15. Personal Information Management

INFO 202 - 20 October 2008
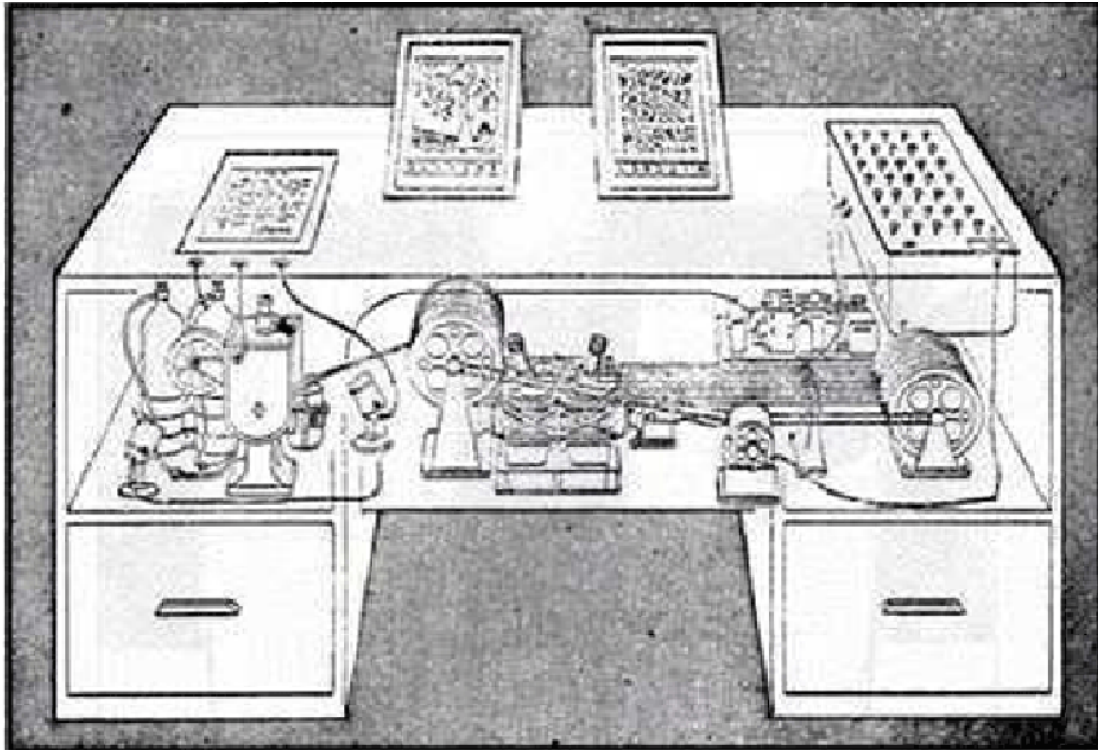
Bob Glushko

Plan for Today's Class

Defining "Personal Information Management"

PIM Technology and Tools

Personal vs Common Information Spaces

PIM Activities and Strategies

Vannevar Bush's Personal Information Manager

Personal Information Management (PIM) - A Fanciful Definition

Personal Information Management is a game of catch

... in which a person tosses their personal information into the future

... in the hope of being able to catch it later

Maybe "later" is "forever" and we hope that someone else will do the catching

PIM - A More Serious Definition

"The practice and the study of the activities that people perform to acquire, organize, maintain, and retrieve information for everyday use"

So we limit PIM to cover actions (or inactions) that are the result of individual choices

It is also personal -- rather than social -- is that we decide on the activities and organizational schemes and carry them out by ourselves

(although it may be part of our PIM strategy to rely partly on others)

Having this discretion about information organization and "making sense of it" to make decisions is a defining characteristic of professional, as opposed to clerical, information work

Why PIM Matters

PIM matters to us as individuals and professionals because better PIM results in better use of our time and attention and ultimately in better quality of life

PIM matters within enterprises because better PIM means increased productivity (in the short term) and better knowledge management (in the long term)

Advances in PIM may also translate to improvements in education and in helping old people "match their mental lifespan to their physical lifespan"

PIM and Prevailing Technology

We can view many inventions as responses to the need for PIM

Because PIM is embedded in user tasks and work context it reflects the prevailing technology support for information work

The personal computer has had the greatest impact (so far) on PIM technology

But as processors and connectivity is increasingly embedded in objects of all kinds information "from and about stuff" will have to be managed and the PC may lose its central role

And replication of digital objects so they can be stored "in the cloud" is becoming common

19th Century PIM Technology Breakthroughs

Six Common Strategies for Personal Digital Archives (Marshall)

Periodic system backups

Copy old "My Documents" directories each time a new PC is acquired

Save "important" files to external storage media (CDs, DVDs, USB drives)

Use web email services and send important files to self as mail attachments

Post photos and videos to social media sites

Save old computers with installed software and associated peripherals and reboot as needed to access old files

The Contemporary PIM Challenge

Vastly more personal information of all kinds

Much of this information is in incompatible digital formats produced and consumed by a variety of devices and applications

These devices and applications may separately provide support for PIM that is collectively incompatible

But our work often requires that information be created, managed, and retrieved in ways that cut across these format and device boundaries

Issues Raised by Multiple and Distributed Copies

People often create multiple versions of the same digital object over time or for different applications

These versions can differ in:

- The type and extent of associated metadata or description
- Representational fidelity or format
- The collection or context in which they are organized
- How the "authoritative" one is identified

Data Unification in PIM

APPROACH	EXAMPLE	OPERATIONS	ENABLES
Standard datatype	Text	Cut/copy/ paste	Unified search
Unified presentation	Window manager	Layout, tile, show, hide	Simultaneous view
Unified namespace		Reference, de-reference	
Grouping	Directories	Group, ungroup	Organize, browse
Metadata	ID3 tags, XML	Annotate, query	Search, organize, browse
Cross-ref	Web links, OLE	Link, follow, embed	Simultaneous view, orient
Relations	RDF, Haystack	Record named relations	Unified search, browse, orient

"Data Unification" Via Standard Datatype Should Remind You Of...
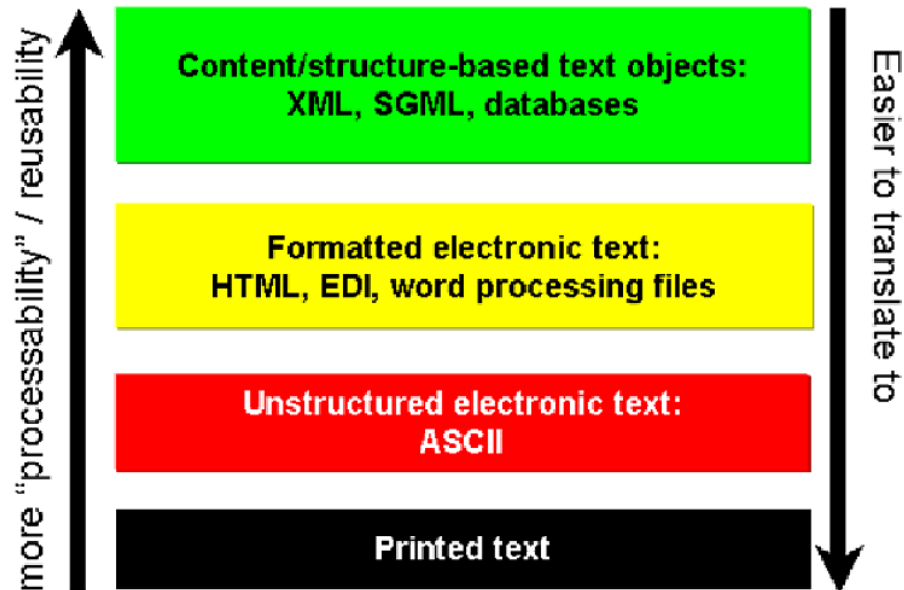
E-Mail and PIM

E-mail was designed as a communications application, but it almost always takes on task and information management functions

A "touch once" strategy assumes that messages can be classified as either "information" or "correspondence"

But this can't be done reliably

Strategies for organizing messages have varying effectiveness and costs

Favorite / Least Favorite PIM Tools and Features

What's your favorite / most used / most usable tool or tool feature for your PIM?

Are your organizational schemes and document models the same, similar, or consistent across the PIM tools you use?

How do your choices affect others in your social or work networks?

"Personal" in Context

We want to focus on "personal" IM by excluding activities that are mandated by the social, institutional, or work context a person is in

But that's impossible if we think that a person's conceptual systems are "embodied" in their perceptions, physical and social experiences, etc...

So PIM is always shaped by the context of work

There may also be conflicts between what people would do as individuals and what they are required to do as employees, project members, etc

The Concept of a Personal Space of Information

All of the information that is nominally under a person's control defines one and only one PSI for that person

So the PSI is an abstract collection of all of the various physical and digital information items

Some of it is located "in the network" and is thus only indirectly controllable, but people tend to think of all the distributed digital storage as being "in one place"

A fundamental challenge is maintaining an integrated view of the PSI

A corollary challenge is reconciling the tension between one's PSI and the need to contribute and use information from common information spaces

Content and Containers in My PSI

Printed Documents In Notebooks

Separate Printed Documents

Files on Personal Computer

Files on Network

Books

Company Email

Newspapers, Magazines & Journals

Business Cards

Personal Email

Web Bookmarks

Desk Phone Voice Mail

Maps

Printed Calendars

Cell Phone Voice Mail

PDA (Files, Phone #s, Calendar)

Desk Drawers

Whiteboard

Cell Phone Phone #s

Bookcase

Post-It

File Cabinet

Tables

Pushed vs Pulled Information

Pushed information comes to our PSI (not as the result of any *immediate* action on our part)

Pulled or *Retrievable* information is what we intentionally seek

Categorizing PIM Activities

Keeping
activities are carried out when you encounter information - they determine the input of information into the PSI

Finding or re-finding
activities within the PSI determine the output of info from the PSI or its use in work

"Curation," maintenance and organization
activities are those taken to organize, re-organize, or integrate the PSI

Decisions to Make on Information Exposure / Receipt

Consume immediately

Ignore

Keep

Keeping and Organization

Does the "to be kept" information item fit into our existing classification scheme within our PSI?

This is often not an easy or binary decision

Is there a scheme for assigning names to (or renaming) information objects?

If we use an existing category or location or naming scheme, is the degree of "fit" acceptable or does it distort the existing organizational or naming scheme?

Do we abstract / summarize / data mine / secure or otherwise process the "to be kept" information?

How much of this "keeping and organization" work could be done automatically?

Does our organization or plan assume or imply future stewardship actions on our part, on anyone else's, or by some automated process?

Where We Keep Information

Action information
- immediately available, organized by task, often in physical formats and organized spatially; often serves as reminders of what to do as much as information to help do it

Personal work files
- accessible, organized spatially or according to conceptual categories, but not intended for use by others

Archive storage
- not as immediately accessible, often files of completed work and possibly organized for use by others

"Pilers" and "Filers" in Malone's Classic Paper

Malone, T. W. (1983). How do people organize their desks?: Implications for the design of office information systems. ACM Transactions on Information Systems, 1, 99-112

- A classic paper that contrasted the strategies of "filing" and "piling" in managing information in offices

- Both files and piles are ways of collecting groups of elements into larger units; files are units where the instances are explicitly titled and arranged in some systematic order, and these structures may themselves to explicitly titled and systematically organized

- Piles tend not to have internal structure, other than access frequency; their spatial location is often the key to finding them

Photo of a Piler's Desk

Photo of a Filer's Desk

PIM and Search

Search can be considered as the unifying or integrating function that cuts applications or information sources to help make sense of your PSI

This is especially true in PC or intranet environments, where "desktop" or "enterprise" search applications can compensate for limited PIM

We make decisions about the likelihood of having to re-find something when we organize it and the expected cost of re-finding it versus finding it "from scratch"

How does "regular" search of the kind you'd use on Google not take advantage of the special properties of Personal information?

Finding vs Re-finding

Finding
implies when no previous experience (or memory of) having the needed information

There is often a stronger emotional dimension in re-finding previously experienced information

Re-finding
is facilitated because people generally know more contextual metadata about previously-found information

Re-encountering
takes place when we find something that we'd forgotten we had; this happens occasionally for physical things, but much less often for digital ones

Kirsh on "Cognitive Overload"

Kirsh points out that "where we work" is a superposition of many specific environments and applications that we move in and out of

Each of these environments and applications has its own cost structure for handling information based on the the tools and resources it makes available

These diverse cost structures result in "computational complexity" for making PIM decisions about keeping and finding information and encourage suboptimal reactive methods rather than careful planning
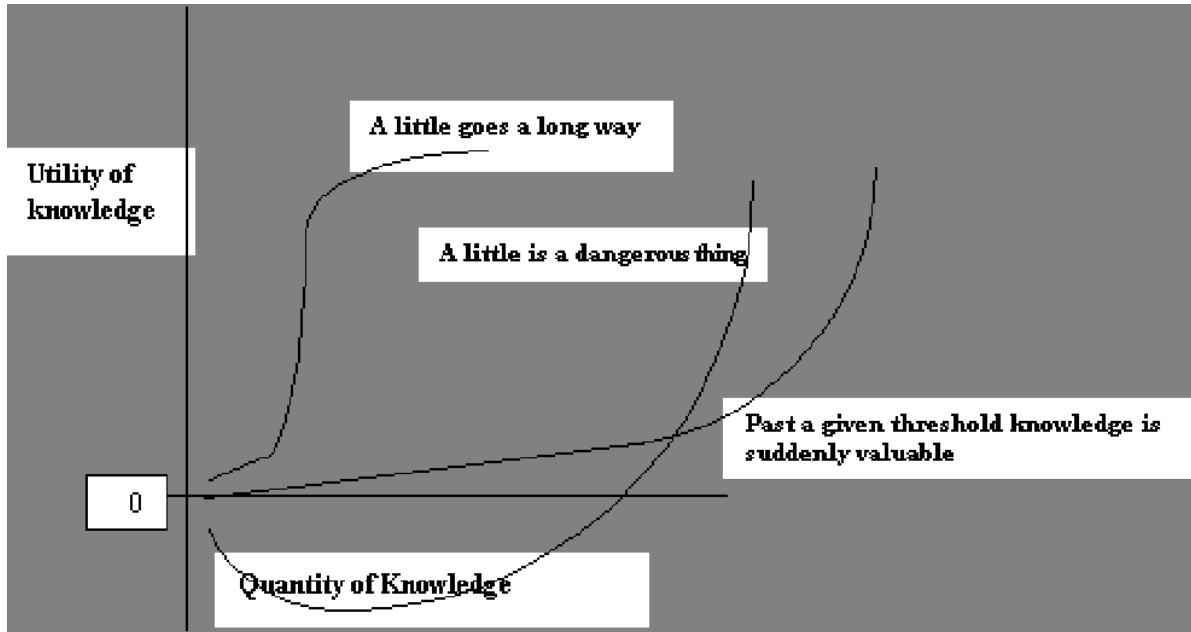
Kirsh on "Information Utility"

People think about and value different kinds of information in different ways - their utility functions are non-monotonic and non-linear

When you are looking for information, can you tell how hard or long it will be before you find it?

How long should you look before you give up?

How valuable will information you don't yet have turn out to be?

Kirsh's Information Utility Functions

A little goes a long way

Utility of knowledge

A little is a dangerous thing

Past a given threshold knowledge is suddenly valuable

0

Quantity of Knowledge

PIM Strategies [1]

The lack of a coherent utility or demand function for information means that different people (or the same person at different times) will follow radically different PIM strategies

Pack Rat or *Blind Accumulation*
-- just save everything, usually spends excessive time filing

Insurer
-- Doesn't keep everything, but creates multiple copies (paper and digital) of information items to maximize re-finding (KFTF)

Surface Clutterer
-- Doesn't keep everything, but strives to keep information accessible, often in spatially organized piles

PIM Strategies [2]

Just-in-Case Learner
-- Spends excessive time consuming information when it arrives so they can always be prepared for some future information need

Just-in-Time Gatherer
-- Ignore all information needs except those needed immediately for current tasks. Maximizes the average value of information items, but some high-value information can't be found this way

Evaluating PIM Strategy

Is it better to pay a small and occasional incremental cost associated with a locally sub-optimal PIM strategy than to pay the certain large costs of an explicitly good strategy?
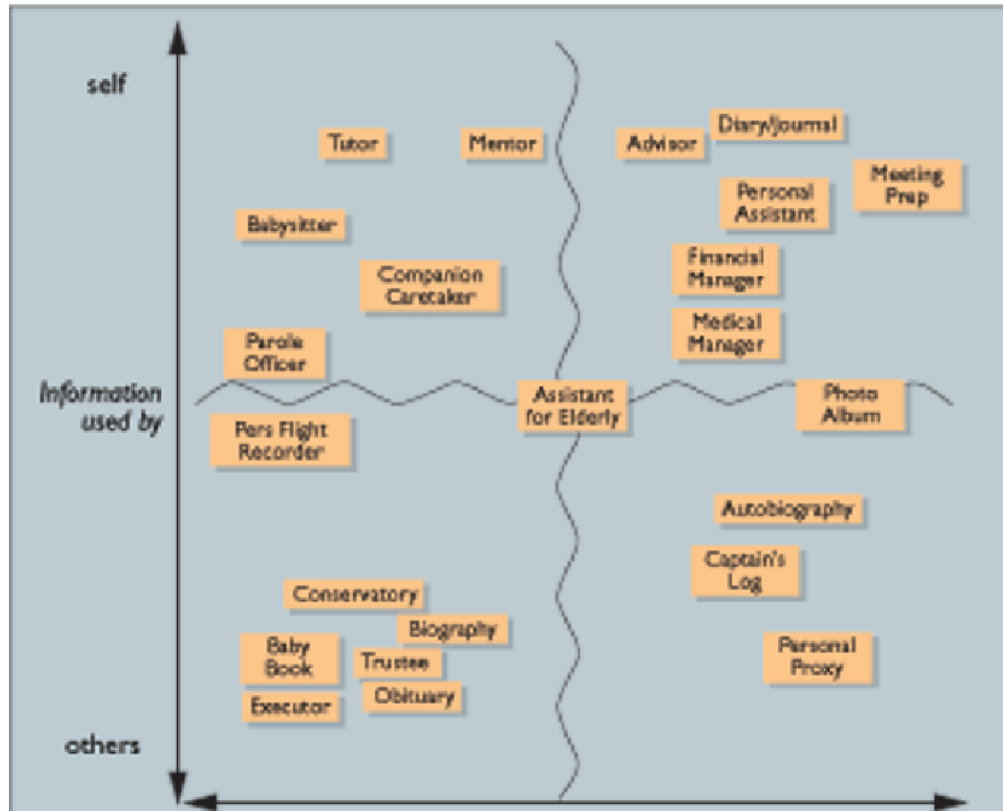
How can you evaluate your management of information in physical formats ?

How can you evaluate your management of information in digital formats?

How does the representation of information change its "cost structure" in using it?

Is this change the same for all "computational agents" ... including people?

A Lifetime of PIM

Readings for INFO Lecture #16

William L. Kuechler, "Business Applications of Unstructured Text," Communications of the ACM, 50(10), 2007, 86-93.

Alexander B. Schwarzman, Hyunmin Hur, Shu-Li Pai, and Carter Glass, "XML-centric workflow offers benefits to scholarly publishers," XML 2004 Conference