Enterprise and Desktop Search

Lecture 5: Desktop Search and Personal Information Management

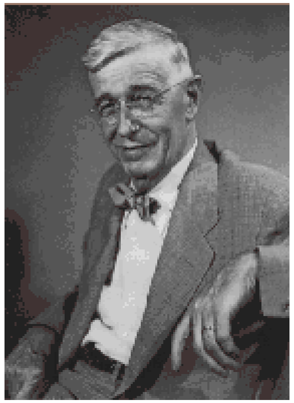
Pavel Dmitriev
Yahoo! Labs
Sunnyvale, CA
USA

Pavel Serdyukov
Delft University of Technology
Netherlands

Sergey Chernov
L3S Research Center
Hannover
Germany

Searching Personal Collections with Memex

Posited by Vannevar Bush in "As We May Think"
The Atlantic Monthly, July 1945

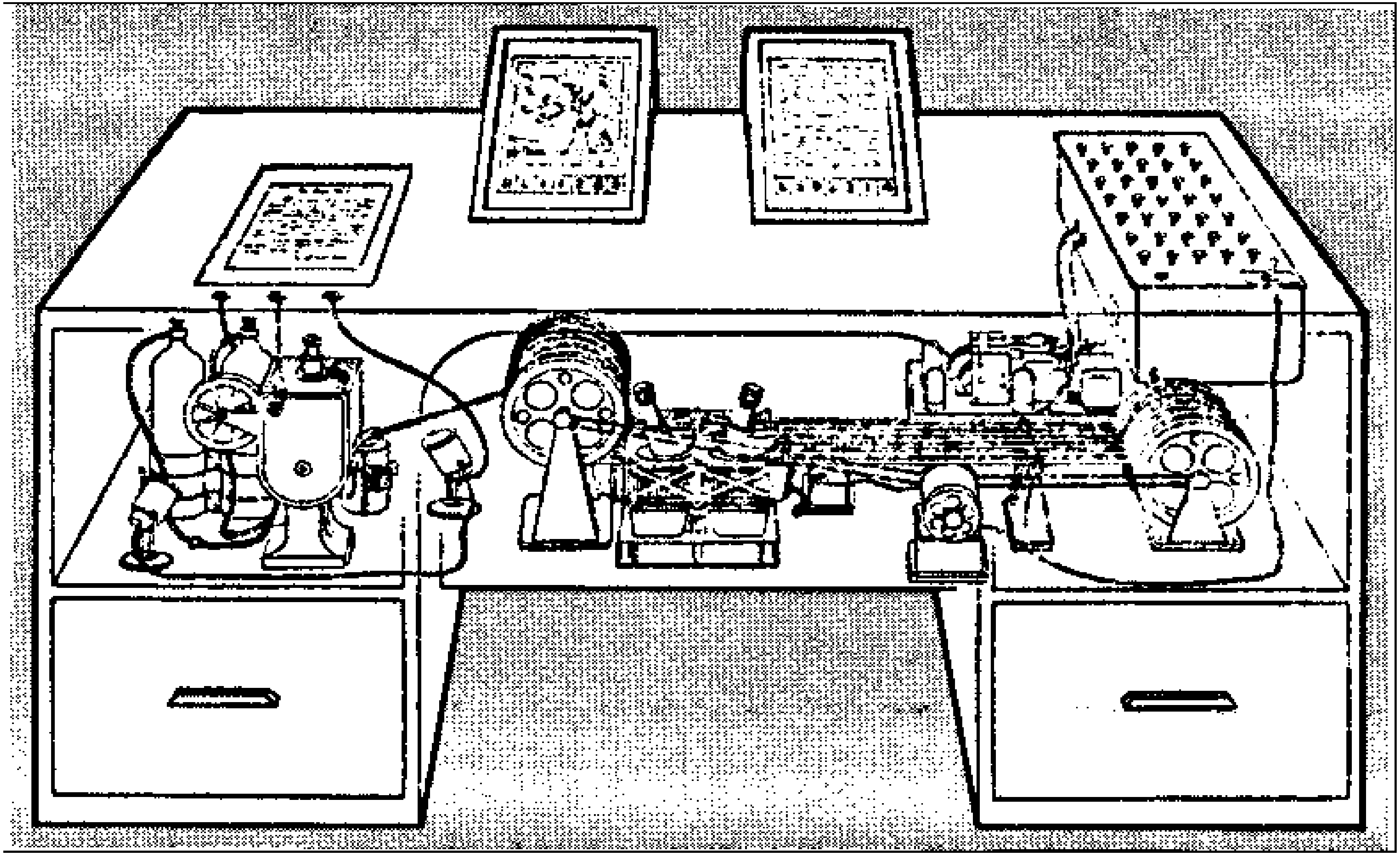


"A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility"

Supports: Annotations, links between documents, and "trails" through the documents

"yet if the user inserted 5000 pages of material a day it would take him hundreds of years to fill the repository, so that he can be profligate and enter material freely"

Sketch of Memex



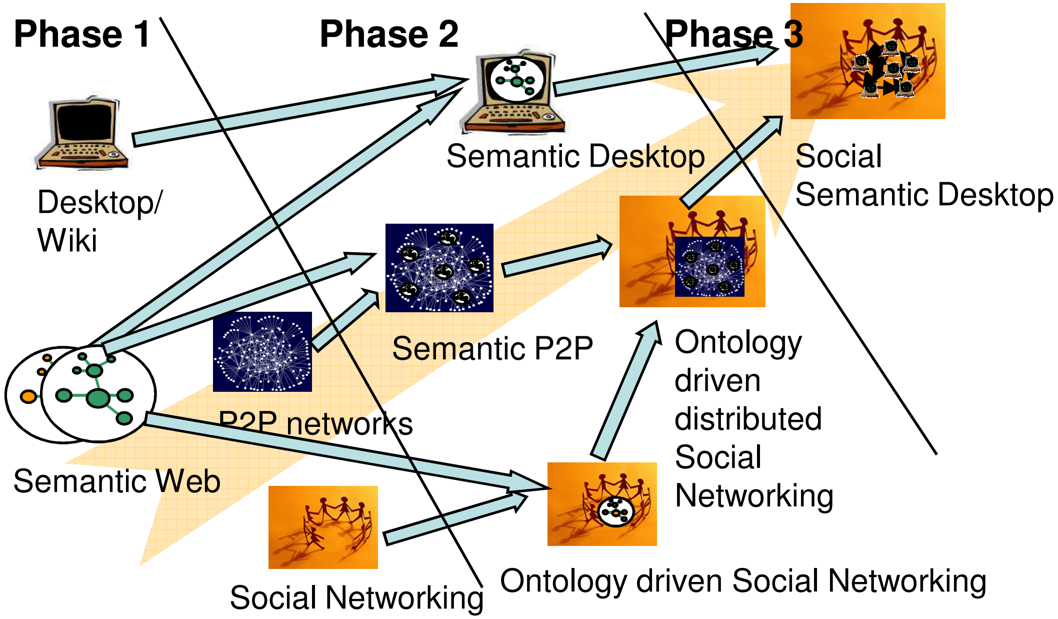
Desktop Search and Personal Information Management

- **Desktop search** is the name for the field of search tools which search the contents of a user's own computer files, rather than searching the Internet. These tools are designed to find information on the user's PC, including web browser histories, e-mail archives, text documents, sound files, images and video.
- **Desktop Search** is a part of a more general field of **Personal Information Management** (**PIM**).
- **Personal Information Management** (**PIM**) refers to both the practice and the study of the activities people perform in order to acquire, organize, maintain, retrieve and use information items such as documents (paper-based and digital), web pages and email messages for everyday use to complete tasks (work-related or not) and fulfill a person's various roles (as parent, employee, friend, member of community, etc.)

Source: Wikipedia

Desktop Search: Motivation

- Why *desktop search*?
  - Size of data on the desktop is big (50k – 500k items) and continously growing
  - Moving towards Social Semantic Desktop
  - Social – communication in a social network
  - Semantic – metadata descriptions and relations



What is Desktop?

- Documents (doc, pdf, ppt, xls, html, txt, …)
- Email
- Calendar
- Instant Messengers (ICQ, Skype, MSN messenger, …)
- Pictures
- Music
- Videos

Desktop Search – Current Status

- Documents on the desktop are not linked to each other in a way comparable to the web
- Simple full text search
  - no personalization
  - no context
  - no ranking possible or too poor
- Metadata enriched search makes use of
  - associations to contexts and activities
  - provenience of information
  - sophisticated classification hierarchies

Spotlight

Windows Search

Differences between Web Search and Desktop Search

- Search on the desktop vs. Search on the Web
  - Re-finding vs. finding
  - Integration across many applications and file formats
  - Users prefer to navigate, not to search
  - Many information types: ephemeral, working, archived
  - Extra sources for ranking improvement:
    - File metadata
    - Usage metadata
    - Folder structure
  - Privacy concerns

Outline

- Today we will talk about:
  - – Modern Desktop Search Engines
  - – Research prototypes
  - – Just-In-Time Retrieval
  - – Context on a Desktop
    - Using context to improve Desktop Search
    - Context Detection
  - – PIM Evaluation

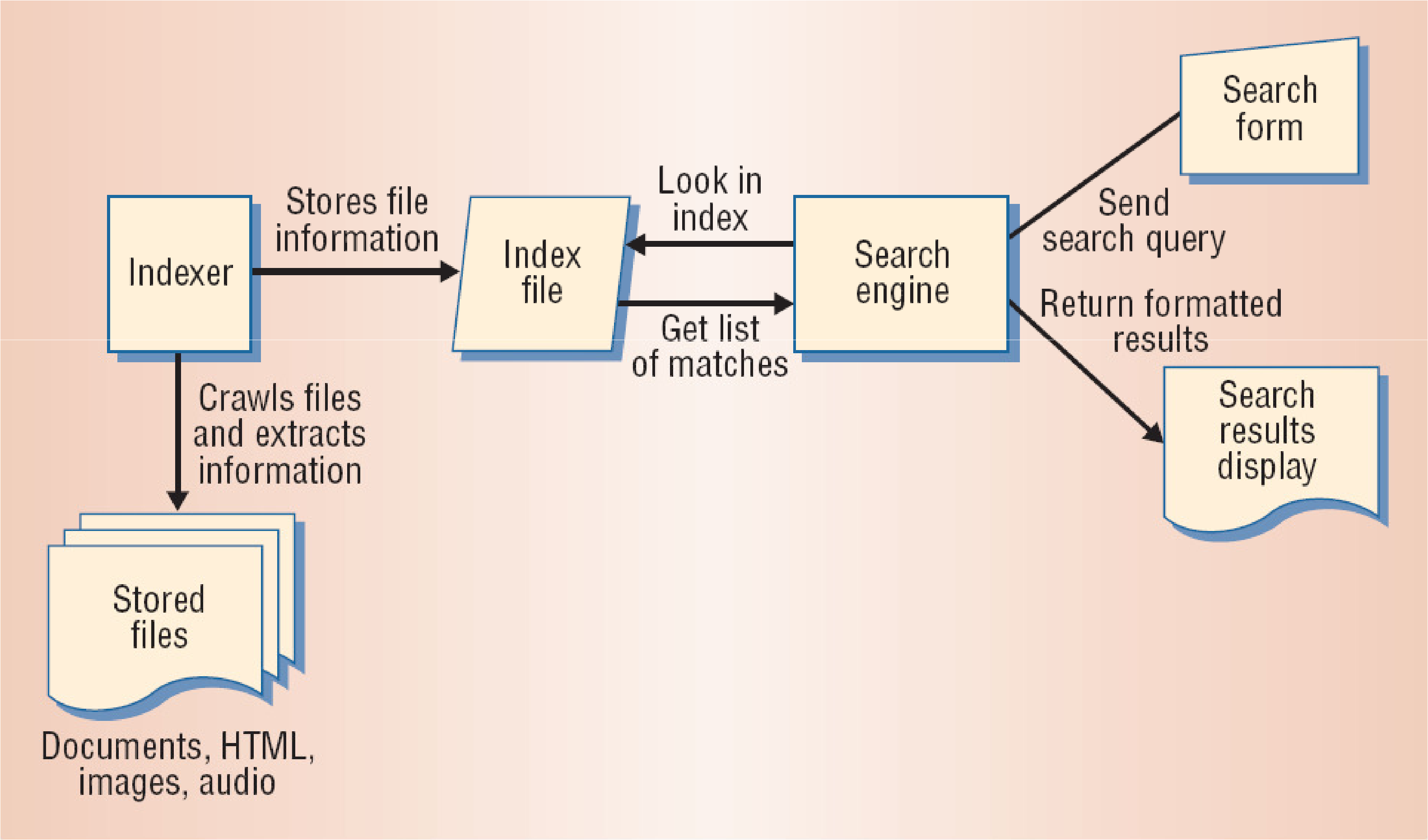
Modern Desktop Search Engines

- Google Desktop (from major web search engine vendor)
- Windows Search (from major OS provider)
- Copernicus (company specialized on DS engines)
- Beagle (open source DS for Linux)
- Yandex (Russian DS)

Some more:

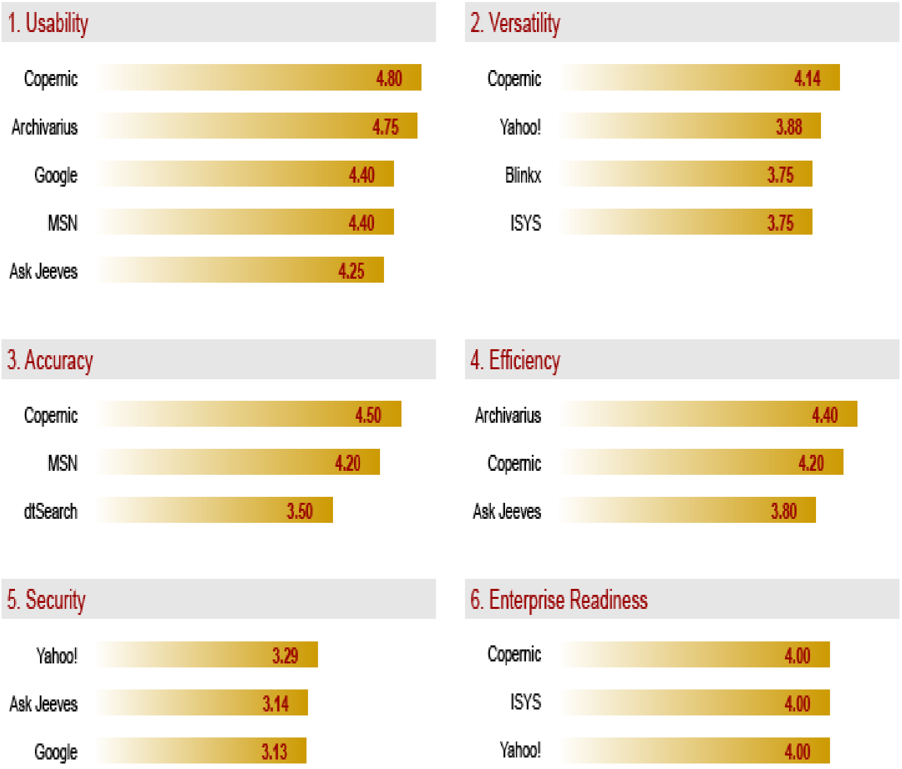
Ask.com, Autonomy, Docco, dtSearch Desktop, Easyfind, Filehawk, Gaviri PocketSearch, GNOME Storage, imgSeek, ISYS Search Software, Likasoft Archivarius 3000, Meta Tracker, Spotlight, Strigi, Terrier Search Engine, Tropes Zoom, X1 Professional Client, etc.

Desktop Search Architecture



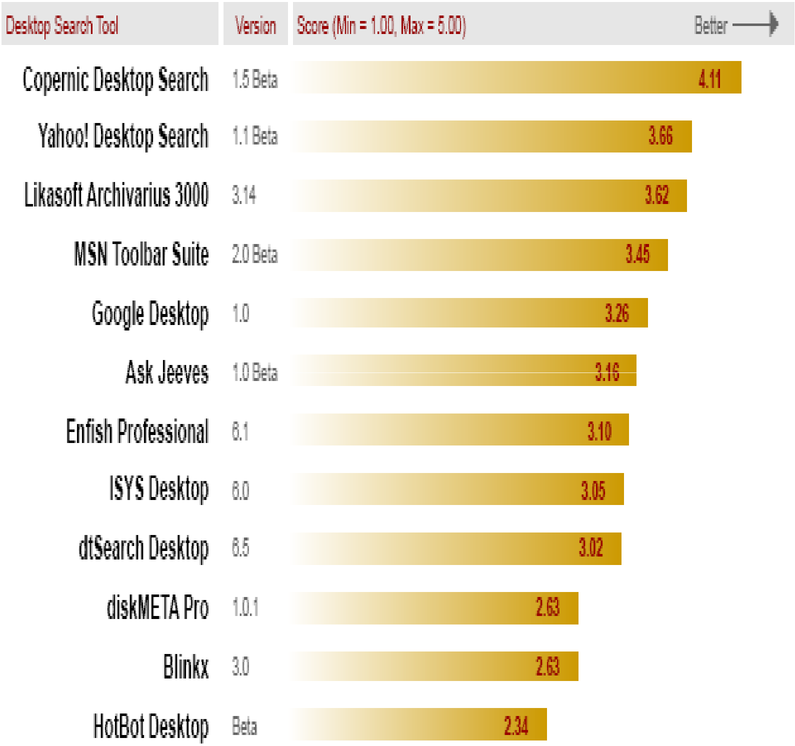
Search Engines Tackle the Desktop,
Bernard Cole, Computer 2005.

Desktop Search Engines in 2005



* Copernic with Coveo, and Yahoo! with X1

Source: UW E-Business Consortium



Source: UW E-Business Consortium

Benchmark Study of Desktop Search Tools, Tom Noda and Shawn Helwig, Technical Report 2005, http://www.uwebi.org/reports/desktop_search.pdf.

Sample Criteria for DS Comparison

Search Format

Plain text

HTML pages stored locally

Microsoft Word (.doc)

Microsoft Excel (.xls)

Microsoft PowerPoint (.ppt)

Rich Text Format (.rtf)

Portable Document Format (.pdf)

Microsoft Outlook email

Microsoft Outlook Express email

Microsoft address books

AOL Instant Messenger

Standard email folder support

Standard news folder support

Browser web history

Browser secure web history

Browser bookmarks

Browser address books

Platform(s)

Windows Vista

Windows XP

Mac OS X

Linux

Mozilla/Firefox

Internet Explorer

Opera

Safari

Languages

Feature

Specifying index location

Incremental indexing

Legacy index by scanning

Engine download size

Install size

Combined local/remote search

Non-anonymous connections

Excluding files

Indexing progress indicator

Recoverable index

File type filtering

Deskbar

Support for compressed files

Support for legacy file formats

Ignoring networked drives

Click to suspend

Click to exit

Opt-in Feature

Default search engine

Web integration

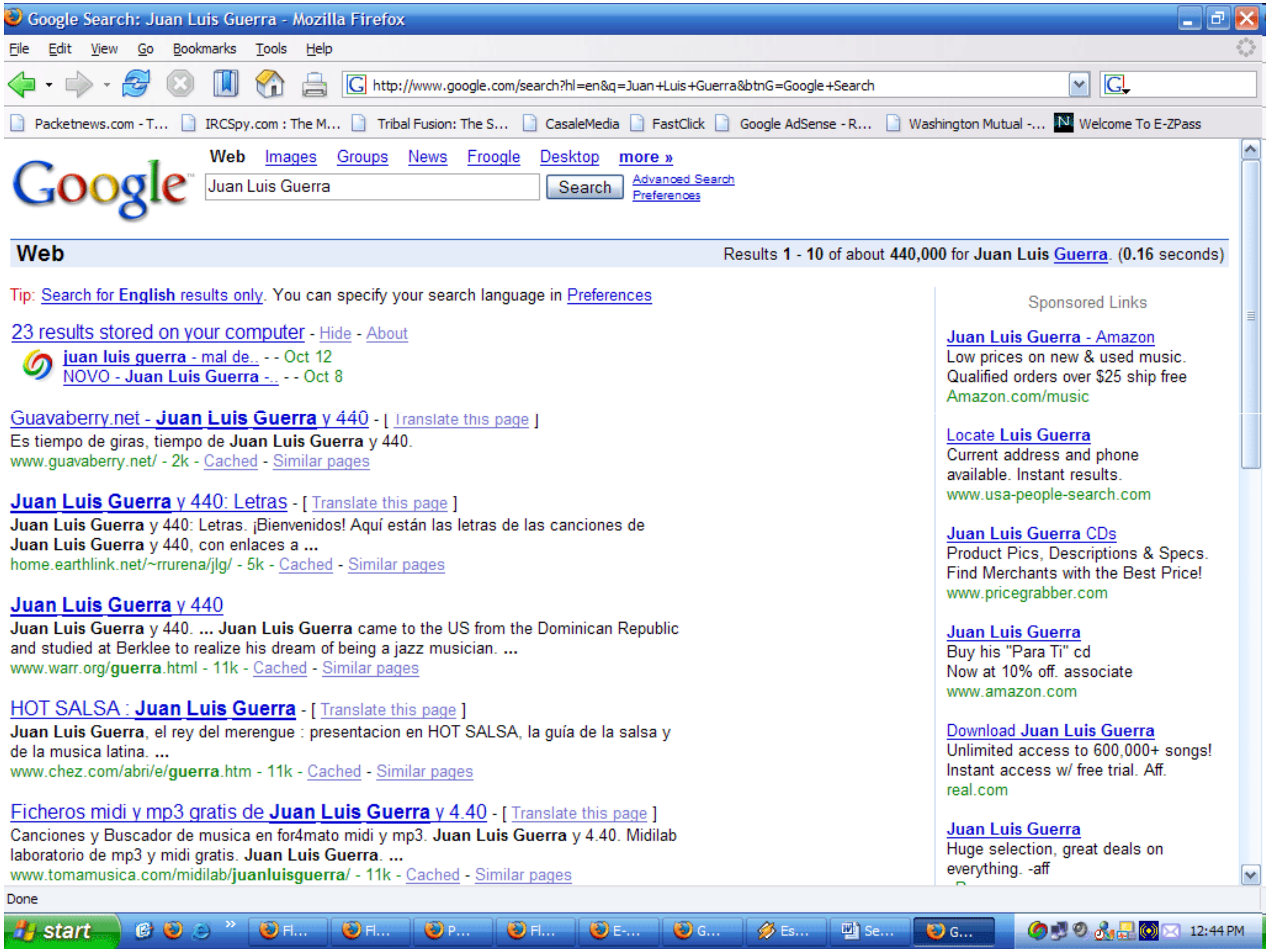
Insecure search

Registration

Engineering feedback

Software updates

Google Desktop Search



Google Search: Juan Luis Guerra - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://www.google.com/search?hl=en&q=Juan+Luis+Guerra&btnG=Google+Search
Packetnews.com - T... IRCSpy.com : The M... Tribal Fusion: The S... CasaleMedia FastClick Google AdSense - R... Washington Mutual -... Welcome To E-ZPass
Web Images Groups News Froogle Desktop more »
Search
Advanced Search
Preferences
Done
start
12:44 PM

Juan Luis Guerra

Web Results **1** - **10** of about **440,000** for **Juan Luis Guerra**. (**0.16** seconds)

Tip: Search for **English** results only. You can specify your search language in Preferences

23 results stored on your computer - Hide - About
juan luis guerra - mal de.. - - Oct 12
NOVO - Juan Luis Guerra -.. - - Oct 8

Guavaberry.net - **Juan Luis Guerra** y 440 - [Translate this page]
Es tiempo de giras, tiempo de **Juan Luis Guerra** y 440.
www.guavaberry.net/ - 2k - Cached - Similar pages

Juan Luis Guerra y 440: Letras - [Translate this page]
Juan Luis Guerra y 440: Letras. ¡Bienvenidos! Aquí están las letras de las canciones de **Juan Luis Guerra** y 440, con enlaces a **...**
home.earthlink.net/~rrurena/jlg/ - 5k - Cached - Similar pages

Juan Luis Guerra y 440
Juan Luis Guerra y 440. **...** **Juan Luis Guerra** came to the US from the Dominican Republic and studied at Berklee to realize his dream of being a jazz musician. **...**
www.warr.org/**guerra**.html - 11k - Cached - Similar pages

HOT SALSA : **Juan Luis Guerra** - [Translate this page]
Juan Luis Guerra, el rey del merengue : presentacion en HOT SALSA, la guía de la salsa y de la musica latina. **...**
www.chez.com/abri/e/**guerra**.htm - 11k - Cached - Similar pages

Ficheros midi y mp3 gratis de **Juan Luis Guerra** y 4.40 - [Translate this page]
Canciones y Buscador de musica en for4mato midi y mp3. **Juan Luis Guerra** y 4.40. Midilab laboratorio de mp3 y midi gratis. **Juan Luis Guerra**. **...**
www.tomamusica.com/midilab/**juanluisguerra**/ - 11k - Cached - Similar pages

Sponsored Links

Juan Luis Guerra - Amazon
Low prices on new & used music. Qualified orders over $25 ship free
Amazon.com/music

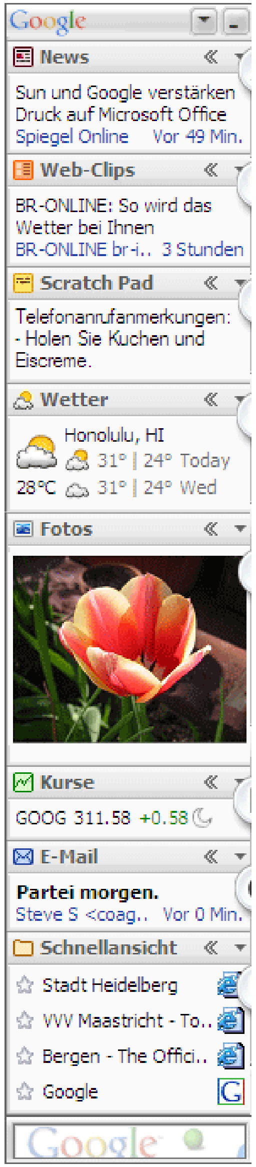
Locate **Luis Guerra**
Current address and phone available. Instant results.
www.usa-people-search.com

Juan Luis Guerra CDs
Product Pics, Descriptions & Specs. Find Merchants with the Best Price!
www.pricegrabber.com

Juan Luis Guerra
Buy his "Para Ti" cd
Now at 10% off. associate
www.amazon.com

Download **Juan Luis Guerra**
Unlimited access to 600,000+ songs! Instant access w/ free trial. Aff.
real.com

Juan Luis Guerra
Huge selection, great deals on everything. -aff



Sun und Google verstärken Druck auf Microsoft Office
Spiegel Online Vor 49 Min.

BR-ONLINE: So wird das Wetter bei Ihnen
BR-ONLINE br-i.. 3 Stunden

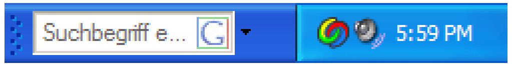
Telefonanrufanmerkungen:
- Holen Sie Kuchen und Eiscreme.

Honolulu, HI
31° | 24° Today
28°C 31° | 24° Wed

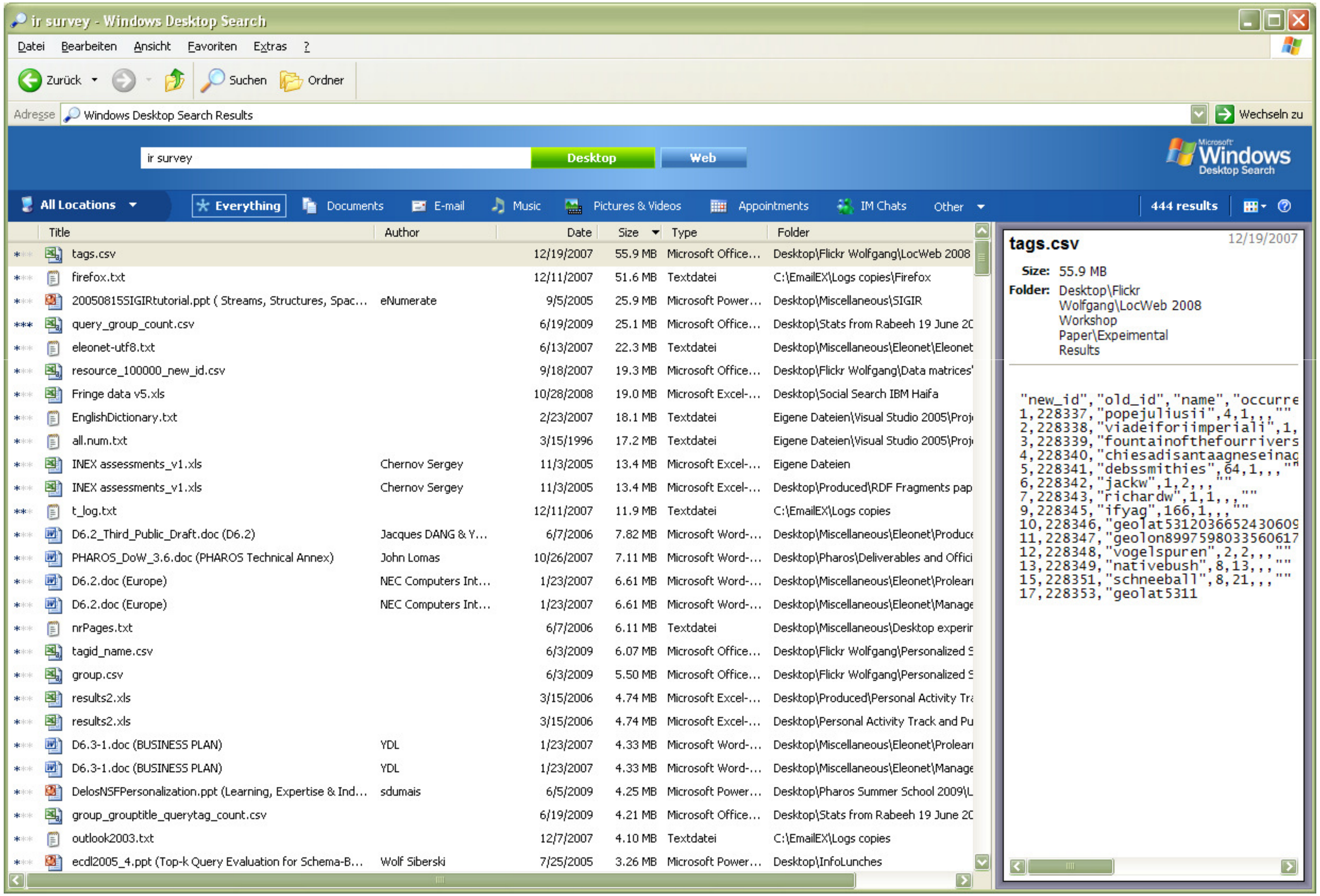
GOOG 311.58 +0.58

Partei morgen.
Steve S <coag.. Vor 0 Min.

Stadt Heidelberg
VVV Maastricht - To..
Bergen - The Offici..
Google



Windows Desktop Search


ir survey - Windows Desktop Search
Datei Bearbeiten Ansicht Favoriten Extras ?
Zurück
Suchen
Ordner
Adresse Windows Desktop Search Results
Wechseln zu
ir survey
Desktop
Web
Microsoft Windows Desktop Search
All Locations
Everything
Documents
E-mail
Music
Pictures & Videos
Appointments
IM Chats
Other
444 results

Title	Author	Date	Size	Type	Folder
tags.csv		12/19/2007	55.9 MB	Microsoft Office...	Desktop\Flickr Wolfgang\LocWeb 2008
firefox.txt		12/11/2007	51.6 MB	Textdatei	C:\EmailEX\Logs copies\Firefox
20050815SIGIRtutorial.ppt (Streams, Structures, Spac...	eNumerate	9/5/2005	25.9 MB	Microsoft Power...	Desktop\Miscellaneous\SIGIR
query_group_count.csv		6/19/2009	25.1 MB	Microsoft Office...	Desktop\Stats from Rabeeh 19 June 20
eleonet-utf8.txt		6/13/2007	22.3 MB	Textdatei	Desktop\Miscellaneous\Eleonet\Eleonet
resource_100000_new_id.csv		9/18/2007	19.3 MB	Microsoft Office...	Desktop\Flickr Wolfgang\Data matrices'
Fringe data v5.xls		10/28/2008	19.0 MB	Microsoft Excel-...	Desktop\Social Search IBM Haifa
EnglishDictionary.txt		2/23/2007	18.1 MB	Textdatei	Eigene Dateien\Visual Studio 2005\Proj
all.num.txt		3/15/1996	17.2 MB	Textdatei	Eigene Dateien\Visual Studio 2005\Proj
INEX assessments_v1.xls	Chernov Sergey	11/3/2005	13.4 MB	Microsoft Excel-...	Eigene Dateien
INEX assessments_v1.xls	Chernov Sergey	11/3/2005	13.4 MB	Microsoft Excel-...	Desktop\Produced\RDF Fragments pap
t_log.txt		12/11/2007	11.9 MB	Textdatei	C:\EmailEX\Logs copies
D6.2_Third_Public_Draft.doc (D6.2)	Jacques DANG & Y...	6/7/2006	7.82 MB	Microsoft Word-...	Desktop\Miscellaneous\Eleonet\Produce
PHAROS_DoW_3.6.doc (PHAROS Technical Annex)	John Lomas	10/26/2007	7.11 MB	Microsoft Word-...	Desktop\Pharos\Deliverables and Offici
D6.2.doc (Europe)	NEC Computers Int...	1/23/2007	6.61 MB	Microsoft Word-...	Desktop\Miscellaneous\Eleonet\Prolean
D6.2.doc (Europe)	NEC Computers Int...	1/23/2007	6.61 MB	Microsoft Word-...	Desktop\Miscellaneous\Eleonet\Manage
nrPages.txt		6/7/2006	6.11 MB	Textdatei	Desktop\Miscellaneous\Desktop experir
tagid_name.csv		6/3/2009	6.07 MB	Microsoft Office...	Desktop\Flickr Wolfgang\Personalized S
group.csv		6/3/2009	5.50 MB	Microsoft Office...	Desktop\Flickr Wolfgang\Personalized S
results2.xls		3/15/2006	4.74 MB	Microsoft Excel-...	Desktop\Produced\Personal Activity Tra
results2.xls		3/15/2006	4.74 MB	Microsoft Excel-...	Desktop\Personal Activity Track and Pu
D6.3-1.doc (BUSINESS PLAN)	YDL	1/23/2007	4.33 MB	Microsoft Word-...	Desktop\Miscellaneous\Eleonet\Prolean
D6.3-1.doc (BUSINESS PLAN)	YDL	1/23/2007	4.33 MB	Microsoft Word-...	Desktop\Miscellaneous\Eleonet\Manage
DelosNSFPersonalization.ppt (Learning, Expertise & Ind...	sdumais	6/5/2009	4.25 MB	Microsoft Power...	Desktop\Pharos Summer School 2009\L
group_grouptitle_querytag_count.csv		6/19/2009	4.21 MB	Microsoft Office...	Desktop\Stats from Rabeeh 19 June 20
outlook2003.txt		12/7/2007	4.10 MB	Textdatei	C:\EmailEX\Logs copies
ecdl2005_4.ppt (Top-k Query Evaluation for Schema-B...	Wolf Siberski	7/25/2005	3.26 MB	Microsoft Power...	Desktop\InfoLunches

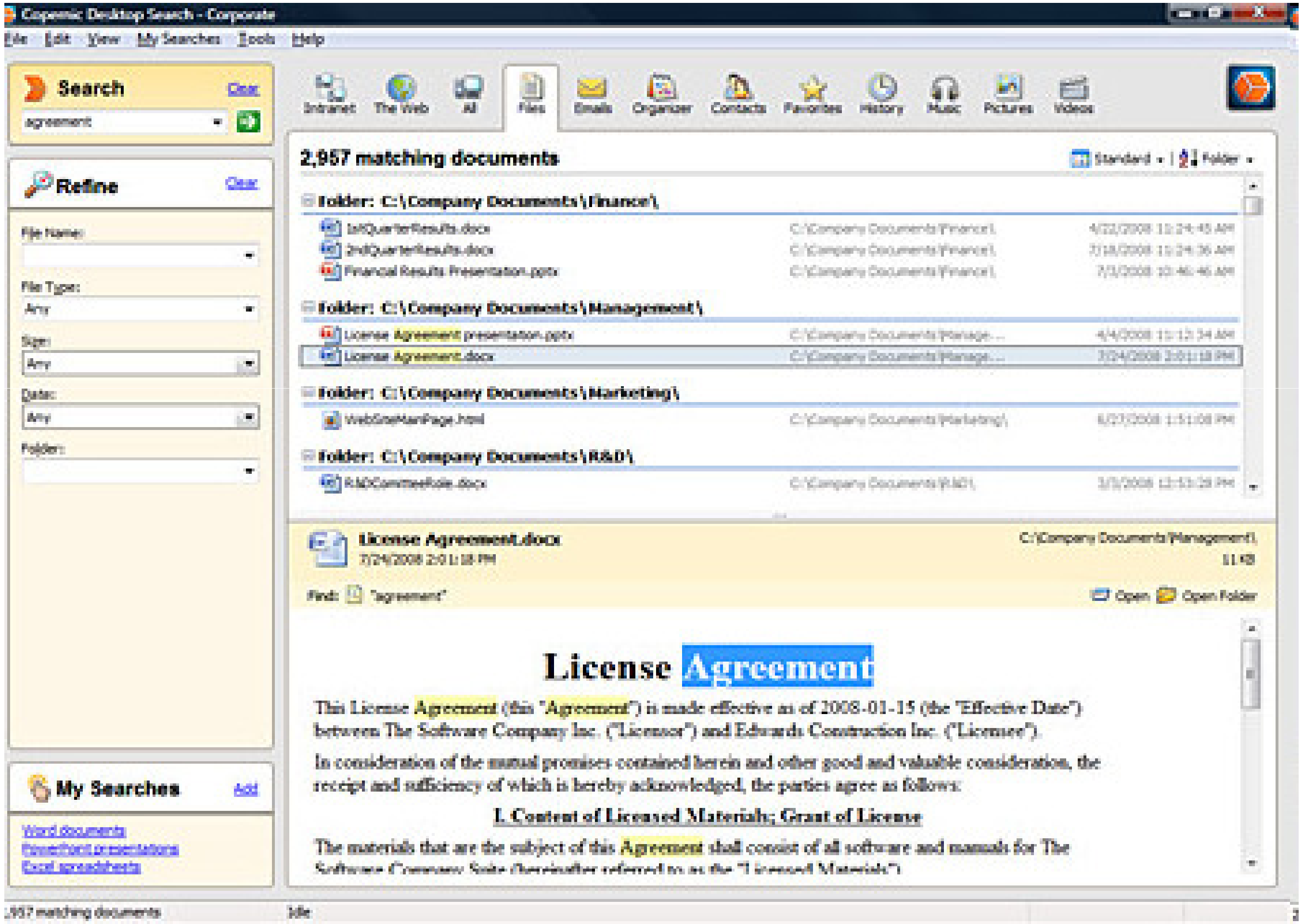
tags.csv 12/19/2007

Size: 55.9 MB

Folder: Desktop\Flickr Wolfgang\LocWeb 2008 Workshop Paper\Expeimental Results

```
"new_id","old_id","name","occurre
1,228337,"popejuliusii",4,1,,,""
2,228338,"viadeiforiimperiali",1,
3,228339,"fountainofthefourrivers
4,228340,"chiesadisantaagneseinac
5,228341,"debssmithies",64,1,,,""
6,228342,"jackw",1,2,,,""
7,228343,"richardw",1,1,,,""
9,228345,"ifyag",166,1,,,""
10,228346,"geolat531203665243060
11,228347,"geolon8997598033560617
12,228348,"vogelspuren",2,2,,,""
13,228349,"nativebush",8,13,,,""
15,228351,"schneeball",8,21,,,""
17,228353,"geolat5311
```

Copernicus Desktop Search



2,957 matching documents

Folder: C:\Company Documents\Finance\

1stQuarterResults.docx C:\Company Documents\Finance\ 4/22/2008 11:24:45 AM
2ndQuarterResults.docx C:\Company Documents\Finance\ 7/18/2008 11:24:36 AM
Financial Results Presentation.pptx C:\Company Documents\Finance\ 7/1/2008 10:46:46 AM

Folder: C:\Company Documents\Management\

License Agreement presentation.pptx C:\Company Documents\Manage... 4/4/2008 11:12:34 AM
License Agreement.docx C:\Company Documents\Manage... 7/24/2008 2:01:18 PM

Folder: C:\Company Documents\Marketing\

WebSiteMainPage.html C:\Company Documents\Marketing\ 6/27/2008 1:51:08 PM

Folder: C:\Company Documents\R&D\

R&DCommitteeRole.docx C:\Company Documents\R&D\ 3/3/2008 12:53:28 PM

License Agreement.docx
7/24/2008 2:01:18 PM
C:\Company Documents\Management\
11 KB

License Agreement

This License Agreement (this "Agreement") is made effective as of 2008-01-15 (the "Effective Date") between The Software Company Inc. ("Licensor") and Edwards Construction Inc. ("Licensee").

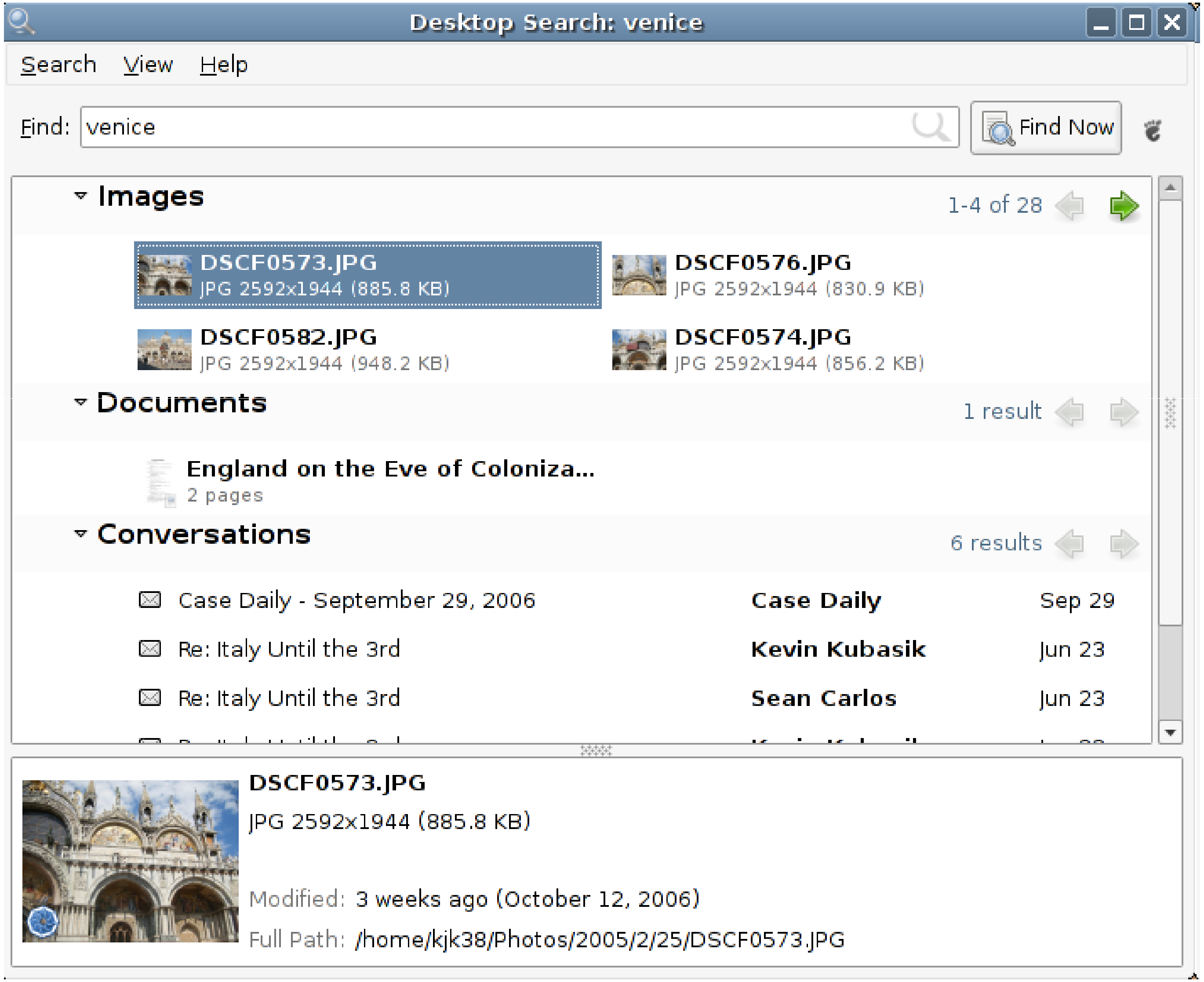
In consideration of the mutual promises contained herein and other good and valuable consideration, the receipt and sufficiency of which is hereby acknowledged, the parties agree as follows:

I. Content of Licensed Materials; Grant of License

The materials that are the subject of this Agreement shall consist of all software and manuals for The Software Company Suite (hereinafter referred to as the "Licensed Materials").

2,957 matching documents

Beagle Desktop Search



Find: venice

Images 1-4 of 28

DSCF0573.JPG JPG 2592x1944 (885.8 KB)	**DSCF0576.JPG** JPG 2592x1944 (830.9 KB)
DSCF0582.JPG JPG 2592x1944 (948.2 KB)	**DSCF0574.JPG** JPG 2592x1944 (856.2 KB)

Documents 1 result

England on the Eve of Coloniza...
2 pages

Conversations 6 results

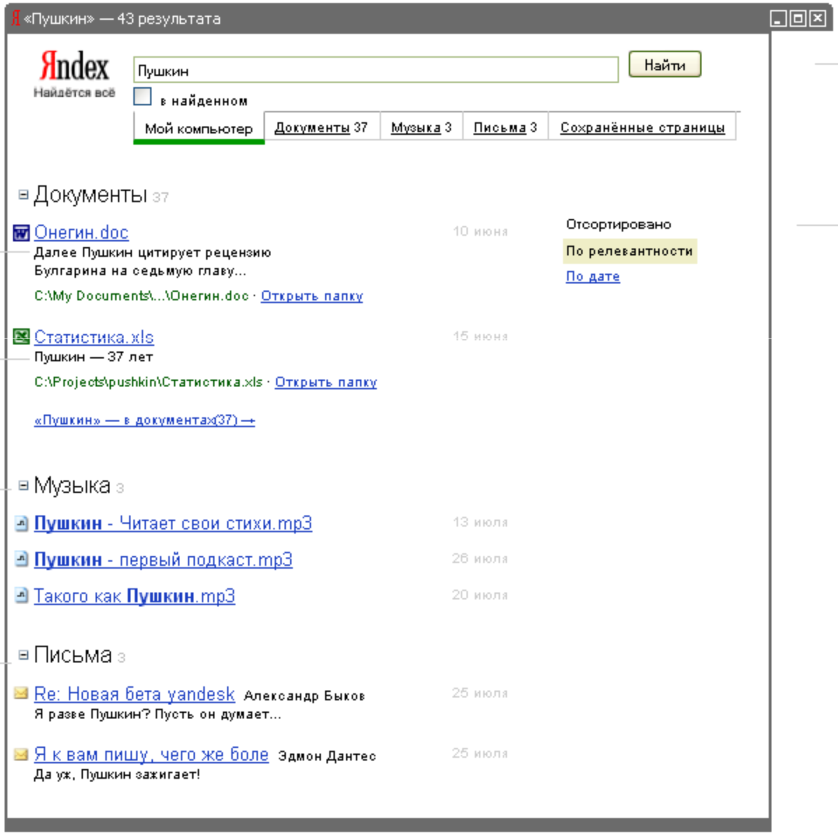
Case Daily - September 29, 2006	**Case Daily**	Sep 29
Re: Italy Until the 3rd	**Kevin Kubasik**	Jun 23
Re: Italy Until the 3rd	**Sean Carlos**	Jun 23

DSCF0573.JPG
JPG 2592x1944 (885.8 KB)

Modified: 3 weeks ago (October 12, 2006)
Full Path: /home/kjk38/Photos/2005/2/25/DSCF0573.JPG

Yandex Desktop Search

Персональный поиск Яндекса — это программа на вашем компьютере, осуществляющая поиск по файлам и письмам с учётом морфологии русского языка.



«Пушкин» — 43 результата

Яndex Найдётся всё

Пушкин [Найти]

□ в найденном

Мой компьютер | Документы 37 | Музыка 3 | Письма 3 | Сохранённые страницы

⊟ Документы 37

Онегин.doc — 10 июня
Далее Пушкин цитирует рецензию
Булгарина на седьмую главу...
C:\My Documents\...\Онегин.doc · Открыть папку

Статистика.xls — 15 июня
Пушкин — 37 лет
C:\Projects\pushkin\Статистика.xls · Открыть папку

«Пушкин» — в документах(37) →

Отсортировано
По релевантности
По дате

⊟ Музыка 3

Пушкин - Читает свои стихи.mp3 — 13 июля

Пушкин - первый подкаст.mp3 — 26 июля

Такого как **Пушкин**.mp3 — 20 июля

⊟ Письма 3

Re: Новая бета yandesk Александр Быков — 25 июля
Я разве Пушкин? Пусть он думает...

Я к вам пишу, чего же боле Эдмон Дантес — 25 июля
Да уж, Пушкин зажигает!

Форма поиска

Управление представлением результатов
Можно выбрать способ группировки и сортировки.

Конфиденциальность
Можно запретить искать в определенных папках или целых дисках.

Цитаты из найденных документов
Повышают информативность результатов поиска.

Группировка результатов по типам
Помогает ориентироваться в большом количестве найденных файлов.

Вы сможете найти ваши письма в Outlook, Outlook Express, Thunderbird и TheBat!

Окно запроса и результатов поиска открывается в обычном браузере

Доступ к Персональному поиску из панели задач

Research prototypes and Semantic Desktops

- Beagle++ (extended open source DS)
- Semex (includes Malleable Schemas)
- Haystack and Magnet (Semantic Web approach)
- Stuff I’ve Seen (Phlat predecessor)
- Phlat (was used as a basis for Windows DS)
- PIA (semantic desktop solution from DB area)

Some more:

Gnowsis, CALO

Next 14 slides are adapted from Wolfgang Nejdl and Raluca Paiu

Beagle++

P.-A. Chirita, S. Costache, W. Nejdl, and R. Paiu. Beagle++ : Semantically enhanced searching and ranking on the desktop. In ESWC 2006.

- Why is it so hard to find what you need on your desktop – “You still use Google even for files stored on your computer?”
- Current desktop search engines use only full text index
- People tend to associate things to certain contexts
- For desktop search we need to support contextual information in addition to full text!
  - Relationships between information items (citations)
  - Relationships based on interactions (email exchange, browsing history)
  - Relationships between different types of items (authorship, publication venues, email sender information, recommendations)
  - Other situational context

Semantically Rich Recommendations in Social Networks for Sharing, Exchanging and Ranking Semantic Context, Stefania Ghita, Wolfgang Nejdl, and Raluca Paiu. In ISWC 2005.

The Beagle++ Toolbox: Towards an Extendable Desktop Search Architecture, Ingo Brunkhorst, Paul - Alexandru Chirita, Stefania Costache, Julien Gaugaz, Ekaterini Ioannou, Tereza Iofciu, Enrico Minack, Wolfgang Nejdl and Raluca Paiu. Technical Report 2006.

Scenario 1: The Need for Context Information

- Alice and Bob are working together in the research group
- Alice is currently writing a paper about searching and ranking on the semantic desktop and wants to find some good papers on this topic, which she remembers she stored on her desktop
- Some time ago Bob sent her a very useful paper on this topic as an attachment to an email, together with some useful comments about its relevance to her new semantic desktop ideas
- ***Will Alice find the paper from Bob when issuing a query on the desktop, using the search terms “semantic desktop” ?***

Context Information is necessary!

- **Problems**:
  - (Mail) Documents sent as attachments lose all contextual information as soon as they are stored on the PC
  - (Web) When searching for a document we downloaded from the CiteSeer repository, we would like to retrieve not only the specific document, but all the referenced and referring papers which we already downloaded as well

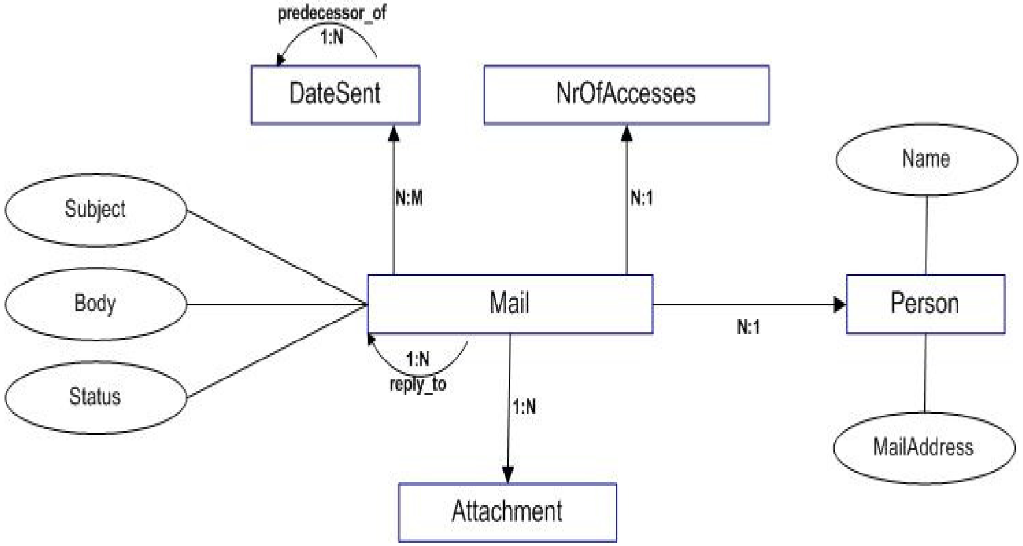
- Current desktop search approaches don’t make use of desktop specific information, especially contextual information, like:
  - **Email** context
  - **Web** context
  - **Publication** context

Representing Context by Semantic Web Metadata

- Metadata for resources can be created by appropriate metadata generators
- Ontologies specify context metadata for:
  - Emails
  - Files
  - Web pages
  - Publications
- Metadata have to be application-independent!

→ Store Metadata as RDF

- generated and used by whatever application you can think of



Beagle++ Layer Architecture

Beagle++ is our extension of the open source Beagle search project, enabling it to exploit context information

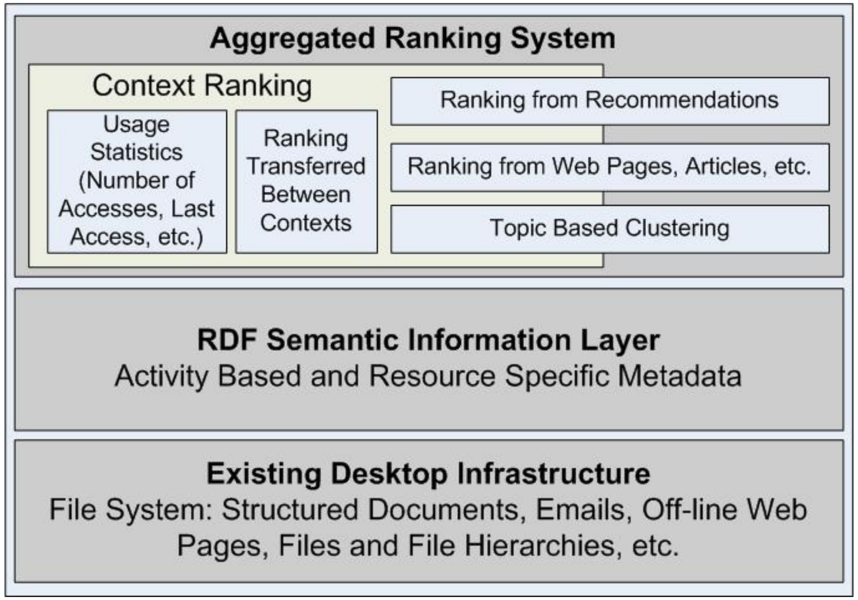
RDF metadata are generated based on ontologies for specific contexts (email, web, etc.)

Indexing and metadata generation on the fly - triggered by events upon occurrence of file system changes (*inotify-enabled linux kernel)*

Benefits:

Context allows us to better organize and find information

Context gives us the possibility to compute the value / importance of resources



Aggregated Ranking System

Context Ranking

Usage Statistics (Number of Accesses, Last Access, etc.)

Ranking Transferred Between Contexts

Ranking from Recommendations

Ranking from Web Pages, Articles, etc.

Topic Based Clustering

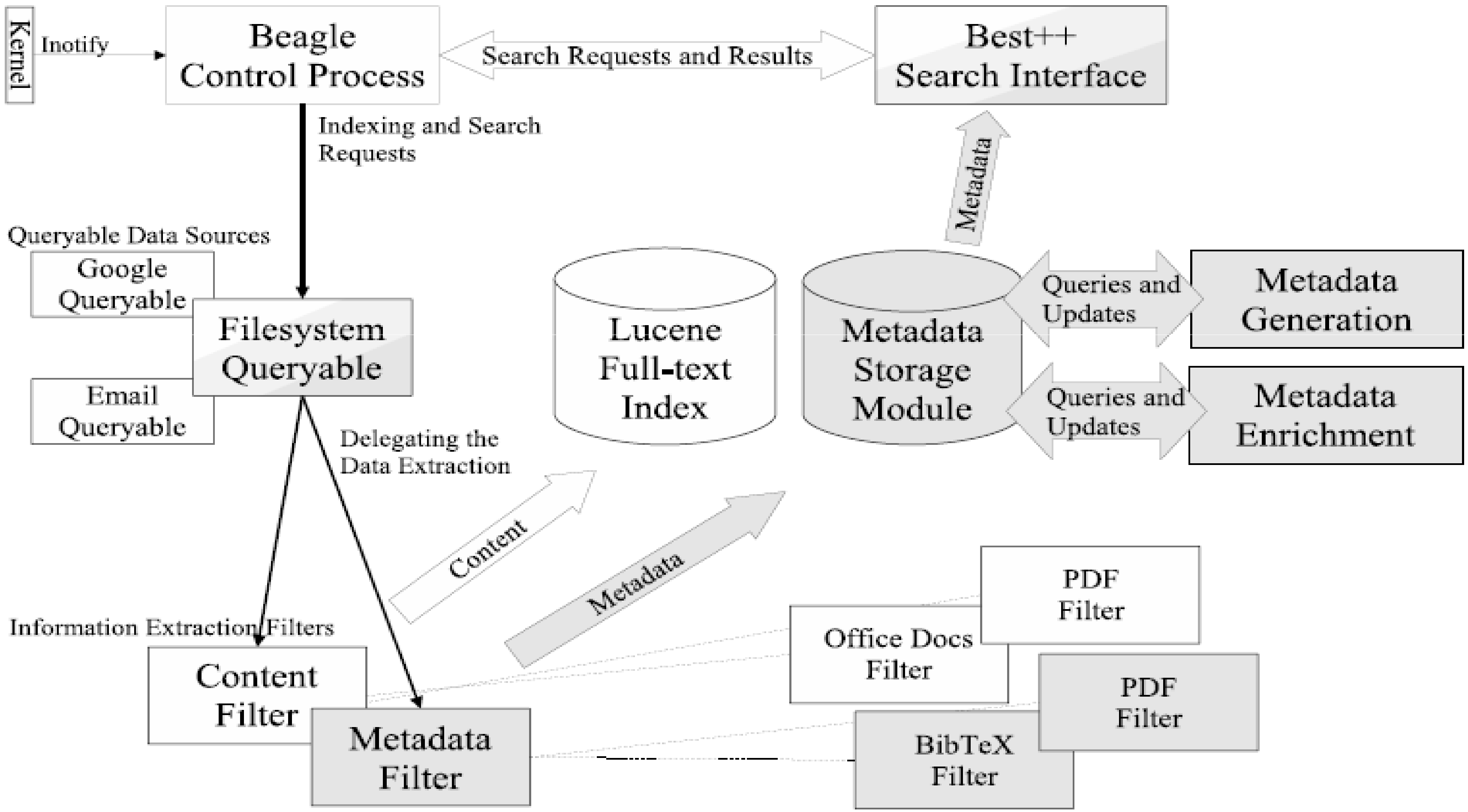
RDF Semantic Information Layer

Activity Based and Resource Specific Metadata

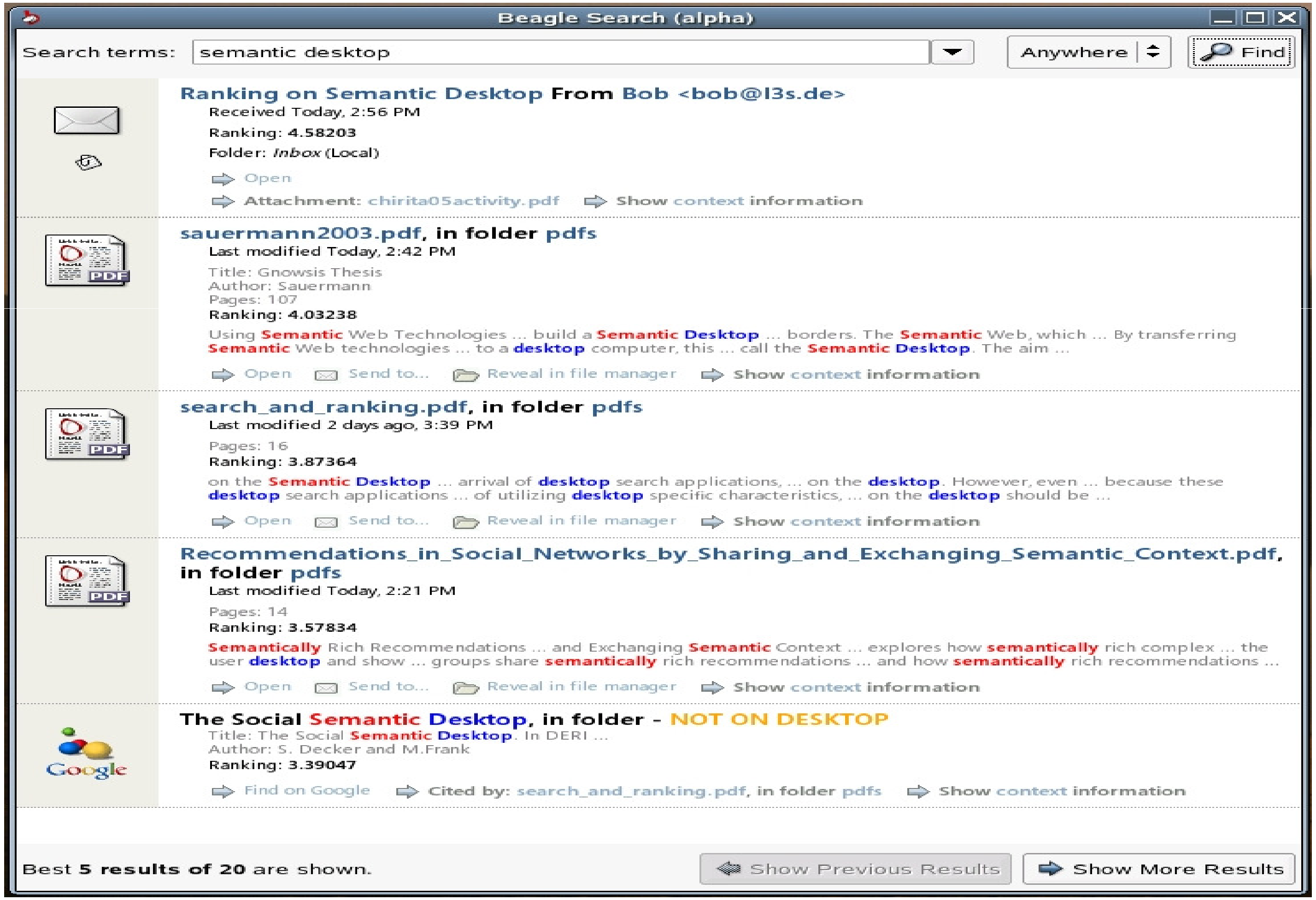
Existing Desktop Infrastructure

File System: Structured Documents, Emails, Off-line Web Pages, Files and File Hierarchies, etc.

Beagle++ Architecture



Beagle++: Find more than documents



Search terms: semantic desktop

Ranking on Semantic Desktop From Bob <bob@l3s.de>
Received Today, 2:56 PM
Ranking: 4.58203
Folder: *Inbox* (Local)

sauermann2003.pdf, in folder pdfs
Last modified Today, 2:42 PM
Title: Gnowsis Thesis
Author: Sauermann
Pages: 107
Ranking: 4.03238
Using Semantic Web Technologies ... build a Semantic Desktop ... borders. The Semantic Web, which ... By transferring Semantic Web technologies ... to a desktop computer, this ... call the Semantic Desktop. The aim ...

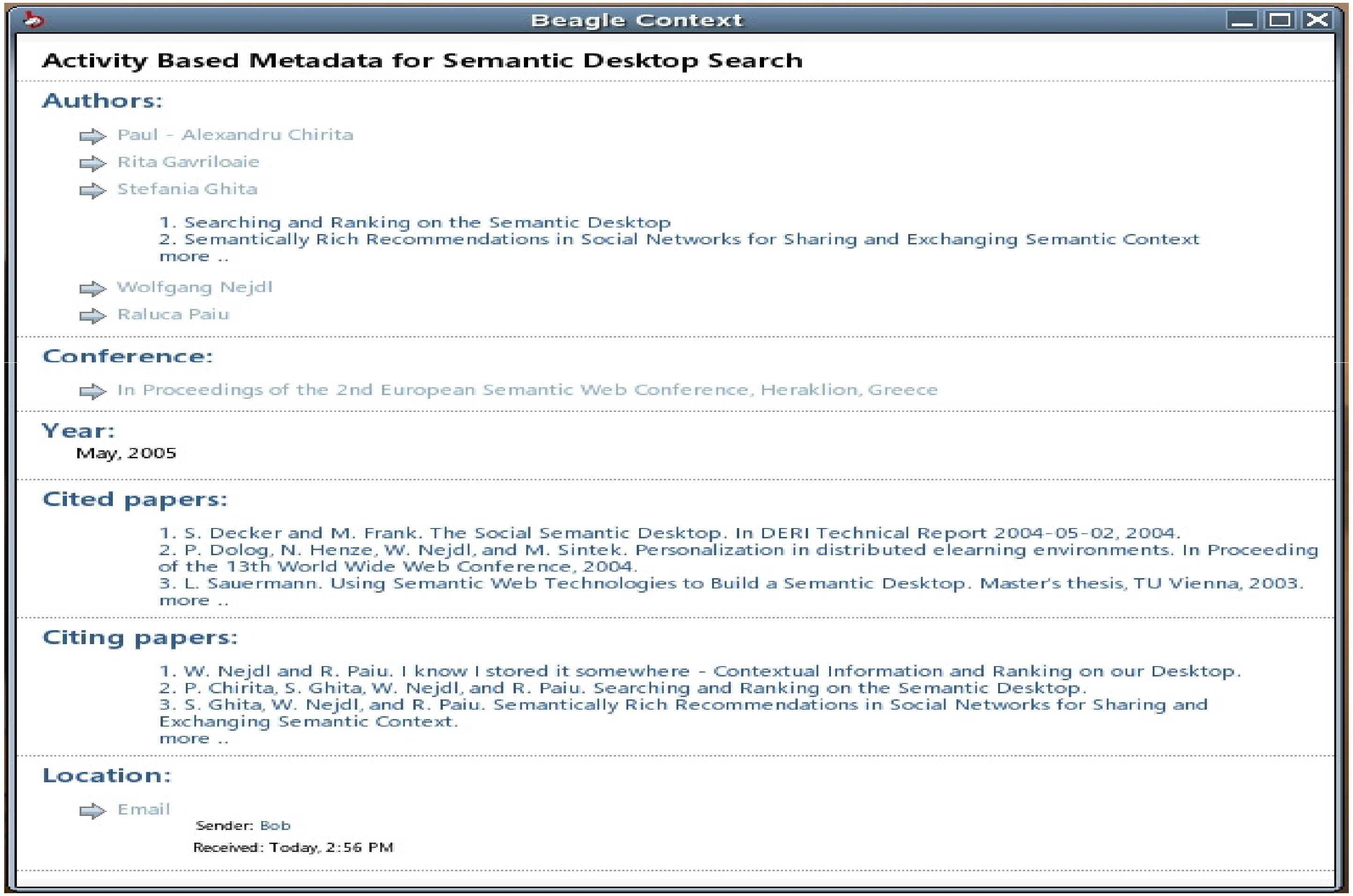
search_and_ranking.pdf, in folder pdfs
Last modified 2 days ago, 3:39 PM
Pages: 16
Ranking: 3.87364
on the Semantic Desktop ... arrival of desktop search applications, ... on the desktop. However, even ... because these desktop search applications ... of utilizing desktop specific characteristics, ... on the desktop should be ...

Recommendations_in_Social_Networks_by_Sharing_and_Exchanging_Semantic_Context.pdf, in folder pdfs
Last modified Today, 2:21 PM
Pages: 14
Ranking: 3.57834
Semantically Rich Recommendations ... and Exchanging Semantic Context ... explores how semantically rich complex ... the user desktop and show ... groups share semantically rich recommendations ... and how semantically rich recommendations ...

The Social Semantic Desktop, in folder - NOT ON DESKTOP
Title: The Social Semantic Desktop. In DERI ...
Author: S. Decker and M.Frank
Ranking: 3.39047

Best **5 results of 20** are shown.

Beagle++: Display additional context



Activity Based Metadata for Semantic Desktop Search

Authors:

- Paul - Alexandru Chirita
- Rita Gavriloaie
- Stefania Ghita

  1. Searching and Ranking on the Semantic Desktop
  2. Semantically Rich Recommendations in Social Networks for Sharing and Exchanging Semantic Context

  more ..
- Wolfgang Nejdl
- Raluca Paiu

Conference:

- In Proceedings of the 2nd European Semantic Web Conference, Heraklion, Greece

Year:

May, 2005

Cited papers:

1. S. Decker and M. Frank. The Social Semantic Desktop. In DERI Technical Report 2004-05-02, 2004.
2. P. Dolog, N. Henze, W. Nejdl, and M. Sintek. Personalization in distributed elearning environments. In Proceeding of the 13th World Wide Web Conference, 2004.
3. L. Sauermann. Using Semantic Web Technologies to Build a Semantic Desktop. Master's thesis, TU Vienna, 2003.

more ..

Citing papers:

1. W. Nejdl and R. Paiu. I know I stored it somewhere - Contextual Information and Ranking on our Desktop.
2. P. Chirita, S. Ghita, W. Nejdl, and R. Paiu. Searching and Ranking on the Semantic Desktop.
3. S. Ghita, W. Nejdl, and R. Paiu. Semantically Rich Recommendations in Social Networks for Sharing and Exchanging Semantic Context.

more ..

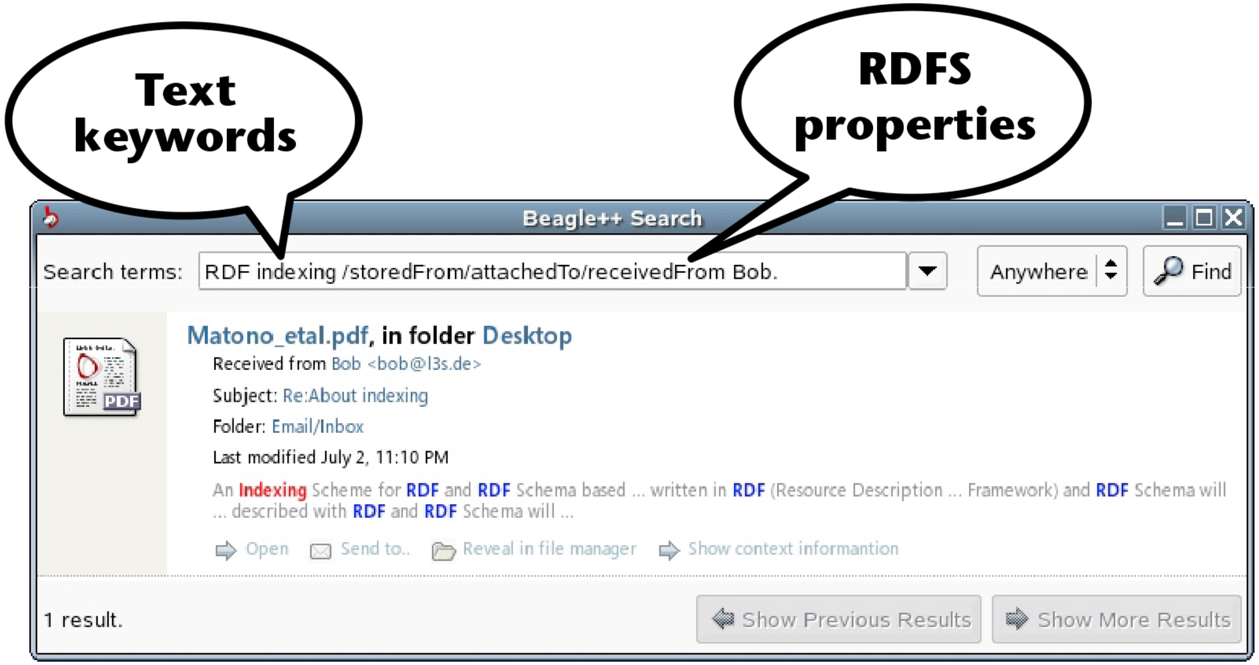
Location:

- Email

  Sender: Bob

  Received: Today, 2:56 PM

Integrating Keyword and Metadata Search



Search terms: RDF indexing /storedFrom/attachedTo/receivedFrom Bob.

Matono_etal.pdf, in folder Desktop

Received from Bob <bob@l3s.de>

Subject: Re:About indexing

Folder: Email/Inbox

Last modified July 2, 11:10 PM

An Indexing Scheme for RDF and RDF Schema based ... written in RDF (Resource Description ... Framework) and RDF Schema will ... described with RDF and RDF Schema will ...

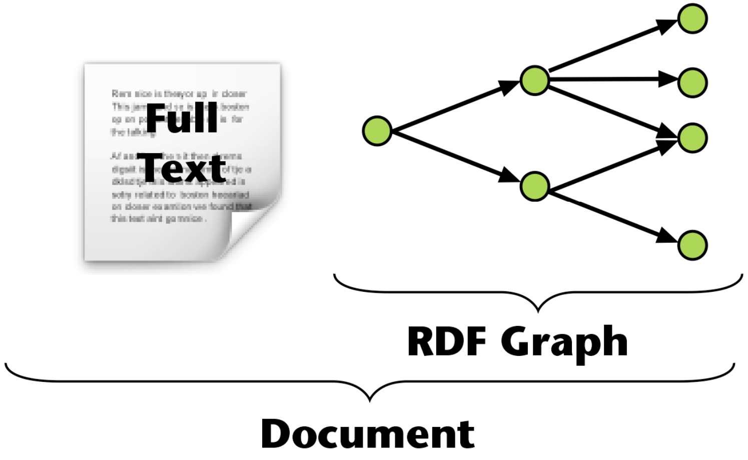
1 result.

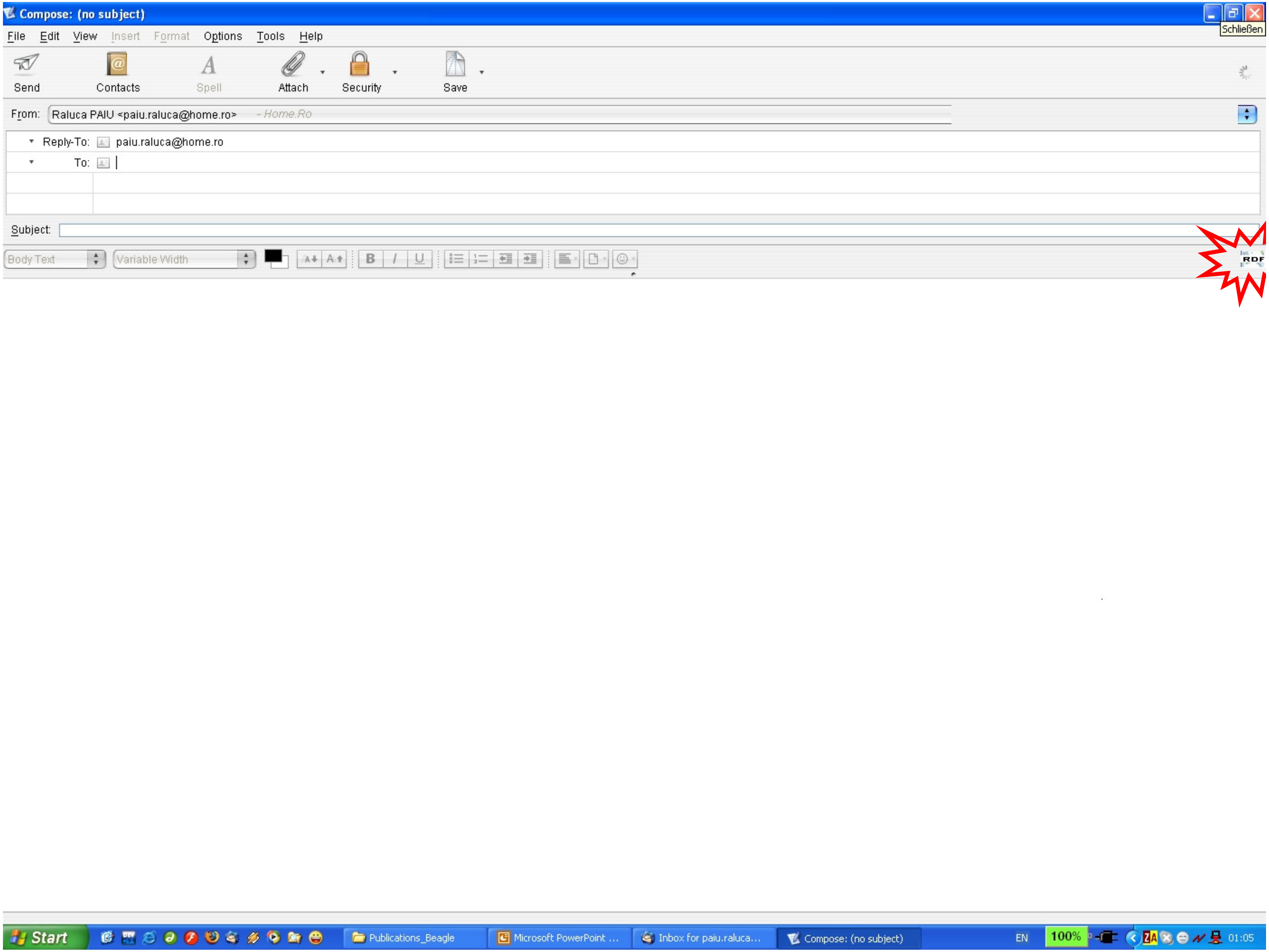
- Search text and metadata on the desktop
- Search efficiently in a user-friendly way
- Simple query language
- No complete schema knowledge necessary

Documents / RDF Fragments

- Metadata stored as RDF graphs, each document has a corresponding RDF fragment
- Extended documents consisting of both full-text and metadata properties
- Query model supports the operator selection, projection and union, intersection and set difference

  - Support for approximate and imprecise metadata queries
  - Separation between metadata statements is ensured by positional indices





From: Raluca PAIU <paiu.raluca@home.ro> - Home.Ro

Reply-To: paiu.raluca@home.ro

To:

Peer-Sensitive ObjectRank [1]

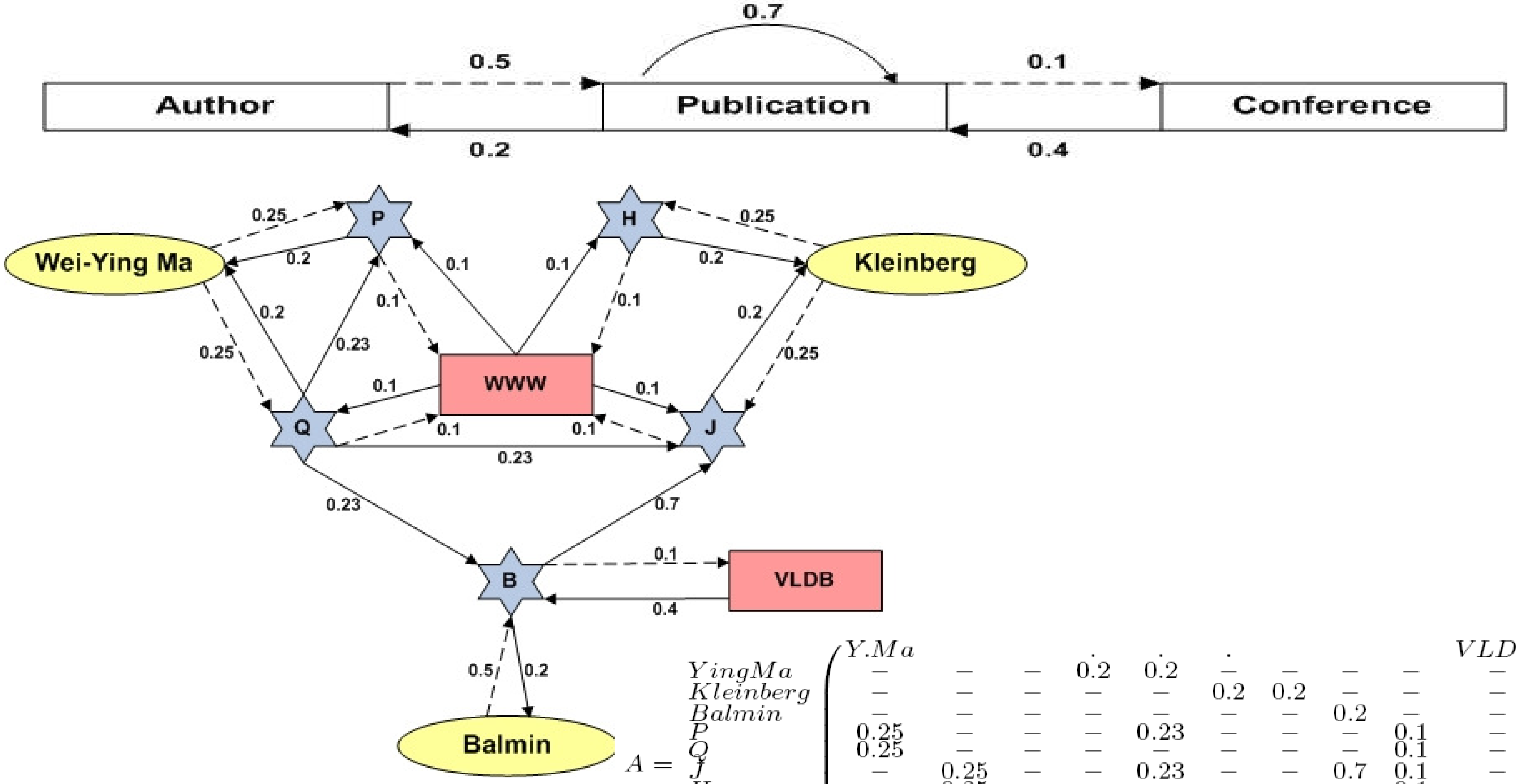
- Step 1: start with PageRank formula – random surfer model

$$\mathbf{r = d \cdot A \cdot r + (1 - d) \cdot e}$$

- d = dampening factor
- A = adjacency matrix
- e = vector for the random jump

- Step 2: distinguish between different kinds of objects
- ObjectRank variant of PageRank

Peer-Sensitive ObjectRank [2]



$$A = \begin{array}{l} \\ YingMa \\ Kleinberg \\ Balmin \\ P \\ Q \\ J \\ H \\ B \\ WWW \\ VLDB \end{array} \left(\begin{array}{cccccccccc} Y.Ma & & & . & . & . & & & & VLDB \\ - & - & - & 0.2 & 0.2 & - & - & - & - & - \\ - & - & - & - & - & 0.2 & 0.2 & - & - & - \\ - & - & - & - & - & - & - & 0.2 & - & - \\ 0.25 & - & - & - & 0.23 & - & - & - & 0.1 & - \\ 0.25 & - & - & - & - & - & - & - & 0.1 & - \\ - & 0.25 & - & - & 0.23 & - & - & 0.7 & 0.1 & - \\ - & 0.25 & - & - & - & - & - & - & 0.1 & - \\ - & - & 0.5 & - & 0.23 & - & - & - & - & 0.4 \\ - & - & - & 0.1 & 0.1 & 0.1 & 0.1 & - & - & - \\ - & - & - & - & - & - & - & 0.1 & - & - \end{array} \right)$$

Peer-Sensitive ObjectRank [3]

- Step 3: Take provenance information into account
- → **Peer-Sensitive ObjectRank**
- Represent different trust in peers by corresponding modifications in the ***e*** vector
- Keep track of the provenance of each resource

$$originates\ (r_i, P_n) = \begin{cases} 1, \text{if } \mathrm{r_i} \text{ is in the initial set of } \mathrm{P_n} \\ 0, \text{otherwise} \end{cases}$$

$$trust(P_i, P_j) \in [0,1], \text{ the trust value of peer } \mathrm{P_i} \text{ for } \mathrm{P_j}$$

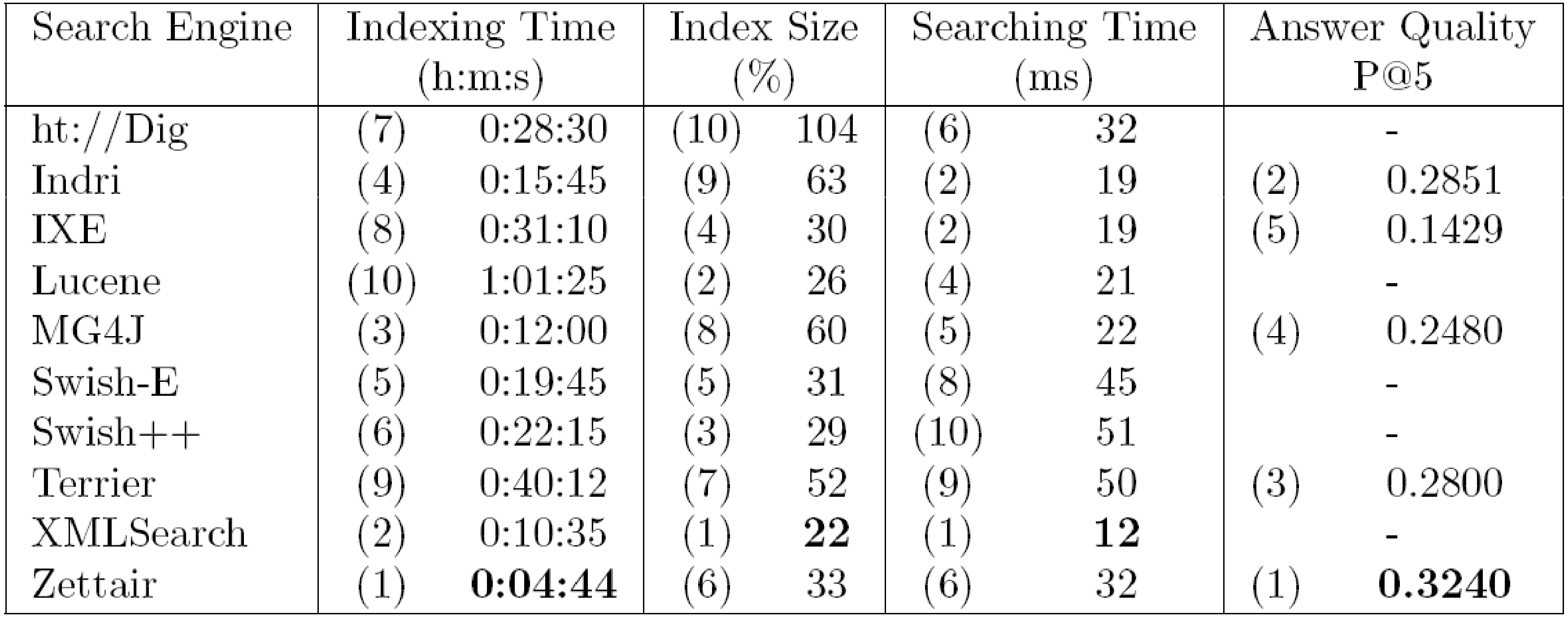
$$e_k(P_i) = \max{}_{j=0}^{N}\{trust(P_{i,} P_j) \cdot originates(r_k, P_j)$$

Beagle++ Demo

Open Source Search Engines

A Comparison of Open Source Search Engines, Christian Middleton and Ricardo Baeza-Yates, Technical Report, 2007 .

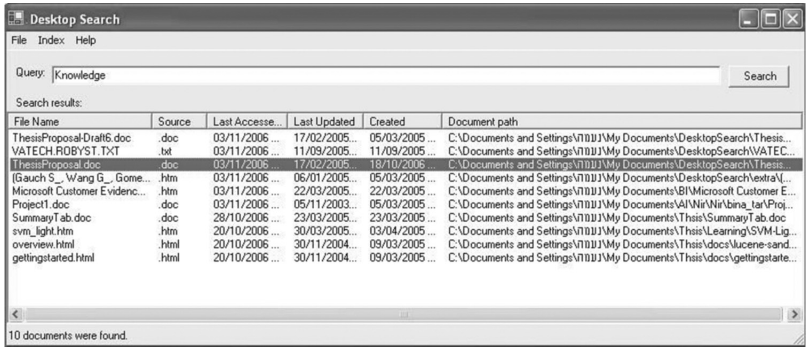
Build your own search engine!



Search Engine	Indexing Time		Index Size		Searching Time		Answer Quality	
	(h:m:s)		(%)		(ms)		P@5	
ht://Dig	(7)	0:28:30	(10)	104	(6)	32		-
Indri	(4)	0:15:45	(9)	63	(2)	19	(2)	0.2851
IXE	(8)	0:31:10	(4)	30	(2)	19	(5)	0.1429
Lucene	(10)	1:01:25	(2)	26	(4)	21		-
MG4J	(3)	0:12:00	(8)	60	(5)	22	(4)	0.2480
Swish-E	(5)	0:19:45	(5)	31	(8)	45		-
Swish++	(6)	0:22:15	(3)	29	(10)	51		-
Terrier	(9)	0:40:12	(7)	52	(9)	50	(3)	0.2800
XMLSearch	(2)	0:10:35	(1)	**22**	(1)	**12**		-
Zettair	(1)	**0:04:44**	(6)	33	(6)	32	(1)	**0.3240**

Selecting an Appropriate Ranking Function

On Ranking Techniques for Desktop Search, Sara Cohen, Carmel Domshlak and Naama Zwerdling, In ACM Transactions on Information Systems 2008.



Query: Knowledge

Search results:

File Name	Source	Last Accesse...	Last Updated	Created	Document path
ThesisProposal-Draft6.doc	.doc	03/11/2006 ...	17/02/2005...	05/03/2005 ...	C:\Documents and Settings\נעמה\My Documents\DesktopSearch\Thesis...
VATECH.ROBYST.TXT	.txt	03/11/2006 ...	11/09/2005...	11/09/2005 ...	C:\Documents and Settings\נעמה\My Documents\DesktopSearch\VATEC...
ThesisProposal.doc	.doc	03/11/2006 ...	17/02/2005...	18/10/2006 ...	C:\Documents and Settings\נעמה\My Documents\DesktopSearch\Thesis...
(Gauch S_, Wang G_, Gome...	.htm	03/11/2006 ...	06/01/2005...	05/03/2005 ...	C:\Documents and Settings\נעמה\My Documents\DesktopSearch\extra\(...
Microsoft Customer Evidenc...	.htm	03/11/2006 ...	22/03/2005...	22/03/2005 ...	C:\Documents and Settings\נעמה\My Documents\BI\Microsoft Customer E...
Project1.doc	.doc	03/11/2006 ...	05/11/2003...	05/03/2005 ...	C:\Documents and Settings\נעמה\My Documents\AI\Nir\Nir\bina_tar\Proj...
SummaryTab.doc	.doc	28/10/2006 ...	23/03/2005...	23/03/2005 ...	C:\Documents and Settings\נעמה\My Documents\Thsis\SummaryTab.doc
svm_light.htm	.htm	20/10/2006 ...	30/03/2005...	03/04/2005 ...	C:\Documents and Settings\נעמה\My Documents\Thsis\Learning\SVM-Lig...
overview.html	.html	20/10/2006 ...	30/11/2004...	09/03/2005 ...	C:\Documents and Settings\נעמה\My Documents\Thsis\docs\lucene-sand...
gettingstarted.html	.html	20/10/2006 ...	30/11/2004...	09/03/2005 ...	C:\Documents and Settings\נעמה\My Documents\Thsis\docs\gettingstarte...

10 documents were found.

Lucene-based DS prototype
19 volunteers.
In total 1219 queries
188 queries had a single result,
916 queries has 2-50 results
115 queries had over 50 results.

Feature	MRR$(\tau, \mathcal{S}_{2-50})$	Feature	MRR$(\tau, \mathcal{S}_{>50})$
SVM	0.54	SVM	0.26
LexOrd	0.53	LexOrd	0.18
Selective	0.5	**Selective**	0.17
UserBest	0.47	AccessDate	0.16
UpdateDate	0.43	UserBest	0.16
Name	0.43	UpdateDate	0.15
AccessDate	0.4	CreateDate	0.12
CreateDate	0.39	Name	0.1
Size	0.39	Path	0.1
Content	0.38	Size	0.08
NormalizedSize	0.36	QueryLog	0.07
Path	0.34	DirRank	0.06
QueryLog	0.34	Content	0.06
DirRank	0.33	NormalizedSize	0.06
Level	0.31	Level	0.03
Random	0.28	*Random*	*0.02*

$$\textsc{Selective}_q(f) \overset{\text{def}}{=} \sum_{\substack{\textsc{Feature} \in \{\textsc{Name}, \textsc{Path},\\ \textsc{Content}, \textsc{QueryLog}\}}} \frac{\textsc{Feature}_q(f)}{nz(\textsc{Feature}_q)}$$

Research prototypes and Semantic Desktops (continues)

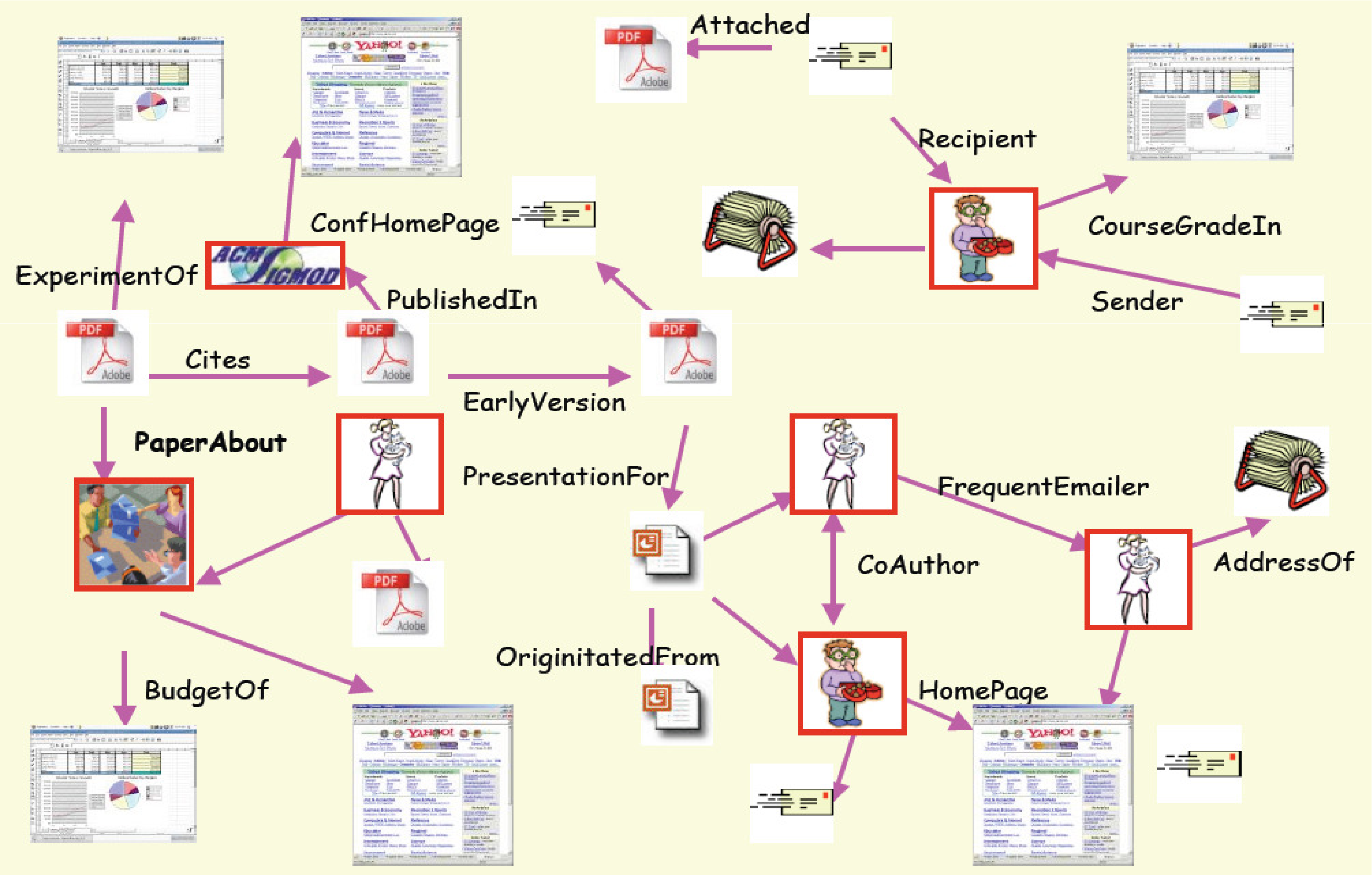
- Beagle++ (extended open source DS)
- Semex (includes Malleable Schemas)
- Haystack and Magnet (Semantic Web approach)
- Stuff I’ve Seen (Phlat predecessor)
- Phlat (was used as a basis for Windows DS)
- PIA (semantic desktop solution from DB area)

Some more:

Gnowsis, CALO

Semex

Personal Information Management with Semex, Yuhan Cai, Xin Luna Dong, Alon Halevy, Jing Michelle Liu, and Jayant Madhavan. In SIGMOD 2005



Semex Features

Slide from Paul Chirita

- Highly database oriented approach
  - Resources connected through *Reference Reconciliation*
  - On-the-fly integration with external sources
  - *Malleable Schemas*

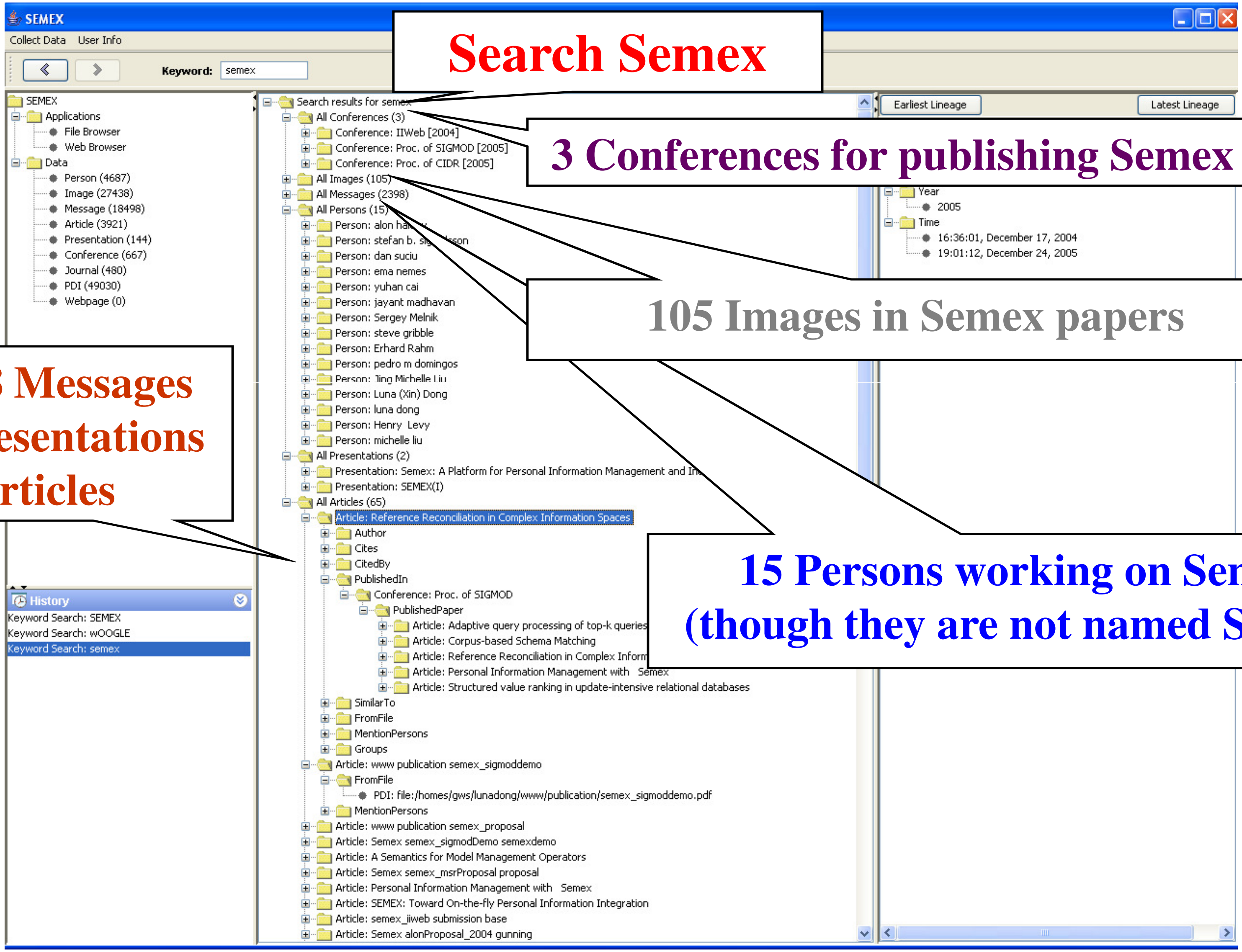
Malleable¤Schemas, Xin Dong and Alon Halevy. In WebDB 2005.

Query Relaxation Using Malleable Schemas
Xuan Zhou, Julien Gaugaz, Wolf-Tilo Balke, Wolfgang Nejdl
Proc. of the SIGMOD Conference (2007)

- Interesting visualization, though a bit too complex for everyday users
- Search
  - Keyword search – IR
  - Domain restricted search (i.e., Organization) – Recent IR
  - Association queries (i.e., triples) – DB
- Less special things, but not very common:
  - Basic PIM ontology used as a *Domain Model*
  - All associations are stored in a database

Slide from Paul Chirita

Semex: Search



Keyword: semex

```
SEMEX
  Applications
    File Browser
    Web Browser
  Data
    Person (4687)
    Image (27438)
    Message (18498)
    Article (3921)
    Presentation (144)
    Conference (667)
    Journal (480)
    PDI (49030)
    Webpage (0)
```

History:

```
Keyword Search: SEMEX
Keyword Search: wOOGLE
Keyword Search: semex
```

```
Search results for semex
  All Conferences (3)
    Conference: IIWeb [2004]
    Conference: Proc. of SIGMOD [2005]
    Conference: Proc. of CIDR [2005]
  All Images (105)
  All Messages (2398)
  All Persons (15)
    Person: alon ha
    Person: stefan b. s      son
    Person: dan suciu
    Person: ema nemes
    Person: yuhan cai
    Person: jayant madhavan
    Person: Sergey Melnik
    Person: steve gribble
    Person: Erhard Rahm
    Person: pedro m domingos
    Person: Jing Michelle Liu
    Person: Luna (Xin) Dong
    Person: luna dong
    Person: Henry  Levy
    Person: michelle liu
  All Presentations (2)
    Presentation: Semex: A Platform for Personal Information Management and In
    Presentation: SEMEX(I)
  All Articles (65)
    Article: Reference Reconciliation in Complex Information Spaces
      Author
      Cites
      CitedBy
      PublishedIn
        Conference: Proc. of SIGMOD
          PublishedPaper
            Article: Adaptive query processing of top-k queries
            Article: Corpus-based Schema Matching
            Article: Reference Reconciliation in Complex Inform
            Article: Personal Information Management with   Semex
            Article: Structured value ranking in update-intensive relational databases
      SimilarTo
      FromFile
      MentionPersons
      Groups
    Article: www publication semex_sigmoddemo
      FromFile
        PDI: file:/homes/gws/lunadong/www/publication/semex_sigmoddemo.pdf
      MentionPersons
    Article: www publication semex_proposal
    Article: Semex semex_sigmodDemo semexdemo
    Article: A Semantics for Model Management Operators
    Article: Semex semex_msrProposal proposal
    Article: Personal Information Management with   Semex
    Article: SEMEX: Toward On-the-fly Personal Information Integration
    Article: semex_iiweb submission base
    Article: Semex alonProposal_2004 gunning
```

```
Year
  2005
Time
  16:36:01, December 17, 2004
  19:01:12, December 24, 2005
```

Search Semex

3 Conferences for publishing Semex papers

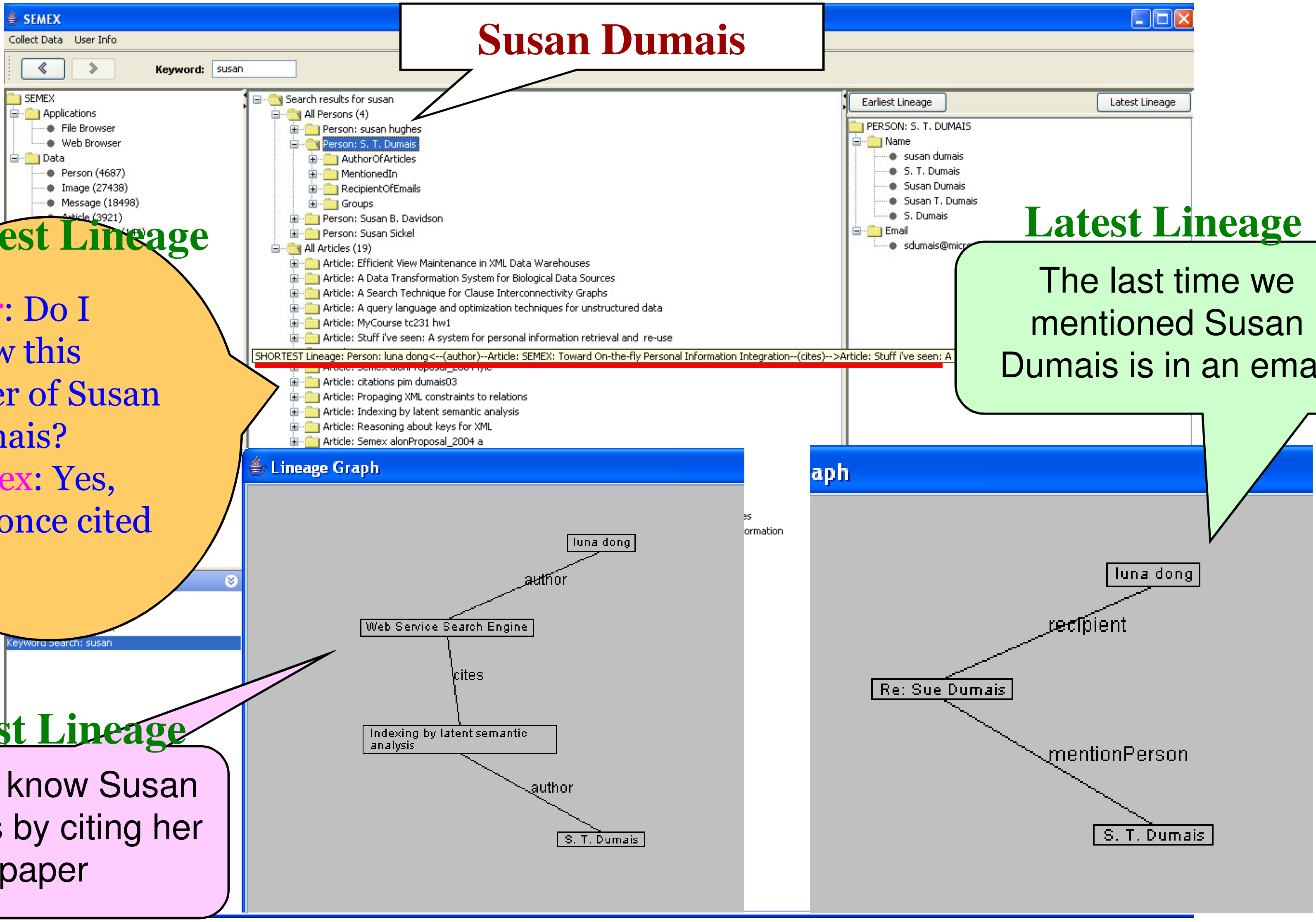
105 Images in Semex papers

2398 Messages
2 Presentations
65 Articles

15 Persons working on Semex
(though they are not named Semex)

Slide from Paul Chirita

Semex: Linkage Vizualization



SEMEX

Collect Data User Info

Keyword: susan

SEMEX
- Applications
  - File Browser
  - Web Browser
- Data
  - Person (4687)
  - Image (27438)
  - Message (18498)
  - Article (3921)

Keyword Search: susan

Search results for susan
- All Persons (4)
  - Person: susan hughes
  - Person: S. T. Dumais
    - AuthorOfArticles
    - MentionedIn
    - RecipientOfEmails
    - Groups
  - Person: Susan B. Davidson
  - Person: Susan Sickel
- All Articles (19)
  - Article: Efficient View Maintenance in XML Data Warehouses
  - Article: A Data Transformation System for Biological Data Sources
  - Article: A Search Technique for Clause Interconnectivity Graphs
  - Article: A query language and optimization techniques for unstructured data
  - Article: MyCourse tc231 hw1
  - Article: Stuff i've seen: A system for personal information retrieval and re-use
  - Article: citations pim dumais03
  - Article: Propaging XML constraints to relations
  - Article: Indexing by latent semantic analysis
  - Article: Reasoning about keys for XML
  - Article: Semex alonProposal_2004 a

SHORTEST Lineage: Person: luna dong<--(author)--Article: SEMEX: Toward On-the-fly Personal Information Integration--(cites)-->Article: Stuff i've seen: A

Earliest Lineage | Latest Lineage

PERSON: S. T. DUMAIS
- Name
  - susan dumais
  - S. T. Dumais
  - Susan Dumais
  - Susan T. Dumais
  - S. Dumais
- Email
  - sdumais@micr

Susan Dumais

Shortest Lineage

User: Do I know this paper of Susan Dumais?
Semex: Yes, you once cited it.

Latest Lineage

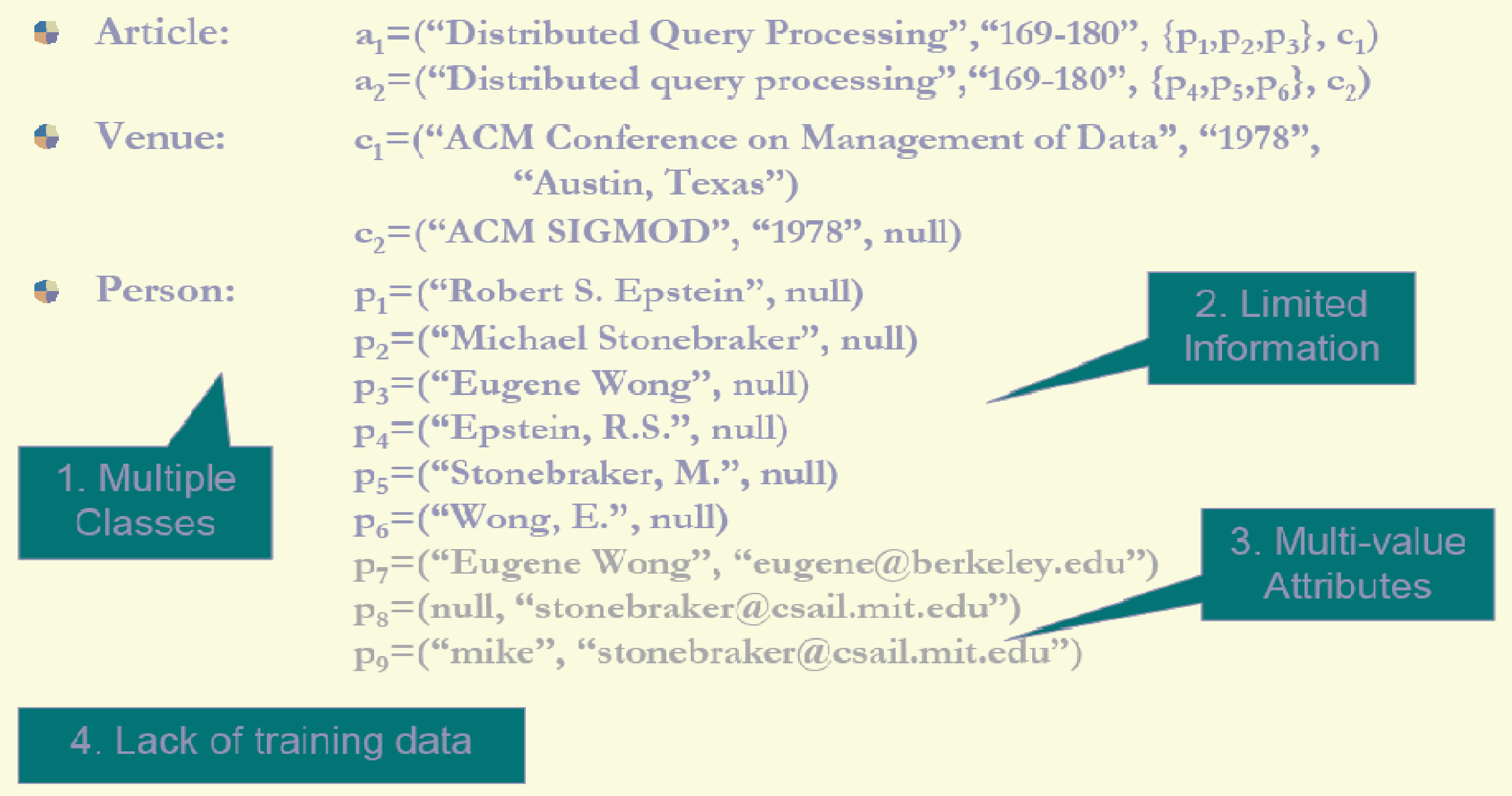
The last time we mentioned Susan Dumais is in an email

Earliest Lineage

I got to know Susan Dumais by citing her paper

Semex: PIM Reference Reconciliation: Challenges

Slide from Paul Chirita



- Article: a_1=("Distributed Query Processing","169-180", $\{p_1,p_2,p_3\}$, c_1)
  a_2=("Distributed query processing","169-180", $\{p_4,p_5,p_6\}$, c_2)
- Venue: c_1=("ACM Conference on Management of Data", "1978", "Austin, Texas")
  c_2=("ACM SIGMOD", "1978", null)
- Person: p_1=("Robert S. Epstein", null)
  p_2=("Michael Stonebraker", null)
  p_3=("Eugene Wong", null)
  p_4=("Epstein, R.S.", null)
  p_5=("Stonebraker, M.", null)
  p_6=("Wong, E.", null)
  p_7=("Eugene Wong", "eugene@berkeley.edu")
  p_8=(null, "stonebraker@csail.mit.edu")
  p_9=("mike", "stonebraker@csail.mit.edu")

1. Multiple Classes

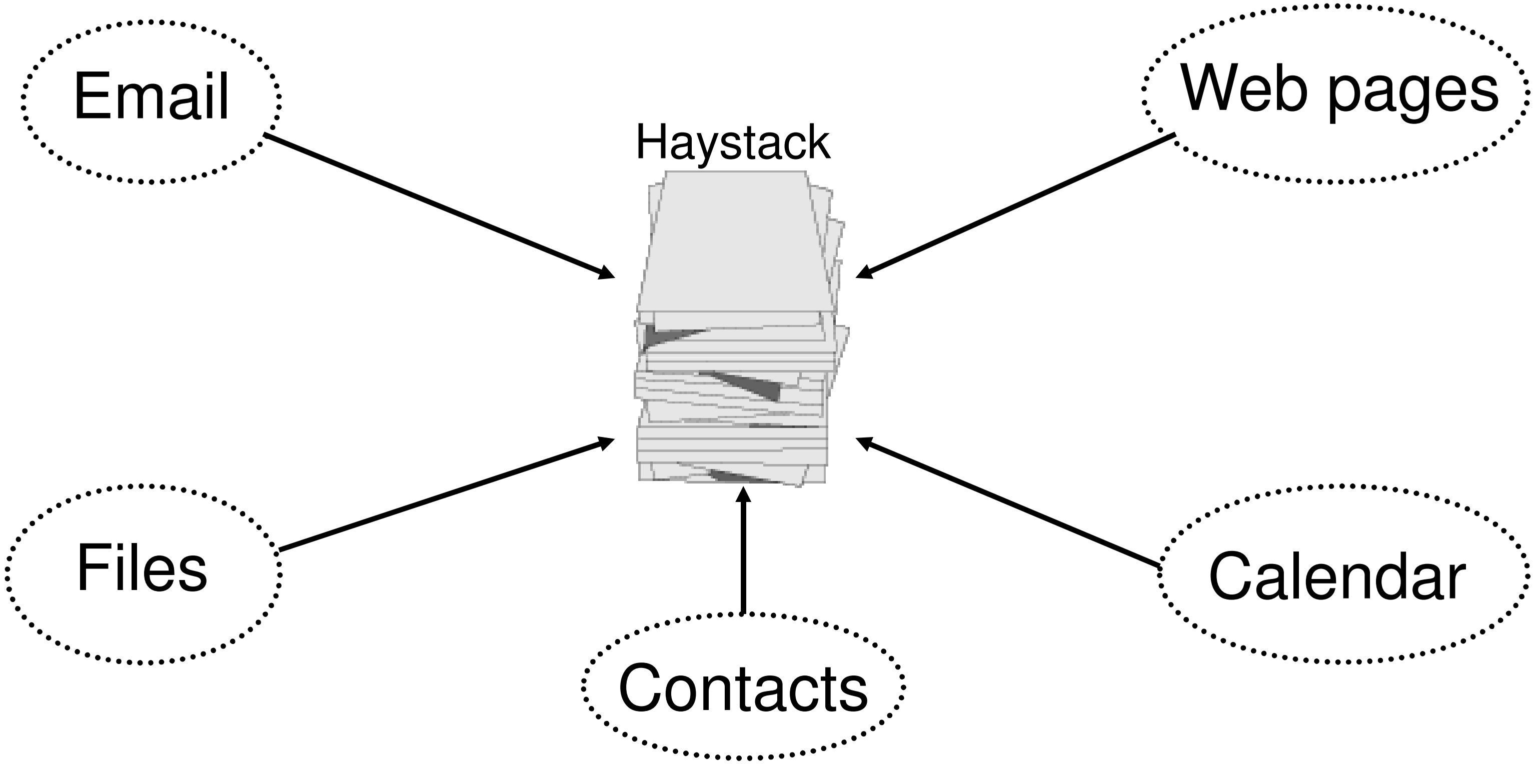
2. Limited Information

3. Multi-value Attributes

4. Lack of training data

Haystack (1)

Haystack: Per-User Information Environment Based on Semistructured Data. David Karger, in "Beyond the Desktop Metaphor" edited by Victor Kaptelinin and Mary Czerwinski. 2007



Email

Web pages

Haystack

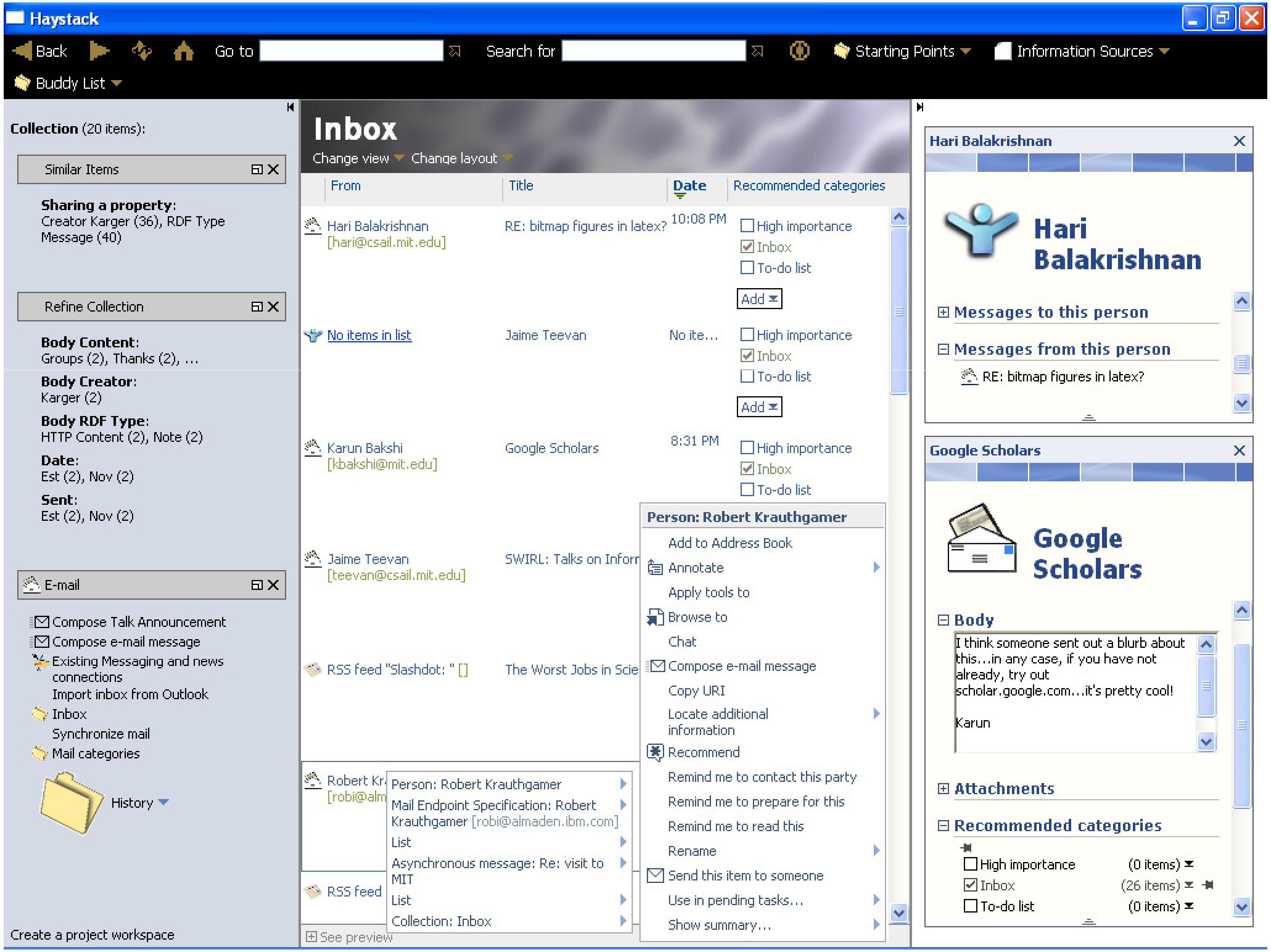
Files

Contacts

Calendar

- Lots of separate info, Haystack stores in central repository.
- Easy to separate info from its form, easy to connect related info.
- Many people could share a single repository

Haystack (2)



Haystack

Back | Go to | Search for | Starting Points | Information Sources | Buddy List

Collection (20 items):

Similar Items

Sharing a property:
Creator Karger (36), RDF Type Message (40)

Refine Collection

Body Content:
Groups (2), Thanks (2), …

Body Creator:
Karger (2)

Body RDF Type:
HTTP Content (2), Note (2)

Date:
Est (2), Nov (2)

Sent:
Est (2), Nov (2)

E-mail

- Compose Talk Announcement
- Compose e-mail message
- Existing Messaging and news connections
- Import inbox from Outlook
- Inbox
- Synchronize mail
- Mail categories

History

Create a project workspace

Inbox

Change view | Change layout

From	Title	Date	Recommended categories
Hari Balakrishnan [hari@csail.mit.edu]	RE: bitmap figures in latex?	10:08 PM	☐ High importance ☑ Inbox ☐ To-do list Add
No items in list	Jaime Teevan	No ite…	☐ High importance ☑ Inbox ☐ To-do list Add
Karun Bakshi [kbakshi@mit.edu]	Google Scholars	8:31 PM	☐ High importance ☑ Inbox ☐ To-do list
Jaime Teevan [teevan@csail.mit.edu]	SWIRL: Talks on Inforr		
RSS feed "Slashdot: " []	The Worst Jobs in Scie		
Robert Kr [robi@alm			
RSS feed			

See preview

Person: Robert Krauthgamer
Mail Endpoint Specification: Robert Krauthgamer [robi@almaden.ibm.com]
List
Asynchronous message: Re: visit to MIT
List
Collection: Inbox

Person: Robert Krauthgamer

- Add to Address Book
- Annotate
- Apply tools to
- Browse to
- Chat
- Compose e-mail message
- Copy URI
- Locate additional information
- Recommend
- Remind me to contact this party
- Remind me to prepare for this
- Remind me to read this
- Rename
- Send this item to someone
- Use in pending tasks…
- Show summary…

Hari Balakrishnan

Hari Balakrishnan

⊞ **Messages to this person**

⊟ **Messages from this person**

RE: bitmap figures in latex?

Google Scholars

Google Scholars

⊟ **Body**

I think someone sent out a blurb about this…in any case, if you have not already, try out scholar.google.com…it's pretty cool!

Karun

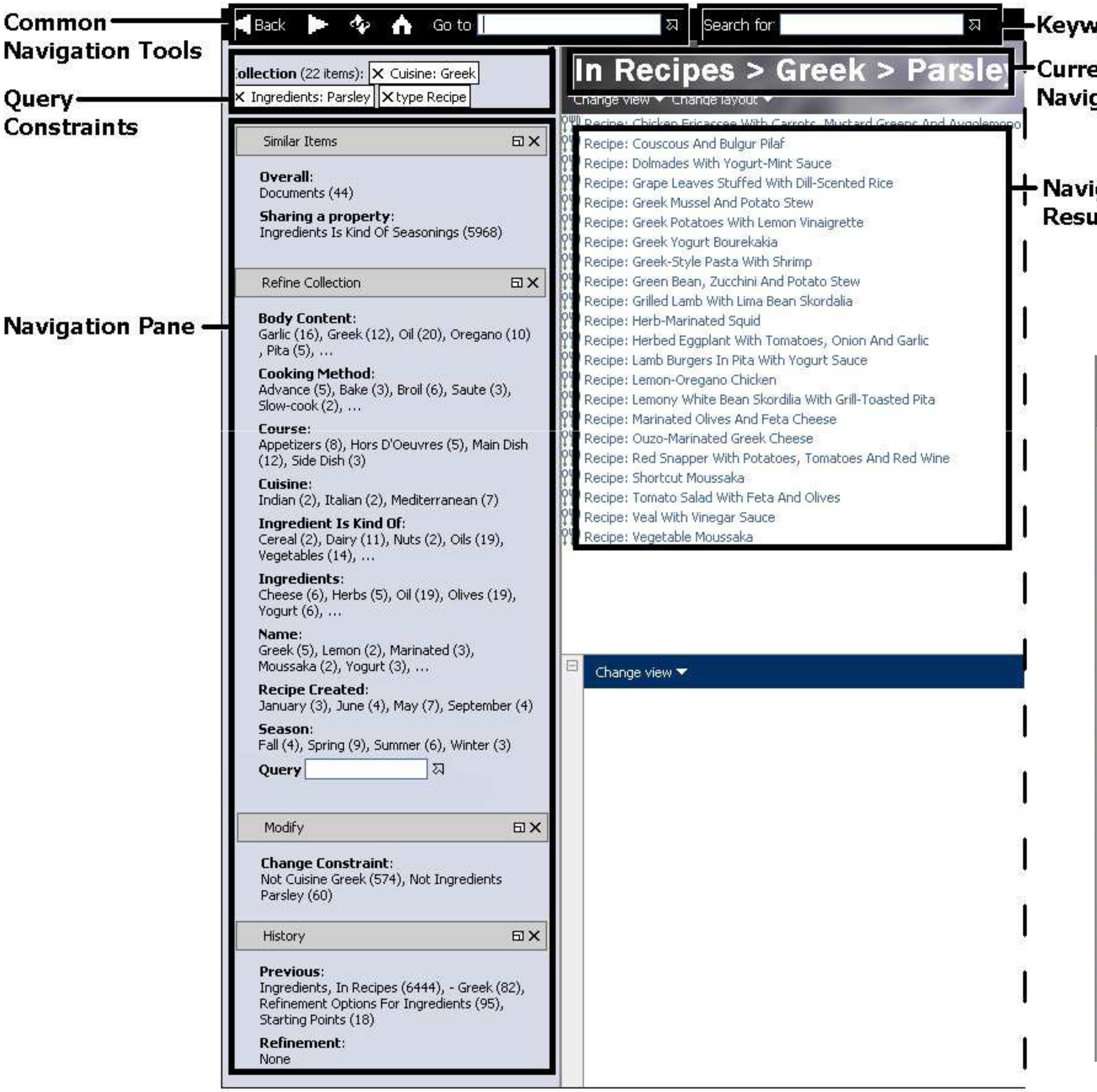
⊞ **Attachments**

⊟ **Recommended categories**

☐ High importance	(0 items)
☑ Inbox	(26 items)
☐ To-do list	(0 items)

Magnet

Magnet: Supporting Navigation in Semistructured Data Environments. Vineet Sinha and David R. Karger, in SIGMOD 2005.



ollection (22 items): ☒ Cuisine: Greek ☒ Ingredients: Parsley ☒ type Recipe

In Recipes > Greek > Parsle

Similar Items

Overall:
Documents (44)

Sharing a property:
Ingredients Is Kind Of Seasonings (5968)

Refine Collection

Body Content:
Garlic (16), Greek (12), Oil (20), Oregano (10), Pita (5), ...

Cooking Method:
Advance (5), Bake (3), Broil (6), Saute (3), Slow-cook (2), ...

Course:
Appetizers (8), Hors D'Oeuvres (5), Main Dish (12), Side Dish (3)

Cuisine:
Indian (2), Italian (2), Mediterranean (7)

Ingredient Is Kind Of:
Cereal (2), Dairy (11), Nuts (2), Oils (19), Vegetables (14), ...

Ingredients:
Cheese (6), Herbs (5), Oil (19), Olives (19), Yogurt (6), ...

Name:
Greek (5), Lemon (2), Marinated (3), Moussaka (2), Yogurt (3), ...

Recipe Created:
January (3), June (4), May (7), September (4)

Season:
Fall (4), Spring (9), Summer (6), Winter (3)

Query

Modify

Change Constraint:
Not Cuisine Greek (574), Not Ingredients Parsley (60)

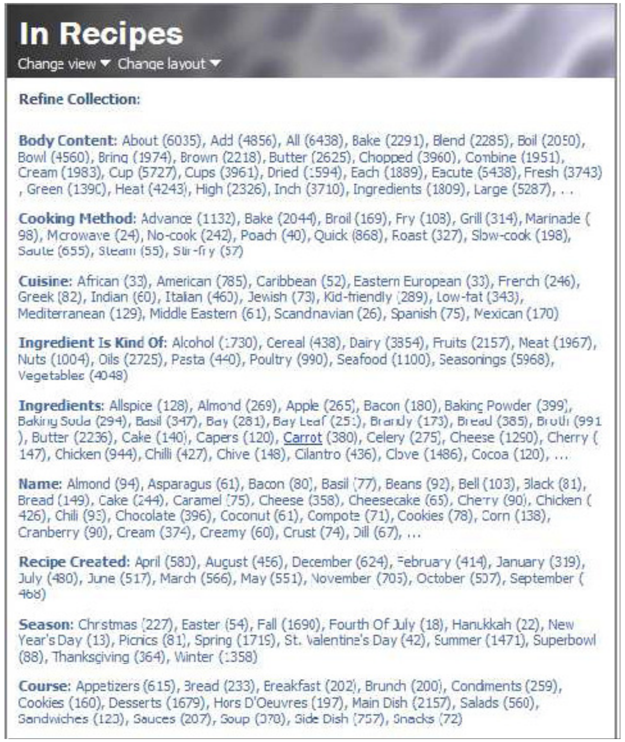
History

Previous:
Ingredients, In Recipes (6444), - Greek (82), Refinement Options For Ingredients (95), Starting Points (18)

Refinement:
None

Recipe: Chicken Fricassee With Carrots, Mustard Greens And Avgolemono
Recipe: Couscous And Bulgur Pilaf
Recipe: Dolmades With Yogurt-Mint Sauce
Recipe: Grape Leaves Stuffed With Dill-Scented Rice
Recipe: Greek Mussel And Potato Stew
Recipe: Greek Potatoes With Lemon Vinaigrette
Recipe: Greek Yogurt Bourekakia
Recipe: Greek-Style Pasta With Shrimp
Recipe: Green Bean, Zucchini And Potato Stew
Recipe: Grilled Lamb With Lima Bean Skordalia
Recipe: Herb-Marinated Squid
Recipe: Herbed Eggplant With Tomatoes, Onion And Garlic
Recipe: Lamb Burgers In Pita With Yogurt Sauce
Recipe: Lemon-Oregano Chicken
Recipe: Lemony White Bean Skordilia With Grill-Toasted Pita
Recipe: Marinated Olives And Feta Cheese
Recipe: Ouzo-Marinated Greek Cheese
Recipe: Red Snapper With Potatoes, Tomatoes And Red Wine
Recipe: Shortcut Moussaka
Recipe: Tomato Salad With Feta And Olives
Recipe: Veal With Vinegar Sauce
Recipe: Vegetable Moussaka

Change view



In Recipes

Refine Collection:

Body Content: About (6035), Add (4856), All (6438), Bake (2291), Blend (2285), Boil (2050), Bowl (4560), Bring (1974), Brown (2218), Butter (2625), Chopped (3960), Combine (1951), Cream (1983), Cup (5727), Cups (3961), Dried (1594), Each (1889), Eacute (5438), Fresh (3743), Green (1390), Heat (4243), High (2326), Inch (3710), Ingredients (1809), Large (5287), ...

Cooking Method: Advance (1132), Bake (2044), Broil (169), Fry (103), Grill (314), Marinade (98), Mcrowave (24), No-cook (242), Poach (40), Quick (868), Roast (327), Slow-cook (198), Saute (655), Steam (55), Stir-fry (57)

Cuisine: African (33), American (785), Caribbean (52), Eastern European (33), French (246), Greek (82), Indian (60), Italian (460), Jewish (73), Kid-friendly (289), Low-fat (343), Mediterranean (129), Middle Eastern (61), Scandnavian (26), Spanish (75), Mexican (170)

Ingredient Is Kind Of: Alcohol (1730), Cereal (438), Dairy (3854), Fruits (2157), Meat (1967), Nuts (1004), Oils (2725), Pasta (440), Poultry (990), Seafood (1100), Seasonings (5968), Vegetables (4048)

Ingredients: Allspice (128), Almond (269), Apple (265), Bacon (180), Baking Powder (399), Baking Soda (294), Basil (347), Bay (281), Bay Leaf (251), Brandy (173), Bread (385), Broth (991), Butter (2236), Cake (140), Capers (120), Carrot (380), Celery (275), Cheese (1290), Cherry (147), Chicken (944), Chilli (427), Chive (148), Cilantro (436), Clove (1486), Cocoa (120), ...

Name: Almond (94), Asparagus (61), Bacon (80), Basil (77), Beans (92), Bell (103), Black (81), Bread (149), Cake (244), Caramel (75), Cheese (358), Cheesecake (65), Cherry (90), Chicken (426), Chili (93), Chocolate (396), Coconut (61), Compote (71), Cookies (78), Corn (138), Cranberry (90), Cream (374), Creamy (60), Crust (74), Dill (67), ...

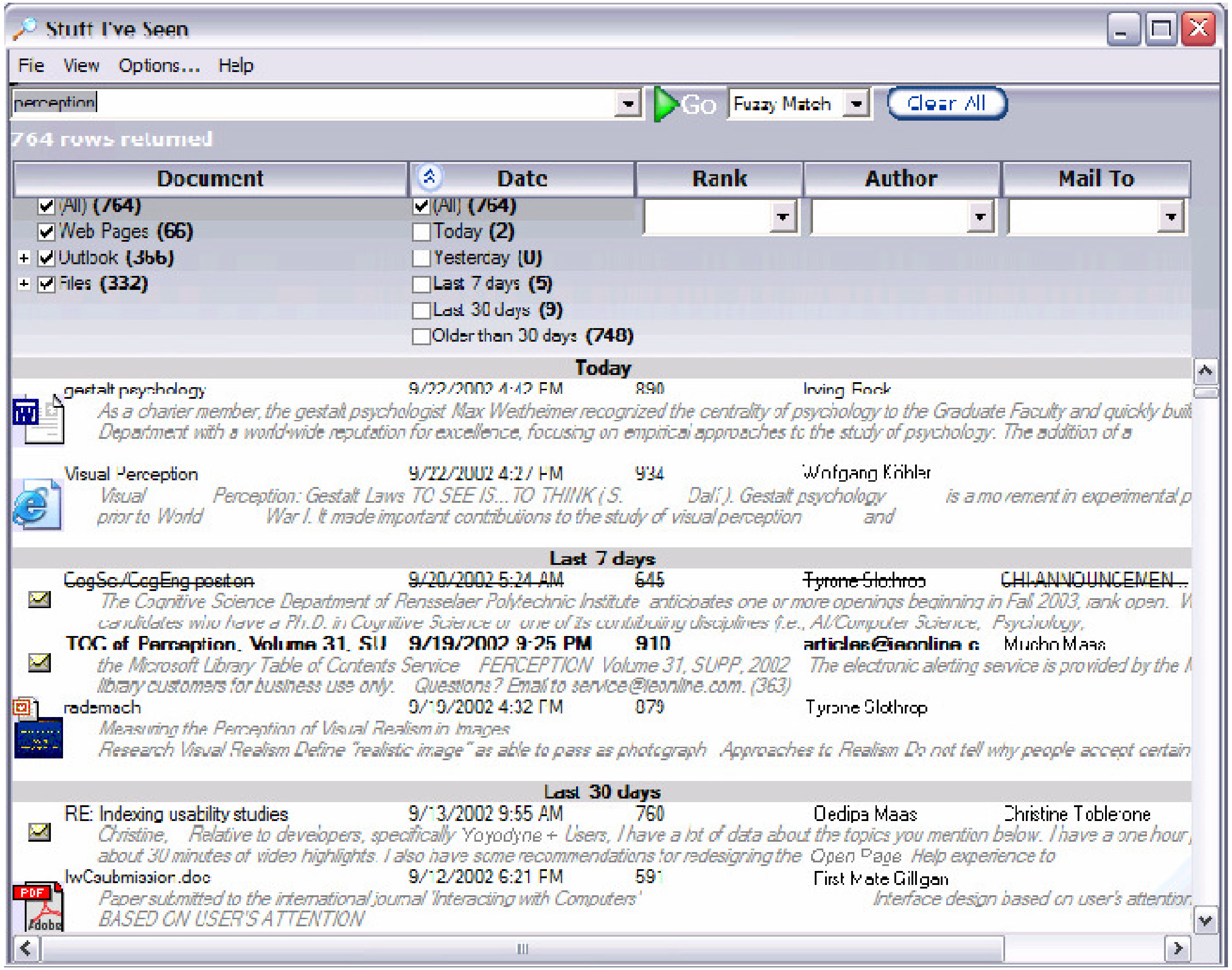
Recipe Created: April (580), August (456), December (624), February (414), January (319), July (480), June (517), March (566), May (551), November (705), October (507), September (468)

Season: Christmas (227), Easter (54), Fall (1690), Fourth Of July (18), Hanukkah (22), New Year's Day (13), Picnics (81), Spring (1719), St. Valentine's Day (42), Summer (1471), Superbowl (88), Thanksgiving (364), Winter (1358)

Course: Appetizers (615), Bread (233), Breakfast (202), Brunch (200), Condiments (259), Cookies (160), Desserts (1679), Hors D'Oeuvres (197), Main Dish (2157), Salads (560), Sandwiches (123), Sauces (207), Soup (370), Side Dish (757), Snacks (72)

Stuff I've Seen (SIS)

S. Dumais, E. Cutrell, J. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins. Stuff i've seen: a system for personal information retrieval and re-use. In SIGIR'03



perception

Fuzzy Match

764 rows returned

Document	Date	Rank	Author	Mail To
☑ (All) (764)	☑ (All) (764)			
☑ Web Pages (66)	□ Today (2)			
+ ☑ Outlook (366)	□ Yesterday (0)			
+ ☑ Files (332)	□ Last 7 days (5)			
	□ Last 30 days (9)			
	□ Older than 30 days (748)			

Today

gestalt psychology — 9/22/2002 4:42 PM — 890 — Irving Rock

As a charter member, the gestalt psychologist Max Wertheimer recognized the centrality of psychology to the Graduate Faculty and quickly buil Department with a world-wide reputation for excellence, focusing on empirical approaches to the study of psychology. The addition of a

Visual Perception — 9/22/2002 4:27 PM — 934 — Wolfgang Köhler

Visual Perception: Gestalt Laws TO SEE IS...TO THINK (S. Dalí). Gestalt psychology is a movement in experimental p prior to World War I. It made important contributions to the study of visual perception and

Last 7 days

~~CogSc/CogEng position — 9/20/2002 5:24 AM — 645 — Tyrone Slothrop — CHI-ANNOUNCEMEN...~~

The Cognitive Science Department of Rensselaer Polytechnic Institute anticipates one or more openings beginning in Fall 2003, rank open. W candidates who have a Ph.D. in Cognitive Science or one of its contributing disciplines (i.e., AI/Computer Science, Psychology,

TOC of Perception, Volume 31, SU — 9/19/2002 9:25 PM — 910 — articles@ieonline.c — Mucho Maas

the Microsoft Library Table of Contents Service PERCEPTION Volume 31, SUPP, 2002 The electronic alerting service is provided by the M library customers for business use only. Questions? Email to service@ieonline.com. (363)

rademach — 9/19/2002 4:32 PM — 879 — Tyrone Slothrop

Measuring the Perception of Visual Realism in Images
Research Visual Realism Define "realistic image" as able to pass as photograph Approaches to Realism Do not tell why people accept certain

Last 30 days

RE: Indexing usability studies — 9/13/2002 9:55 AM — 760 — Oedipa Maas — Christine Tobleone

Christine, Relative to developers, specifically Yoyodyne + Users, I have a lot of data about the topics you mention below. I have a one hour about 30 minutes of video highlights. I also have some recommendations for redesigning the Open Page Help experience to

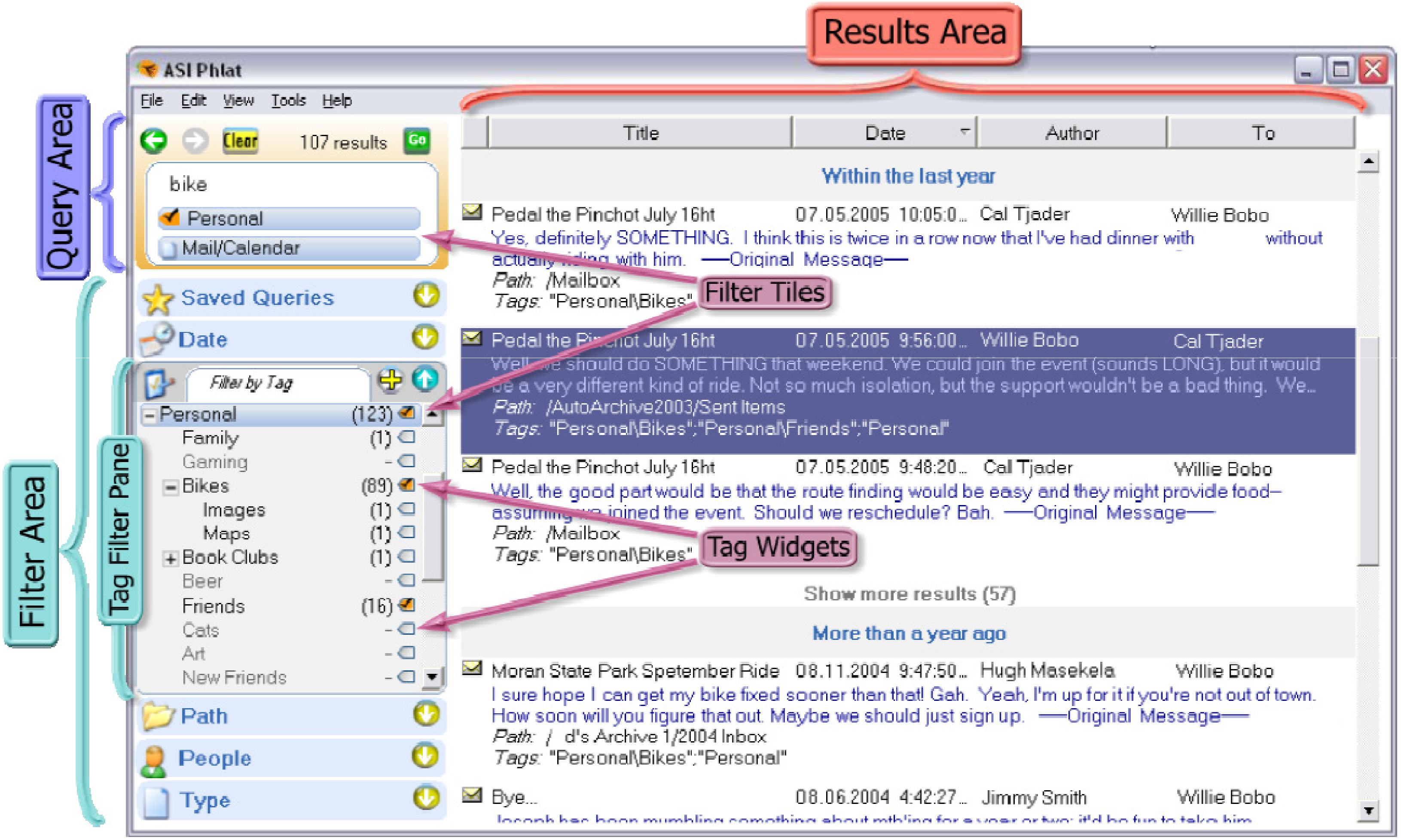
IwCsubmission.doc — 9/12/2002 6:21 PM — 591 — First Mate Gilligan

Paper submitted to the international journal 'Interacting with Computers' Interface design based on user's attention BASED ON USER'S ATTENTION

Phlat

E. Cutrell, D. Robbins, S. Dumais, and R. Sarin. Fast, Flexible Filtering with phlat. In CHI '06

http://research.microsoft.com/en-us/downloads/0cdb50f3-ccf6-4198-b874-4643791d4dc4



− Personal	(123)
Family	(1)
Gaming	-
− Bikes	(89)
Images	(1)
Maps	(1)
+ Book Clubs	(1)
Beer	-
Friends	(16)
Cats	-
Art	-
New Friends	-

Within the last year

Pedal the Pinchot July 16ht 07.05.2005 10:05:0... Cal Tjader Willie Bobo
Yes, definitely SOMETHING. I think this is twice in a row now that I've had dinner with without actually riding with him. —Original Message—
Path: /Mailbox
Tags: "Personal\Bikes"

Pedal the Pinchot July 16ht 07.05.2005 9:56:00... Willie Bobo Cal Tjader
Well, we should do SOMETHING that weekend. We could join the event (sounds LONG), but it would be a very different kind of ride. Not so much isolation, but the support wouldn't be a bad thing. We...
Path: /AutoArchive2003/Sent Items
Tags: "Personal\Bikes";"Personal\Friends";"Personal"

Pedal the Pinchot July 16ht 07.05.2005 9:48:20... Cal Tjader Willie Bobo
Well, the good part would be that the route finding would be easy and they might provide food– assuming we joined the event. Should we reschedule? Bah. —Original Message—
Path: /Mailbox
Tags: "Personal\Bikes"

Show more results (57)

More than a year ago

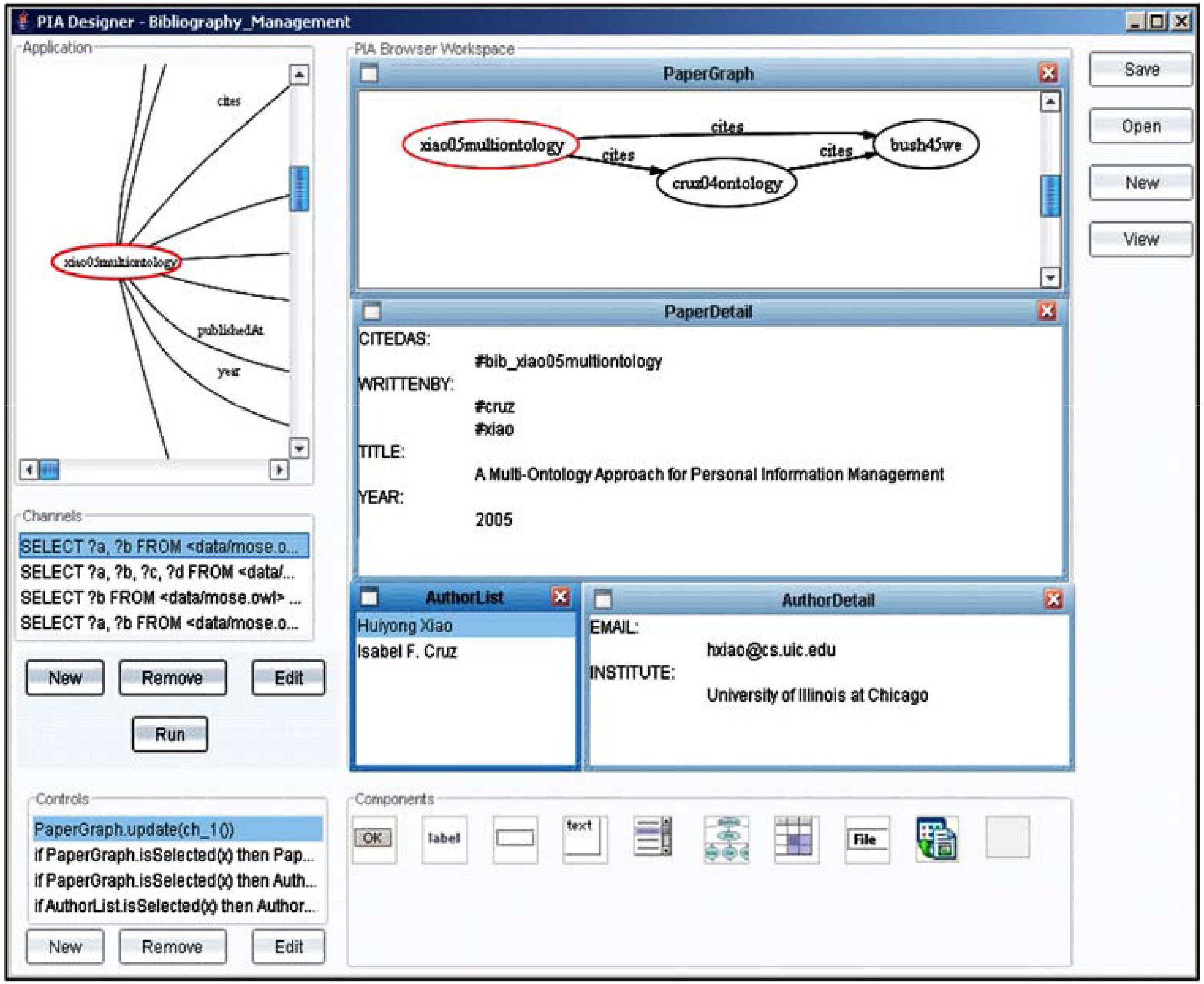
Moran State Park Spetember Ride 08.11.2004 9:47:50... Hugh Masekela Willie Bobo
I sure hope I can get my bike fixed sooner than that! Gah. Yeah, I'm up for it if you're not out of town. How soon will you figure that out. Maybe we should just sign up. —Original Message—
Path: / d's Archive 1/2004 Inbox
Tags: "Personal\Bikes";"Personal"

Bye... 08.06.2004 4:42:27... Jimmy Smith Willie Bobo
Joseph has been mumbling something about mtb'ing for a year or two; it'd be fun to take him

Figure 2. The Phlat interface with a query of a single keyword and two filters.

Phlat is written in Microsoft Visual C# and uses the Windows Desktop Search indexing and search engine

Personal Information Application



Channels:

```
SELECT ?a, ?b FROM <data/mose.o...
SELECT ?a, ?b, ?c, ?d FROM <data/...
SELECT ?b FROM <data/mose.owl> ...
SELECT ?a, ?b FROM <data/mose.o...
```

PaperDetail:

CITEDAS:
#bib_xiao05multiontology
WRITTENBY:
#cruz
#xiao
TITLE:
A Multi-Ontology Approach for Personal Information Management
YEAR:
2005

AuthorList:

Huiyong Xiao
Isabel F. Cruz

AuthorDetail:

EMAIL:
hxiao@cs.uic.edu
INSTITUTE:
University of Illinois at Chicago

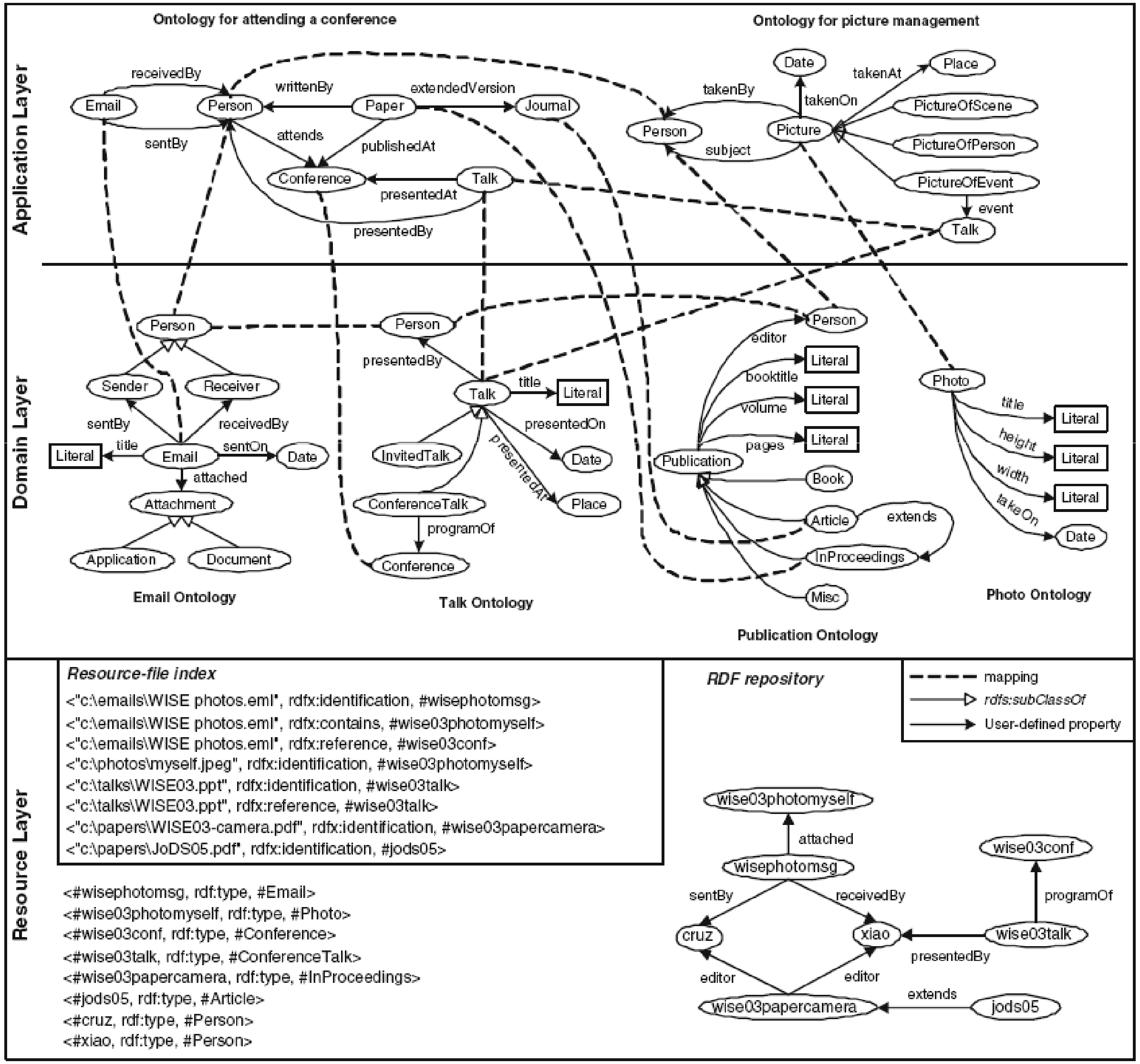
Controls:

```
PaperGraph.update(ch_1())
if PaperGraph.isSelected(x) then Pap...
if PaperGraph.isSelected(x) then Auth...
if AuthorList.isSelected(x) then Author...
```

A layered framework supporting personal information integration and application design for the semantic desktop, Isabel F. Cruz, Huiyong Xiao, in VLDB Journal 2008

Using RDQL (RDF Data Query Language)

PIA: Ontology

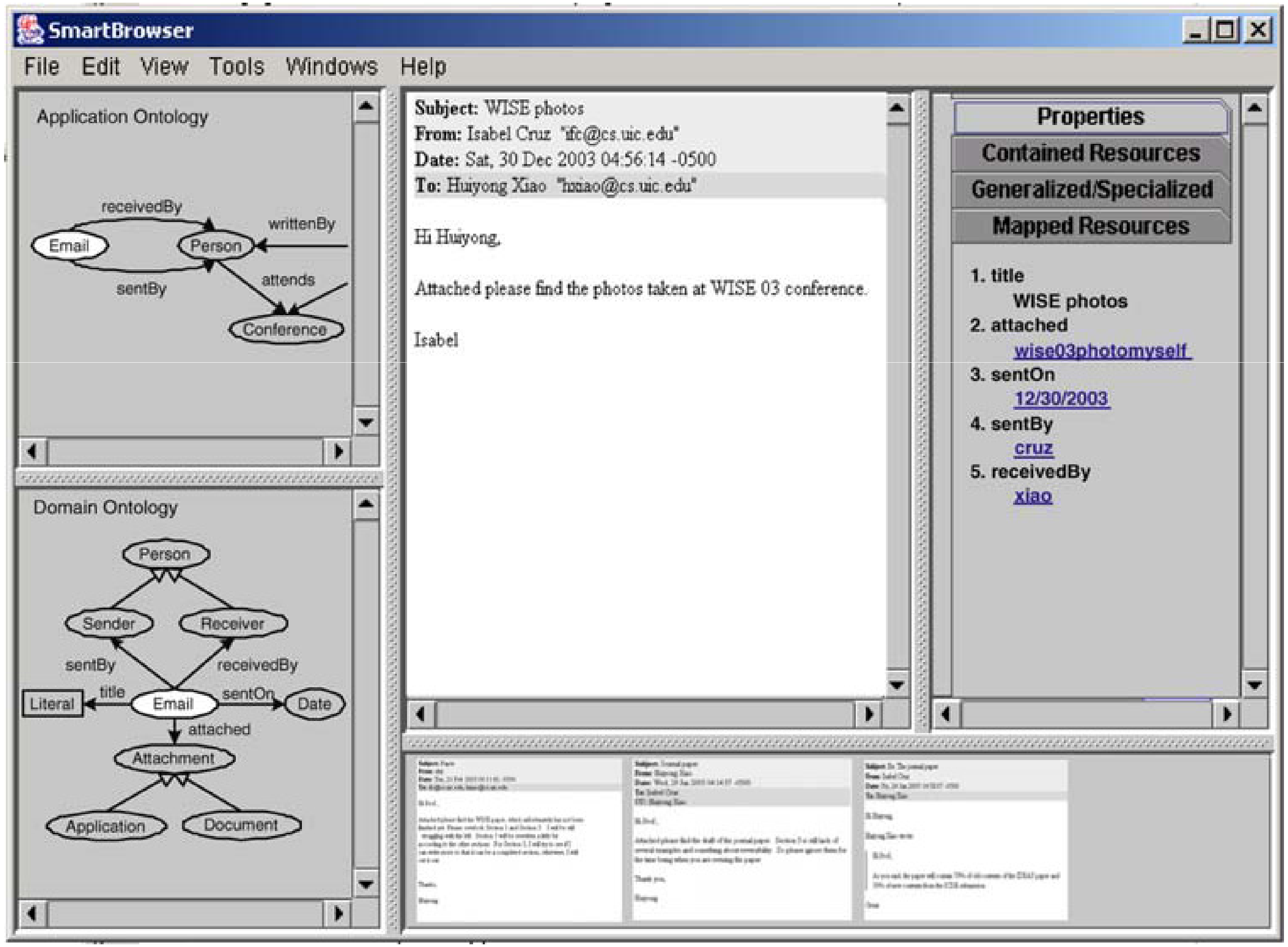


Resource-file index

<"c:\emails\WISE photos.eml", rdfx:identification, #wisephotomsg>
<"c:\emails\WISE photos.eml", rdfx:contains, #wise03photomyself>
<"c:\emails\WISE photos.eml", rdfx:reference, #wise03conf>
<"c:\photos\myself.jpeg", rdfx:identification, #wise03photomyself>
<"c:\talks\WISE03.ppt", rdfx:identification, #wise03talk>
<"c:\talks\WISE03.ppt", rdfx:reference, #wise03talk>
<"c:\papers\WISE03-camera.pdf", rdfx:identification, #wise03papercamera>
<"c:\papers\JoDS05.pdf", rdfx:identification, #jods05>

<#wisephotomsg, rdf:type, #Email>
<#wise03photomyself, rdf:type, #Photo>
<#wise03conf, rdf:type, #Conference>
<#wise03talk, rdf:type, #ConferenceTalk>
<#wise03papercamera, rdf:type, #InProceedings>
<#jods05, rdf:type, #Article>
<#cruz, rdf:type, #Person>
<#xiao, rdf:type, #Person>

PIA: Smart Browser



Subject: WISE photos
From: Isabel Cruz "ifc@cs.uic.edu"
Date: Sat, 30 Dec 2003 04:56:14 -0500
To: Huiyong Xiao "hxiao@cs.uic.edu"

Hi Huiyong,

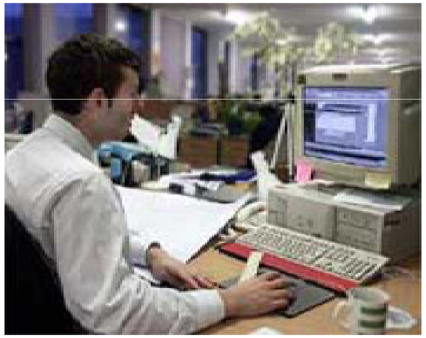
Attached please find the photos taken at WISE 03 conference.

Isabel

1. title
   WISE photos
2. attached
   wise03photomyself
3. sentOn
   12/30/2003
4. sentBy
   cruz
5. receivedBy
   xiao

Just-In-Time Retrieval

- "**Just-in-time Information –** Proactively offering a user information that is highly relevant to what s/he is currently focused on" (Pattie Maes)



JIT Approaches

- Watson
- Remembrance Agent
- Jimminy

All approaches aim to suggest relevant information snippets when the user writes a document or an email

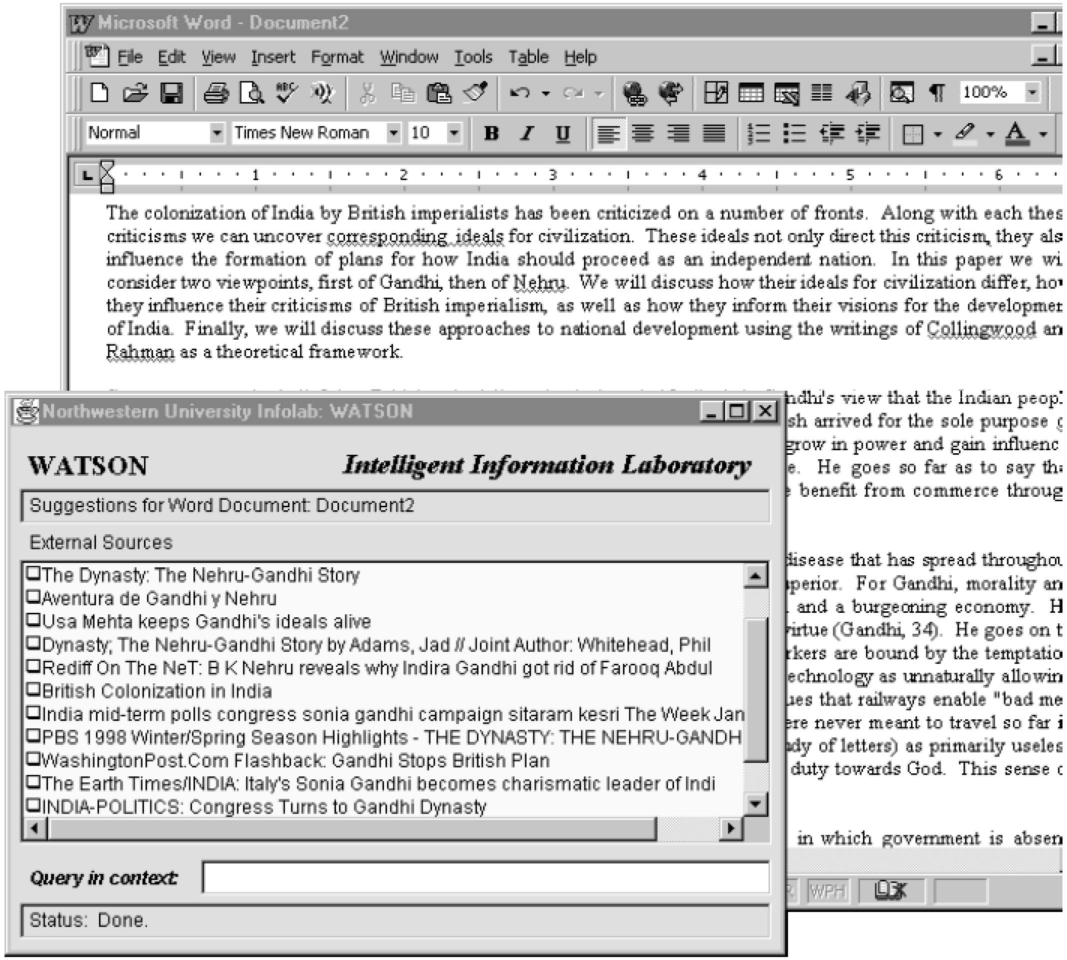
Some more:

QUESCOT, MarginNotes, Letizia, WordSieve, CALVIN, Kenjin

WATSON

J. Budzik and K. J. Hammond. User interactions with
everyday applications as context for just-in-time
information access. In IUI '00

- supports just-in-time access to task-relevant information
- a system gathers contextual information as a text of the document the user is manipulating
- proactively retrievs documents from distributed information repositories
- Potential problems:
  - managing interruptions
  - ranking suggestions


Microsoft Word - Document2
File Edit View Insert Format Window Tools Table Help
100%
Normal
Times New Roman
10
B I U
WPH

The colonization of India by British imperialists has been criticized on a number of fronts. Along with each thes
criticisms we can uncover corresponding ideals for civilization. These ideals not only direct this criticism, they als
influence the formation of plans for how India should proceed as an independent nation. In this paper we wi
consider two viewpoints, first of Gandhi, then of Nehru. We will discuss how their ideals for civilization differ, ho
they influence their criticisms of British imperialism, as well as how they inform their visions for the developme
of India. Finally, we will discuss these approaches to national development using the writings of Collingwood an
Rahman as a theoretical framework.

ndhi's view that the Indian peop
sh arrived for the sole purpose
grow in power and gain influenc
e. He goes so far as to say th
benefit from commerce throug

disease that has spread throughou
perior. For Gandhi, morality an
and a burgeoning economy. H
virtue (Gandhi, 34). He goes on t
rkers are bound by the temptatio
echnology as unnaturally allowin
ues that railways enable "bad me
ere never meant to travel so far i
dy of letters) as primarily useles
duty towards God. This sense

in which government is absen

Northwestern University Infolab: WATSON

WATSON ***Intelligent Information Laboratory***

Suggestions for Word Document: Document2

External Sources

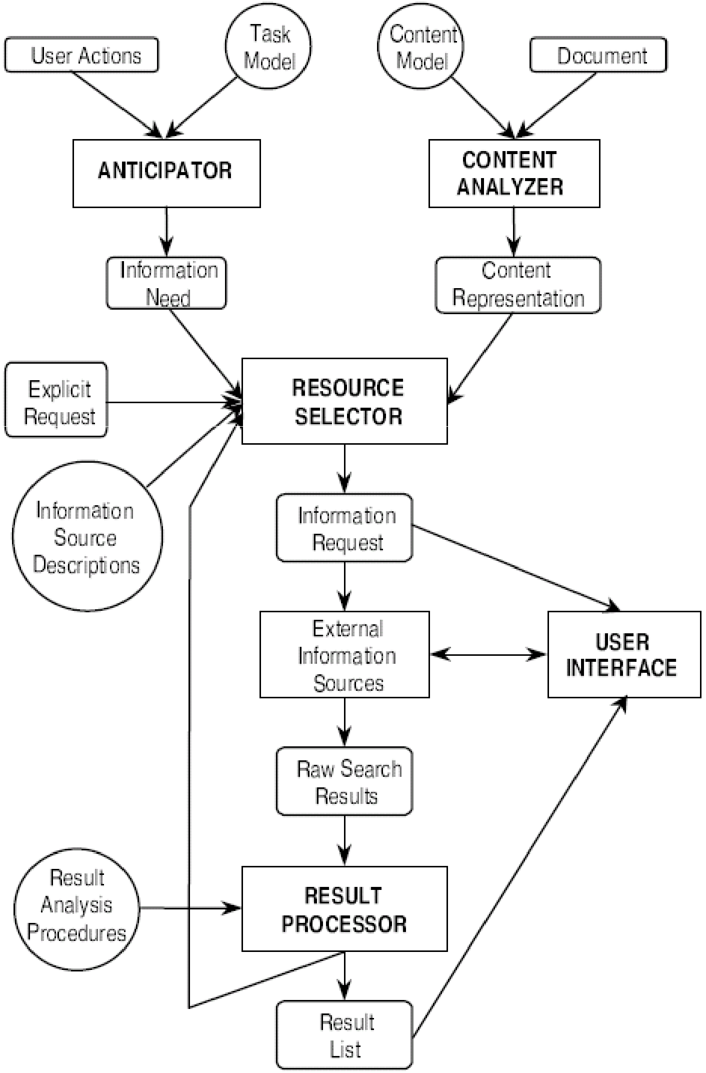
- □The Dynasty: The Nehru-Gandhi Story
- □Aventura de Gandhi y Nehru
- □Usa Mehta keeps Gandhi's ideals alive
- □Dynasty; The Nehru-Gandhi Story by Adams, Jad // Joint Author: Whitehead, Phil
- □Rediff On The NeT: B K Nehru reveals why Indira Gandhi got rid of Farooq Abdul
- □British Colonization in India
- □India mid-term polls congress sonia gandhi campaign sitaram kesri The Week Jan
- □PBS 1998 Winter/Spring Season Highlights - THE DYNASTY: THE NEHRU-GANDH
- □WashingtonPost.Com Flashback: Gandhi Stops British Plan
- □The Earth Times/INDIA: Italy's Sonia Gandhi becomes charismatic leader of Indi
- □INDIA-POLITICS: Congress Turns to Gandhi Dynasty

Query in context:

Status: Done.

Figure 2: Watson is suggesting documents as a user is writing a paper.

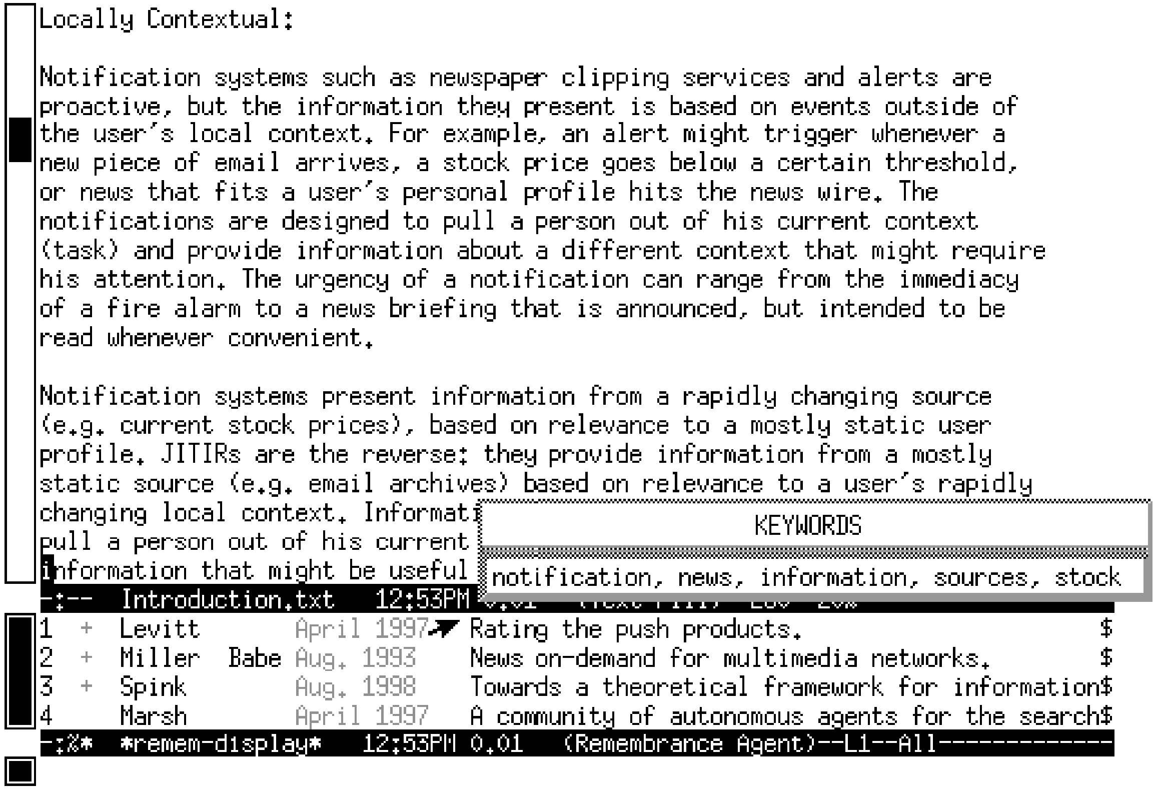
Watson Architecture



Remembrance Agent (RA)

- Remembrance Agent (‘96) / RADAR later for Word

Rhodes, B. and Starner, T. The Remembrance Agent: A continuously running information retrieval system, in *PAAM’96*



```
Locally Contextual:

Notification systems such as newspaper clipping services and alerts are
proactive, but the information they present is based on events outside of
the user's local context. For example, an alert might trigger whenever a
new piece of email arrives, a stock price goes below a certain threshold,
or news that fits a user's personal profile hits the news wire. The
notifications are designed to pull a person out of his current context
(task) and provide information about a different context that might require
his attention. The urgency of a notification can range from the immediacy
of a fire alarm to a news briefing that is announced, but intended to be
read whenever convenient.

Notification systems present information from a rapidly changing source
(e.g. current stock prices), based on relevance to a mostly static user
profile. JITIRs are the reverse: they provide information from a mostly
static source (e.g. email archives) based on relevance to a user's rapidly
changing local context. Informati
pull a person out of his current
information that might be useful
-:--  Introduction.txt   12:53PM
1  +  Levitt       April 1997   Rating the push products.                    $
2  +  Miller  Babe Aug. 1993    News on-demand for multimedia networks.      $
3  +  Spink        Aug. 1998    Towards a theoretical framework for information$
4     Marsh        April 1997   A community of autonomous agents for the search$
-:%*  *remem-display*   12:53PM 0.01   (Remembrance Agent)--L1--All-------------
```

```
KEYWORDS
notification, news, information, sources, stock
```

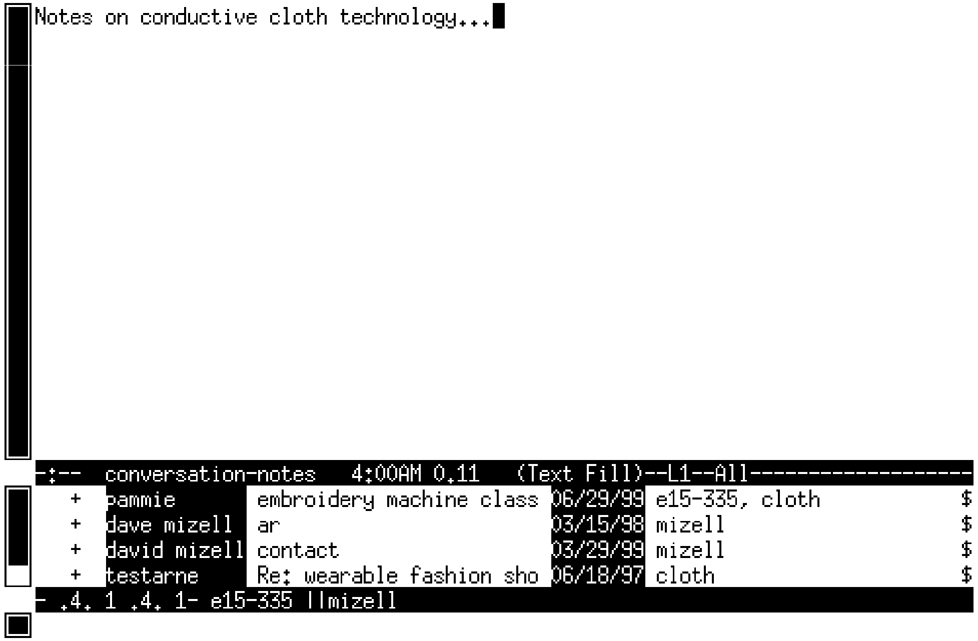
Jimminy

B. J. Rhodes. Just-in-time information retrieval. PhD thesis, 2000.

Rhodes, B., The Wearable Remembrance Agent: a system for augmented memory, in *Personal Technologies: Special Issue on Wearable Computing, 1997.*

- “Jimminy provides information based on a person's physical environment: her location, people in the room, time of day, and subject of the current conversation”

- “Processing is performed on a shoulder-worn “wearable computer,” and suggestions are presented on a head-mounted display.”



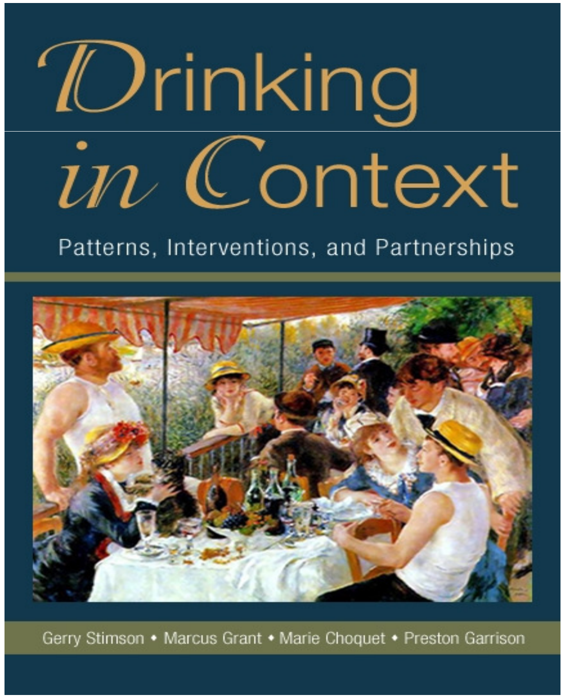
```
Notes on conductive cloth technology...









-:--  conversation-notes   4:00AM 0.11   (Text Fill)--L1--All------------------
   +  pammie       embroidery machine class 06/29/99 e15-335, cloth             $
   +  dave mizell  ar                       03/15/98 mizell                     $
   +  david mizell contact                  03/29/99 mizell                     $
   +  testarne     Re: wearable fashion sho 06/18/97 cloth                      $
- .4. 1 .4. 1- e15-335 ||mizell
```

Slide from
Stefania
Costache

What is context?

- Synonyms for context: (user/application) environment, situation, state, scenario, task, …

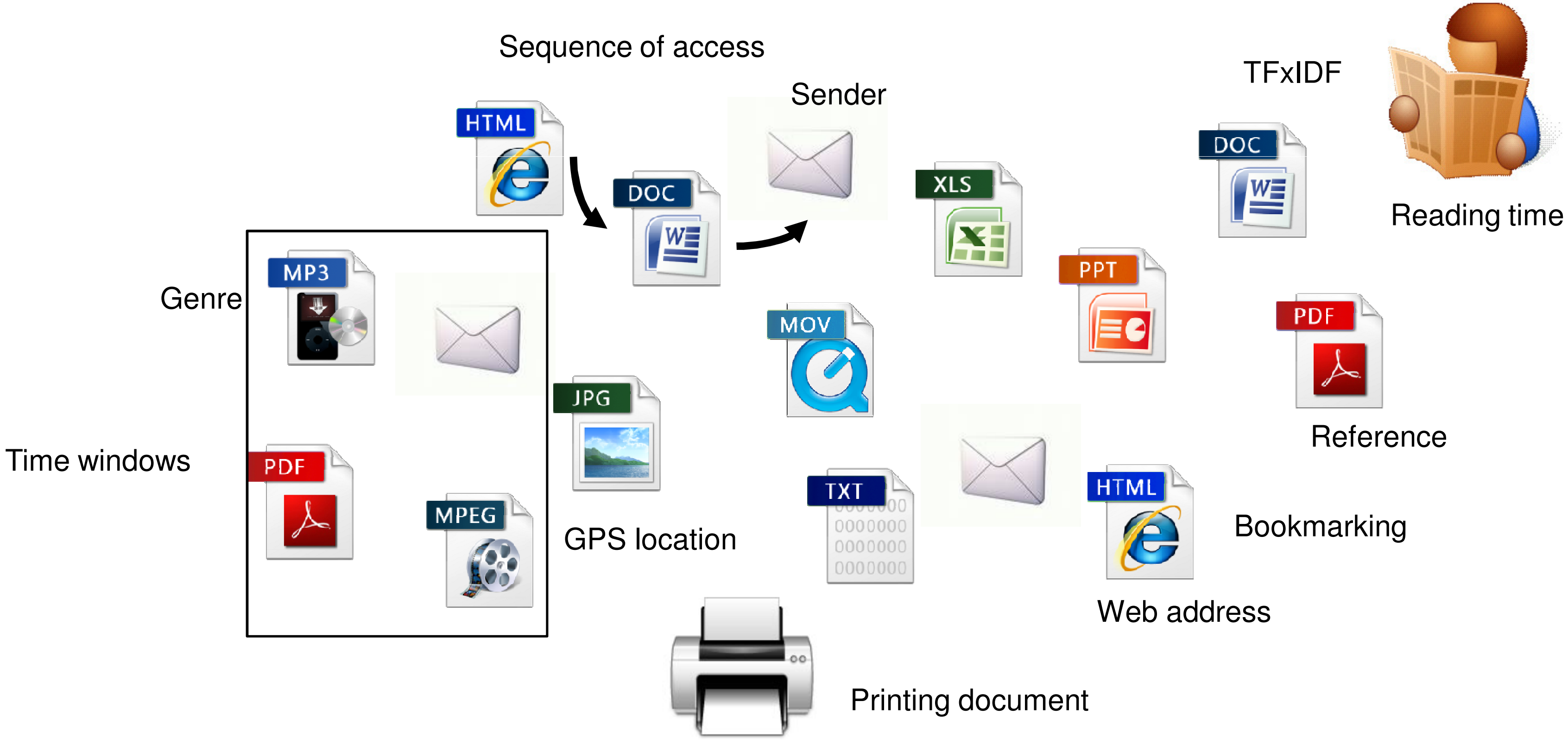
- Elements of context:
  - Location
  - People
  - Activities (tasks)
  - Time of day, season, temperature
  - Objects and changes to objects
  - Emotional state
  - Focus of attention


Drinking
in Context
Patterns, Interventions, and Partnerships
Gerry Stimson • Marcus Grant • Marie Choquet • Preston Garrison

Context on a Desktop

Resource as context

Interaction with resource as context



Sequence of access

Sender

TFxIDF

Reading time

Genre

Time windows

GPS location

Reference

Bookmarking

Web address

Printing document

Using Context to Improve Desktop Search

- Connections (**HITS and PageRank on File traces**)
- Confluence (**HITS and PageRank on File traces and Window focus**)
- SeeTrieve (**TFIDF variant on text snippets graph**)
- Method by P.Chirita and W. Nejdl, (**PageRank on File traces**)

Connections

C. A. N. Soules and G. R. Ganger. Connections: using context to enhance file search. In SOSP '05

- Tracing file system calls
- Temporal relationships between files
- Used to reorder content search results

- Relation window of N seconds
- Number of occurrences of a sequence of files

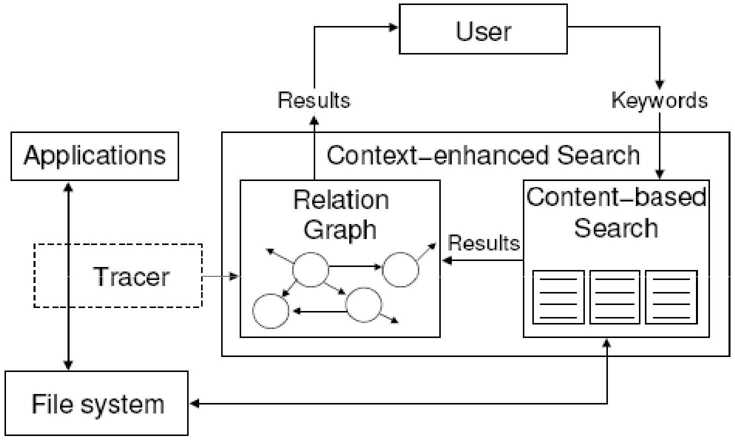


Figure 1: *Architecture of Connections.* Both applications and the file system remain unchanged, as the only information required by Connections can be gathered either by a transparent tracing module or directly from existing file system interfaces.

Confluence

K. A. Gyllstrom, C. Soules, and A. Veitch. Confluence: enhancing contextual desktop search. In SIGIR '07

Activity put in context: Identifying implicit task context within the user’s document interaction, Karl Gyllstrom, Craig Soules, Alistair Veitch, IIiX 2008

Confluence is an extension to ***Connections***

- ***Confluence*** records *window focus* events within the *GUI,* which are generated each time the user activates a different application window. These events are used to infer *task*.
- Contextual relationships can be used to augment traditional search methods with additional, conceptually related files that do not match the text query.
- *For example, if documents A and B are frequently accessed at similar points in time, this suggests a task commonality. Searches that return "A" now return "B“ as well.*

SeeTrieve

K. Gyllstrom and C. Soules. Seeing is retrieving: Building information context from what the user sees. In IUI '08

- A personal document retrieval and classification system
- Considers only the text presented to the user.
- Identifies information about the task associated with a document.

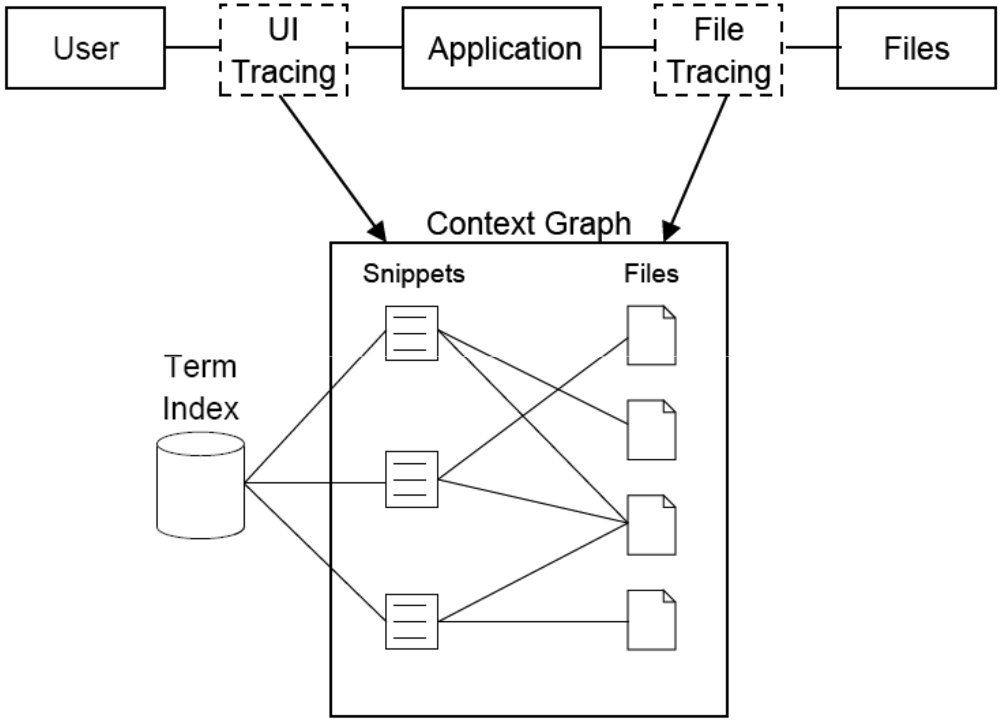
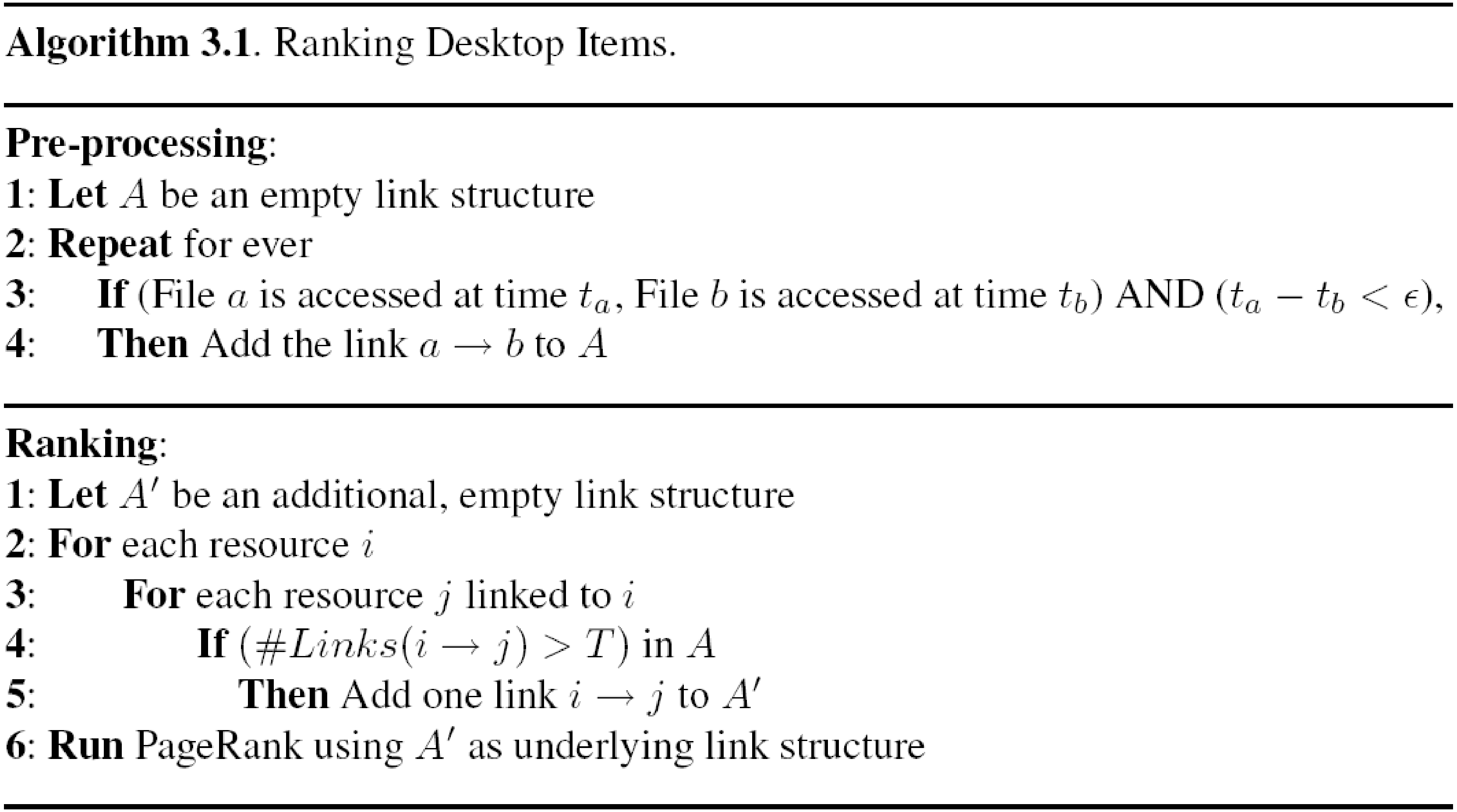


Figure 1. ***SeeTrieve*** **architecture.**

Method by P. Chirita and W. Nejdl

Analyzing User Behavior to Rank Desktop Items. Paul-Alexandru Chirita, Wolfgang Nejdl. In SPIRE 06



Algorithm 3.1. Ranking Desktop Items.

```
Pre-processing:
Let $A$ be an empty link structure
Repeat for ever
    If (File $a$ is accessed at time $t_a$, File $b$ is accessed at time $t_b$) AND ($t_a - t_b < \epsilon$),
    Then Add the link $a \rightarrow b$ to $A$

Ranking:
Let $A'$ be an additional, empty link structure
For each resource $i$
    For each resource $j$ linked to $i$
        If ($\#Links(i \rightarrow j) > T$) in $A$
            Then Add one link $i \rightarrow j$ to $A'$
Run PageRank using $A'$ as underlying link structure
```

Context Detection

- Lumiere (**Bayesian User Models**)
- Nepomuk (**K-Medoids and TFIDF**)
- TaskTracer and TaskPredictor (**Naïve Bayes/SVM**)
- SWISH (**Probabilistic Latent Semantic Indexing**)
- CAAD (**GaP probabilistic model**)

Some more:

QUESCOT, EPOS, MyLifeBits, Lifestreams

Lumiere

E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelse. The lumiere project: Bayesian user modeling for inferring the goals and needs of soft. In UAI'98

Goal:

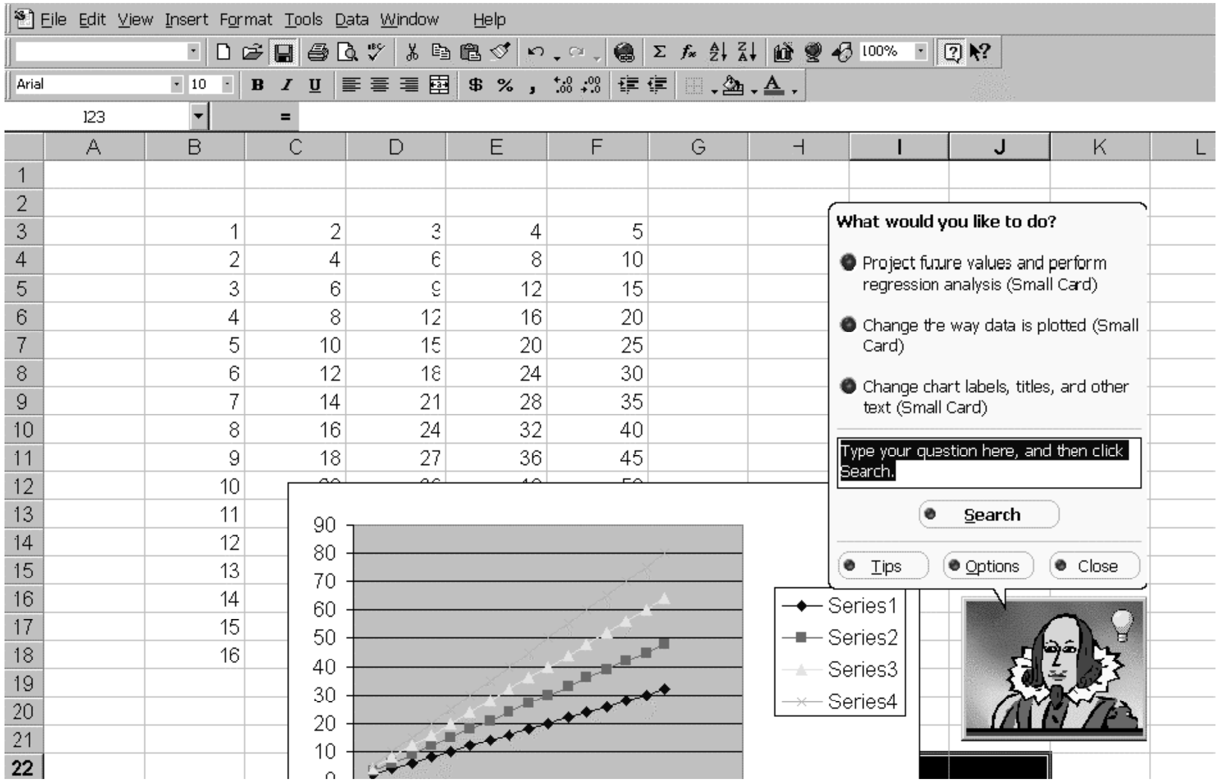
- help assistant for MS Office 97
- predict if help is needed, if yes, what is the problem?

Tools:

- Bayesian User Models

Lessons learned:

- advise capabilities are of limited utility
- recommendations can be annoying



	A	B	C	D	E	F
1						
2						
3		1	2	3	4	5
4		2	4	6	8	10
5		3	6	9	12	15
6		4	8	12	16	20
7		5	10	15	20	25
8		6	12	18	24	30
9		7	14	21	28	35
10		8	16	24	32	40
11		9	18	27	36	45
12		10				
13		11				
14		12				
15		13				
16		14				
17		15				
18		16				
19						
20						
21						
22						

What would you like to do?

- ● Project future values and perform regression analysis (Small Card)
- ● Change the way data is plotted (Small Card)
- ● Change chart labels, titles, and other text (Small Card)

Type your question here, and then click Search.

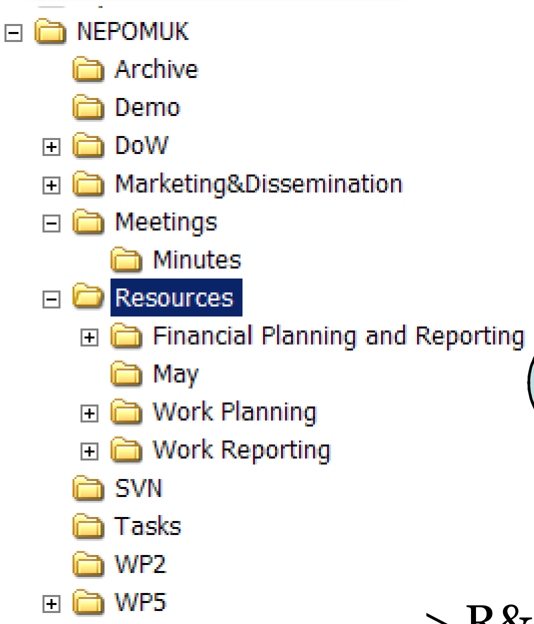
● Search

● Tips ● Options ● Close

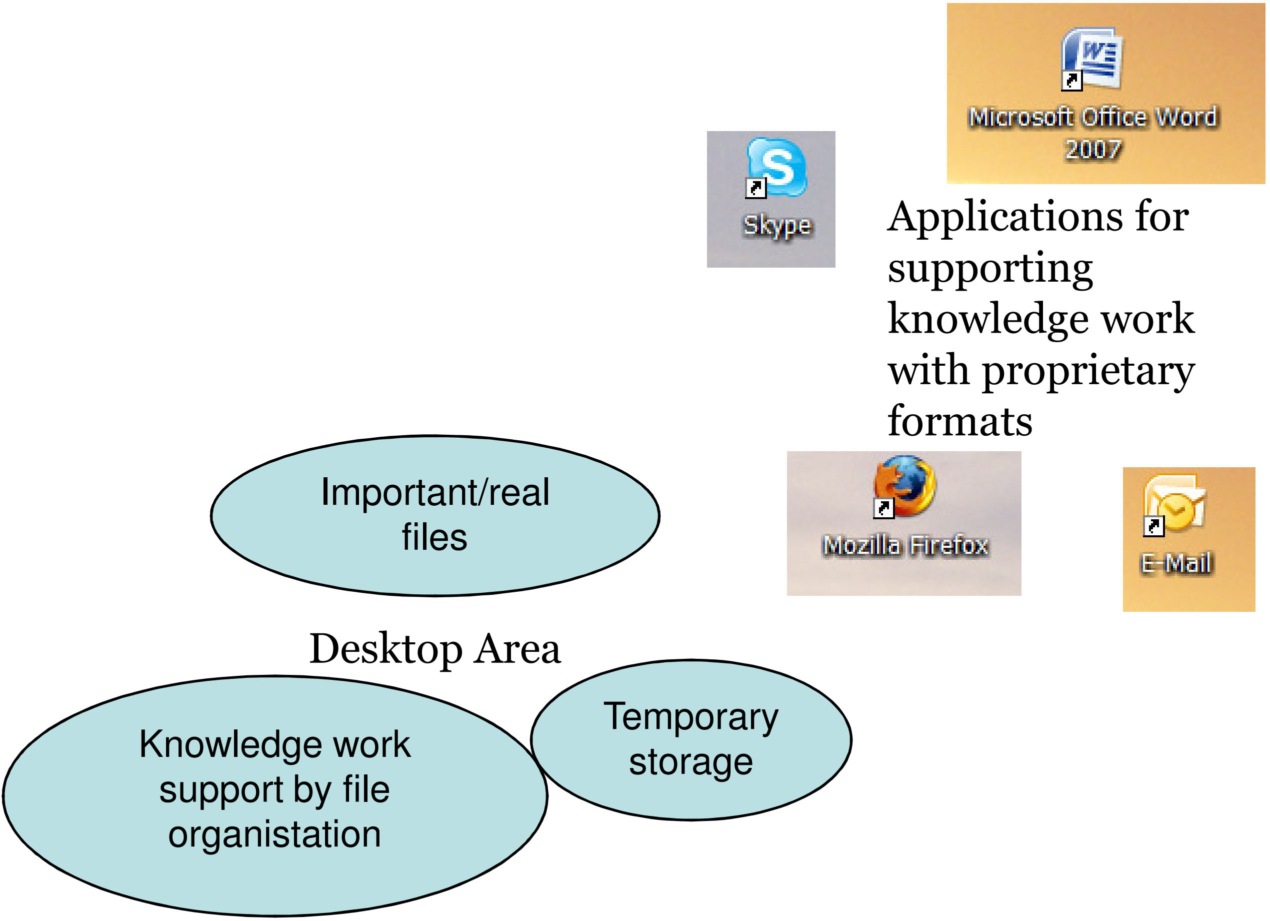
Nepomuk (1)

Current desktop

More or less organized folder hierarchy



- NEPOMUK
  - Archive
  - Demo
  - DoW
  - Marketing&Dissemination
  - Meetings
    - Minutes
  - Resources
    - Financial Planning and Reporting
    - May
    - Work Planning
    - Work Reporting
  - SVN
  - Tasks
  - WP2
  - WP5



Applications for supporting knowledge work with proprietary formats

Important/real files

Desktop Area

Knowledge work support by file organisation

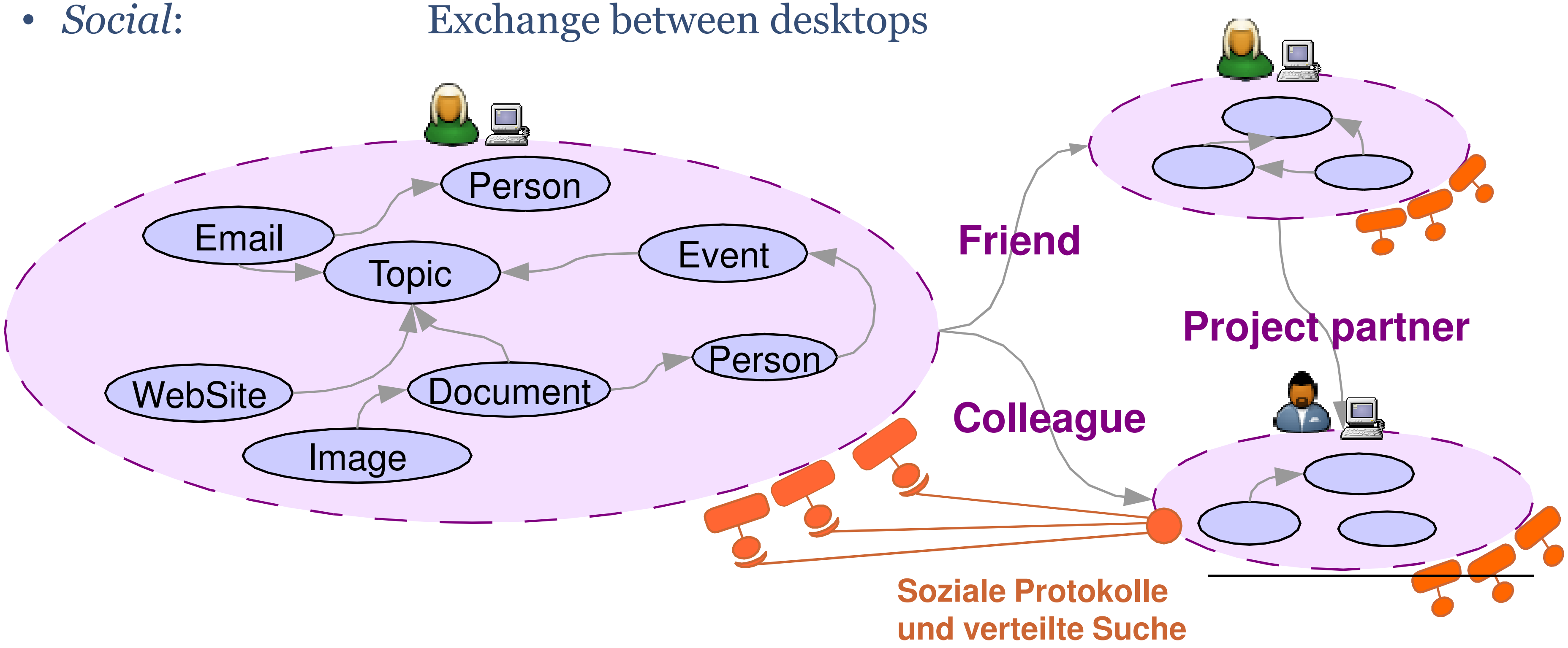
Temporary storage

-> R&D in Personal Information Management (PIM)

Nepomuk (2)

Desktop with Nepomuk

- *Semantic Desktop*: Information layer on top of the desktop content (personal semantic web) allowing machines to process information and provide intelligent services
- *Social*: Exchange between desktops



Nepomuk (3)

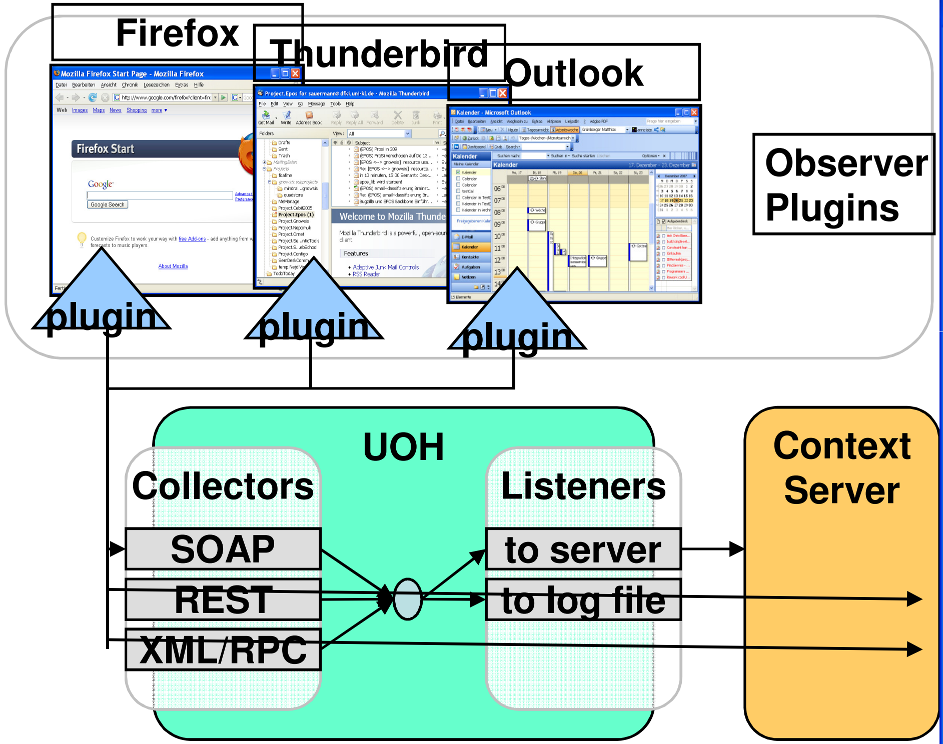
P. A. Chirita, J. Gaugaz, S. Costache, and W. Nejdl. Desktop context detection using implicit feedback. In PIM 2006.

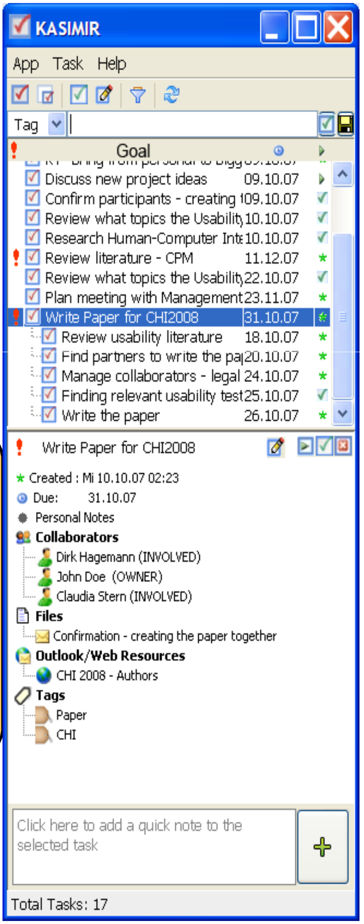
Goal:

- task-based document clustering

Tools:

- mixture of TFxIDF and K-Medoids clustering





Goal	
Discuss new project ideas	09.10.07
Confirm participants - creating t	09.10.07
Review what topics the Usability	10.10.07
Research Human-Computer Inte	10.10.07
Review literature - CPM	11.12.07
Review what topics the Usability	22.10.07
Plan meeting with Management	23.11.07
Write Paper for CHI2008	31.10.07
Review usability literature	18.10.07
Find partners to write the pa	20.10.07
Manage collaborators - legal	24.10.07
Finding relevant usability test	25.10.07
Write the paper	26.10.07

Write Paper for CHI2008

Created : Mi 10.10.07 02:23

Due: 31.10.07

Personal Notes

Collaborators

- Dirk Hagemann (INVOLVED)
- John Doe (OWNER)
- Claudia Stern (INVOLVED)

Files

- Confirmation - creating the paper together

Outlook/Web Resources

- CHI 2008 - Authors

Tags

- Paper
- CHI

Click here to add a quick note to the selected task

Total Tasks: 17

The final goal is

CONTEXT-AWARE INFORMATION RETRIEVAL

TaskTracer and TaskPredictor

J. Shen, L. Li, T. G. Dietterich, and J. L. Herlocker. A hybrid learning system for recognizing user tasks from desktop activities and email messages. In IUI'06

Goal:

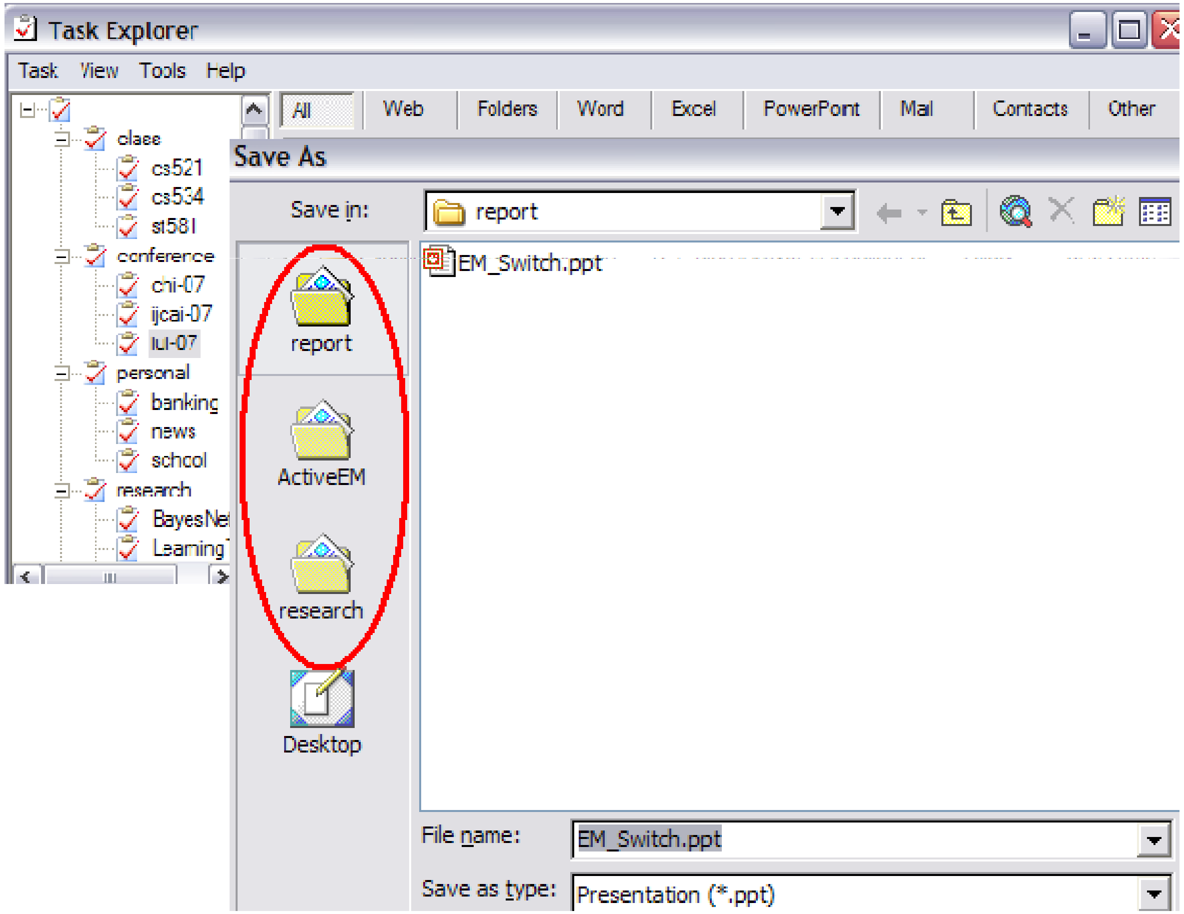
- **associate resources with user activities**

Tools:

- **adaptive file open/save dialog box**
- **Naïve Bayes/SVM classifiers for task prediction**

Lessons learned:

- **precision is about 80%**
- **data is very noisy, users forget to change a task**



Save in: report

EM_Switch.ppt

File name: EM_Switch.ppt

Save as type: Presentation (*.ppt)

SWISH

N. Oliver, G. Smith, C. Thakkar, and A. C. Surendran. Swish: semantic analysis of window titles and switching history. In IUI '06

Goal:

- task-based windows clustering for intelligent interfaces

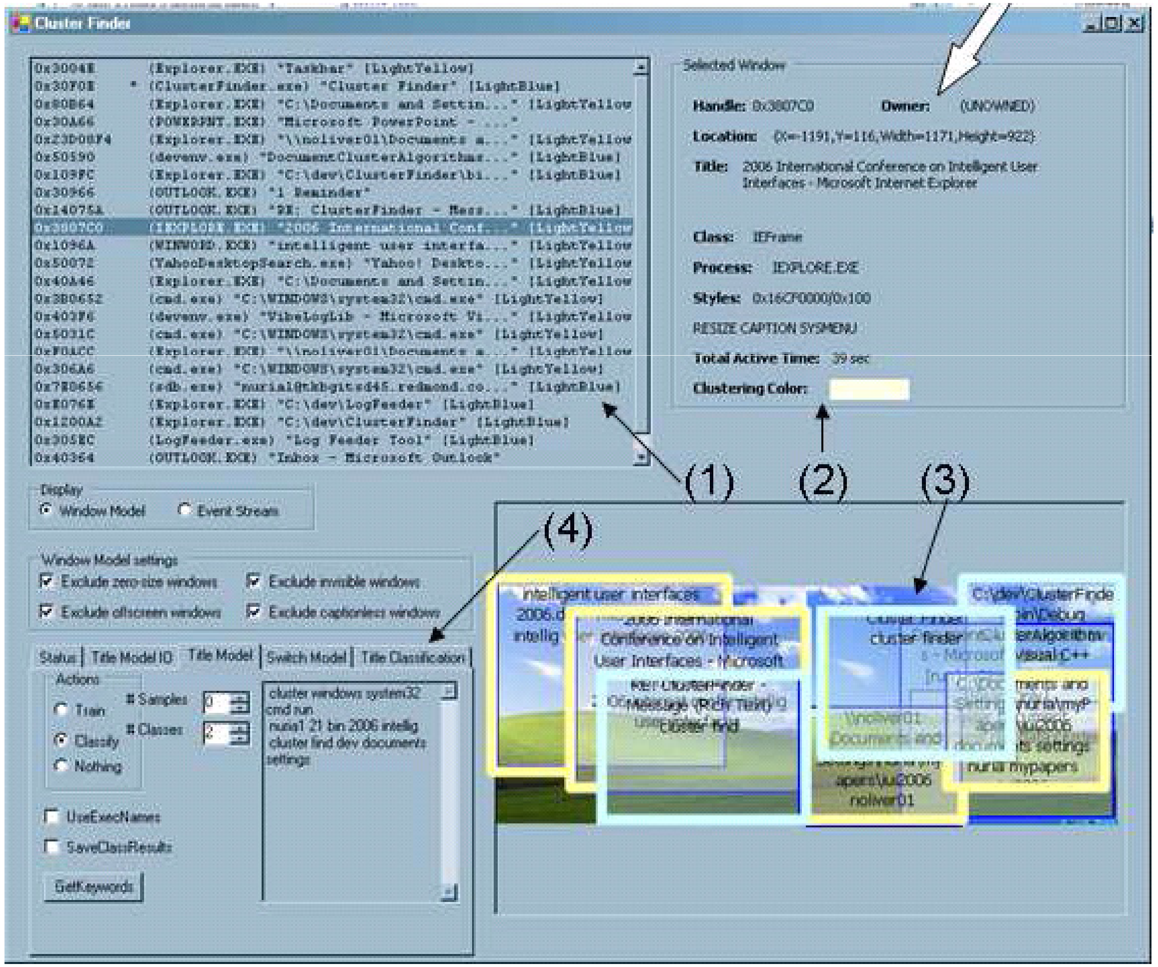
Tools:

- unsupervised learning: Probabilistic Latent Semantic Indexing

Lessons learned:

- precision is about 70%

- data is very noisy due to occasional windows' switches


Cluster Finder
(1)
(2)
(3)
(4)
intelligent user interfaces
2006 International Conference on Intelligent User Interfaces - Microsoft
Cluster Finder
cluster finder
C:\dev\ClusterFinder\bin\Debug
Visual C++
\\noliver01\Documents and Settings
nuria mypapers
papers\iui2006
noliver01

```
0x3004E      (Explorer.EXE) "Taskbar" [LightYellow]
0x30F0E    * (ClusterFinder.exe) "Cluster Finder" [LightBlue]
0x80B64      (Explorer.EXE) "C:\Documents and Settin..." [LightYellow
0x30A66      (POWERPNT.EXE) "Microsoft PowerPoint - ..."
0x23D00F4    (Explorer.EXE) "\\noliver01\Documents a..." [LightYellow
0x50590      (devenv.exe) "DocumentClusterAlgorithms..." [LightBlue]
0x109FC      (Explorer.EXE) "C:\dev\ClusterFinder\bi..." [LightBlue]
0x30966      (OUTLOOK.EXE) "1 Reminder"
0x14075A     (OUTLOOK.EXE) "RE: ClusterFinder - Mess..." [LightBlue]
0x3807C0     (IEXPLORE.EXE) "2006 International Conf..." [LightYellow
0x1096A      (WINWORD.EXE) "intelligent user interfa..." [LightYellow
0x50072      (YahooDesktopSearch.exe) "Yahoo! Deskto..." [LightYellow
0x40A46      (Explorer.EXE) "C:\Documents and Settin..." [LightYellow
0x3B0652     (cmd.exe) "C:\WINDOWS\system32\cmd.exe" [LightYellow]
0x403F6      (devenv.exe) "VibeLogLib - Microsoft Vi..." [LightYellow
0x6031C      (cmd.exe) "C:\WINDOWS\system32\cmd.exe" [LightYellow]
0xF0ACC      (Explorer.EXE) "\\noliver01\Documents a..." [LightYellow
0x306A6      (cmd.exe) "C:\WINDOWS\system32\cmd.exe" [LightYellow]
0x7E0656     (sdb.exe) "nuria1@tkbgitsd45.redmond.co..." [LightBlue]
0xE076E      (Explorer.EXE) "C:\dev\LogFeeder" [LightBlue]
0x1200A2     (Explorer.EXE) "C:\dev\ClusterFinder" [LightBlue]
0x305EC      (LogFeeder.exe) "Log Feeder Tool" [LightBlue]
0x40364      (OUTLOOK.EXE) "Inbox - Microsoft Outlook"
```

Selected Window

Handle: 0x3807C0 **Owner:** (UNOWNED)

Location: {X=-1191,Y=116,Width=1171,Height=922}

Title: 2006 International Conference on Intelligent User Interfaces - Microsoft Internet Explorer

Class: IEFrame

Process: IEXPLORE.EXE

Styles: 0x16CF0000/0x100

RESIZE CAPTION SYSMENU

Total Active Time: 39 sec

Clustering Color:

Display

● Window Model ○ Event Stream

Window Model settings

☑ Exclude zero-size windows ☑ Exclude invisible windows

☑ Exclude offscreen windows ☑ Exclude captionless windows

Status | Title Model IO | Title Model | Switch Model | Title Classification

Actions

○ Train

● Classify

○ Nothing

Samples 0

Classes 2

cluster windows system32
cmd run
nuria1 21 bin 2006 intellig
cluster find dev documents
settings

☐ UseExecNames

☐ SaveClassResults

GetKeywords

CAAD

T. Rattenbury and J. Canny. Caad: an automatic task support system. In CHI '07

Goal:

- task-based windows clustering

Tools:

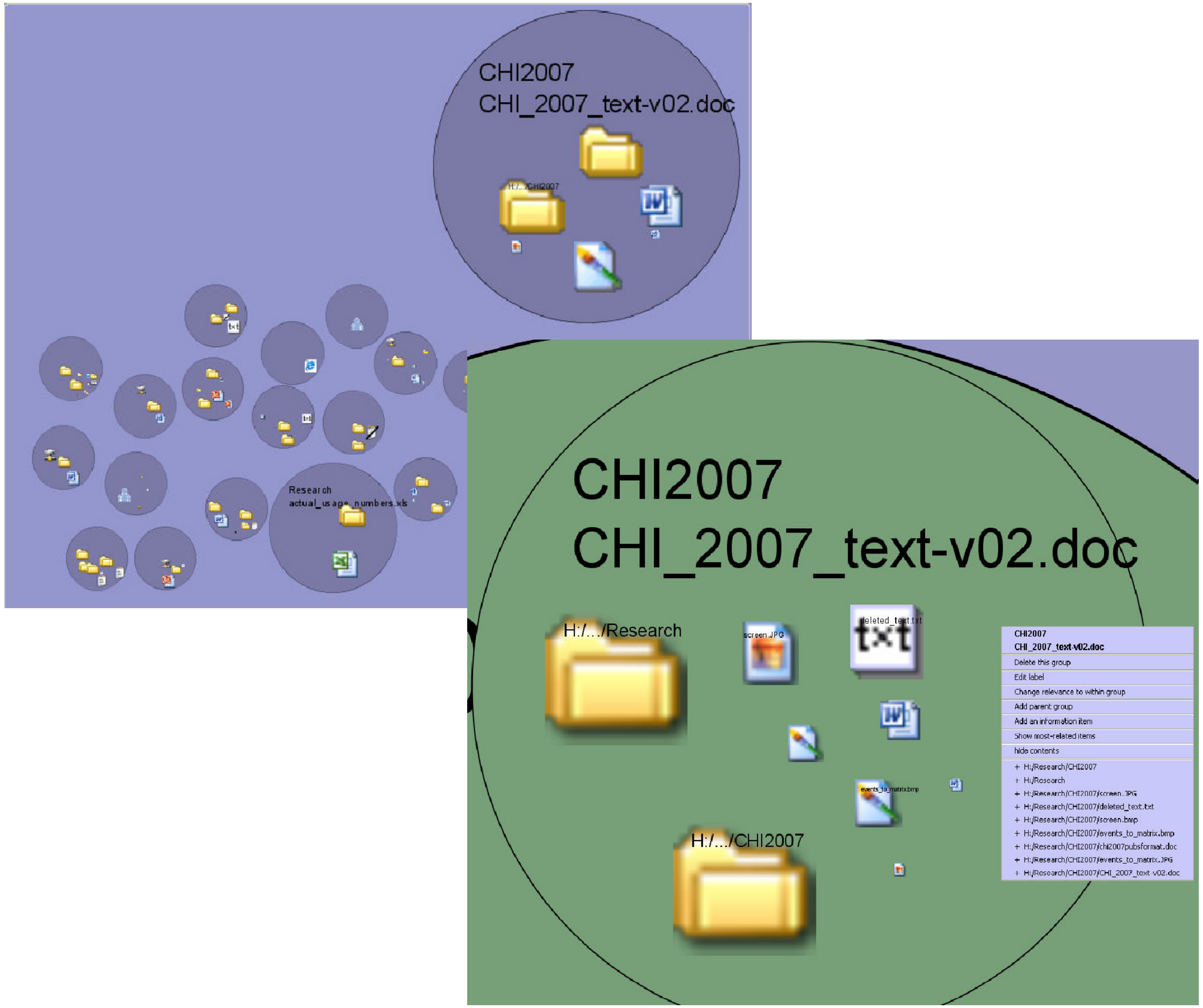
- GaP probabilistic model for Context Structures

- concatenated filenames for labels

Lessons learned:

- relevance is useless, if novelty is important or information changes quickly

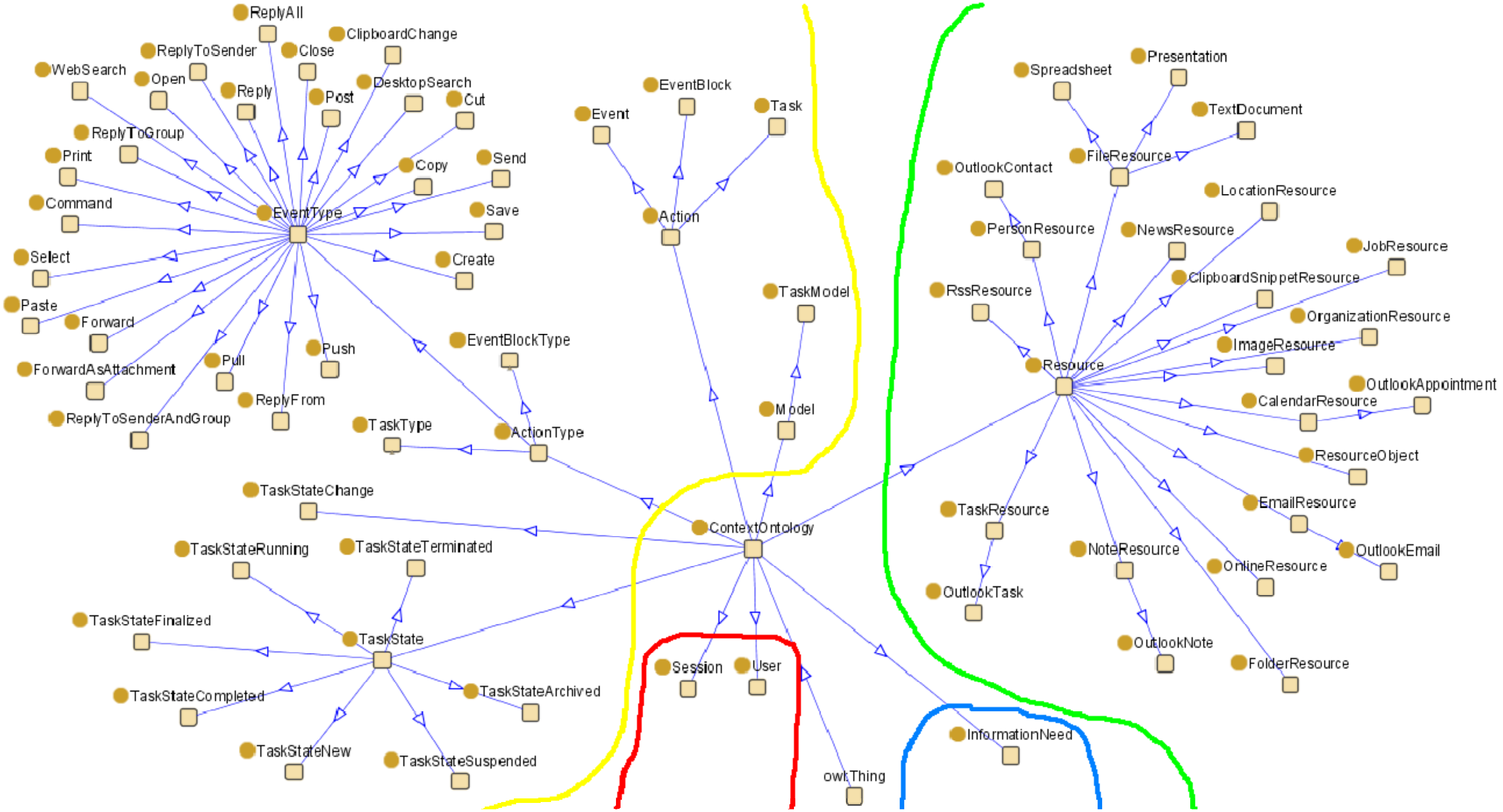
- user models are too broad or too narrow



UICO

UICO: An Ontology-Based User Interaction Context Model for Automatic Task Detection on the Computer Desktop. Andreas S. Rath, Didier Devaurs, Stefanie N. Lindstaedt. In CIAO 2009.

- Ontology-based user interaction context model (UICO) automatically derives relations between the model's entities and automatically detects the user's task



Current State

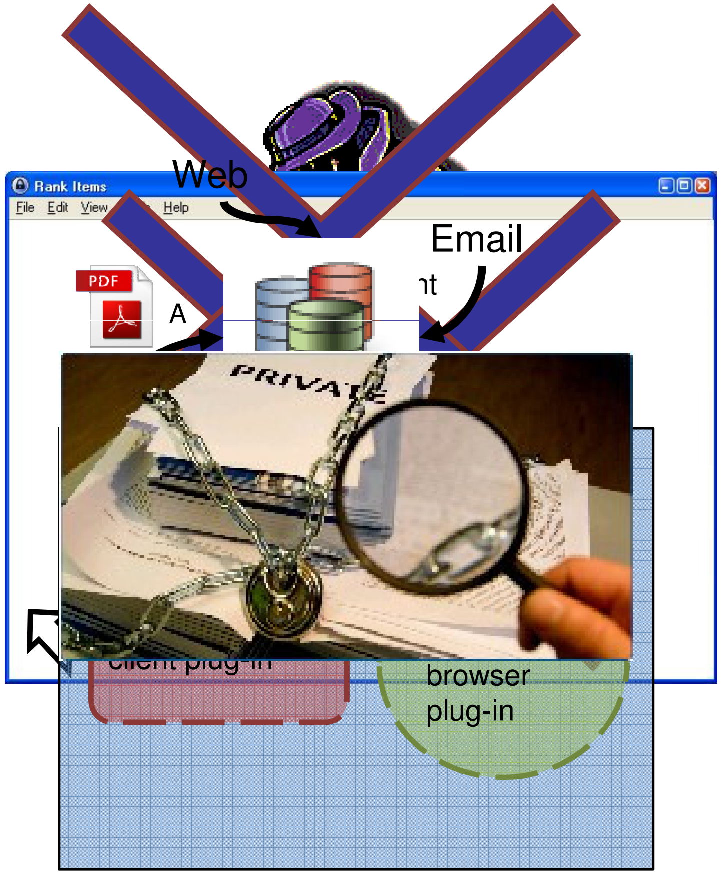
- Automatic Task Detection is under active development
  - most publications are within 2006-2009 time interval
  - no perfect solution so far

- Task Detection is based on machine learning
  - Naïve Bayes, PLSI, SVM

- Training data is missing
  - Activity-Logging can be used for data gathering

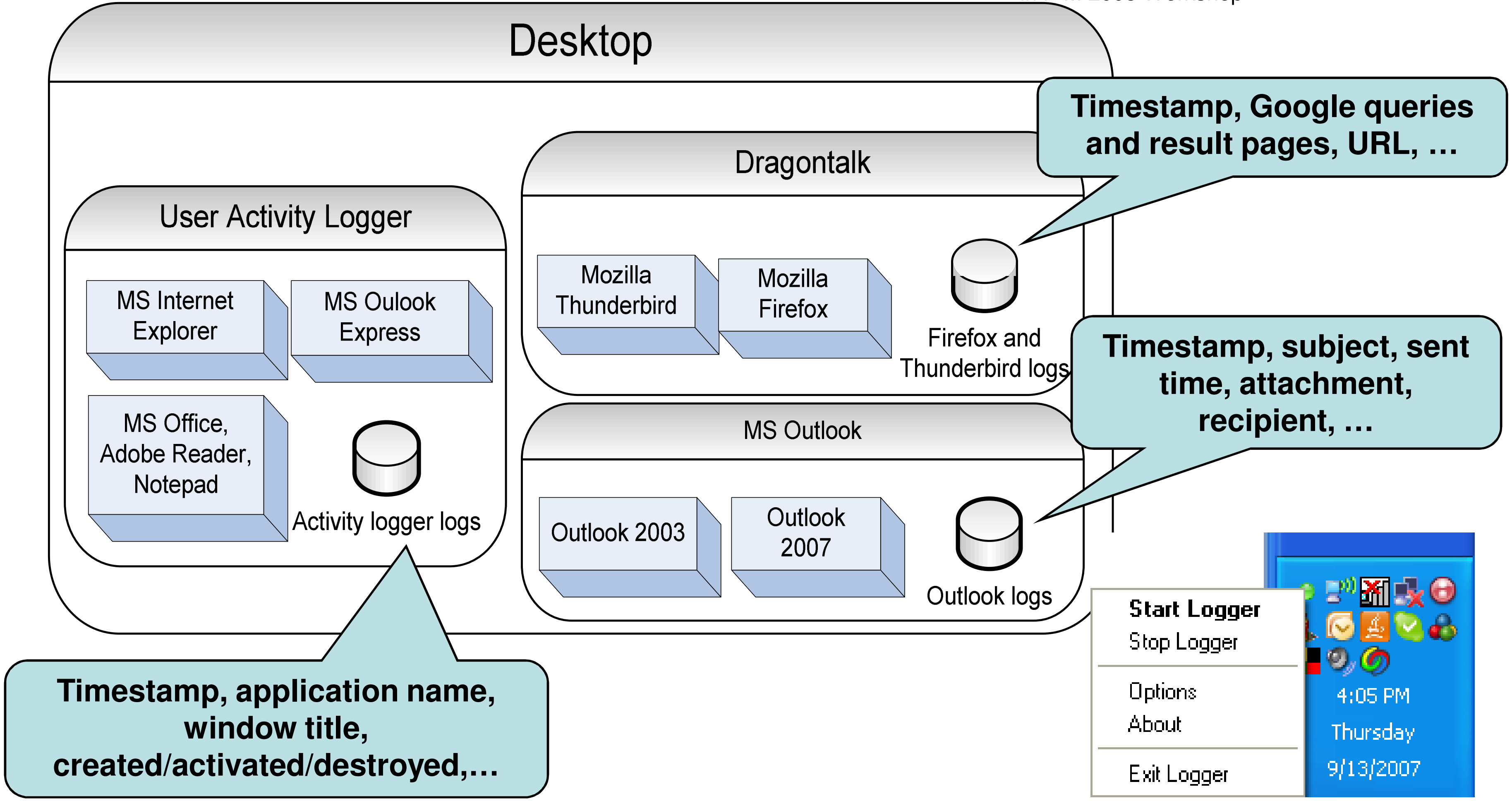
Towards Requirements for Logging Desktop

- **Automatic**
- **Cross-application**
- **Implicit Feedback**
- **Privacy preserving**
- **Extensible**



Desktop Logging Framework

Sergey Chernov, Gianluca Demartini, Eelco Herder, Michal Kopycki, and Wolfgang Nejdl. Evaluating Personal Information Management Using an Activity Logs Enriched Desktop Dataset in PIM 2008 Workshop



Desktop

User Activity Logger

MS Internet Explorer

MS Oulook Express

MS Office, Adobe Reader, Notepad

Activity logger logs

Dragontalk

Mozilla Thunderbird

Mozilla Firefox

Firefox and Thunderbird logs

MS Outlook

Outlook 2003

Outlook 2007

Outlook logs

Timestamp, Google queries and result pages, URL, …

Timestamp, subject, sent time, attachment, recipient, …

Timestamp, application name, window title, created/activated/destroyed,…

Supported notifications

Notification
W
D
Ic
H
L

N	
T	
B	
F	
P	
Submit Web form	Firefox

Email	
E	
Email (receive, reply, delete, move, print)	Tl
Address book entry (create, modify, delete)	Tl
Email folder (create, rename, delete)	Tl

Instant Messen	
Conversation (start, active, finish)	MSN,

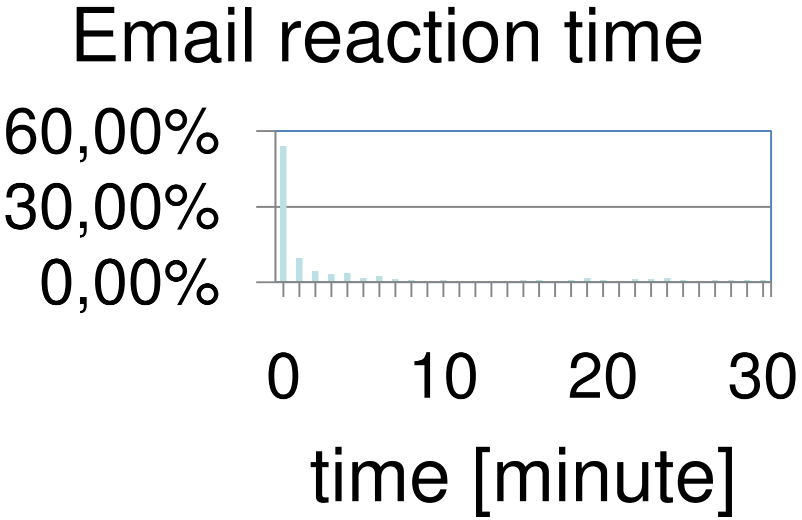
Collected Data

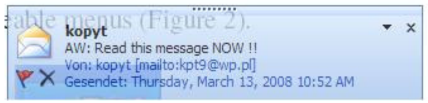
- 21 participants
- Average of 170 active logging days
- 2,828,706 Events
- Average of 2,815 distinct emails per user
- Average of 9,337 distinct URLs per user
- Average of 902 events per user per day
- Average 5 hours of active interaction per user per day

A glimpse into user behavior (1)

Sergey Chernov, Gianluca Demartini, Eelco Herder, Michal Kopycki, and Wolfgang Nejdl. Evaluating Personal Information Management Using an Activity Logs Enriched Desktop Dataset in PIM 2008 Workshop

Instant reader

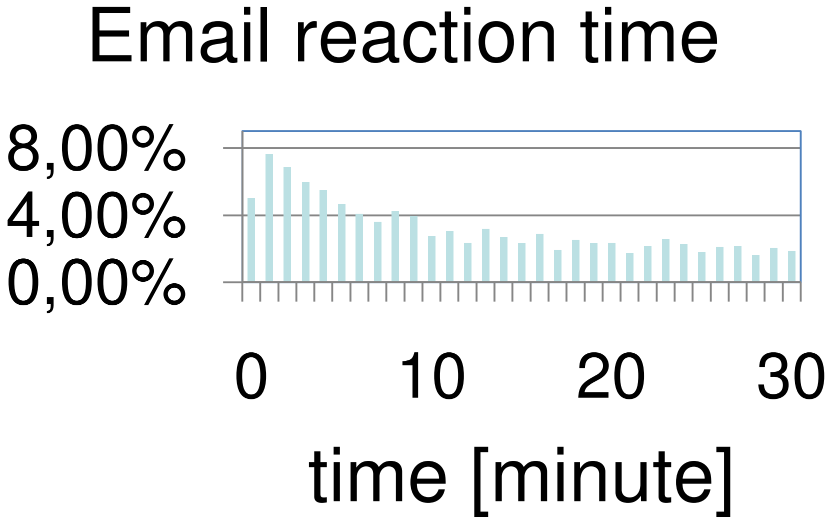


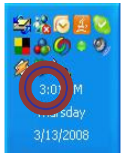

able menus (Figure 2).

kopyt
AW: Read this message NOW !!
Von: kopyt [mailto:kpt9@wp.pl]
Gesendet: Thursday, March 13, 2008 10:52 AM

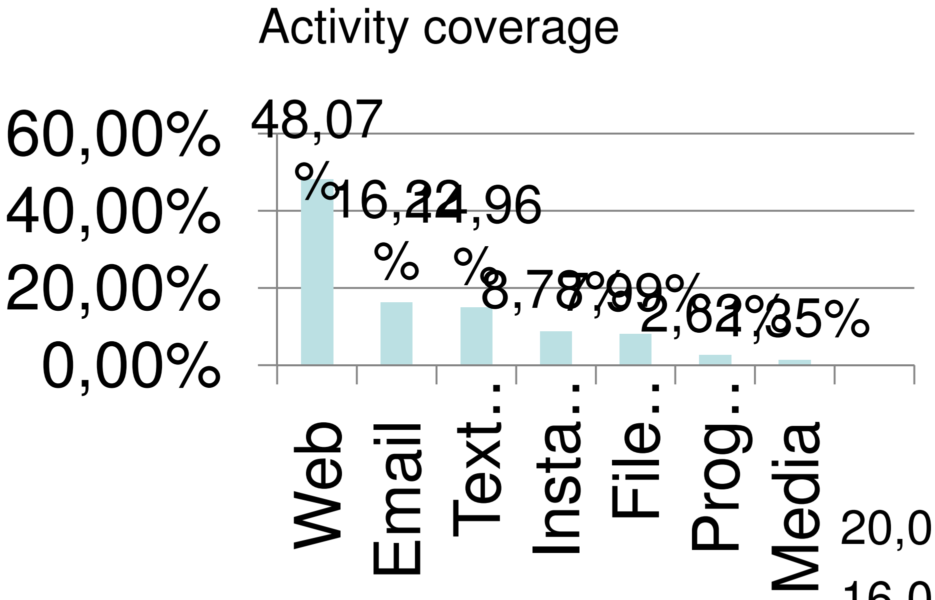
11:31 AM

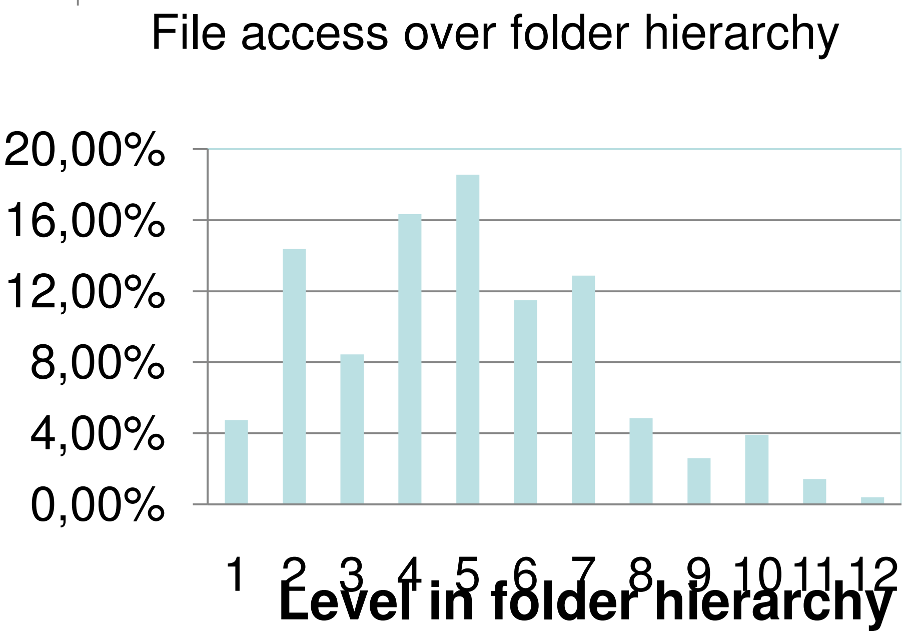
Moderate reader




3:0 M
rsday
3/13/2008

A glimpse into user behavior (2)





Evaluation

Understanding What Works: Evaluating PIM Tools. Diane Kelly and Jaime Teevan. In “Personal Information Management” edited by William Jones and Jaime Teevan, 2008.

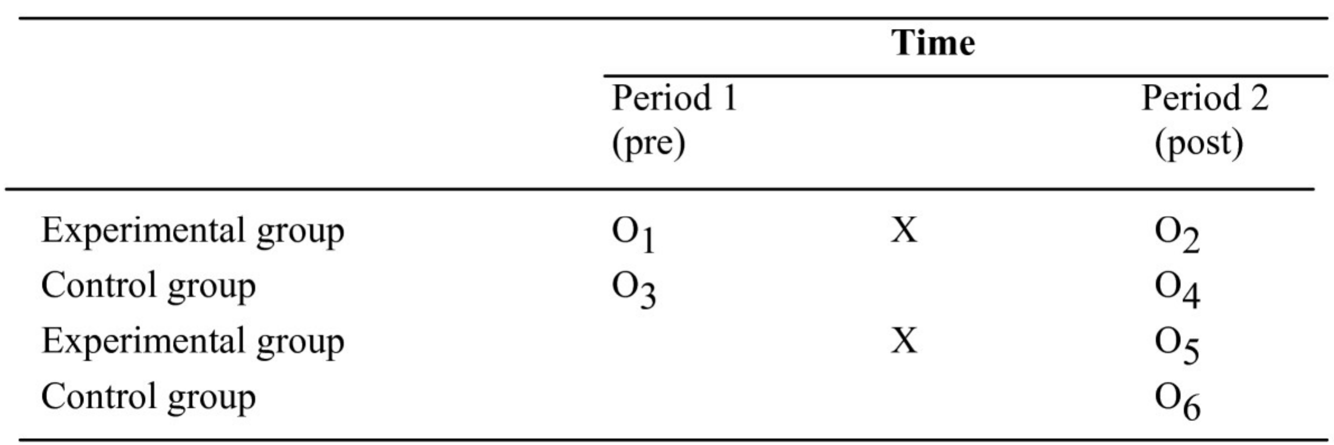
- Evaluation frameworks:
  - Naturalistic (one-time evaluation in a natural environment with own data)
  - Longitudinal (studies over extended period of time with measurements at fixed points)
  - Case study (in-depth picture of few individuals behavior)
  - Laboratory (controlled scenarios)
- Could and should be combined with each other
- Challenges:
  - Lack of control over environment (unpredictable interactions)
  - Appropriate time intervals and study duration
  - Narrow scope of evaluation task

Evaluation Components: Participants, Collections, Tasks

- Participants
  - Compared to Web Search: harder to recruite, data is too sensitive, prototype must be more robust, more involvement is required, limited generalization, using “personas” – simulated users
- Collections
  - Users should provide their own data, it is a mixture of documents, photos, emails, contacts, etc.
- Tasks
  - Tasks are broad, user-centric and situation-specific
  - Different granularity level (doing email vs. search for a piece of text in email)
  - Different types of tasks (planning a travel, reading the news, finding information about X)

Evaluation Components: Baselines

– Solomon four group design



	Time		
	Period 1 (pre)		Period 2 (post)
Experimental group	O_1	X	O_2
Control group	O_3		O_4
Experimental group		X	O_5
Control group			O_6

– O: Observation. X: Intervention

– Caveat: *Trained Incapacity* – users create unique ways of using tools that the original designers may not have intended.

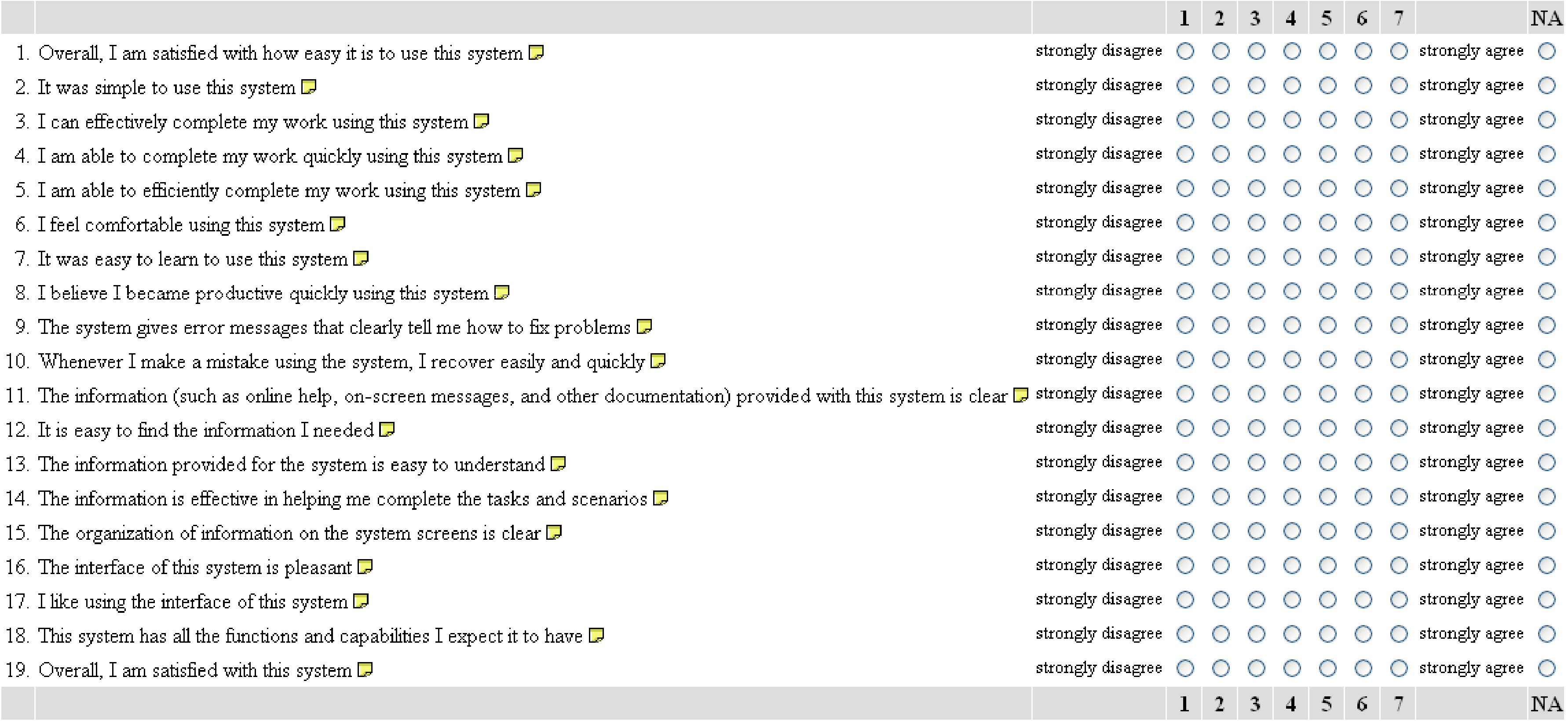
Evaluation Components: Measures

- Measures could be defined in two ways:
  - Nominal – what is it? (Learnability is defined by a grade on a 5-point Likert scale)
  - Operational – how exactly it should be measured? (Learnability is a length of time it takes for a user to learn to use an interface)

- Standard usability measures:
  - Effectiveness, Efficiency, Satisfaction, Usefulness, Ease of use, Ease of learning

- Usability measures in PIM context:
  - Performance (recall/precision), Adoption and Use, Flow, Quality of Life

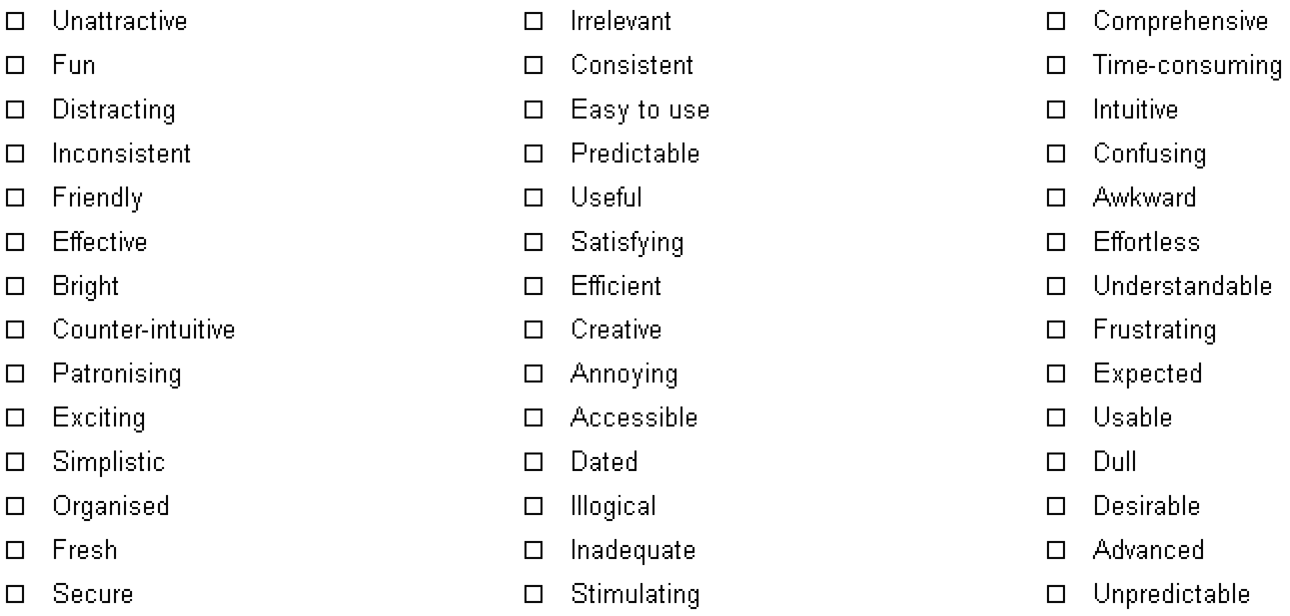
Usability Questionnaire Example 1



			1	2	3	4	5	6	7		NA
1.	Overall, I am satisfied with how easy it is to use this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
2.	It was simple to use this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
3.	I can effectively complete my work using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
4.	I am able to complete my work quickly using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
5.	I am able to efficiently complete my work using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
6.	I feel comfortable using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
7.	It was easy to learn to use this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
8.	I believe I became productive quickly using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
9.	The system gives error messages that clearly tell me how to fix problems	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
10.	Whenever I make a mistake using the system, I recover easily and quickly	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
11.	The information (such as online help, on-screen messages, and other documentation) provided with this system is clear	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
12.	It is easy to find the information I needed	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
13.	The information provided for the system is easy to understand	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
14.	The information is effective in helping me complete the tasks and scenarios	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
15.	The organization of information on the system screens is clear	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
16.	The interface of this system is pleasant	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
17.	I like using the interface of this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
18.	This system has all the functions and capabilities I expect it to have	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
19.	Overall, I am satisfied with this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
			1	2	3	4	5	6	7		NA

Usability Questionnaire Example 2

Step 1: Read over the following list of words. Considering the product you have just used, tick those words that best describe your experience with it. You can choose as many words as you wish.



- □ Unattractive
- □ Fun
- □ Distracting
- □ Inconsistent
- □ Friendly
- □ Effective
- □ Bright
- □ Counter-intuitive
- □ Patronising
- □ Exciting
- □ Simplistic
- □ Organised
- □ Fresh
- □ Secure
- □ Irrelevant
- □ Consistent
- □ Easy to use
- □ Predictable
- □ Useful
- □ Satisfying
- □ Efficient
- □ Creative
- □ Annoying
- □ Accessible
- □ Dated
- □ Illogical
- □ Inadequate
- □ Stimulating
- □ Comprehensive
- □ Time-consuming
- □ Intuitive
- □ Confusing
- □ Awkward
- □ Effortless
- □ Understandable
- □ Frustrating
- □ Expected
- □ Usable
- □ Dull
- □ Desirable
- □ Advanced
- □ Unpredictable

Step 2: Now look at the words you have ticked. Circle five of these words that you think are most descriptive of the product.

Summary and Challenges

- Desktop Search research just started ☺
- Main future directions are:
  - Logging of user activities and creating context-aware DS
  - Integration of metadata and fulltext search in personal repositories
  - Building social semantic desktop - collaboration, recommendation and knowledge sharing functionalities should extend basic information access on the desktop
  - Better understanding of user needs
  - Seamless integration of search and browsing behavior

We are hiring!

- Relevant Areas
  - Search and Information Retrieval
  - Information and Concept Extraction
  - Data Mining and Statistical Analysis
  - User Interface Engineering and Interaction Design
  - Semantic Technologies and Web 2.0
  - Multimodal Communication and Analysis
  - Social Software for Technology Enhanced Learning

- Phd and PostDoc positions
  - See handouts or http://www.l3s.de/web/page23g.do

- 6-months internships for Master Students
  - Send your CV (1-3 pages) and Research Statement (1-2 pages) to Prof. Wolfgang Nejdl (nejdl@L3S.de) or most relevant person from L3S

  - Further questions – come and ask now or write to chernov@L3S.de

References: Research DS prototypes

- A layered framework supporting personal information integration and application design for the semantic desktop, Isabel F. Cruz, Huiyong Xiao. In VLDB Journal 2008.

- S. Dumais, E. Cutrell, J. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins. Stuff i've seen: a system for personal information retrieval and re-use. In SIGIR 2003.

- E. Cutrell, D. Robbins, S. Dumais, and R. Sarin. Fast, Flexible Filtering with phlat. In CHI 2006.

- P.-A. Chirita, S. Costache, W. Nejdl, and R. Paiu. Beagle++ : Semantically enhanced searching and ranking on the desktop. In ESWC 2006.

- Semantically Rich Recommendations in Social Networks for Sharing, Exchanging and Ranking Semantic Context, Stefania Ghita, Wolfgang Nejdl, and Raluca Paiu. In ISWC 2005.

- The Beagle++ Toolbox: Towards an Extendable Desktop Search Architecture, Ingo Brunkhorst, Paul - Alexandru Chirita, Stefania Costache, Julien Gaugaz, Ekaterini Ioannou, Tereza Iofciu, Enrico Minack, Wolfgang Nejdl and Raluca Paiu. Technical Report 2006.

References: Just-In-Time Retrieval

- J. Budzik and K. J. Hammond. User interactions with everyday applications as context for just-in-time information access. In IUI 2000.

- Rhodes, B. and Starner, T. The Remembrance Agent: A continuously running information retrieval system. In PAAM 1996.

- B. J. Rhodes. Just-in-time information retrieval. PhD thesis, 2000.

- Rhodes, B., The Wearable Remembrance Agent: a system for augmented memory. in Personal Technologies: Special Issue on Wearable Computing, 1997.

References: Context-based DS

- C. A. N. Soules and G. R. Ganger. Connections: using context to enhance file search. In SOSP 2005.

- K. A. Gyllstrom, C. Soules, and A. Veitch. Confluence: enhancing contextual desktop search. In SIGIR 2007.

- Activity put in context: Identifying implicit task context within the user’s document interaction, Karl Gyllstrom, Craig Soules, Alistair Veitch. In IIiX 2008.

- K. Gyllstrom and C. Soules. Seeing is retrieving: Building information context from what the user sees. In IUI 2008.

- Analyzing User Behavior to Rank Desktop Items. Paul-Alexandru Chirita, Wolfgang Nejdl. In SPIRE 2006.

References: Context Detection Tools

- E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelse. The lumiere project: Bayesian user modeling for inferring the goals and needs of soft. In UAI 1998.

- P. A. Chirita, J. Gaugaz, S. Costache, and W. Nejdl. Desktop context detection using implicit feedback. In PIM 2006.

- J. Shen, L. Li, T. G. Dietterich, and J. L. Herlocker. A hybrid learning system for recognizing user tasks from desktop activities and email messages. In IUI 2006

- N. Oliver, G. Smith, C. Thakkar, and A. C. Surendran. Swish: semantic analysis of window titles and switching history. In IUI '06

- T. Rattenbury and J. Canny. Caad: an automatic task support system. In CHI 2007.

- UICO: An Ontology-Based User Interaction Context Model for Automatic Task Detection on the Computer Desktop. Andreas S. Rath, Didier Devaurs, Stefanie N. Lindstaedt. In CIAO 2009.

- Sergey Chernov, Gianluca Demartini, Eelco Herder, Michal Kopycki, and Wolfgang Nejdl. Evaluating Personal Information Management Using an Activity Logs Enriched Desktop Dataset. In PIM 2008.