Lamming, M., Flynn, M. Does a video diary help recall? *People and Computers VII* Cambridge University Press, Cambridge *257 – 269* (1992).
6. Fertig, S., Freeman, E., and Gelernter, D. Lifestreams: An alternative to the desktop metaphor. *Proceedings of CHI’96*, 410-411.
7. Gemmell, J. Bell, G., Lueder, R. MyLifeBits: a personal

database for everything, *Communications of the ACM, January 2006, Vol. 49, No. 1* (2006).

*8.* Horvitz, E., Dumais, S., Koch, P. Learning predictive models of memory landmarks. In *Proceedings of the CogSci 2004: 26th Annual Meeting of the Cognitive Science Society*, Chicago, USA, August 2004 (2004).

9. Lamming, M., Newman, W. Activity-based information retrieval: technology in support of personal memory.

10. Newman, W., Eldridge, M., Lamming, M. PEPYS: Generating autobiographies by automatic tracking. *ECSCW* Amsterdam, The Netherlands *175 – 188* (1991).

11. Ringel, M., Cutrell, E., Dumais, S., Horvitz, E. Milestones in time: the value of landmarks in retrieving information from personal stores. *Proceedings of Interact* (2003).

12. The RFID Ecosystem Project – University of Washington, CSE. http://rfid.cs.washington.edu/.

*13.* Tulving, E. Elements of episodic memory. Oxford University Press (2004).

*14.* Tulving, E., Thomson, D. Encoding specificity and retrieval processes in episodic memory. *Psychological Review 80,* 352-373 (2004).

*15.* Welbourne, E., Khoussainova, N., Letchner, J., Li, Y., Balazinska, M., Borriello, G., Suciu, D. Cascadia: A System for Specifying, Detecting, and Managing RFID Events. MobiSys 2008.