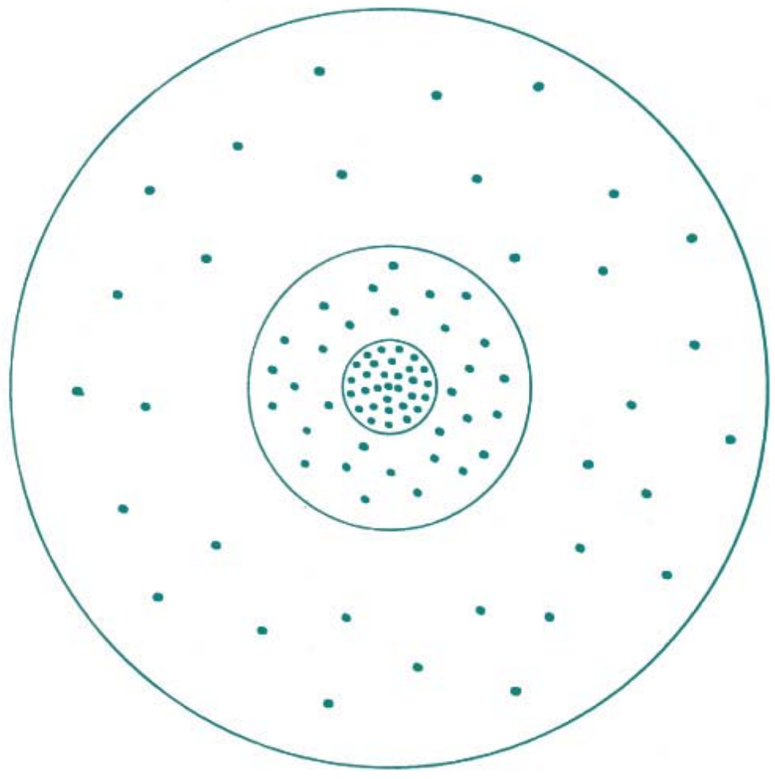
NATURAL CLUSTERING OF INFORMATION

Marcia J. Bates
Department of Information Studies
UCLS
Los Angeles, CA 90095-1520

Socially generated information distributes naturally in Bradford rings. Number of information units is equal in each ring, while areas of rings are in the relation of $1 : n : n^2$ …where $n \approx 5$.



From the side:

NSF Personal
Information Management
Workshop

Information School
University of Washington
January 27-29, 2005

RESEARCH QUESTIONS:

Bradford has been found in a wide variety of information contexts.

Is it also operative in personal collections?

If so, what are the implications for system design, education, etc.?

References:

Bradford, S. C. (1948) *Documentation.* London: Crosby Lockwood.

Bates, M. J. ”Speculations on Browsing, Directed Searching, and Linking in Relation to the Bradford Distribution,” Emerging Frameworks and Methods: Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (CoLIS4), Edited by Harry Bruce, Raya Fidel, Peter Ingwersen, and Pertti Vakkari. Greenwood Village, CO: Libraries Unlimited, 2002, pp. 137-150.

Ben Bederson

Interfaces for Staying in the Flow

University of Maryland, Human-Computer Interaction Laboratory

What?

To help people focus on their information management without feeling disrupted by the interface. What kinds of interfaces and combinations of features enable people to feel "in the flow"? We think of flow as that hard-to-describe feeling you get when you are fully engaged and can concentrate on the task at hand to the exclusion of all else.

Why?

Flow is important because without it, we feel frustrated and slow as computer users. When the interface gets in the way, our thought processes can be affected. Computers should ***augment*** *our capabilities, not interfere with them.*

Our personal information is so important, and is accessed so frequently, and often during other tasks. It is crucial in this domain more than others to have interfaces that are "invisible" to the user. Ideally, a PIM should give you the information you want just before you ask for it…

How?

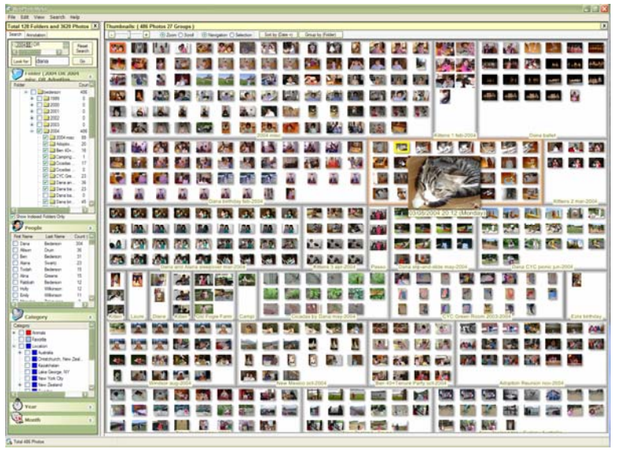
By focusing on the interface, and the ease with which people can author, annotate, search and browse ***quickly*** *in an* ***integrated*** *environment.*

So far, we have concentrated on "manual" interfaces that are responsive to user's commands. Under what circumstances will predictive or adaptive interfaces support flow better?

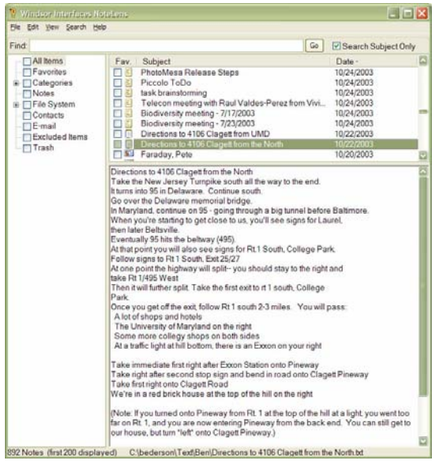
Results so far:

Our tools are good enough to use (we rely on them for our own PIM, and are commercializing them). Some qualitative studies show strong user engagement – but how to ***measure*** *flow remains an open question.*

Sample systems designed to support "flow":



PhotoMesa image browser
www.photomesa.com

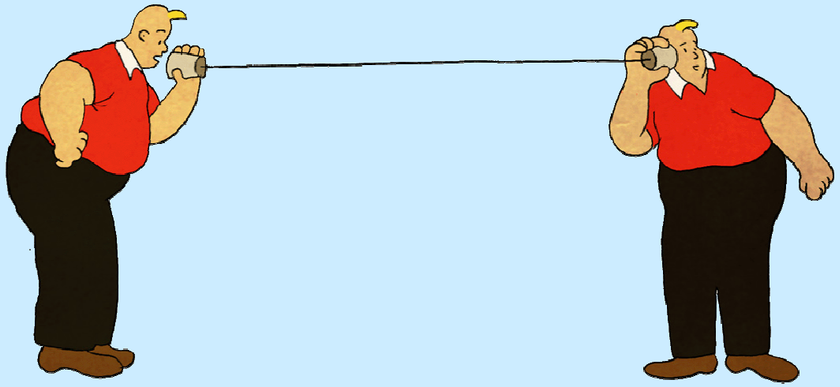


NoteLens note-taking tool
www.notelens.com

The User-subjective Approach to PIM Systems Design

Ofer Bergman, Tel-Aviv University

Personal Information Management (PIM) is unique in that the same person who stores the information is the one who later on retrieves it. Therefore, PIM can be thought of as a communication between a person and him/herself at different points in time.



The user-subjective approach (Bergman, Beyth-Marom & Nachmias, 2003) proposes that PIM systems should capture subjective attributes implied by users' interaction with the information, in order to help them retrieve it in the future. Such subjective attributes are: the project that the user is working on, the subjective importance of the information, and the context in which the user is working with the information.

The User-Subjective Principles

Derived from the user-subjective approach are three generic principles for PIM systems design:

The subjective relevancy principle proposes that the subjective importance of information should determine its visual salience and accessibility: Relevant items should be visible and accessible while irrelevant items should not distract the user's attention.

The subjective context principle suggests that information should be retrieved and viewed by the user in the same context in which it was previously used in order to help the user recall it.

The subjective classification principle states that all information items (such as documents, emails, bookmarks, and to-do lists) related to the same project should be stored, or at least retrieved, together regardless of their technological format.
I will be happy to should you *MyProjects*, an implementation of this principle.

Research

A research was conducted to test the user-subjective approach. The research asked whether PC users used subjective attributes in their daily work. Research methods included in-depth interviews (N=20) and a questionnaire (N=84). Results showed that:

1. Participants used subjective attributes whenever the PIM system design suggested it:
   - Stored files of different format in same folders according to their project.
   - Used desktop and "history" to retrieve subjectively relevant files.
2. Participants tended not to use subjective attributes when the PIM system design did not suggest it:
   - Tended not to store emails and favorites with files according to their project.
   - Used various alternative ways to archive files of subjectively low relevancy, as this activity is not supported by the system design.
3. Participants showed positive attitude towards various user-subjective applications designs.

Future research includes creation and testing of user-subjective interfaces.

Reference

Bergman, O., Beyth-Marom, R., and Nachmias, R. (2003). The user-subjective approach to personal information management systems. *Journal of the American Society for Information Science and Technology 54*(9): p. 872-878. Available at URL: http://muse.tau.ac.il/publications/81.pdf.

Who?	Rick Boardman rickb@google.com	Present Affiliation:	Google	Research performed at:	Imperial College London

Project WorkspaceMirror: investigating the utility of sharing folder structures between PIM-tools

What/Why?

- Many people struggle to manage their personal info
- Popular design route = integration
  - oLimitation 1: little X-tool data to guide design
  - oLimitation 2: design evaluation is rare
- Need for systematic investigation of integration as a design strategy for improving PIM support

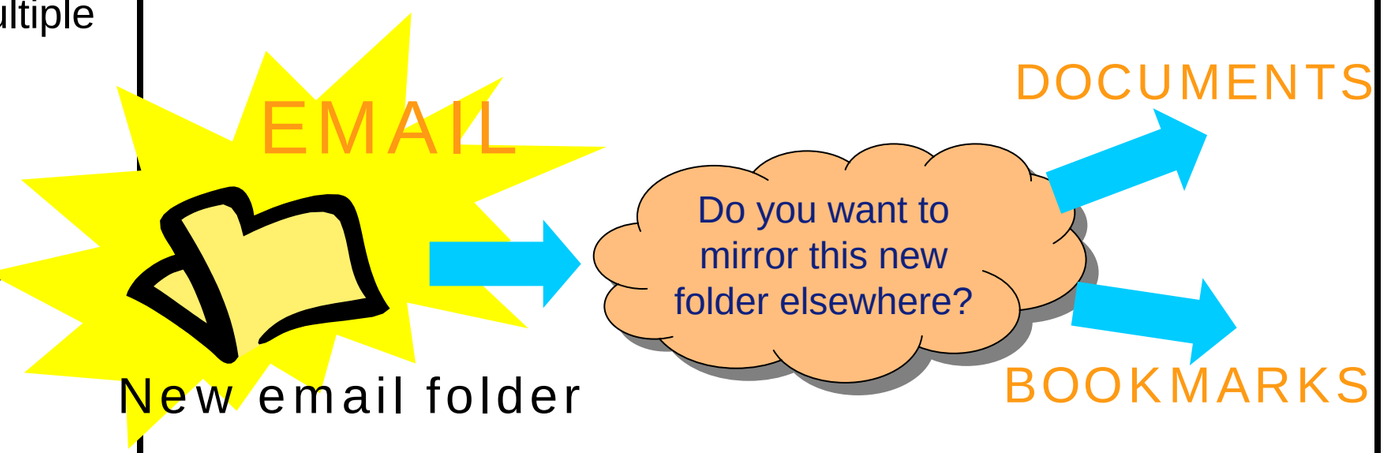
How? User-centered Design Methodology – 3 Steps:

- **Step 1: Cross-tool Exploratory Study** - How do people manage files, email, bookmarks? ⇒ Observations of **folder overlap** for many users, i.e. many folders appear in multiple tools (highest between files/email, av. 7.6)
- **Step 2: Design of WorkspaceMirror** - an integration mechanism which allows the user to share one folder structure between their files, email and bookmarks
  - oDriven by observation of **folder overlap** in Step 1
  - oIncremental design ⇒ easier to evaluate
- **Step 3: Field trial to evaluate design** – see right

Results from Field Trial (8 users, av. 69 days)

- Did users find WorkspaceMirror useful?
  - o 57 folder creations mirrored (max 26, min 0, av. 7)
  - o No deletes or moves (little maintenance performed)
  - o Mirrored folders were "shallow" – average depth 1.83
    ⇒ **Sharing is most useful for top-level folders**
  - o Trade-off between resulting consistency between tools, and users' needs for organizational flexibility
    ⇒ **Lesson: may sometimes be best <u>not</u> to integrate**
- Methodological lesson: PIM evaluation is hard due to indiv diffs, existing habits, need for trust, + PIM is discretionary!

WorkspaceMirror: Example interaction



More detail: See my PhD Thesis (Google for "rick boardman phd")

People:	Harry Bruce, William Jones, Susan Dumais	Institution:	University of Washington
Project:	***Keeping Found Things Found***		

What?

- *How do people manage information that they find or encounter and plan to use at a later point in time*
  - *Studied in relation to information located or encountered on the World Wide Web*

Why?

- *People regularly locate, encounter or acquire information that they know they will want to use again.*
- *People need to organize and manage the information that they use for work, fun, and everyday tasks.*

How?

- ***Keeping Study***
  Goals – (1) To understand the diversity of keeping and leaving methods that people use to manage Web information; (2) To analyze the underlying function of observed keeping methods
- ***Re-finding Study***
  Goal - To explore the various methods that people use for re-finding information previously located or encountered on the World Wide Web
- ***Surveys (2)***
  Goals – (1) To determine, from a larger sample, if the list of observed keeping methods is complete and how frequently people use each method; (2) To find out why certain keeping methods are used
- **Folders use study**
  Goal – To observe the role of folders in the organization of project-related information

Results

- *Individuals use **keeping** methods when they identify information as useful and take action to make the information accessible in the future*
- *Individuals "do nothing" or **leave** information when they judge the information is useful (and that it will be useful in the future), but also that it can be located again easily*
- *People often use several keeping methods – choosing these methods according to their functionality and the purpose that the information kept is likely to serve in the future*
- *No one keeping method provides an individual with every function that he or she might need*
- *People report using on average just over 5 methods for keeping web information at least once per week.*
- *Roughly 2/3rds of the time that people return to a web site they do so using one of three methods – direct entry, search again or hyperlink from another web site – these methods require no prior overt act of keeping.*
- *People are generally successful at re-finding information*
- *People keep masses of information that they never use - large "stacks of good intentions" for web references, paper documents, etc. they mean to read "some day" but never do*
- *People forget to use information they have kept until it is too late.*
- *People go to great lengths to arrange and highlight information so that they can see the things that matter first*
- *People have "information closets" – esp. in digital form*
- *People complain about having too many organizations - for email, e-documents, paper, web references. Some people go to great lengths to consolidate these organizations.*

Participants: *Researchers; Information professionals; Managers*

People:	Tiziana Catarci	Affiliation:	Dipartimento di Informatica e Sistemistica Università di Roma "La Sapienza", Italy
Project:	*TIM: Task-centered Information Management*		

What?

Definition of an innovative approach to the management of personal information and services that builds on two key issues:

1. ontology-based information classification
   a. "semantic save" of documents
   b. ontology visualization and direct manipulation
   c. personalization
2. task-oriented interaction
   a. learning the way in which the user carries on typical tasks
   b. task-based environment providing high level e-services

Expected Results:

1. "Personal" ontology definition and manipulation language
2. Ontology visualization and interaction mechanisms
3. Data integration framework
4. User profiler and matchmaker
5. Implementation of the "semantic save" layer
6. Formal task-definition language
7. User-intent inferencing algorithm
8. Interaction service definition and composition language

Why?

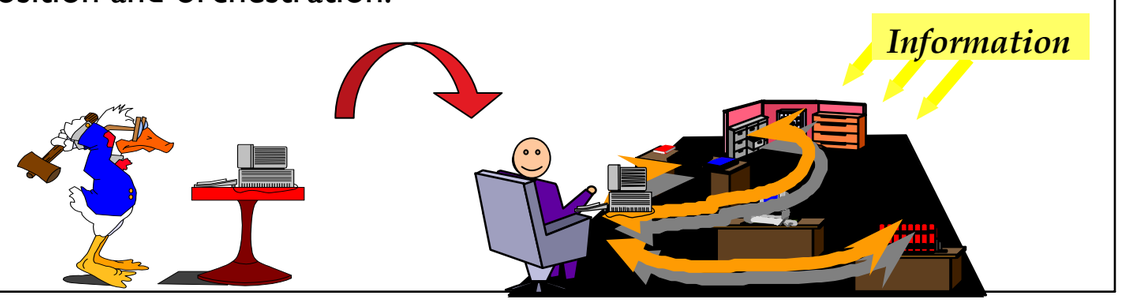
In order to manage (instead of being overwhelmed) the huge amount of personal information available in disparate forms on several devices it is necessary to move on from the traditional WIMP interfaces (that are around since the early 80's). In these environments the user initiates precise actions and the system reacts to them. Such an interaction is completely operational and the execution of tasks through a sequence of actions is completely in charge of the user.

In the TIM approach, the user has at her/his disposal an integrated task-oriented environment instead of a series of separate tools and functionalities, where not only semantically related information are linked in the system, e.g. a picture of a person and a document s/he wrote, but also the operations that are needed to accomplish typical tasks are almost automatically carried on by the system that learns from the users' execution patterns.

How?

A complete definition of this new approach to human-computer interaction and information management requires investigation of several research topics, including.:

- data integration (e.g., reference reconciliation, integration with external data, format heterogeneity, data quality);
- ontology management (e.g., ontology creation and maintenance, ontology access and visualization, ontology alignment) ;
- personalization (e.g., user profiling, matchmaking)
- formal (machine-readable) task languages (e.g., refinement calculus);
- machine learning techniques (of the so-called "user intent inference" type);
- service composition and orchestration.



Before TIM

After TIM

People:	Mary Czerwinski, MSR VIBE team	Affiliation:	Microsoft Research
Projects:	*GroupBar, Scalable Fabric, VibeLogger and Digital Memories*		

What?

Now that we can track events across applications, the phone, your location, even your body parameters (like cognitive effort, heart rate, etc.), and collect photos or video of these events, what useful problems can we solve?

Three Low-hanging Fruit Scenarios:

1) Reacquisition of task contexts after interruptions
2) Summarization of task-based activity for review or billable hours
3) Budgeting your time and priorities via smart system guidance

Results so far:

Two of our systems, GroupBar and Scalable Fabric, helped users keep track of projects through manual project creation and maintenance. Still both were better than existing tools (e.g., folders and the taskbar).

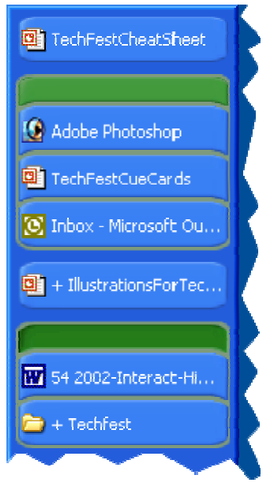
Personal Vibelogger is at a preliminary stage but attempts to automatically identify tasks (groups of related documents) and patterns of behavior for visualization, summarization and eventually intelligent assistance in prioritizing work habits.

Why?

We should be able to improve productivity for information workers and even consumers. Results can also go a long way toward Digital Memories research (i.e., how do I detect interesting events in my digital life, package them appropriately and store or present them in interesting ways, whether for business or pleasure?).

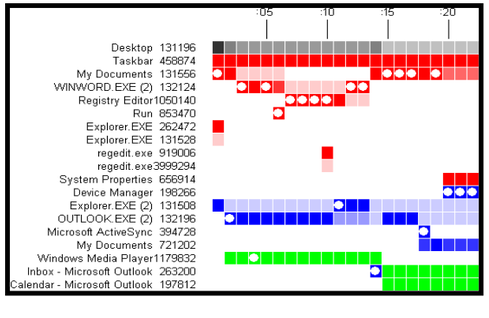
How?

Building systems and evaluating real usage of them.



groupbar

scalable fabric



vibe log

People:	Susan Dumais, Edward Cutrell, Raman Sarin	Affiliation:	Microsoft Research
Project:	***Stuff I've Seen***		

What/Why?

- Support information re-access ("stuff you've seen")
- Provide unified access to heterogeneous information from different sources (e.g., email, contacts, files, web pages, pictures, music, IM, RSS, etc.)
- Enable fast and flexible information access
- Develop rich user interfaces, since it's your stuff

How?

- Developed and deployed a prototype
  - More than 3000 users over 2 years
- Features:
  - Fast search over full-text and meta-data; flexible query iteration; rich visualizations; automatic updates; etc.
- Evaluated using many techniques:
  - Free-form feedback; questionnaires; structures interviews; extensive log data; laboratory studies; etc.

Results so far:

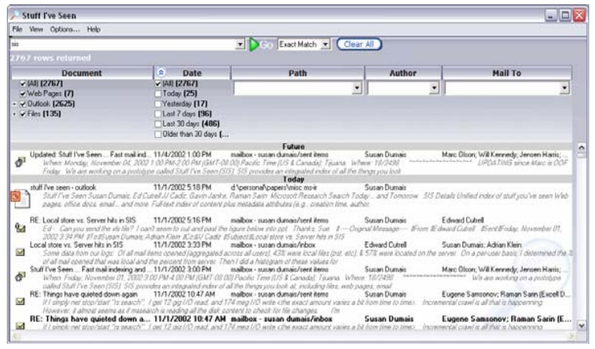
- Very positive reactions
- Short queries, quick iteration
- People and time especially important
- Some people stop filing
- More details – SIGIR 2002, Interact 2002, SIGIR 2003

Next steps:

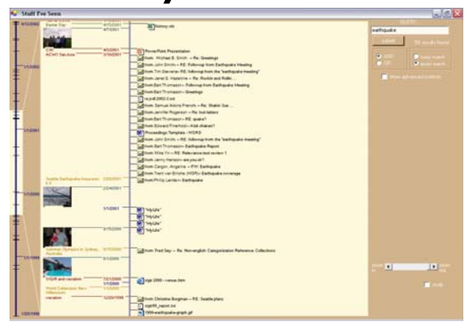
- The end of filing, as you know it -- Phlat
- Contextualized search
  - Search is *not* the end-goal
  - Search in the context of ongoing work, not as a separate app
- Personalized search

Try (something like) it: ***http://beta.toolbar.msn.com***

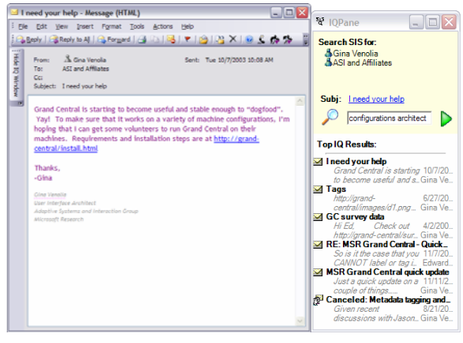
Stuff I've Seen



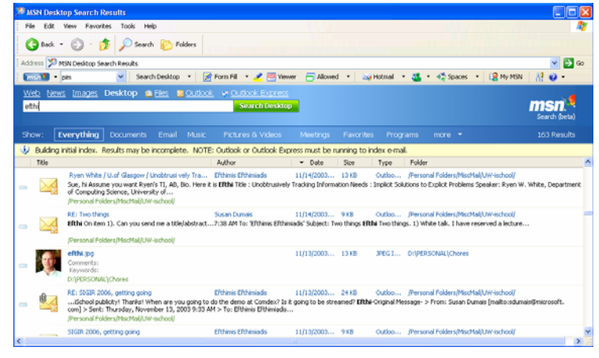
Memory Landmarks



Implicit Queries



MSN Toolbar



People:	**Thomas Erickson**	**Affiliation:**	**IBM T.J. Watson Research Center**
Project:	***The tension between public and private in personal information management***		

My Aims

I am designer, and my principal interest is in designing systems that gracefully mesh with the often turbulent flows of activity that constitute people's lives.

With regard to PIM, I am interested in the tension between private and public. While personal information may be managed for private instrumental ends—such as enabling contact with others or supporting reflection—it is also often shared with others, and in that mode it supports more public activity, and also serves a symbolic role by reinforcing (or not) the impression of its owner as organized, efficient, insightful, etc.

What I've Done

My best known PIM work is a self study of the long term design and use of a personal electronic notebook. Drawing upon 4 years of daily use, I observed that the notebook's usage model diverged radically from my expectations as a result of the incidental inclusion of messaging functionality. The possibility that much of the notebook's content might be 're-used' in messages, or shared in other ways, qualitatively changed what I recorded.

(Since this work, I've carried out a variety of studies of activity management in various contexts, but the work on Proteus seems to me to be most relevant to PIM.)

Comment

It strikes me that the emergence of blogs as a new genre in CMC is a recapitulation, albeit in a new key, of the productive tension between private and public.

Results

Erickson, T. The Design and Long-Term Use of a Personal Electronic Notebook: A Reflective Analysis. The Proceedings of CHI '96.

> "Probably the most significant impact of [the Notebook] is a sort of synergy among the activities of note making, reference, and messaging. Messaging drives this synergy. I write notes more often—and my notes are of higher quality—because I know that I am more likely to email them to others. Because the quality of my notes is higher, I reference (and reuse) them more when I'm writing a relevant message or paper. Also, the increased quality means that I am more likely to understand them when I look back at them after six months. In turn, the increased frequency of reference also drives note making and messaging: the more use I get out of them, the more effort I'm willing to put into them."

In more recent work — developing shared text-based conversation spaces for workgroups — a consistent pattern that has occurred across most of the more than 2-dozen plus deployments is that members create individual 'rooms' (see "Offices" in the Places list, below right) that are used to manage and display personal information relevant to other work group members (e.g. schedules, home phone numbers) and as places to leave personal notes…

People:	Mike Franklin (w/ Kurt Brown, Wei Hong, Allison Woodruff and others)	Affiliation:	UC Berkeley and Intel Research Berkeley
Project:	Home is Where the Bits are: Data Management for the Digital Home		

What?

- The home is becoming increasingly ***data-intensive***.
- We are building a flexible, unified data management infrastructure for the home.

Why?

- Crucial data (your kid's 1st birthday party photos, your financial records, your music, your correspondence, …) is unprotected, on disparate devices, not backed up, not easily searchable…
- New, incompatible data sources such as media, sensors, etc. are being added to the mix constantly.
- Current infrastructure is piecemeal --- no way to build applications that utilize data from all devices in the home.

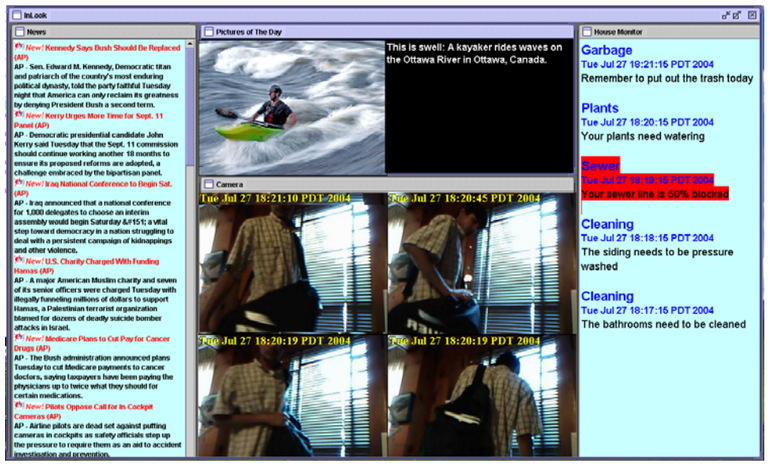
How?

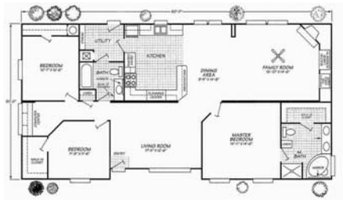
A Three-pronged approach:

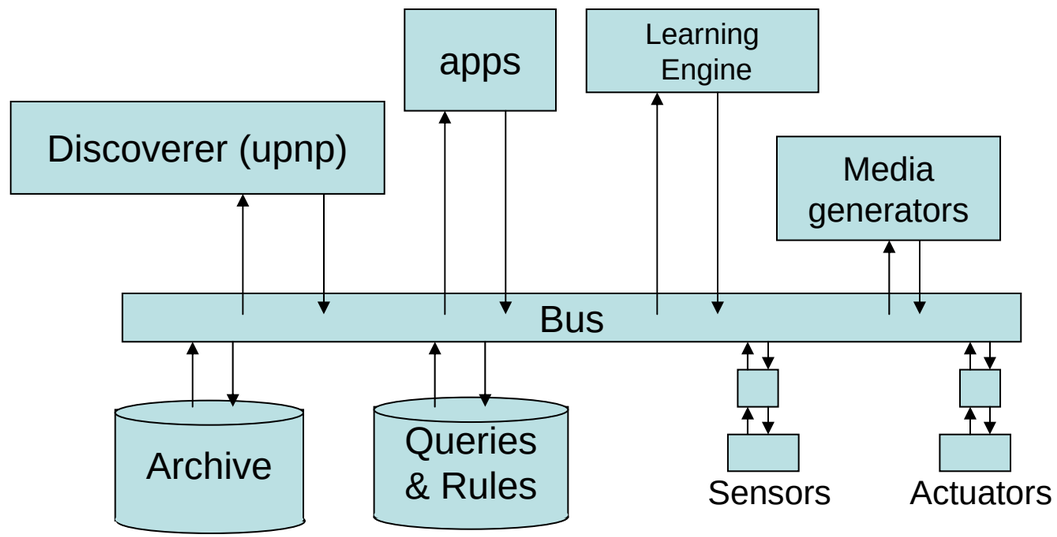
- Ethnographic studies of data and device usage in the home.
- Development of software infrastructure including: schema, event engine, learning algorithms, storage, backup, …
- The "data furnace" – a home server and common repository for all data in the home including shared metadata. Mostly self-maintaining, hands-off operation, upgradable for 25 yrs, not too expensive, in the basement (like your furnace!)

Results so far:

- Initial Architecture and Schema Development.
- Mock-up and brief trial deployment of "Inlook" home information portal.







People:	Douglas Gage	Affiliation:	XPM Technologies
Project:	***Episodic Memory***		

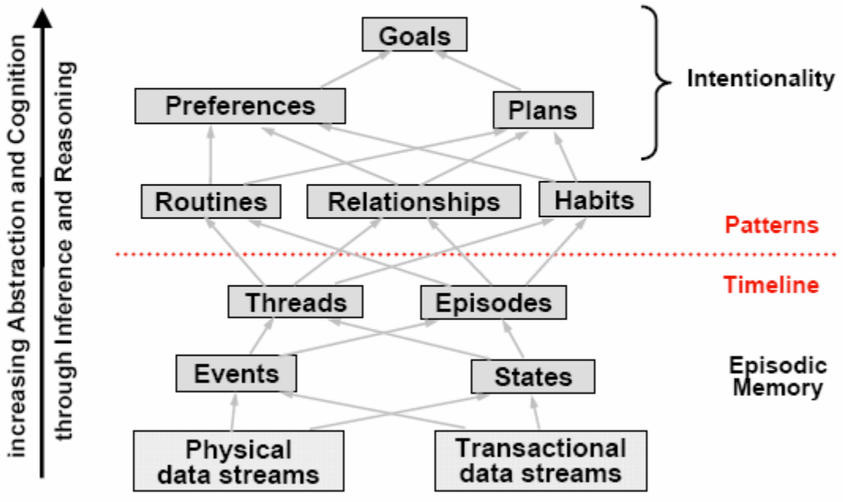
What?

Allow a person or cognitive system to recall and understand his/ her/ its experience in and interactions with the world, while maintaining security and privacy

How?

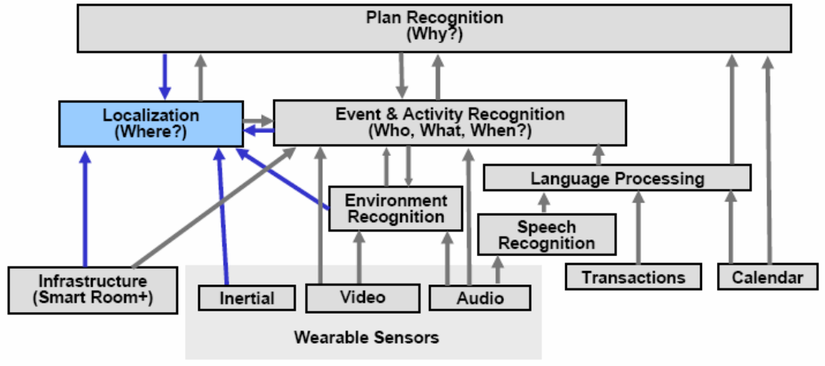
Process and correlate sensor (physical) data streams with other available (transactional) data sources to (nearly-)automatically generate a personal timeline at a useful level of abstraction

Creating an index over a flow of experience



1

Example: alternative paths to determining "Where am I?"



Infrastructure: "Where does the xyz system say I am?"
- GPS, cellphone, 802.11 signal strength, emplaced RFIDs, "smart rooms"

Motion: "Where have I moved to?"
- Dead reckoning: optical flow, pedometer, compass,…
- Need to know where I was

Recognition: "Do I recognize this place?"
- Statistical signature matching
- Need to "know" the place
- Strong feature recognition (e.g., signs)

Plan: "Where did I intend to be?"
- Plan recognition from calendar, email, past history

2

People:	Jim Gemmell, Gordon Bell & Roger Lueder	Affiliation:	Microsoft Research
Project:	*MyLifeBits*		

What?

MyLifeBits is a lifetime store of everything. It is the fulfillment of Vannevar Bush's 1945 Memex vision including full-text search, text & audio annotations, and hyperlinks. There are two parts to MyLifeBits: an experiment in lifetime storage, and a software research effort. The experiment is to digitize as much of Gordon Bell's life as possible. The software research has led to the development of a suite of software tools based on relational database storage.

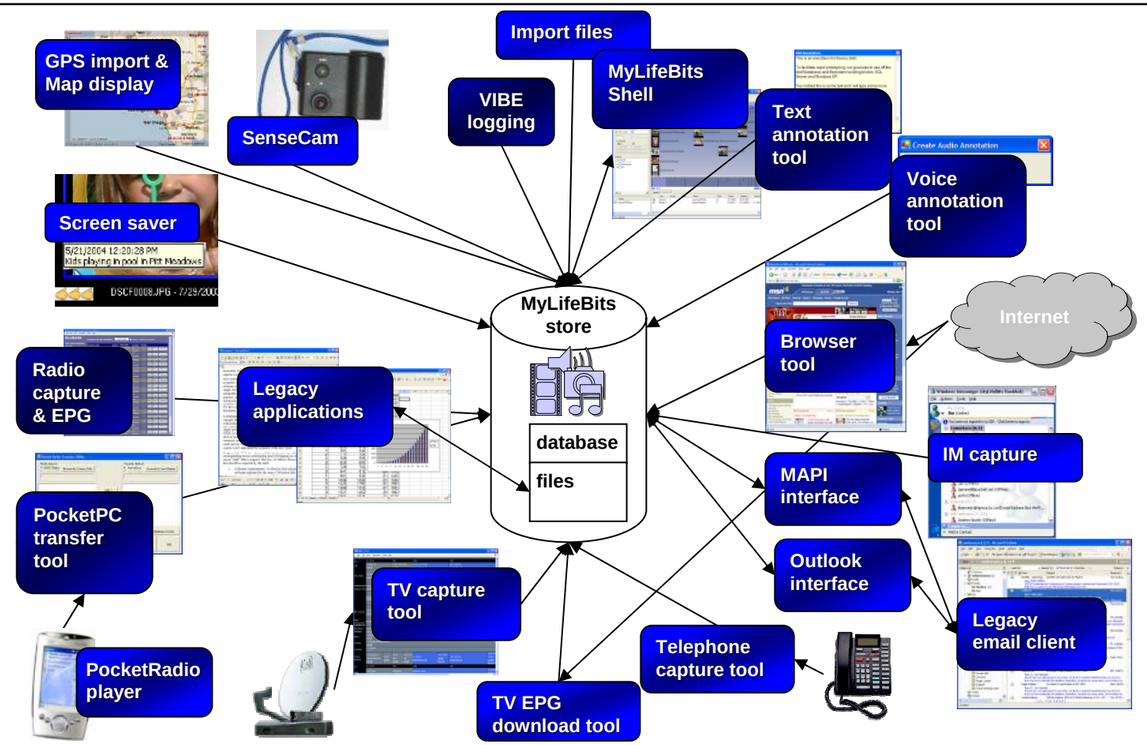
Results so far:

The experiment: Gordon Bell has captured a lifetime's worth of articles, books, cards, CDs, letters, memos, papers, photos, pictures, presentations, home movies, videotaped lectures, and voice recordings and stored them digitally. He is now paperless, and is beginning to capture phone calls, IM transcripts, television, and radio.

Our software is in use by our team, and a new version of the UI is under preparation for user testing. We are active in promoting the CARPE (Capture, Archival and Retrieval of Personal Experiences) research community, including leading the first workshop at ACM Multimedia 2004.

How?

The MyLifeBits software leverages SQL server to support: hyperlinks, annotations, reports, saved queries, pivoting, clustering, and fast search. MyLifeBits is designed to make annotation easy, including gang annotation on right click, voice annotation, and web browser integration. It includes tools to record web pages, IM transcripts, radio and television. The MyLifeBits screensaver supports annotation and rating. We are beginning to explore features such as document similarity ranking and faceted classification. We have collaborated with the WWMX team to get a mapped UI, and with the SenseCam team to digest and display SenseCam output. For papers and more details see www.mylifebits.com



People:	Jonathan Grudin and Eric Horvitz	Affiliation:	Microsoft Research, Adaptive Systems & Interaction
Project:	*Sharing Personal Information Online: What Do We Share With Whom, and How Can This Be Better Managed*		

What?

More personal information is represented digitally, primarily to enable easier sharing of it. Understanding who people want to share what with, and when, and helping them manage this, is a focus of our research.

Why?

More of the personal information we create and work with is digital, but the tools we have for managing it are crude.. Typically we must specify access settings when creating an object and remember to change them as circumstances evolve; most often we do not, and permission is more restrictive or permissive than would be optimal.

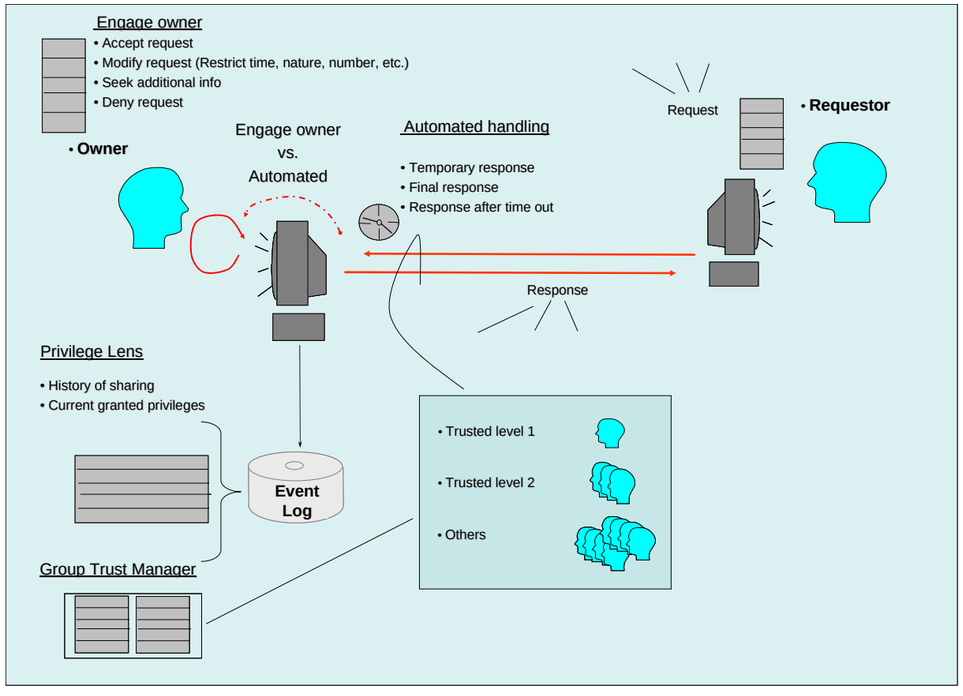
How?

We have conducted studies aimed at identifying patterns of preferences and individual differences in personal information sharing. We also created and are testing access control mechanisms that may allow people to make decisions about sharing in appropriate contexts

Results so far:

Sharing preferences do cluster for objects and people in different categories. We have found indications that responses to a few questions could serve as the basis for presenting a modifiable set of information sharing defaults. People respond favorably to our experimental interfaces for access control.

Graphic:



People:	**Jacek Gwizdka** www.gwizdka.com	Affiliation:	University of Toronto: Psych, MIE, KMDI
Projects:	**Personal Information Management for Future Actions**		

What/Why?

People keep information "in the loop" with *intent* to *later* perform actions on this information.

There is a need to :

- Understand how people manage this kind of info
  - How are the actions expressed in PIM?
- Develop tools to better support PIM for action
  - Internal representation of actions
  - External representation of actions
- Understand individual differences between people that affect PIM strategies

Results so far:

Email Strategy	"cleanrs" Transfer info out of email	"keepers"
Flexibility of closure	low	high

- Identified a range of email strategies:

process now, limit, encode more info, accumulate

- Found individual differences affecting email behavior
- Designed and evaluated electronic engineering notebook (Gwizdka, et al. HFES 1998)
  - abstract ontology vs. own concrete terminology: abstract is reused; own is trusted; coverage of concepts better in free-from UI than in fixed-form
- Designed and evaluated email UI – WebTaskMail (Gwizdka, PhD dissertation 2004, CASCON 2002) – pending task dates presented graphically
  - subjective preference for visual UI; visual UI faster for finding dates
  - individual differences (e.g., low short-term memory slower and sort more often)

How?

Questionnaires & field-studies to characterize issues and strategies

Design and develop tools & techniques to address issues

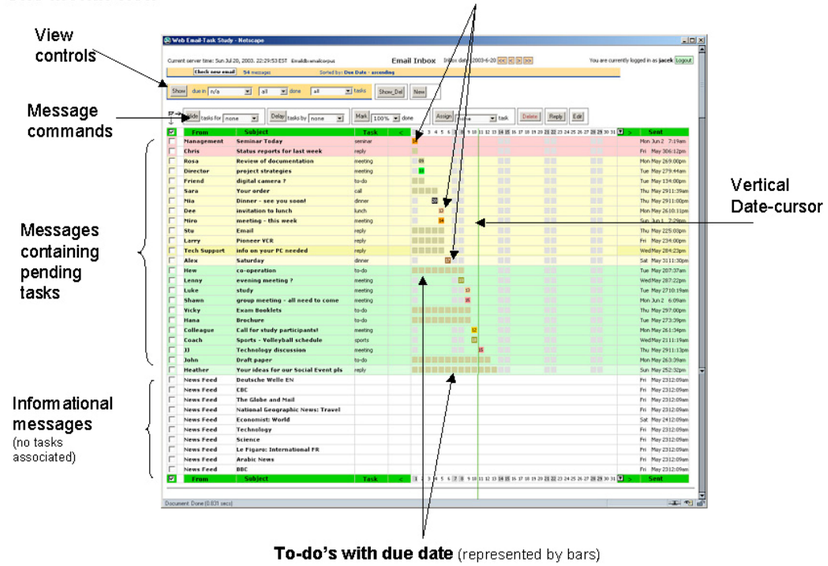
- Metadata / ontologies
- User Interface
  - interaction mechanisms for adding metadata
  - visual representation of information

Evaluate with users (in the field and/or lab)

EEN
(Newton & Java clients; Prolog server)

WebTaskMail (Full-Visual)
One month view

Events (e.g. meetings) with fixed start- and end-dates
(represented by squares. Time of day is color-coded, hour is shown in text)



WebTaskMail (web mail "task-centric" client)

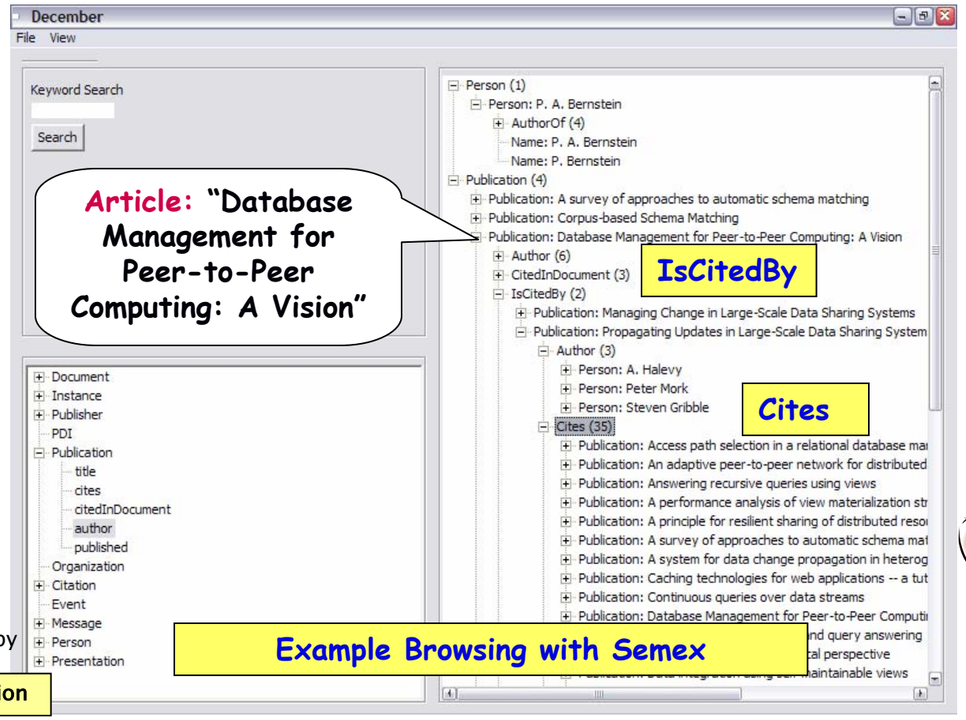
People:	Alon Halevy and Xin (Luna) Dong	**Institution:**	University of Washington
Project:	*Semex: a Platform for Personal Information Management and Integration*		

Questions I Can't Answer (But the answers are on my desktop)

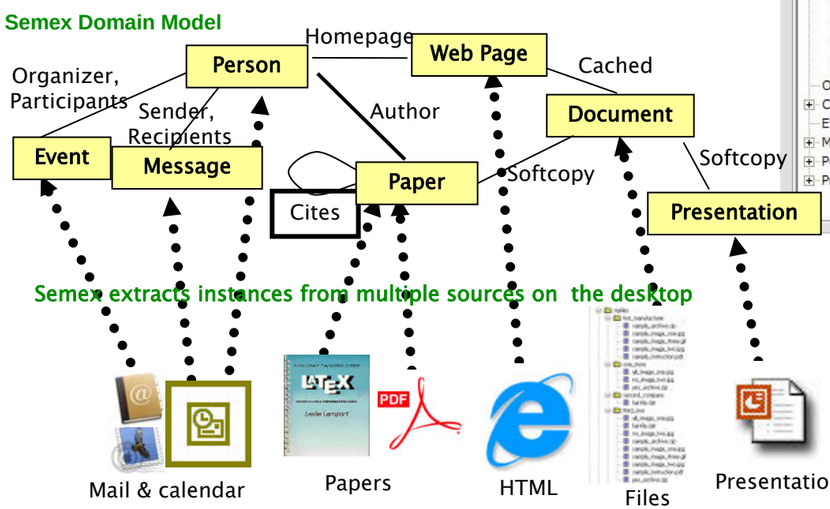
- Find my VLDB04 paper; and the PowerPoint (maybe in an attachment).
- Find emails from my Californian friends.
- Which Widom paper did I cite in my VLDB04 paper?
- What quarter was Mary in my class and what grade did she get?
- Which experiment did I run with *NF1* and which emails discussed them?

The Goals of Semex

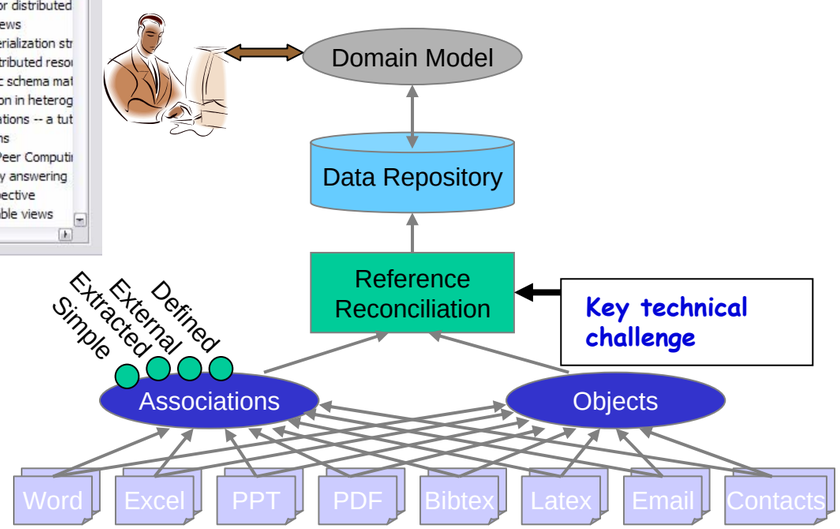
- Enable semantic browsing of the information on one's desktop:
  - How can we create a '**AHA!**' experience with a PIM system? *Should be good enough to be indispensable!*
- Given a knowledge base of one's personal information, how can we leverage it to **increase productivity** in other tasks?
  - First example: enable *on-the-fly* information integration: enable transient integration tasks by non-technical users.



Semex [Semantic Explorer]



System Architecture



People:	William Jones, Charles F. Munat, Harry Bruce	*Affiliation*:	The Information School, University of Washington

Project: The Universal Labeler™: Plan the Project and Let Your Information Follow

What?

- A single, unified scheme of "labeling" to organize your electronic documents, email messages, web references, etc.
- A Project Planner. Use Microsoft Word to structure and order your thoughts. Headings = folders = labels.
- "Remind me by", "Due by" and other project-related properties integrate information and time management (using Microsoft Outlook).
- A part of the Keeping Found Things Found project.

Why?

People keep information to get things done, for example, to complete projects ("finish a course", "re-model a house", etc.).

Project-planning involves problem-solving: A person's conceptualization of a project can often be characterized as a hierarchy of subprojects and tasks.

Project structure can organize related information. Project structure made explicit can aid not only in planning but also in the organization of the information needed for the project.

How?

- An "un-application" approach that stresses integrations with existing applications: the file manager, the calendaring and email application, the word processor, etc.
- "Label With…" for in-context creation of shortcuts.
- "Label Properties…" to set "Due by" and other properties.

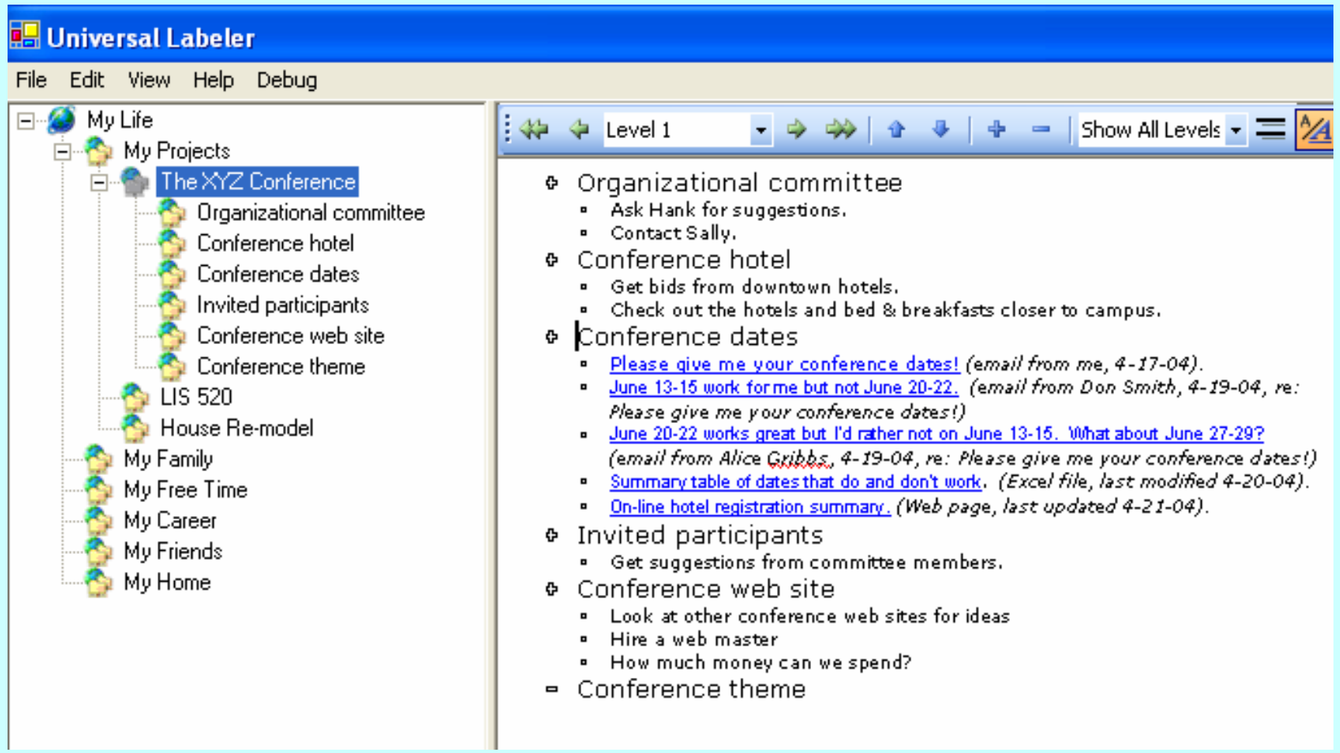
"Drag & Link" to make connections.

Good Results So Far…

- "My Life" and "My Projects" give a useful new slant: folders are labels representing various roles, priorities, activities and projects in a person's life.
- "Label With…" and "Label Properties…" support integration.
- "Top-down thinkers" may find the Planner especially useful.

Next steps…

- Fix problems with "Drag & Link" and the synchronization of views.
- Support "Life Organizers" and the use/re-use of structure.
- Support the Life Cycle for personal information: "reference libraries", archival…



People:	David Karger	Affiliation:	MIT CSAIL
Project:	*Haystack: Per-User Information Environments*		

What?

A system emphasizing end-user control over

- *What information objects are worth recording*
- *what relationships/attributes are worth recording*
- *how to present/edit/navigate those relationships*
- *how to gather into coherent, task-focused workspaces*

A "semantic web browser" letting end user work with arbitrary semistructured information

An end-user application development environment

Why?

To retrieve well, must record properties user remembers

- *Every user remembers different things*
- *Application developers cannot keep up*

Individuals' structured knowledge can be shared

How?

One unified semantic-network data model (RDF)

Recursive rendering architecture: e.g. to display email

- *Lookup author, subject, notes, body*
- *Recursively display each in some layout*

Views, operations are user-customizable data(in RDF)

All visible objects "alive" to context menus, drag & drop

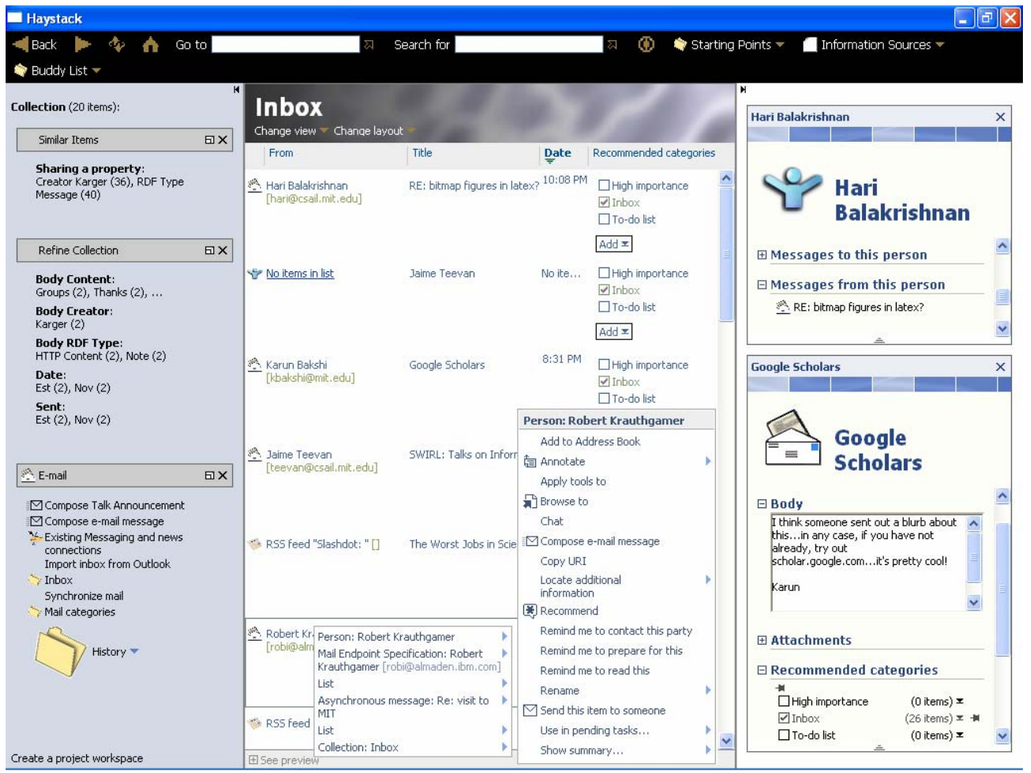
Results so far:

Prototype system available for download at
http://haystack.csail.mit.edu/

Handles numerous data types such as

- *Email, RSS Feeds*
- *Documents, File systems*
- *Music, Photographs*
- *Collections*

under assorted natural views



People:	**Nicholas J. Belkin & Diane Kelly**	Affiliation:	**Rutgers University & University of North Carolina**
Project:	***Contextual Personalization of Support for Information Seeking***		

What?

- Develop implicit, behavior-based techniques for making sense of, and applying the data that users produce each time they seek information to PIM
- Understand how task type, topic familiarity, and other aspects of information seeking context affect PIM and how it can be unobtrusively identified and incorporated into PIM
- Develop mixed method approaches for studying PIMs which incorporate elements of naturalistic, longitudinal and experimental methods

Why?

In order to personalize support for a person's information seeking both locally (on one's own, or one's group's systems) and generally (in external information resources) -- the general idea is to tailor any information seeking episode to the specific context in which it occurs.

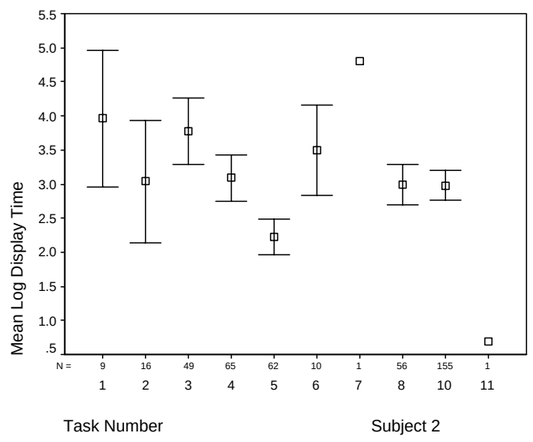
How?

We completed a 14-week study of the information seeking and use activities of seven users. We used client- and server-side software to monitor users' online and offline activities, and supplemented this data with data collected during weekly meetings where users identified personal tasks and topics, classified the web pages that they viewed into these task and topic classes and judged their usefulness.

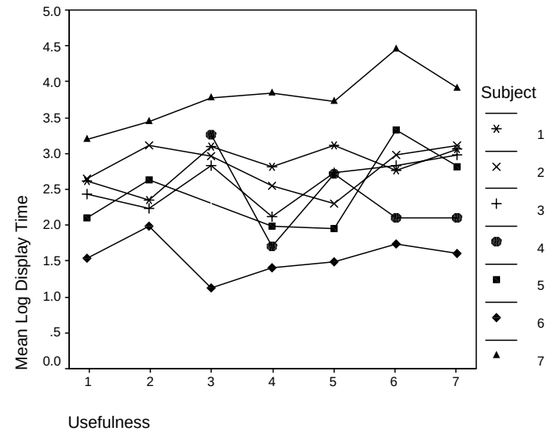
Results so far:

- Commonly used behaviors for inferring document usefulness, such as display time, vary according to user and context
- Task type and topic familiarity are important contextual factors that affect behavior and how behavior-based approaches can be used for PIM
- Users had few problems identifying tasks and topics and classifying the web pages that they viewed into these classes
- Traditional clustering algorithms performed poorly when compared to users' personal web page classifications

Display Time & Task



Display Time & Usefulness



Information and the Quality of Life

David M. Levy * UW Information School

Issues

- Information overload
- Fragmentation
- Busyness
- Speedup

Center for Information and the Quality of Life

- Living laboratory
- Research center
- Teaching center
- Clearinghouse
- Thinktank /policy center

See www.ischool.washington.edu/iql

People:	David Maier, Lois Delcambre, Sun Murthy	Affiliation:	Portland State University, Computer Science
Project:	**Superimposed Information Management**		

What?

- Support "enhancement" of in-situ information sources via an overlay of superimposed information: linking, annotation, excerpts, recombination
- How can we support construction of applications that help create and manage superimposed information?
- Supported by several NSF grants: IRI 9509955, IIS 9817492, EIA 9983518

Results so far:

- SPARCE middleware for marking and accessing content and context for wide range of base information types: XML, PDF, Word, Excel, Video, …
- Prototypes of a variety of superimposed applications: SLIMPad, Superimposed Schematics, RIDPad
- To be used in new NSDL project: Superimposed Tools for Active Arrangement and Elaboration of Educational Resources (w/ E. Fox & L. Cassel)

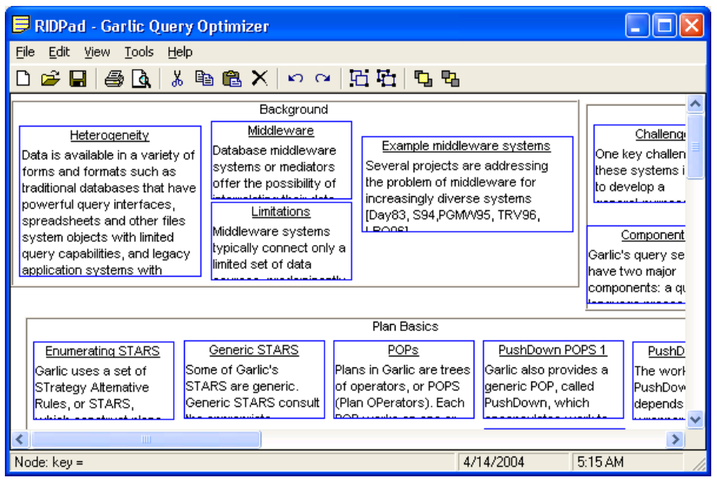
Why?

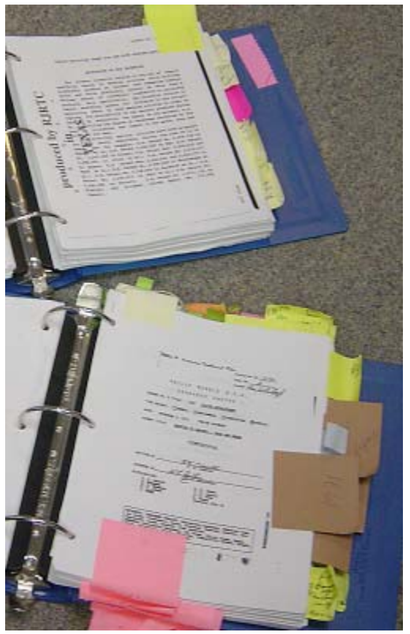
- Finding and keeping documents is only part of most information-intensive tasks. People need to analyze, judge, organize, mark, highlight, extract the information they find in those documents.
- Being able to capture the results of those activities in a digital form can help with the task at hand and future tasks involving the same information.
- Trying to ***Reuse Attention***

How?

- To get ideas for superimposed applications, collaborate with people doing tasks involving large information spaces: clinician exploring patient medical records, chemist analyzing tobacco document archive; forester responding to decision appeals
- Build prototype superimposed applications on top of common middleware layer; keep adding new capabilities to middleware; expand base information types supported.

Graphic:





People:	Cathy Marshall (w/Sara Bly and other collaborators)	Affiliation:	Microsoft
Project:	*Personal Digital Libraries: Making the leap to stuff-centered computing*		

What?

Help transform computers into portable personal digital libraries by filling in missing pieces that go beyond identified needs (e.g. search, readability):

- Interactivity – personalizing digital library materials by annotating, clipping, gathering, triaging, or otherwise interacting with them;
- Encounter and re-encounter – the opposite of searching, the serendipitous side of browsing; and
- Sustainability – archiving at the personal level.

Why?

We're taking a long view of how we keep and use a heterogeneous body of personal materials beyond particular tasks and technologies.

We hope to have an immediate impact on OS-level development efforts and an eventual effect on policy as well as technology.

Of particular note is our emphasis on encountered information, everyday reading, and digital lifespan.

How?

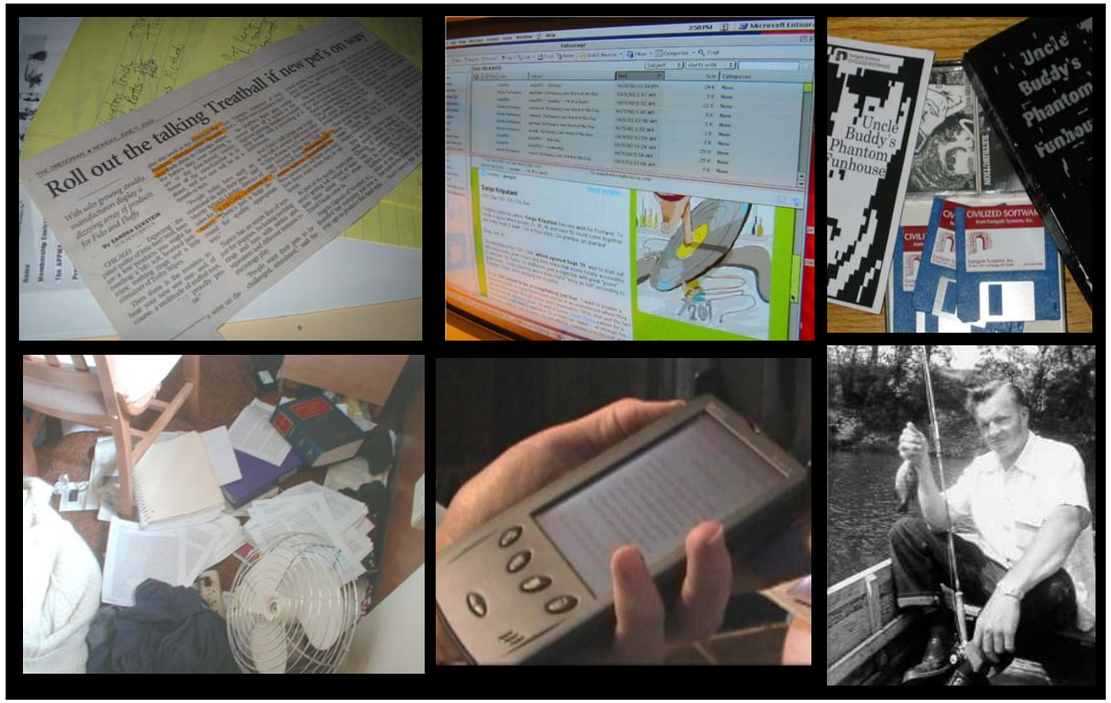
Ethnographic approach to understand a broad range of readers

- Focus on everyday reading;
- Across multiple genres and forms; and
- Over extended periods of time.

Working with product teams to fold the implications into software design and policy decisions.

Results so far:

1. *Over the long term, the most important effect of artifacts of interaction (e.g. annotation, bookmarks) is to form a personal geography.*
2. *Search and SIS exist in counterpoint to encounter/re-encounter.*
3. *In addition to problems associated with institutional archiving, personal sustainability will require us to cope with problems of heterogeneity, distributed data, personalization (e.g. annotations), opacity of document formats, difficulty anticipating value, and security trade-offs.*



http://www.csdl.tamu.edu/~marshall

People:	**Manuel A. Pérez-Quiñones**, Robert Capra, Mary Pinney, Alyssa Sams, Pardha Pyla, Chandresh Chhatpar	Institution:	Center for HCI, VT
Project:	*Multiplatform User Interface Support for PIM*		

What?

- *Information pervasiveness* – multiple devices (laptops, cell phones, pdas, mp3 players, web devices) with access to information sources
- *What are the choices of devices and why? What are the costs of stoping a task on one device and continue on another?*
- *Do all devices need to be able to access and manipulate information? Under what circumstances are "access only" devices appropriate (e.g. iPod and MS SPOT watches)?*
- *How is information organized for future use (a.k.a. keeping things found)?*

Why?

- *Device convergence has impact on PIM tasks*
- *Need to evaluate how different products and tools affect tasks*
- *Understand day-to-day user activities better*

How?

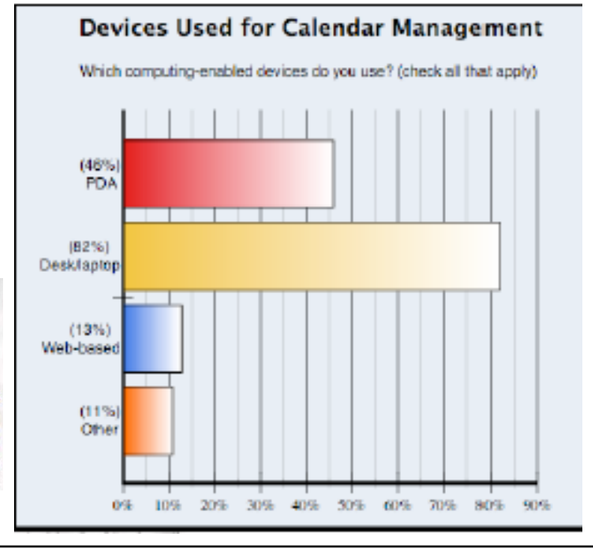
- ***Controlled lab studies** to explore information type effect on information refinding*
- ***Observation study** to explore news browsing behavior*
- ***Online surveys and interviews** to learn about day to day activities (98 surveyed and 15 interviewed for calendar use)*

Results so far:

- *Refinding behavior depends on task & information type*
- *Calendar management SW is complemented with paper products*
- *News browsing behavior is influenced more by information architecture than by user interface layout or platform capabilities*
- *Early development of framework to evaluate cognitive cost of task migration*

Computer Science
@VIRGINIA TECH

CENTER FOR human computer INTERACTION



People:	Wanda Pratt, William Jones, Raya Fidel, Kent Unruh, Andrea Civan, Meredith Skeels	Affiliation:	Information School and Biomedical & Health Informatics, University of Washington
Project:	***Personal Health Information Management for Breast Cancer Patients***		

What?

Focusing on breast cancer patients:

1. Create a descriptive model of the work that patients do to manage their personal health information
2. Develop new technology that helps patients manage their personal health information
3. Evaluate the effectiveness of our new technology in helping patients manage their personal health information, participate in their own health care, and maintain their daily activities

Why?

Without good personal health information management, patients can become:

- Overwhelmed by their information
- Anxious about that information,

which can lead to:

- Increased stress
- Difficulty maintaining desired daily life activities
- Lack of engagement in their health care
- Poor decision making

How?

1. Use cognitive work analysis to study patient work
2. Use participatory design to develop tools that support patients
   a. Organizing health & personal information into different views
   b. Integrating information across applications
   c. Sharing designated health & personal information with others

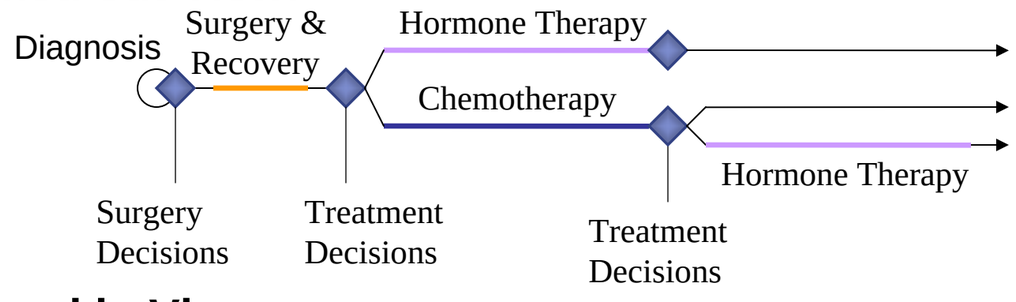
Results so far:

Patient Strategies:

- Create personal health information collections, organized by:
  - Chronological Overviews
  - Event Views (by clusters of significant events, tasks, bundles)
  - People Views (by people)
  - Phase of Care Views (include different organizational units)
  - Predicted Use-Space (all the info needed in a space, task lists)
- Depend on friends and family for help => need to share info

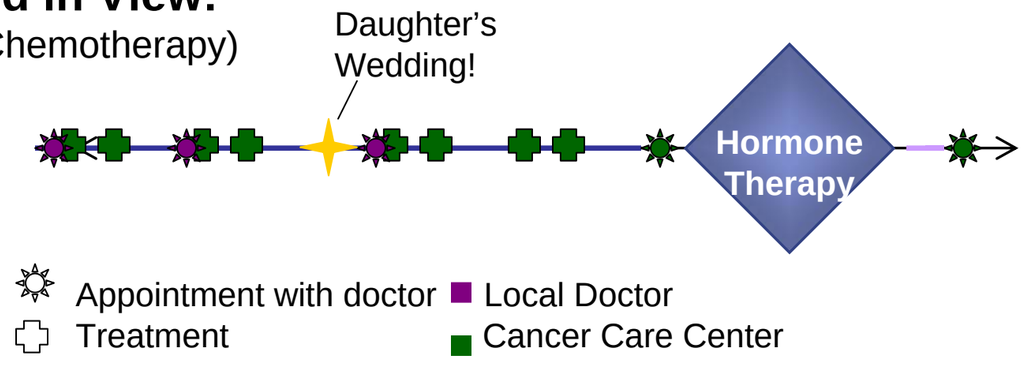
Example Phase of Care Views:

Zoomed out View:



Zoomed in View:

(during Chemotherapy)



People:	Brian H. Ross	Affiliation:	Dept. of Psychology & Beckman Institute, University of Illinois
Project:	Learning and using concepts		

What?

How do people learn and use concepts in multiple tasks (with different goals)? Related to PIM are issues of:

- *How are concepts used for a variety of purposes?*
- *How do people organize concepts in real-world domains and deal with issues of cross-classification?*
- *How do people access relevant information to accomplish a goal?*
- *How are actions incorporated into conceptual representations?*

Why?

The concepts people have influence their understanding and use of new technologies.

A clearer idea of how concepts are learned and used in these various activities may provide a better idea of how to design particular applications as well as how to integrate various activities.

How?

Cognitive Psychology/Cognitive Science Approach:

Study of people learning new concepts in lab settings, in problem solving situations, and in simple virtual reality environments.

Examination of real-world concepts (foods, social categories) and how they are organized and used in inferences and explanations.

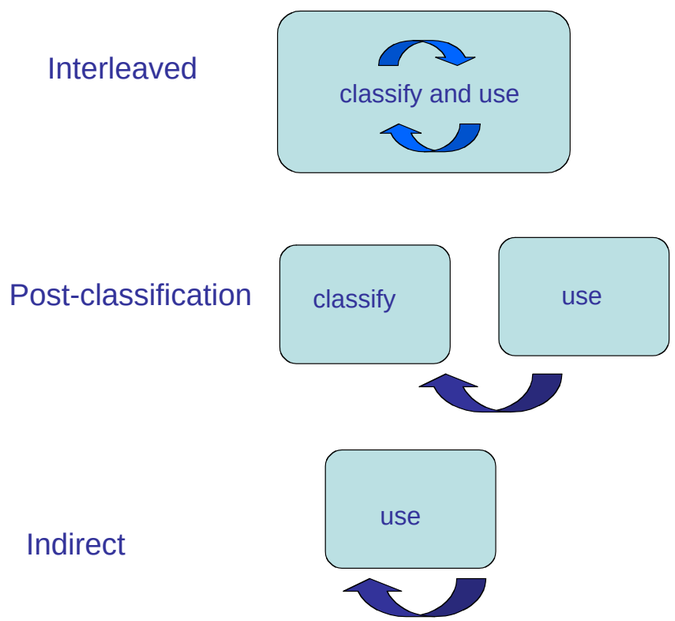
Results so far:

The use of a concept during learning has large influences on what is learned and what is available for other uses.

- *Concept evolves over extended use(s).*
- *Some types of learning lead to more flexible uses.*

Cross-classification of real-world concepts – coherence of concepts influences what knowledge is used for inferences and explanations.

Concept learning is influenced by how the concept is used even in simple learning cases.



THE RE:SEARCH ENGINE

Helping People Return to Information on the Web

Jaime Teevan * CSAIL, Massachusetts Institute of Technology * teevan@csail.mit.edu

Results: GPS

Garmin eTrex GPS
With waterproof exterior (yellow) th…

GPS For Dummies
By Joel McNamara, this book introd…

Magellan SporTrak Color GPS
Features a large 240x160 pixel, high …

Microsoft Streets and Trips 2005
CD ROM with accurate destination r…

Geological and Planetary Sciences
CalTech's GPS division. Academic …

A past search for "GPS" produces some search results.

Things people remember about search results:
- Clicked results
- Anomalous results
- First and last result

What is remembered is affected by:
- Task being performed
- Elapsed time

PAST

Search results change over time as search engines update their indices to reflect the current state of information.

Results: GPS

Garmin Legend GPS
New version released, replaces eTrex...

New Super Improved GPS System
Just out on the market, this is the bes…

Timex Speed and Distance GPS
Triathlon watch and workout compa…

Microsoft Streets and Trips 2005
CD ROM with accurate destination r…

Magellan SporTrak Color GPS
Features a large, high resolution scre…

Later, the same search produces different results.

The Re:Search Engine recalls the original results for repeated searches and incorporates new information, changing only what's not remembered.

PRESENT

Results: GPS

Garmin eTrex GPS
With waterproof exterior (yellow) th…

Garmin Legend GPS
New version released, replaces eTrex...

New Super Improved GPS System
Just out on the market, this is the bes…

Magellan SporTrak Color GPS
Features a large, high resolution scre…

Geological and Planetary Sciences
CalTech's GPS division. Academic …

The Re:Search Engine merges past and present results.

How people return to information:
- Prefer to take small, local steps over large leaps
- Use contextual knowledge
- Delay specification of information need

On the Web, the information environment often changes.

Benefit: Changes provide new information
Drawback: Changes make returning to information hard!

The drawback can be mitigated if the environment *appears* static. Because many changes go unnoticed, new information can be incorporated with old invisibly by preserving only what the user remembers of the original information.

People:	Steve Whittaker s.whittaker@shef.ac.uk	Affiliation:	University of Sheffield, UK
Project:	Characterising and Addressing Information Overload		

What?

People have too much new, working and archival information to deal with.
They need strategies and systems for:
- filtering new incoming information
- managing working information related to current tasks
- filing useful non-current information.

A general problem, not just email, but voicemail, IM, paper, phone

Why?

A pervasive user problem that is getting worse as connectivity increases.
The research should produce:
Deeper understanding of user problems with overload.
New systems
- *short term* -> new interfaces (email, voicemail, browsers, filing systems)
- *long term* -> new types of systems that give users greater control over their information, rather than being under control of their inbox. Inbox currently controls/manipulates activities, we need to redesign it to reflect users' tasks and interests

How?

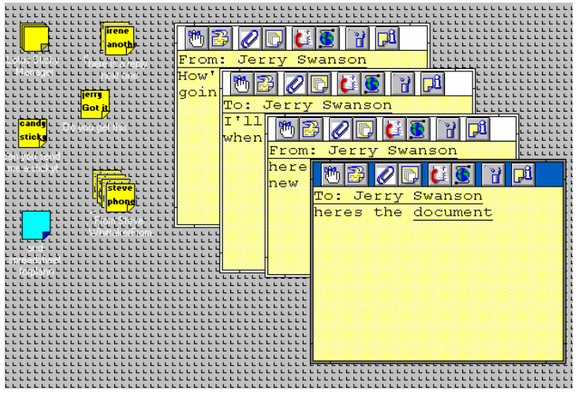
Iterative user-centric system design.
- ethnographic user studies to characterize user problems/strategies
- design novel systems (interfaces and algorithms) to address problems
- users evaluate resulting systems
- iterate

Results so far:

Empirical studies have identified general overload problems and user strategies in email, voicemail and paper management, including
- dealing with new information
- managing working information/tasks - reminding about tasks, collating information related to tasks, (note strong individual differences in strategies)
- filing useful non-current information (Whittaker and Hirschberg, 2001, Whittaker and Sidner, 1996, Whittaker et al., 1998, 2002, Whittaker, in press)

Built and evaluated two novel systems constructed to address these problems
- TeleNotes – email client task-centric/filter ephemeral information
- ContactMap – social interface – access and organize email/information by people and projects (Whittaker et al., in press)

TeleNotes



ContactMap