

Personal digital collections

Background-Research

Z

hdk

Zürcher Hochschule der Künste

Interaction Design, Zurich University of the Arts
Bachelor Project | 6th Semester | Spring 2011

Students

Alain Groeneweg
Philippe Meier

alain.groeneweg@zhdk.ch
philippe.meier@zhdk.ch

Mentors

Prof. Jürgen Späth
Magnus Rembold

juergen.spaeth@zhdk.ch
magnus.rembold@zhdk.ch

Table of contents

LITERATURE

- p.03 Universal knowledge and the deluge of information
- p.04 Classification
- p.04 Taxonomy
- p.04 Conclusion
- p.05 Personal digital archiving
- p.05 Creation / Retaining data
- p.05 File versions and replication
- p.06 Relevance of data
- p.07 Summary

PROJECTS

- p.08 Project 1 - "Snapshots of Provence"
- p.11 Project 2 - "Milestones in Time"
- p.13 Project 3 - "Papers"
- p.15 Project 4 - "Fidgt"
- p.17 Project comparison

CONCLUSIONS

- p.18 Summary
 - p.18 Concept refinements
-

Universal Knowledge and the deluge of information

Human beings have always tended to structure information in order to understand things, giving them the ability to learn and communicate in a better way. Thanks to evolutionary psychologists we start to understand why we naturally have a need for “curating our cultural memory”.

“Information systems are named as essential tools for survival and adaption”¹

Without this adaption and creation of cultural artifacts, the survival of human seems impossible. Another huge information deluge occurred about 5000 years ago, where the alphabet writing came into existence. In this case people started to live in even bigger settlements because of agricultural progress and better environmental conditions. In commercial activities the alphabet was probably first used, in order to keep track of trade.

Eventually, those documents evolved into government records, religious texts, and historic books. The growing alphabetic output gave rise to the first institutional ‘knowledge bureaux’, impersonal social entities like temples, governments and schools that came into being in purpose of collecting, organizing and distributing a growing store of collectively generated data.²

With the invention of the press by Gutenberg, new forms of organizing information was again necessary due to the fact of its overwhelming mass. Among others, first scientific methods were the consequences. The first attempt of an Encyclopedia with the goal of storing all human knowledge was initiated by Diderot. Historic and current forms of Encyclopedias reflect the spirit and condition of an age very well. But the effort to encounter more and more information is diametrically opposed to the longing for absolute knowledge.

“Universal Knowledge will never happen, we will just have better ways of finding what’s around. “Universal knowledge” sounds like there is an end and that we’ll get there. But it’s one of these things that, if you get very close, it has already moved further. It’s a moving target.”³

When it was proclaimed that the Library contained all books, the first impression was one of extraordinary happiness. All men felt themselves to be the masters of an intact and secret treasure. There was no personal or world problem whose eloquent solution did not exist in some hexagon. The universe was justified, the universe suddenly usurped the unlimited dimensions of hope.... As was natural, this inordinate hope was followed by an excessive depression. The certitude that some shelf in some hexagon held precious books and that these precious books were inaccessible, seemed almost intolerable.⁴

“We are living in the midst of a digital dark age”⁵

“Who controls the past controls the future. Who controls the present controls the past.”⁶

Nowadays the digital age has completely changed the definition of space, societies and communication. The deluge of data challenges us in terms of how we interact with the new environment. How do we know which information is relevant to us and how can we organize it?

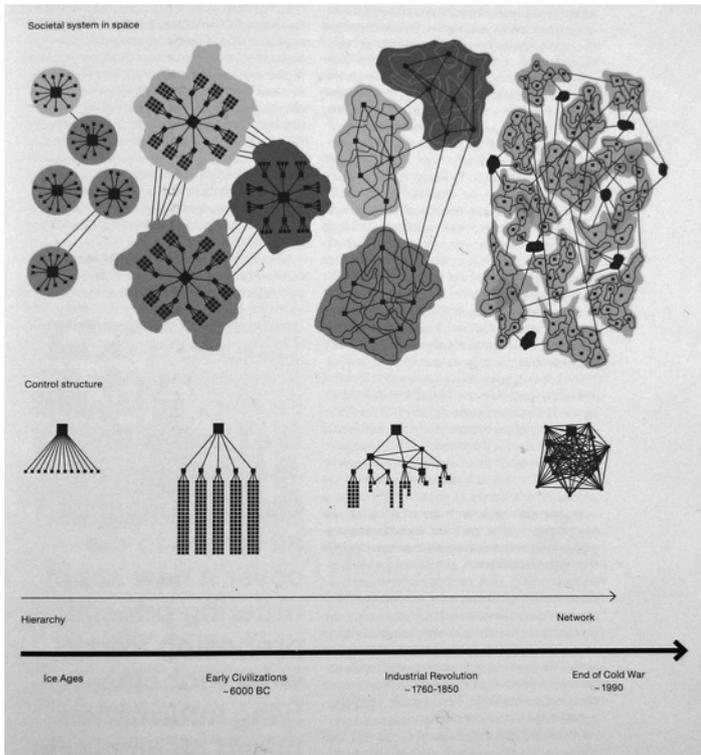


Figure 1 – Designing Universal Knowledge, 2009, p.113

On the graph above you can see how social structures have gained complexity over time. The link between social structures, forms of communication, the arise of cultural artifacts and new information systems can be made.

Based on Alex Wright

(a writer and information architect for The New York Times):

We know a major shift in paradigm 25-35'000 years ago. Here you can find a fundamental leap in human-based information-systems. Change in environmental conditions forced human beings to start new forms of living together. It was the beginning of the Ice Age. To survive in that time, it was necessary to stay in groups to be able to hunt down big animals. Also the warmth of more people living together in caves was necessary. Those small evolving social structures needed methods of communication and shared knowledge. Drawings on the wall of caves documented and taught others.



ARCHIVES

¹ Alex Wright in “Gerlinde Schuller’s Designing Universal Knowledge“, 2009, p.111

² Alex Wright in “Gerlinde Schuller’s Designing Universal Knowledge“, 2009, p.112

³ Richard Saul Wurman in “Gerlinde Schuller’s Designing Universal Knowledge“, 2009, p.109

⁴ The Library of Babel – Jorge Luis Borges, The Library of Babel, 1941

⁵ T. Kuny, „A Digital Dark Ages? Challenges in the Preservation of Electronic Information,“ International Preservation News, 17 (May 1998).

⁶ George Orwell, Nineteen Eighty-Four, 1949

Classification

Different kinds of information and different amounts of data need different organizational structures. The main goal in our opinion, is to find an item fast and in a consistent way. When you look at libraries they have established content-specific hierarchical structures. These taxonomies indicate the amount, the content-specialty and the media-variety.

A classification is a spatial, a temporal or a spatio-temporal segmentation of the world. A classification system is a set of boxes either metaphorical or literal into which things can be put into enabling some kind of bureaucratic work or knowledge production. In an abstract ideal sense a classification system exhibits the following properties:¹

- There are consistent, unique classificatory principles in operation.
- The categories are mutually exclusive.
- The system is complete.

Taxonomy

Taxonomy is the practice and science of classification. Typically taxonomies contain hierarchical parent-child relationships. In such an inheritance relationship, the subtype by definition has the same properties, behaviors, and constraints as the supertype plus one or more additional properties, behaviors, or constraints. For example: car is a subtype of vehicle, so any car is also a vehicle, but not every vehicle is a car. Therefore a type needs to satisfy more constraints to be a car than to be a vehicle.²

*common parameters in organizing data:
location, alphabet, time, category, hierarchy*

*possible additional parameters:
personal events, relevant files at that time.
other saved data at that time, format,
recipients of files (e.g. email), file versions...*

Richard Saul Wurman on

“Services which distinguish between junk and high quality information”:
...“There are existing services but people still going to google, because it's comfortable. There will be some breakthroughs. There is always somebody who breaks through the obvious and the obvious will be challenged and somebody will do the opposite of it.”³

Conclusion

Mostly when we think of archiving, we think of a relatively high amounts of data being stored. Archives often connote a certain institutional background. Results of scientific research dominate the way how data has and will be stored. The archiving-taxonomies are mostly based on parameters in forms of terms and dates. If we think of individuals digital-data-collections one can find more diverse structural systems. Every individual has his or her very special way of storing data. But here we encounter a certain limitation with our operational systems. What we can say is that archival-techniques itself is something we can manage. The main problem is how we access items. The infrastructure and environment of storage is eminent in order to find relevant information. We tried to omit the word knowledge intentionally. We believe that the term knowledge is often used ambivalently. If we think of archiving, the archive items are always information not knowledge. Knowledge is nothing physical. Information will emerge to knowledge when dealing with it, if you can link them with existing experience, if you can communicate it. Your personal motivation and interest are key-factors of doing so. We also think that personally stored items have a higher potential containing personal relevant information because the chances are higher that you have already dealt with them.

Especially if you start to develop a special way in structuring your data, it tells you something about your personal relation to that file.

In the upcoming section we try to focus on personal storing styles, to find out more about information relevance.



ORGANIZING DATA

PERSONAL
RELEVANCE

¹ Geoffrey C. Bowker and Susan Leigh Star in “Gerlinde Schuller’s Designing Universal Knowledge”, 2009, p.30

² www.wikipedia.org, 2011 <<http://en.wikipedia.org/wiki/Taxonomy>>

³ Richard Saul Wurman in “Gerlinde Schuller’s Designing Universal Knowledge”, 2009, p.109

Personal Digital Archiving

Digital-data and its personal value over time.

During our lives we store huge amounts of data. Personally stored digital artifacts always carry a certain personal value. Some files have more value and others have less, but how can you tell which data is important to you at a certain time?

As we look upon our personal digital data we realize that there are huge differences in those artifacts. The differences are where they are stored, the kind of format they have and in which context they are embedded. Common folder structures result in insufficient organisation.

There is a huge gap in storing, usage and finding files. Recently produced files are often more relevant to us because our short term memory allows us to make direct associations. A good example is the current files lying on our desktops. When you look at them you will likely be able to recognize what they contain and in which context you created them. But this does not reflect their importance. In my case there are a lot of recently produced project files lying on my desktop which I sent to a customer. These original files were generated and exported through a creation program. The desktop in my case therefore just gives a brief overall view of what I did recently but the original files reside in complex folder structures. Right now the files on my desktop have a certain relevance but they won't be used again because the original versions still exist. If there is too much data on my current desktop files or it doesn't seem worth saving elsewhere, then those will be the first items deleted.

This is an example of how the importance of the context of an item can be completely change. Versions and copies of files are spread all over different systems and devices. The one file on my desktop was only relevant to me from the time I created and worked with it and until it was sent to the customer. Afterwards the file lost its context and became worthless to me. Indicating the relevance of current items seems to be a big challenge. However, we believe that in future every file will indicate more information about its relevance to its current user.

Creation / Retaining Data

Due to the massive creation and distribution of data in the last 20 years, we have to ask ourselves what happens to the produced data. Storage places are getting cheaper and more and more services are providing storage. With every new device data seems to get distributed to even more places.

For example people don't have to delete any of their emails anymore because the specific services offer almost unlimited free storage. The process of selection of preserving only relevant data seems to have disappeared. Even in back up situations people just migrate their current system to another one without selection. The loss of data is naturally rated as more dangerous than not finding the data anymore. As we have seen, digital data is created and stored with little regard to long-term preservation. Archiving should start with its creation when thinking of storage and long-term preservation.

Digital Data is fragile and can easily be corrupted, copied or altered. With the same ease and speed that digital data is created you can also lose track of them. Regarding long-term preservation there are even more problems which will occur. File types or software get replaced by "better" and new ones and new software probably will emerge so one could lose access to their old files as well. The challenge here is to explicitly or absolutely put a value on your personal files, created by oneself or found and gathered.

"Some types of information, such as multimedia, are so closely linked to the software and hardware technologies that they cannot be used outside these proprietary environments"¹

*"The greatest temptation is to keep almost everything and sort it out later"*²

File Versions and Replication

"Different copies are put in different places for different reasons, and all of these copies are not equal."³

Cathrine Marshall talks in her article of two properties of distributed storage:

1. There is a need to keep track of all of the various copies of any specific item.
2. It is vital to record the provenance of each item.

The original stays on your computer but are you sure you will find it again? With your digital findings we got similar problems. You save a copy from the web. Can you retrieve files, once stored, from the same source again? E.g. in case of not finding it anymore in your archive or in order to find its creator, its former context and different versions of the file.



THE GREAT
TEMPTATION

PERSONAL VALUES
AND DIGITAL DATA

¹ T. Kuny, „A Digital Dark Ages? Challenges in the Preservation of Electronic Information,“ International Preservation News, 17 (May 1998).
<<http://www.ifla.org/IV/ifla63/63kuny1.pdf>>

² Catherine C. Marshall "Rethinking Personal Digital Archiving Part 2" D-Lib Magazine, March/April 2008, p.2
<<http://www.dlib.org/dlib/march08/marshall/03marshall-pt2.html>>

³ Catherine C. Marshall "Rethinking Personal Digital Archiving Part 2" D-Lib Magazine, March/April 2008, p.5
<<http://www.dlib.org/dlib/march08/marshall/03marshall-pt2.html>>

Relevance of data

“While it is important to be able to decode, render, and interact with a lifetime’s worth of digital objects, and crucial to develop repositories that are trusted, robust storage for these digital objects, it is not enough to stop there.”¹

We would like to give brief summaries on the following question which Catherine C. Marshall also thought about¹:

1. What should we keep?
2. Where should we put it?
3. How should we maintain it?
4. How will we find it again?

1. What should we keep?

Catherine C. Marshall claims as we did that because of smaller devices and increased storage places, people tend to keep all their collected or produced data. But she also demonstrates the need of peoples to control their digital belongings. People want the ability to find their files depending on their value. She mentions the “value spectrum”: ...“only small number of items that are valuable in a noteworthy way and to be able to care for them accordingly – and to delete other items in a way that makes them irretrievable.”

The most problematic items are everyday items such as holiday photos and emails to friends. She calls them “life souvenirs”.

Marshall: “How can we ensure that an adequate number of these things survive and are findable without overwhelming an individual or making her feel that she’s lost control of what she has?”

Marshall introduces three types of indicators that have bearing on an individual item’s value but the stability of value is not guaranteed:

The source. importance of the origin, how did the individual get it, could it have created emotional value to the user?

Actions. modification by the user, usage count

Disposition. re-distribution, storing-location, shared data

Marshall mentions the context of items:

“Context makes it easier to distinguish between items that are valuable and items that have simply accumulated.”...“The trick is being able to tell the difference between diminished value and benign neglect: the item is still valuable; it just has been forgotten or not handled in awhile.

...it will be necessary to automatically capture and store new kinds of metadata instead of relying on an individual’s explicit assessment of the item’s value. Intrinsic metadata may be automatically collected based on user activity, device properties, or environmental sensors.”

2. Where should we put it?

According to Marshall, individuals store their data in a lot of different places. She mentions the uncertainty among people causing them to not be aware of how safe a storage place is and where privacy issues occur.

“Storage decisions are often based on the functionality, security, and access offered by external sites.”

Here we claim that, based on the uncertainty, and also because of the not-knowing where a file has been stored, the deluge of data could even accelerate. The way of thinking: “I don’t know if I have a back-up of this file, therefore I will upload it to another service.”

“People in our study lost digital belongings not through technological catastrophe, but rather through minor negligence”²

“Individuals should be able to find out where an item is and how many copies they have of it.”³ Another important topic is the version of a file:

„Different Copies are put into different places for different reasons and all of these copies are not equal. After a while it seems impossible to keep track, which version is the original or which is just a copy or version of it. The origin of the file is a important indicator.“³

„It is clear that we need to develop better mechanisms for recording an item’s provenance, maintaining it over time, and presenting it when the object is recovered from long-term storage.“³

3. How should we maintain it?

The migration of data is a fundamental issue.

What happens with data over time?

Data-types change over time and “embracing” formats come into existence (e.g. pdf). Data is depending on Software.

Migration is the most common reason for data-loss.

“Finally, it should come as no surprise that curatorial tools should take advantage of distinctions in genre. Photos should be tended as photos, records as records, and movies as movies. Curatorial tools and standard digital formats should vary with genre and with the individual’s commitment to maintaining the collection. This variability in curatorial requirements supports a model of distributed storage and access – it is likely that different storage venues will provide different tools for managing the material stored there.”⁴

Cathrine C. Marshall talks here about using curatorial tools which should all be made individually based on the characteristics of the media it is curating.

When she says tools does she think of independent programs?

In our opinion there is a need of taking into account the different media but it could still be embedded into better tools for supervising these different formats. In the same way with the spread of very small programs for everything you could loose track and the lost consistence of an interface could weaken the usability. Here we will try to create a tool which is not bound to a certain format but to the context and interest of the user.

¹ Catherine C. Marshall “Rethinking Personal Digital Archiving Part 2” D-Lib Magazine, March/April 2008, p.1 <<http://www.dlib.org/dlib/march08/marshall/03marshall-pt2.html>>

² C.C. Marshall, F. McCown, and M.L. Nelson, „Evaluating Personal Archiving Strategies for Internet-based Information. “ Proceedings of Archiving 2007, Arlington, Virginia, May 21-24, 2007, Society for Imaging Science and Technology, Springfield, VA, 2007, pp. 151-156. <<http://arxiv.org/abs/0704.3647>>.

³ K. Muniswamy-Reddy, D. A. Holland, U. Braun, and M. Seltzer, „Provenance-Aware Storage Systems.“ In Proceedings of the 2006 USENIX Annual Technical Conference, Boston, MA, June 2006. <http://www.usenix.org/events/usenix06/tech/full_papers/muniswamy-reddy/muniswamy-reddy.pdf>.

⁴ Catherine C. Marshall “Rethinking Personal Digital Archiving Part 2” D-Lib Magazine, March/April 2008, p.11 <<http://www.dlib.org/dlib/march08/marshall/03marshall-pt2.html>>

4. How will we find it again?

According to Marshall we need more defined and limited digital space to find data: "One thing the personal computer and subsequent services have failed to offer is a stable sense of digital place. What I mean is, the digital equivalent to the box under the bed..., a place where valuables are kept, You have no need to remember what's in the box, just that it's where the valuables are."

Virtual geography as an approach to know where the valuables are: "This needn't be an actual geography, but rather a set of digital places that form a virtual geography, a mode of remembering where things are. Digital geographies may follow in a long tradition of mind maps and their ilk.¹ They should permit a sense of differentiated places and permanent landmarks. Re-encountering digital data is another significant action to have access to your personal archives:

Cathrine C. Marshall on venues for re-encounter:

"Creating venues for re-encounter of lost or forgotten material is part and parcel of supporting access to personal digital archives."

"There have already been many efforts at creating visualizations to support browsing of extensive collections of personal material.² Many of the most effective ones use intrinsic properties of the material such as chronology or geographical location.³

There's considerable room for creativity in creating visualizations. ...there is also a need for methods that create time-based visualizations that cut across different media types."⁴

According to Cathrine C. Marshall, tools to find the duplicates and choose among all the data should be considered a valuable approach to find the relevant data.

(E.g. to find out if a photograph is a copy, a version or the Original file.)

Cathrine C. Marshall's Conclusion:

What we want... is a combination of services and mechanisms that will make it possible to designate which of our digital belongings are the most valuable; to organize the rest of them into tractable archives that reflect the items' value; and not spend extra time taking care of them.

...once people have made a few copies of a treasured item, they feel reasonably secure about its fate. For example, when participants describe storing photos on Flickr, they no longer see a reason to create additional backups of the files. Why – once the photos are shared and tucked away locally – would an individual take the time and bandwidth (or possibly bear the expense) to put files in yet another place?

The answer is straightforward:

it is more important to know what we have and where we've put it than it is to centralize all of our stuff into a single repository.

Summary

We have all amassed digital information. Only few people had digital artifacts twenty years ago. But today we are in the midst of an age, where more and more personal belongings become digital. We create and collect digital data with a certain implicitness and this behaviour is getting more ubiquitous without even knowing what kind of data you collect and create. Common computers, cameras, audio-recorders, smart-phones and the World Wide Web is getting enhanced with embodied functionalities. It's true that it's easy to perform benign neglect and follow "lazy preservation" practices like Cathrine Marshall calls it. The Terms "keeping information" and "archiving" are merging. In the digital context archiving already begins with creation. Most people want to know that they have possible control of their digital belongings. We need to know what and where we got it. How many copies exist? It's not enough to rely on our photos, which we uploaded on flickr or elsewhere in the web.

People assume that with the digital revolution they have better access to information. But because of the sheer amount of data and formats it's a big challenge to find relevant information. File-types and software-dependencies cause problems in long-term preservation. Files are bound to their software and could become unreadable with the change of technologies. It needs a lot of care in the field of data-migration, where most of the files-losses occur. Since it is so trivial to make copies of digital information, in long-term, desktop-searches and common ways of storage-solutions won't be sufficient anymore. There must be novel ways of how to access files. It is more like re-encountering than searching and finding them. Time is a factor that comes in to game where location and value of files are no longer permanent.

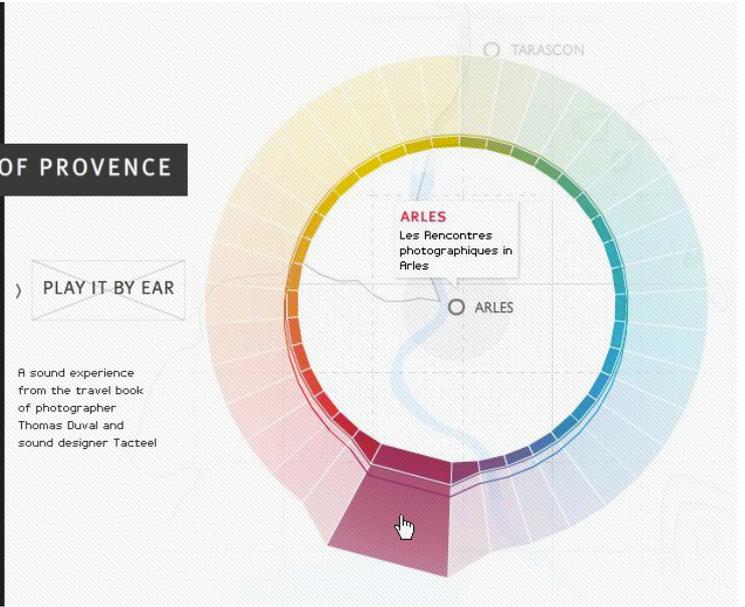
¹ E. Churchill and J. Ubois, "Designing for Digital Archives," *Interactions*, 15, 2 (March/April, 2008). <<http://doi.acm.org/10.1145/1340961.1340964>>.

² A. Perer, B. Shneiderman, and D. W. Oard, 2006, "Using Rhythms of Relationships to Understand Email Archives," *Journal of the American Society for Information Science and Technology*, 57, 14, pp. 1936-1948. <[doi:10.1002/asi.20387](http://doi.org/10.1002/asi.20387)>.

³ A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd, "Time as Essence for Photo Browsing Through Personal Digital Libraries," *Proceedings of JCDL 2002 (Portland, OR, July 14-18 2002)*, pp. 326-335. <<http://doi.acm.org/10.1145/544220.544301>>.

⁴ M. Ringel, E. Cutrell, S. Dumais, E. Horvitz, "Milestones in time: The value of landmarks in retrieving information from personal stores." <<http://research.microsoft.com/en-us/um/people/horvitz/landmark.pdf>>

Relevant Projects



PROJECT 1

Snapshots of Provence

<http://www.snapshotsofprovence.com/>

Abstract

This web application was initiated by Bouches-du-Rhône tourism. The website shows the diversity of land and people in an atmospheric way. The visitor can navigate his own way through the site and becomes an explorer. Different media like photography, maps, videos and sound illustrate daily activities and are building an overall impression of the Provence, the Camargue and Marseille.

Use

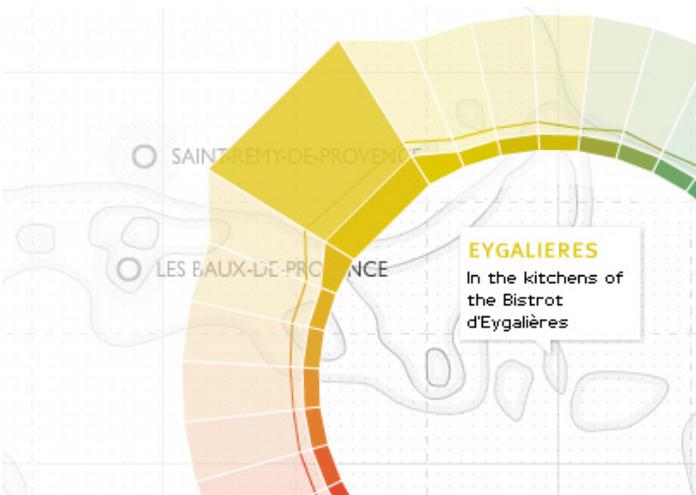
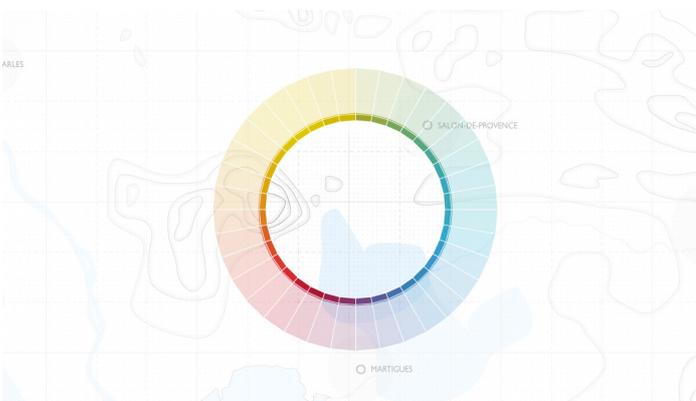
When you open up the page you can choose between two possibilities to explore the Provence through the website's interface. You can either explore the Provence with a map where you get an audio-visual feedback of selected locations or a photography/movie archive. With the map version a „wheel“ in the center of the screen is the main-navigation. Based on the angle of where your mouse pointer lies, the next location in this direction becomes visible. In this case you don't have a lot of influence on what you are seeing and hearing, but you get a sound experience from different places by switching the location...



EXPLORING DATA



ACCOUSTIC & VISUAL EXPERIENCE



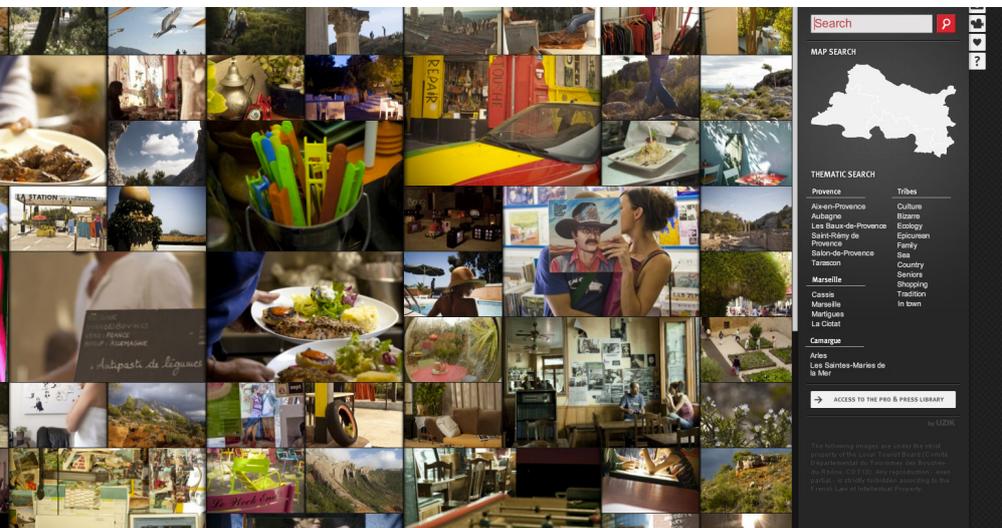


Depending on the everyday sounds you actually hear of the place you decide which location you want to see in detail or you choose it just because you are interested in the specific location by name.

When clicked on the location, you switch to a full view where photographs illustrate the place.

On top of the photograph you can open specific information of that place. After a while the page switches to the next photograph allocated to the same location. You get a good overview on the bottom of the page of how long a photograph is visible and how many there are.

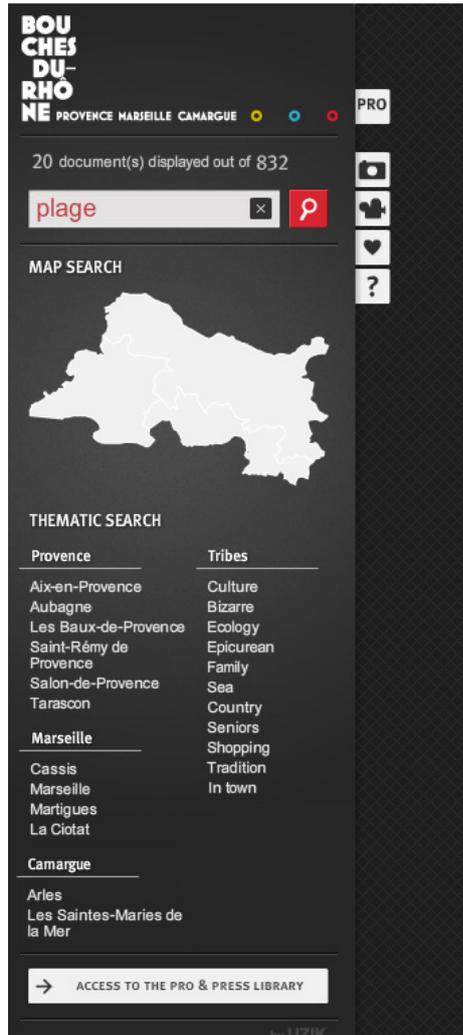
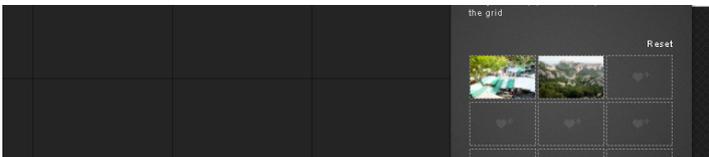
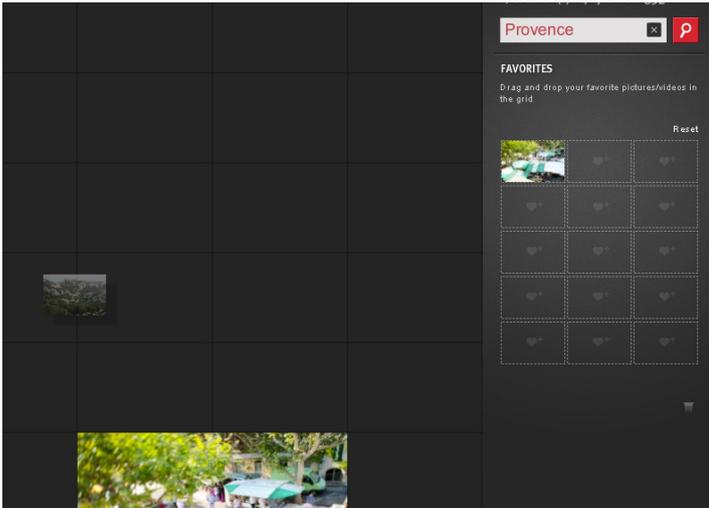
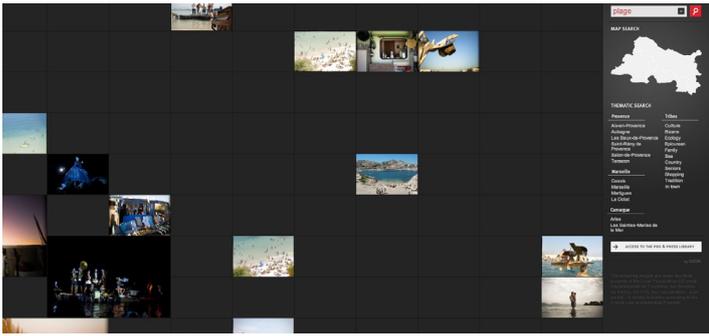
You can also switch manually to other photographs.



Another part of the application lies in a photograph/movie archive. Here you can browse more specifically and more target oriented enabling more impressions of a certain area.

Through a mini-map you can narrow down your results or you can type any relevant term into the search-field.

TARGET-BROWSING VS. CONSUMING



Relevant pictures can be saved as my favourites on the right panel.



COLLECTING

EMERGING EMPATHY
AS USER-MOTIVATION

Relevance to our Project

If we would claim that our project is about daily consumers then this website is a good example of where you are able to browse randomly and the same time can be aim oriented. The creators obviously very clearly separated those two different ways of browsing. The other thing we analysed in this website was the fact that if you are browsing with a certain aim (through filtering) there is the possibility to store your personal favorites. If you are browsing target oriented then the probability of finding relevant information is higher and you therefore want to save the findings more often. On the other hand the aimless browsing could lead to unexpected findings but then it's more about the overall experience giving one an impression of the subject. Basically the user has a lot of ways to interact with this application and the motivation of this website lies in exploring the Provence. Because we are experienced we believe that the atmospheric user-experience leads to having fun when using this website. You will like to spend a lot of time and you will probably go back to get new impressions.

Critics

The browsing-experience has a very high granularity where the user has a lot of freedom to choose how to explore the different locations in his own way. The website creates an atmosphere where you almost feel as if you have become a citizen of these viewed locations and a part of the scenery. The sound design plays an important role in letting generating this atmosphere. Nevertheless some functions are not that obvious. For example if you want to save an image to your favourites at first you don't know that you can drag the items into the favorite collection. We also see a weakness in the system. Do you really want to save these images or movies? You always have to go back to this website to see your favourites again but wouldn't it be nice to save these pictures or movies locally on your desktop knowing where they came from and what their context was? What we found good was that when you filter information you always see what happens. You always have an immediate feedback which means the process of your action have been made visible.

PROJECT 2

Milestones in Time: The Value of Landmarks in retrieving information from Personal Stores.

Meredith Ringel, Edward Cutrell, Susan Dumais, Eric Horvitz
Stanford University, 353 Serra Mall, Stanford, CA, USA
Microsoft Research, One Microsoft Way, Redmond, WA, USA

<<http://research.microsoft.com/en-us/um/people/horvitz/landmark.pdf>>

Original abstract

We describe the design and analysis of timeline visualizations for displaying the results of queries on an index of personal content.

The visualization was built on top of a personal search engine that provides a unified index of all the information a user has seen, including web pages, email, and documents. Results of searches are presented with an overview-plus-detail timeline visualization. A summary view shows the distribution of search hits over time, and a detailed view allows for inspection of individual search results. In a user study, we explore the value of extending a basic time view by adding public landmarks (holidays and important news events) and personal landmarks (photos and important calendar events).

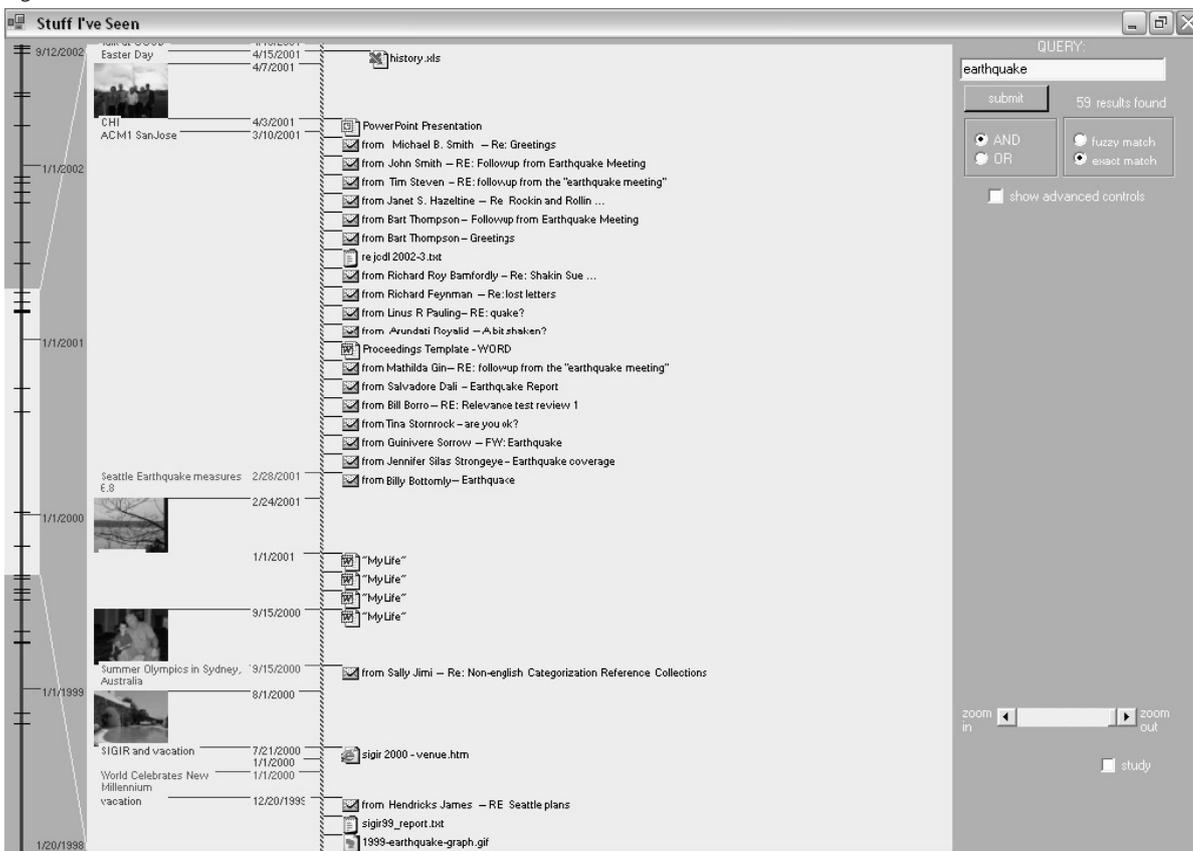
Goal of research

The goal of this research was to find out how „episodic memories“ influence the associations to certain files and how they could help one to remember the content and location of files. The Episodic memory is the memory of autobiographical events (times, places, associated emotions, and other contextual knowledge) that can be explicitly stated.¹ Studies have also shown that people forget a lot of computer based tasks that they did months before. Showing them images and videos helped them enormously to recall more details of those tasks.

Prototype

In Figure 1 you can get an idea how they implemented the prototype. On the righthand side a test person entered “earthquake” because he remembered the files that he wanted to find were stored around that time an earthquake took place. The overview area at the left shows a timeline with hash marks representing the distribution of the search results over time. The highlighted region of the overview timeline corresponds to the segment of time displayed in the detail view. On the detailed timeline, landmarks from news headlines, holidays, calendar appointments, and digital photographs have been set. To the right of the landmarks, you can see details of individual search results (represented by icons and titles).

Figure 1



¹ http://en.wikipedia.org/wiki/Episodic_memory, February 2011

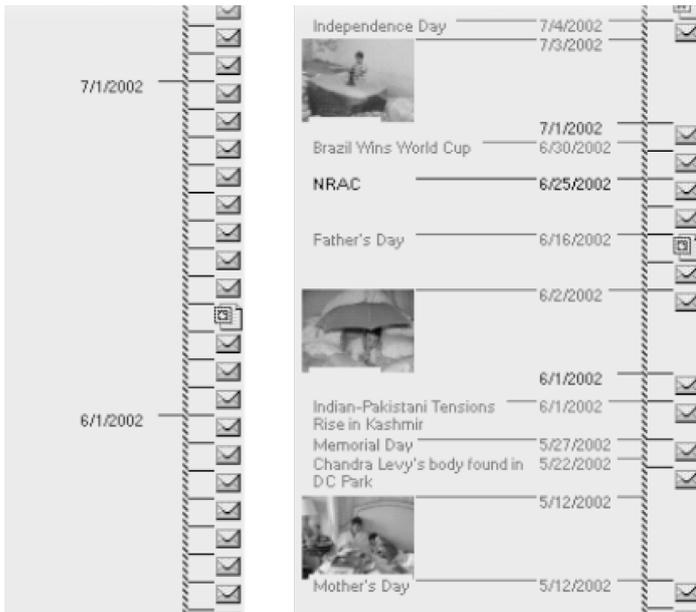
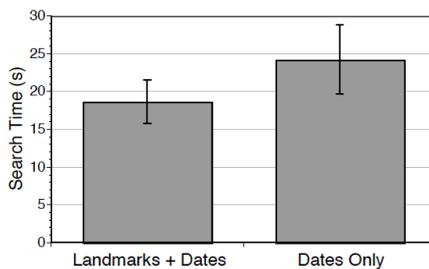


Figure 2 (a)

(b)

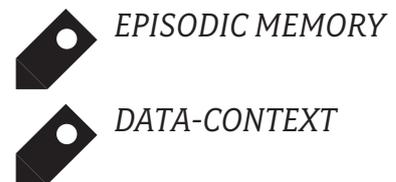
Twelve Microsoft employees from the ages twenty-five to sixty took part in the studies. They merged a file with 110 Messages into each participants Email -Managementsystem. The participants completed a series of search tasks based on date-only or landmark searches. The difference of the interface is visible in Figure 2.

Figure 2:
 (a) "Dates" condition displays only dates on the left of the timeline.
 (b) "Landmarks + Dates" condition has additionally a timeline that displays landmarks (holidays, news headlines, calendar, appointments, and personal photographs)



Results of the studies

Media search time with the use of landmark events, displayed on the timeline (left) was significantly faster than the search time where only dates were used (right).



Relevance to our Project

We claimed that the users brain is triggered in finding specific information. Taking the theory of the „Episodic memory“ as a foundation to gain knowledge of user-behaviour this verified our idea of creating a more visual approach to archiving. How does a user make associations with plain information? The reasearch project illustrates that there is a need for different approaches to access data and gives small directives where such systems could lead and how we could recall the context of data out if our episodic memory.

Critics

The research team concluded that added context helps find digital items. Unfortunately the user tests were only made with added context of news and public events. It would be nice to see a prototype which also covers personal events. The researchers mentioned: "We believe that this work demonstrates the potential value of adding global and personal context to the presentation of search results, as well as suggesting directions for future study." An avenue for future study is to explore the value of different types of events — e.g., running separate "personal landmarks" and "public landmarks" conditions in addition to the two conditions explored here." In addition, there are opportunities for investigating more generally when timeline-centric views are most useful for finding target results of interest. The usage of timeline views will be positively influenced by the addition of landmarks corresponding to past events when searching for data.

PROJECT 3

Papers

<http://mekentosj.com/papers/>

Abstract

Papers is a native application to manage scientific papers and articles. They compare their application with itunes but it's made for PDFs. Since there is a canonicalized format such as the pdf-format the whole science world collects a lot of pdf files. Because of the amount of files scientists tend to collect it gets hard for them to manage their collection. With this application one has the possibility of many features to manage the readables. "Papers" is made for desktop-computers, iPad and iPhone.

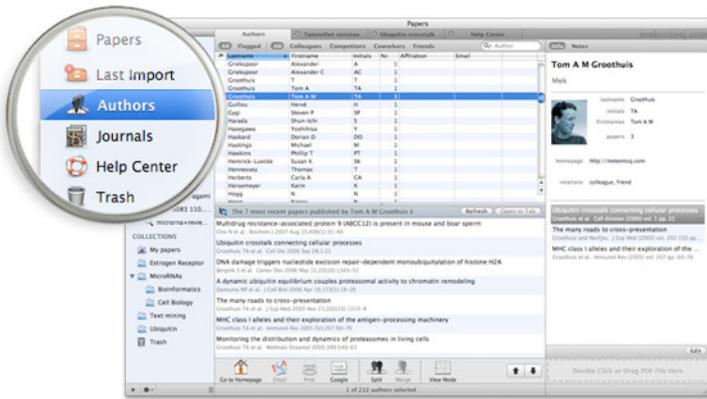
Use

Papers is based on a centralized storage place. (But you have still the possibility to keep at an individual place.) You can manage, read, share, search, maintain and create playlist of your collection. There are functions which integrates the collection in workprocesses. For example, when you are writing a paper and have to cite to other papers, there is a feature where you can do that with automated tools.



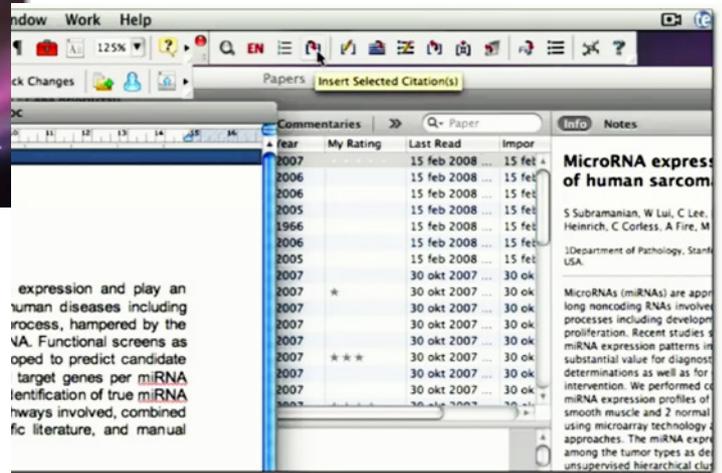
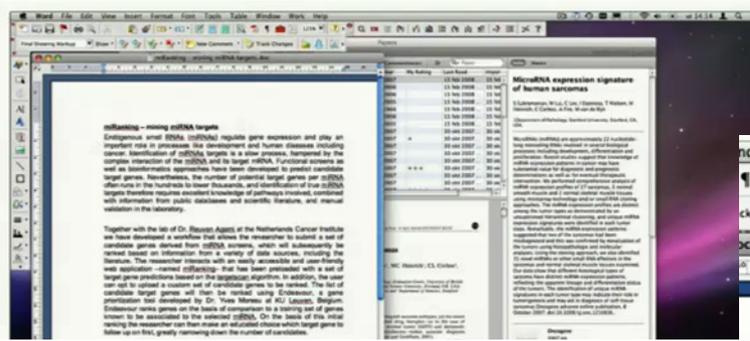
**CANONALISED
DATAFORMAT**

**EXT. DATABASE
ENHANCED**

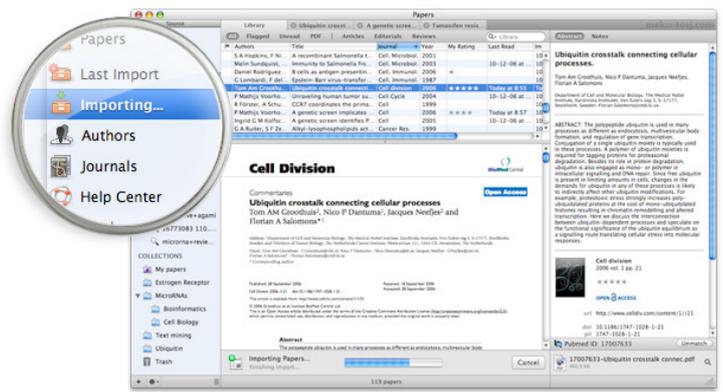
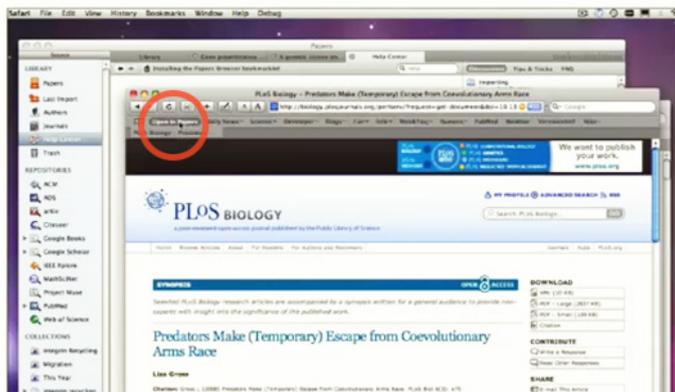


There are a lot of similarities to iTunes such as browsing the collection. For example you can browse your collection depending on authors. In a sidebar you get more information about the author and you have the possibility to see more articles and papers related to this author. So you therefore have the option of adding and reading more papers that you didn't have or knew of before. Therefore "Papers" is connected to different common scientific databases such as Google Scholar, Web of Science, JSTOR and many more.

Another function of "Papers" is that you are able to search and browse web content out of the application itself. You can choose which external science database you want to search for articles. You get a list of found articles which is sortable by different keywords. The findings-list is enhanced by showing more information about the items in the sidebar while you are scrolling through the content.



One interesting feature is that the application is integrated into work processes. While you are writing a paper in Microsoft Word, you have the possibility to refer to external papers. With "papers" you can choose between different editing tools and you are still able to use the features of "papers". It lets you choose the "Viewer" which you are most comfortable with.



The application is able to gather metadata automatically. You have the possibility to add a webbrowser-extension for doing so. When you browse for papers or articles in your webbrowser, you can simply click on the bookmark and the found paper opens in the application. The application recognises the paper automatically and will gather the metadata related to the paper. Therefore you can just add the paper to your collection with all the implied information.

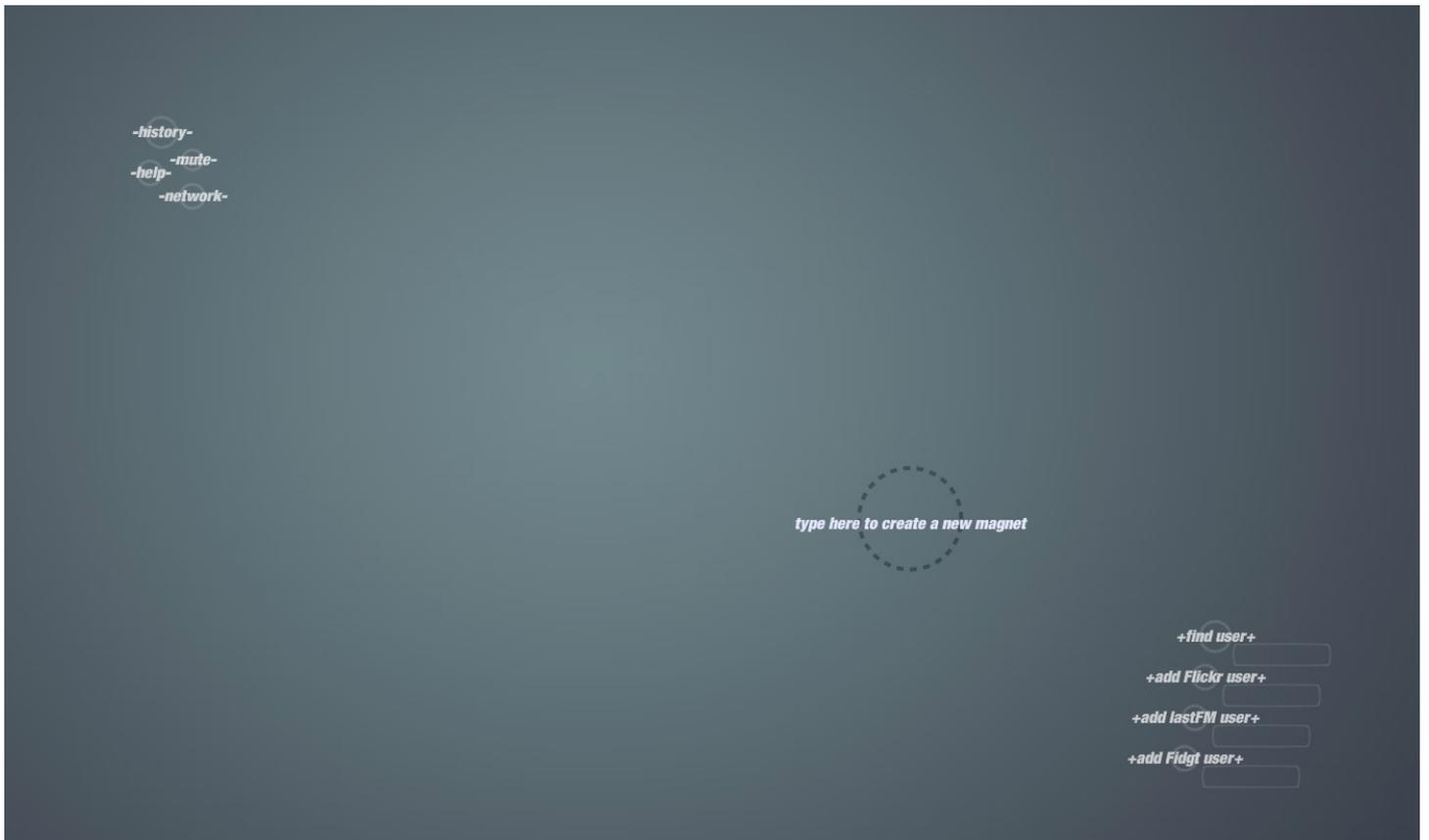
WORKFLOW INTEGRATION
AUT. METADATA GATHERING

Relevance to our Project

This application truly meets the needs of its purpose. We realize that we can't implement such an overall service in our project. Therefore we'll only pick the best features as inspiration for our concept. This application is a good example of how to try to embrace other tools and combine them in order to get a persistent use in formats.

Critics

The "Papers" interface provides a clean and "easy to use" workflow for searching, sorting, downloading and reading articles. This application uses systematically structured external databases. It proves that such systems depend on existing qualities like funded databases. The difficulties we see are that you have a large amount of features spread all over other system, which in turn could lead to uncertainties in its use.



PROJECT 4

Fidgt

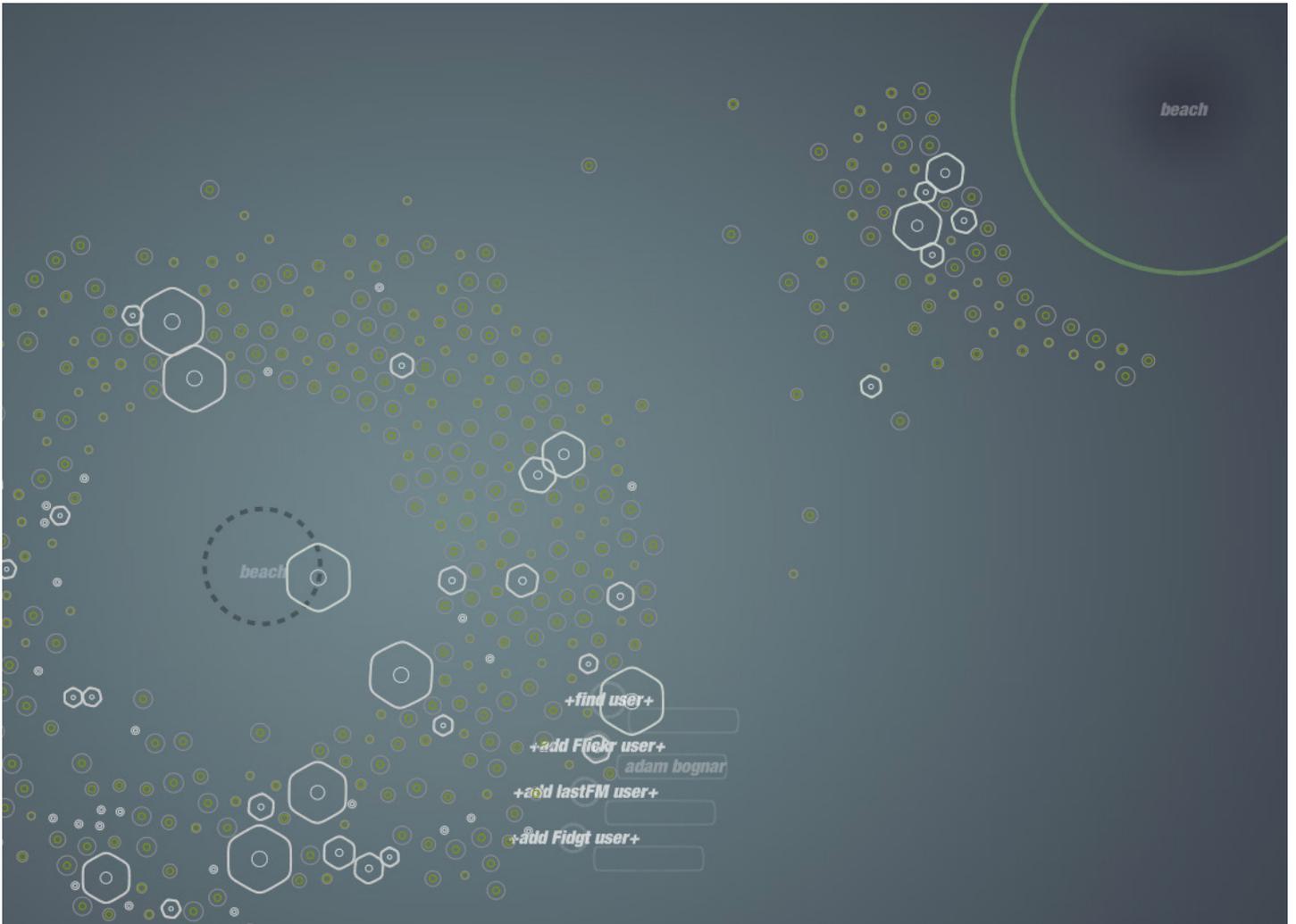
<http://www.fidgt.com/visualize>

Abstract

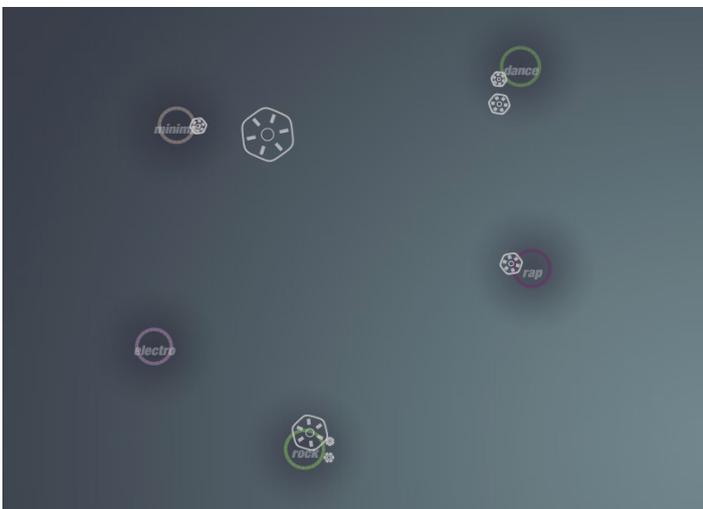
Fidgt is a visualization tool where you can filter and display contents and users of different network services. The program starts up with a minimal setup. You can create new magnets or add users of lastFM or flickr to the board. When creating new magnets you enter a term or topic which will drop into the interface as a new circular element. With the magnets you can filter the network of a user. All the friends out of a users network, depending on their match, of metadata will be attracted to the magnet containing the previously entered term.



VISUAL FILTERS



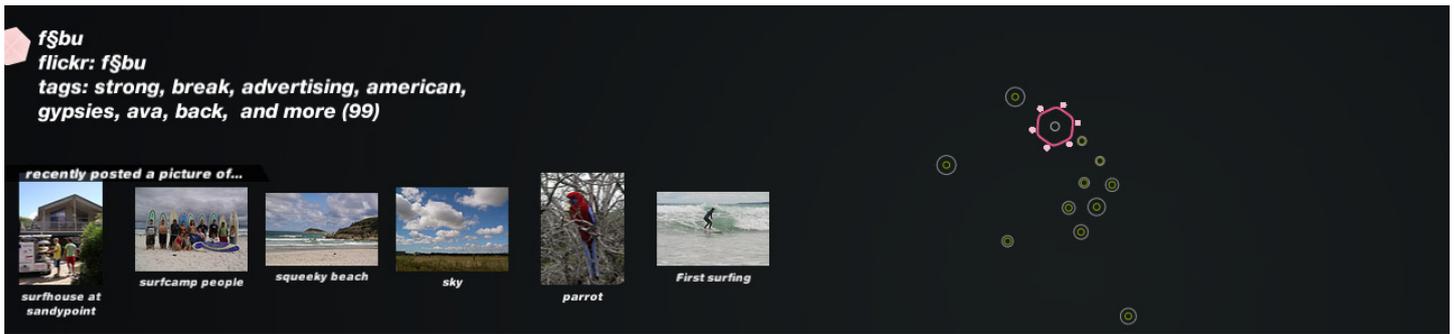
The screenshot above displays, that one flickr user and one magnet has been added. The whole network of this user is now visible: The hexagon symbolises the other user in this network and the greenish circles are metatags. You can't click on the green ones and there is no explanation why you can't. As you can see peoples with similar "metatags" are moving towards the magnet „beach“.



**FLUID, NATURAL
DESIGN**

**INDIVIDUAL
ORGANISATION**

Because the user is defining his filter-interest, every fdgt-user gets a totally personalized overview. You always know what your filtering criterias are because they are visible until you delete the the matching magnet. Through the fluid design the user is able to organize in his own manner: On the left you can see how a couple of different magnets with their aggregated relevant data have been arranged.



Above you can see a screenshot of fidget's content visualization. In the image above on the right hand side you can see that a user was clicked on and is therefore now colored purple. The most relevant tags and an excerpt of his photographs become visible on the left hand side.

**USABILITY VS
DATA-AMOUNT**

Relevance to our Project

Fidget is an unusual and visual way of how one could browse through content. The user decides how data should be arranged. Every process you do is made visible: You add a new magnet which then becomes a part of your system. You rearrange magnets and you immediately get a feedback of what is going on. Our intention is to create an environment which gives a visual overview of personally high valued data whereby the process of what you did and where you did it will also be important.

Critics

The fact that the user has good feedback of his interactions has been made very well in our opinion. You understand quickly without much explanation what the tool is doing. The system is limited and causes difficulties if you add a huge amount of data to it. It won't be comprehensible anymore. But usually you don't look at as much as ten last.fm accounts with all their friends at the same time. Fidget is more of a playful approach to browsing than an everyday usable tool and therefore we don't see a greater benefit in the field of organizing and having access to your files. Another weakness we encountered was how you browse content. You only see a limited amount of data and will have to switch to the original system to have a better and more complete experience.

Relevant Projects - Comparison

As we experienced with the "Snapshots of Provence" project, the user gets a choice how he wants to access the information. In the "digital Landmarks" project, the user was challenged by search tasks with two different interfaces, date-only or landmark-enhanced. Accessing information can be done in different ways depending on the goal. In the Landmark-project we saw that you can save time by enhancing timecentric search by landmarks such as public events. These public events seem to trigger the "episodic memory" of the subjects. On the other hand in the Snapshots-project the experience is enhanced by not only giving a visual but also an audio context. Creating an atmosphere where you almost feel as if you have become a citizen of these viewed locations and a part of the scenery. Emotional factors have been taken into consideration to encounter an informative space which we also would like to pursue.

The "Papers" application provides a clean "easy to use" workflow for searching, sorting, downloading, and reading PDF articles. It allows you to use your collection within different context and applications. "Papers" turns out to be an intelligent front-end solution to access linked databases. These scientific databases for written information of scientific papers and articles have clear standards and are consistent. Therefore recognizing differences in the articles or scientific papers. Metadata is gathered automatically, which augments the application to a powerful tool when managing scientific readables. Another tool we found is Fidget. It is an experimental way to visualize data of any users from platforms like flickr or last.fm. It will display their network e.g. friends with a magnet where you can add keywords to filter relevant contents of the network. From this project we would like to adopt the users freedom to organize and display data. The interaction is comprehensible due to a constant feedback of its process.

Summary

The digital revolution mystifies a better access to knowledge. Archival techniques themselves are manageable. The main problem is how we access information. The infrastructure and environment of storage places are eminent in order to find relevant information. Information will only emerge to knowledge when dealing with it. If you can link it to an existing experience you can communicate it. Your personal motivation and interest are key-factors of doing so. We create and collect digital data with a certain implicitness. This behavior is getting more and more ubiquitous: You possibly don't know what kind of data you collect and produce. The diversity of devices and the pervasive use in everyday context is leading to „lazy“ forms of data-preservation. This causes an uncontrolled and ungraspable user experience in the context of handling digital data. Most people want to know that they have possible control of their digital belongings. We should possibly know what and from where we got it from. How many copies exist, how many versions, what the originals are and how the personal relevance of this data changed over time? Through our background research we encountered systems, which augmented the successful locating of files by adding news and personal events to a file-browsing-interface. It has been proven that digital-files can be found easier when you are looking for something in a certain time frame while displaying the previously mentioned additional context. A reason for this phenomenon is called the „episodic memory“. Another result of our research is how strong digital content and the use of it are bound to each other. The site „snapshot of Provence“ proves that an atmospheric environment, which generates emotional values, leads to a sustainable and joyful use. File types and software dependencies cause problems in long-term preservation. Files are bound to their software and could get unreadable with the change of technologies. A lot of care is needed in the field of data migration where most of the files losses occur. Since it is so trivial to make copies of digital information. Regarding long-term preservation, desktop searches and common ways of storage solutions won't be sufficient anymore. There must be novel ways of how one accesses files whereby the individual user has to be taken into account.

Concept refinement based on the background-research findings:

Common systems, which encounter the deluge of data deal insufficiently with the value of its content. Digital items are often hierarchically leveled to one visual equality. We will try to enable individuals to give them a defined space of their valuables. Because people are spreading / storing their personal digital data all over different places for different reasons, the time, the provenance, the target and personal context should be taken into account. How can we gather intrinsic information of specific files? Individuals should be motivated to add information to their valuables and external context should also augment these valuables. A „virtual geography“ could be an interesting approach to help the user get an orientation of his personal digital belongings. How can we track files which have been spread all over different storage places? A great challenge is to deal not only with one specific set of data like canonicalised formats like PDF but also with different formats and still meet the media specific needs.