



Intelligente Suchtechniken für Personal Information Management

Wissenschaftlicher Vortrag
M.Sc. Zoltán Fiala

Technische Universität Dresden
Fakultät Informatik

Dresden, den 19.02.2007

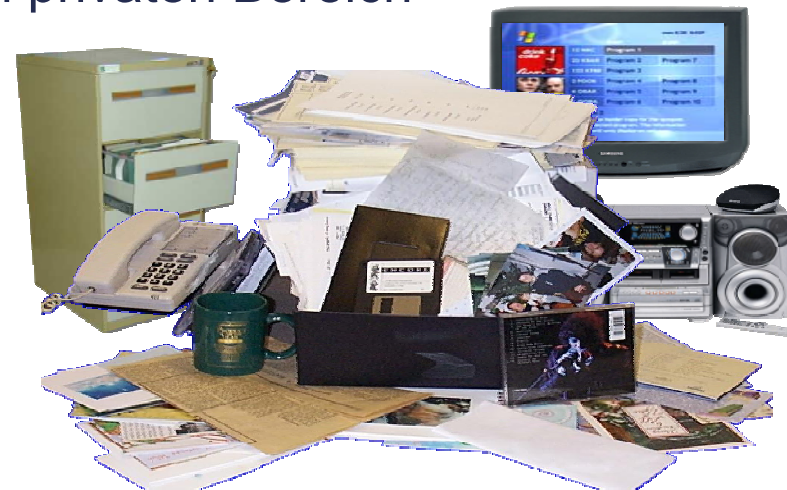


Gliederung

- Personal Information Management (PIM)
 - Besonderheiten der Suche in PIM-Systemen
- Suchverfahren und -systeme für PIM
 - Semantic Desktop Systeme
 - Neuartige User Interfaces für PIM-Suche
- Aktuelle Herausforderungen
- Zusammenfassung und Ausblick

Motivation

- Zunehmende Digitalisierung unseres privaten Lebens
- Verbreitung von Computern, Aufnahme- und Abspielgeräten
- Aktuelle Speichergeräte: steigende Kapazitäten, sinkende Preise
- Stetig wachsende Datenmengen im privaten Bereich
 - *Office-Daten:*
 - Emails, Formulare, Verträge etc.
 - *Kommunikation:*
 - Instant Messaging, Telefonate
 - *Private Medienkollektionen:*
 - Fotos, Musik, Videos
- MyLifeBits [Gemm+06]: pers. Archive im dreistelligen GB-Bereich
- Forderung nach effektiver Verwaltung persönlicher Daten



[Gemm+06]

Personal Information Management (PIM)

- Personal Information Management
 - Als Begriff erstmals 1988 von Lansdale genannt [Lans 88]
 - Subsummiert Tätigkeiten um die Verwaltung persönlicher Daten

“Personal Information Management (PIM) is intended to support the activities we, as individuals, perform to order our daily lives through the acquisition, organization, maintenance, retrieval, and sharing of information.” [Teev+06]
 - Die wichtigsten PIM-Aktivitäten nach Barreau [Barr 95]
 1. Akquisition von PI
 2. Organisation, Ablage, Speicherung
 3. Verwaltung (Backup, Archivierung)
 4. Suche und Retrieval
 5. Ausgabe, Präsentation, Verwendung

Besonderheiten der PIM-Suche

- Suchziele in klassischen „*Information Storage And Retrieval*“ (ISAR) Systemen nach Rose und Levinson [Rose+04]
 - **Navigationsuche:** der Nutzer hat ein bestimmtes Dokument im Kopf, dessen Ort er allerdings vergessen hat
 - **Informationssuche:** Suche nach neuen Informationen in einem bestimmten Thema (Lernen, Lokalisieren, Frage beantworten)
 - **Ressourcensuche:** Zugriff auf eine „nicht-informative“ Ressource bzw. Dienst (Download, Online-Banking)
- PIM-Suche
 - Fokus auf bekannte Objekte: „search into our past“ [Chir+05]
 - in erster Linie *Navigations-* bzw. *Ressourcensuche*
 - „ich weiss, dass ich das Dokument irgendwo hingelegt habe, aber ich weiss nicht genau wo“

Besonderheiten der PIM-Suche

- Geringere Datenmengen im persönlichen Bereich
- Information Fragmentation
 - Gesuchte Ressourcen über verschiedene Applikationen verstreut
 - Oft eigene Datenschemata, Verwaltungs- und Suchfunktionen
 - Parallele Hierarchien von Verzeichnissen, Mails, Aufgaben
- Keine professionellen Nutzer
- PI sind vor allem persönlich!
 - Suche nach Informationen, die primär vom Suchenden selbst abgespeichert wurden
 - Reiche assoziative und kontextuelle Informationen vorhanden [Cutr+06, Berg+03]

Besonderheiten der PIM-Suche

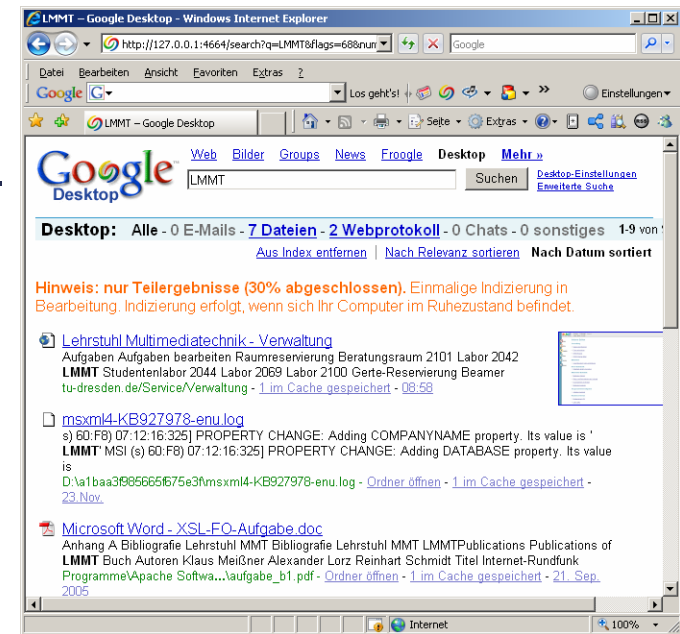
- PIM bzw. PIM-Suche ist stark durch Kontextualität geprägt
- Man möchte Informationen in demselben Kontext wiederfinden, in dem man sie vorher genutzt hat [Berg+03]

“Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application ...” [Dey01]

- Wichtigste Kontextinformationen: Personen, Zeit, Ort und Aktivitäten (Projekte), Situationen
 - „Ich suche Emails, die ich von meinem *Chef* während der *Fußball-WM* bekommen habe“
 - „Die gesuchte Datei habe ich in *Ungarn* auf der *WWW Tagung* erstellt und später zusammen mit meinem *Kollegen X* editiert.“

Existierende PIM-Suchsysteme

- Desktop Search Systeme
 - Von allen namhaften Suchmaschinen-Herstellern angeboten
 - Google Desktop, Yahoo Search, Windows Desktop Search
 - Auf Text-basierte Suche beschränkt
 - Kaum Verwendung von Metadaten
 - Eingeschränkte Filtermöglichkeiten, wenig Interaktivität
 - Fehlende Kontextinformationen und Assoziationen
 - Nicht integriert mit anderen PIM-Systemen



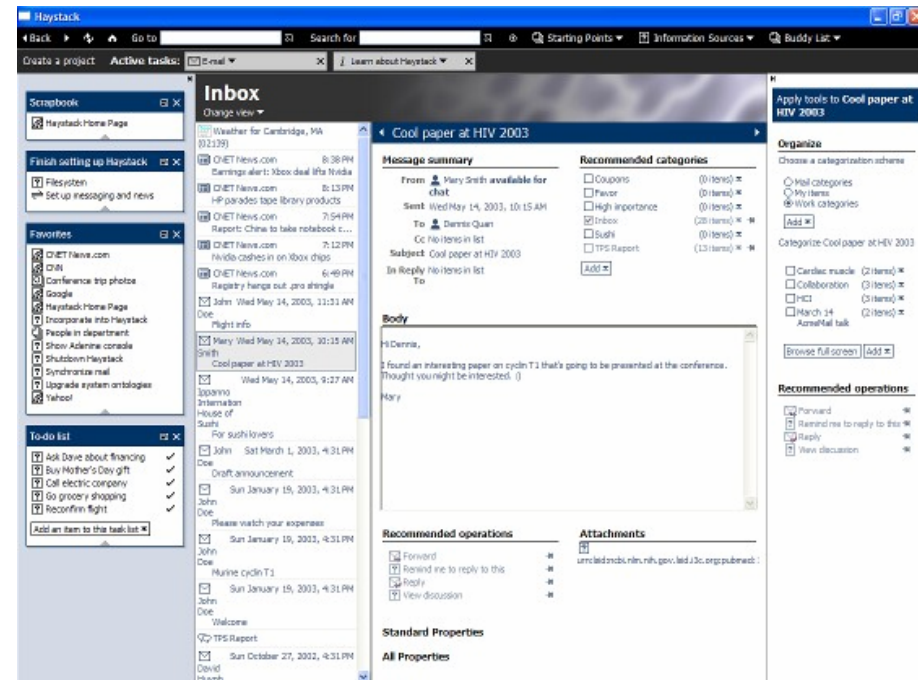
Semantic Desktops

■ „Semantic Desktop“ Lösungen

- Extrahieren, verwalten, visualisieren und verwerten semantische und kontextuelle Assoziationen bzw. Metadaten für PI
- Kombinieren Desktop-Systeme mit Semantic Web Technologien
- Entweder eigenständige PIM-Systeme oder als PIM-Middleware

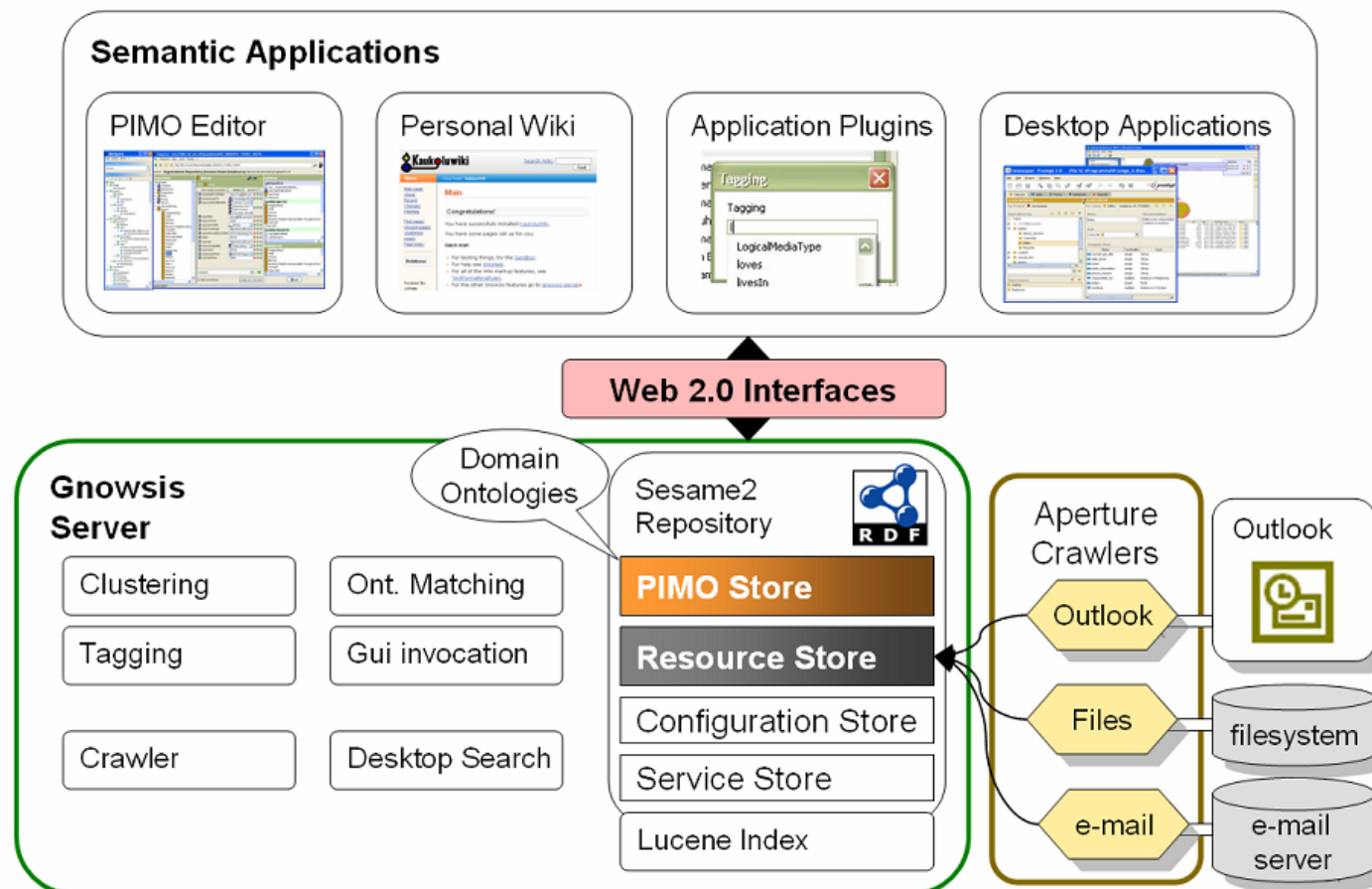
■ Existierende Lösungen

- Gnowsisis [Saue+06]
- Beagle++ [Chir+05]
- Social Semantic Desktop [Deck+04]
- Haystack → [Karg+05]



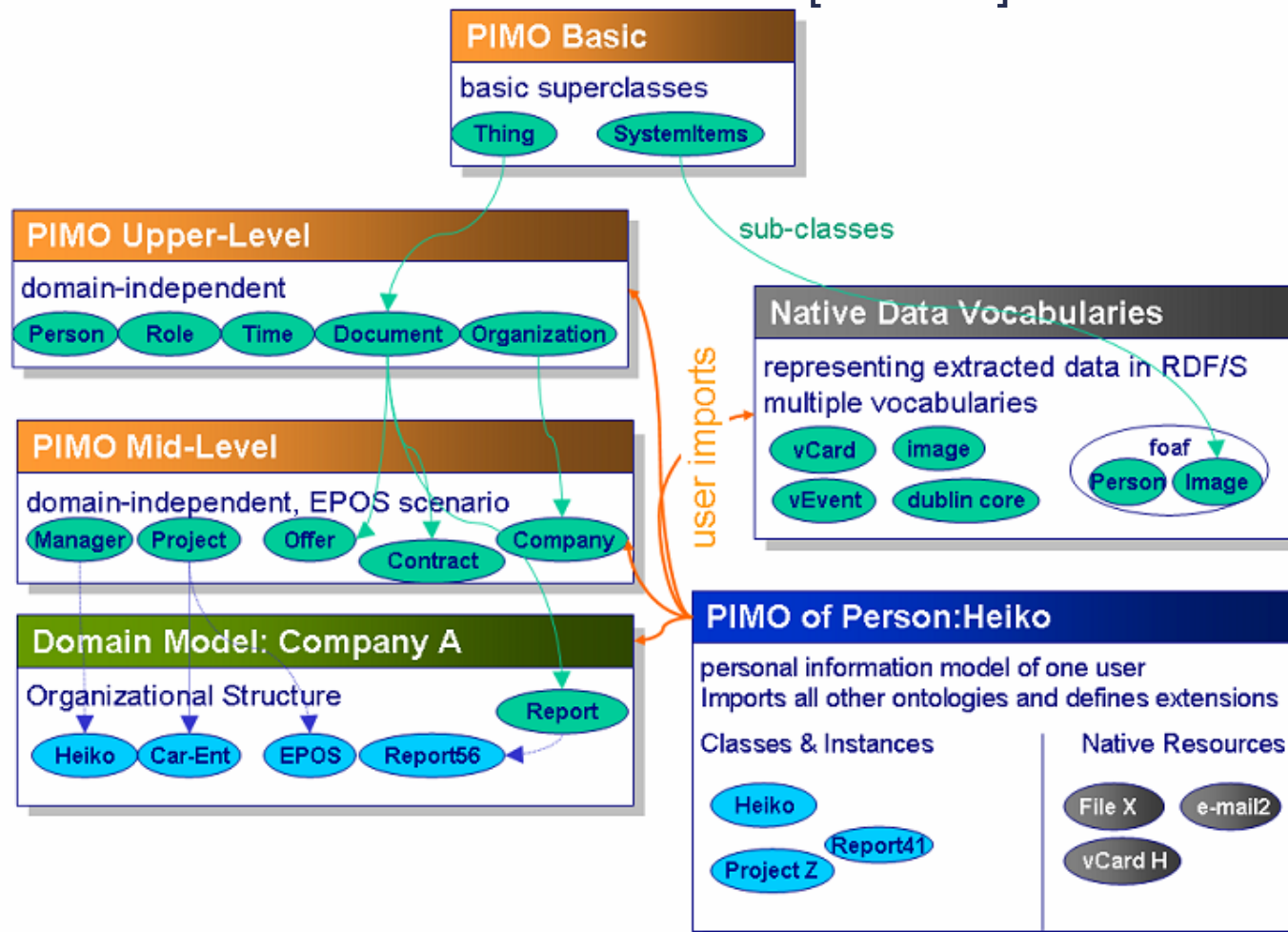
Semantic Desktops – Architektur

- Typische Semantic Desktop Architektur
 - Gnowsis Semantic Desktop [Saue+06]



Semantic Desktops – Datenmodellierung

- Semantische Modellierung von PI
 - PIMO: Personal Information Model [Saue06]



Semantic Desktops – Datenakquisition

- Medien- bzw. anwendungsspezifische Metadaten über Filter
- Explizit vom Nutzer eingegeben (Tagging)
- Zusätzliche Metadaten von Tagging-Sites
 - Flickr, del.icio.us oder BibSonomy
- Weitere kontextuelle Assoziationen ableitbar
 - Dokumente, die immer nacheinander bearbeitet werden
 - Dokumente im selben Ordner oder mit gleichen Namen
- Reasoning-Mechanismen des Semantic Web
 - Z.B. die Association Rules von Beagle++ [Chir+05]



```
from(File, Name) ← stored_as(Attachment, File),  
has_attachment(Mail, Attachment),  
from(Mail, MailAddress),  
belongs_to(MailAddress, Person), name(Person, Name)
```

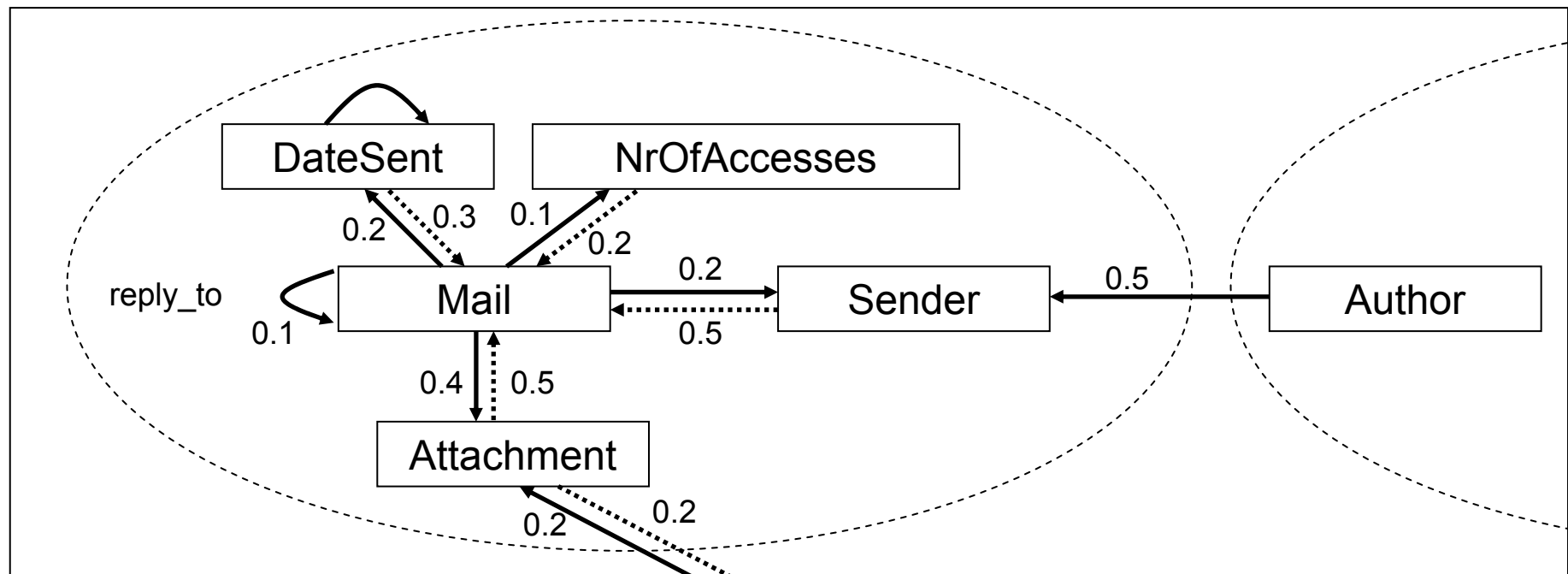
Semantic Desktops – Suchmöglichkeiten

- Suche über Text, Merkmale, Metadaten, Assoziationen
- Kombination strukturierter Anfragesprachen und IR-Verfahren
- Semantische Datenbanken und Frameworks
 - Sesame, KAON, Jena Semantic Web Framework
 - Mächtige Query-Sprachen: RQL, SeRQL, SPARQL
- Merkmalsuche in medialen Daten
 - Bis auf Bildsuche im Desktop-Bereich wenig verbreitet [Yee+03]
- Volltextsuche in Dokumenten und Metadaten
 - Automatische Clusteringverfahren
 - Thesaurus-/Synonymensuche
 - Meistens Volltextsuche über invertierte Indexlisten
 - Gewichtung typischerweise über Termhäufigkeit



Semantic Desktops – Ranking

- Gewichtung bzw. Ranking
 - Inhaltsbasierte Gewichtung gibt keinen Aufschluß über Assoziationen
 - Kontext-basiertes Ranking als Erfolgsfaktor von Suchmaschinen
 - Beagle++: Kontextbasiertes Ranking mit ObjectRank [Chir+05]

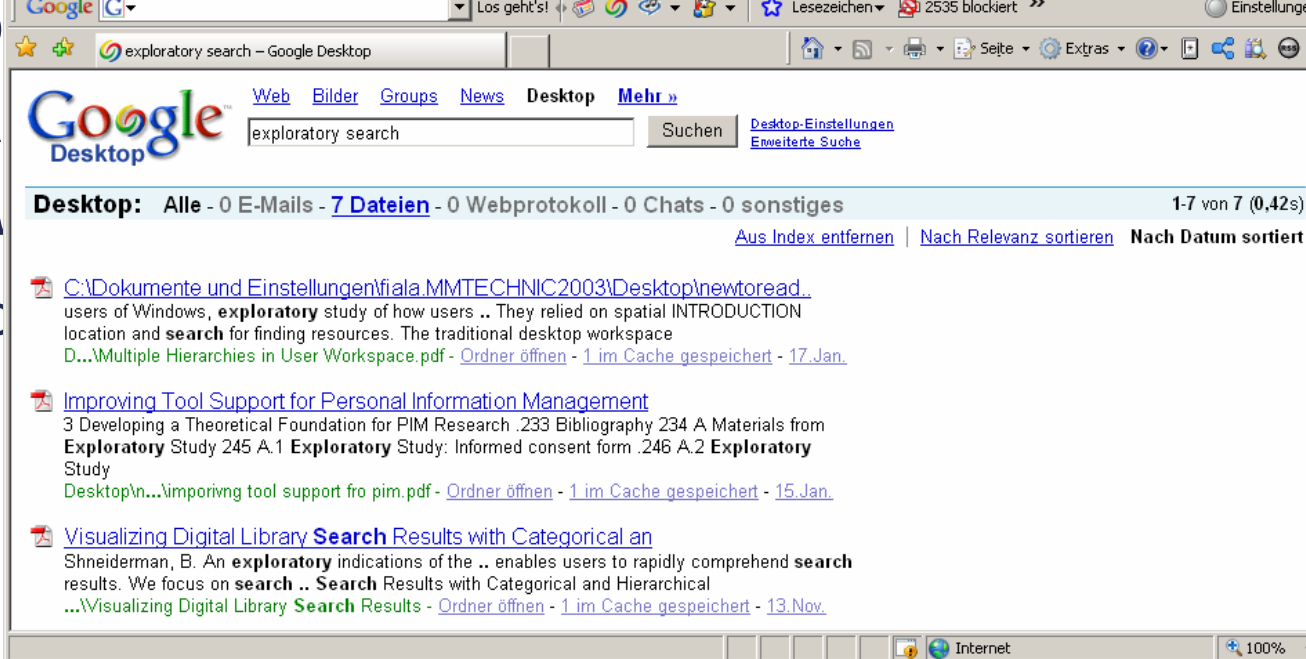


UIs für PIM-Suche

- Klassische Suchinterfaces
 - Meist Texteingabe über Textfelder, eventuell Filter-Checkboxen
 - Ziel: möglichst exakte (perfekte) Ergebnisse liefern
 - *Teleporting*: „Sprung“ zur gesuchten Ressource [Teev+04]

- Studie [Teev+04]

- Desktop
- Anzeigen
- Web
- Nach
- Suchen
- Navigieren
- Ergebnisse



exploratory search – Google Desktop – Windows Internet Explorer

http://127.0.0.1:4664/search?q=exploratory+search&flags=68&num=10&s=12cZdtSE_owi5Ab:Fbk9b8EjmT4

Google Desktop

exploratory search

Suchen

Desktop: Alle - 0 E-Mails - 7 Dateien - 0 Webprotokoll - 0 Chats - 0 sonstiges 1-7 von 7 (0,42s)

Aus Index entfernen | Nach Relevanz sortieren | Nach Datum sortiert

C:\Dokumente und Einstellungen\fiala.MMTECHNIC2003\Desktop\newtoread...
users of Windows, **exploratory** study of how users .. They relied on spatial INTRODUCTION
location and **search** for finding resources. The traditional desktop workspace
D:\Multiple Hierarchies in User Workspace.pdf - Ordner öffnen - 1 im Cache gespeichert - 17.Jan.

Improving Tool Support for Personal Information Management
3 Developing a Theoretical Foundation for PIM Research .233 Bibliography 234 A Materials from
Exploratory Study 245 A.1 **Exploratory** Study: Informed consent form .246 A.2 **Exploratory**
Study
Desktop\...improving tool support fro pim.pdf - Ordner öffnen - 1 im Cache gespeichert - 15.Jan.

Visualizing Digital Library Search Results with Categorical an
Shneiderman, B. An **exploratory** indications of the .. enables users to rapidly comprehend **search**
results. We focus on **search** .. Search Results with Categorical and Hierarchical
...Visualizing Digital Library Search Results - Ordner öffnen - 1 im Cache gespeichert - 13.Nov.

(≈3%)

en
mulieren

UIs für PIM-Suche

The screenshot displays the ASI Phlat browser interface. The main window is titled "ASI Phlat" and contains a search results area. The search results are organized into sections: "Within the last year" and "More than a year ago". The results are displayed in a table with columns for Title, Date, Author, and To. The first section, "Within the last year", contains three results. The second section, "More than a year ago", contains one result. The interface also features a "Query Area" on the left side, which includes a search bar with the text "bike" and a "Go" button. Below the search bar are "Saved Queries" and "Date" filters. The "Filter Area" is located on the left side of the interface, featuring a "Tag Filter Pane" with a list of tags such as "Personal", "Family", "Gaming", "Bikes", "Images", "Maps", "Book Clubs", "Beer", "Friends", "Cats", "Art", and "New Friends". The "Tag Filter Pane" also includes a "Filter by Tag" section with a list of tags and their counts. The "Filter Area" also includes "Path", "People", and "Type" filters. The "Results Area" is the main area where the search results are displayed. The interface is annotated with several labels: "Query Area" (blue bracket), "Filter Area" (green bracket), "Tag Filter Pane" (green bracket), "Results Area" (red bracket), "Filter Tiles" (pink label), and "Tag Widgets" (pink label). The "Filter Tiles" label points to the "Personal" tag in the "Filter by Tag" section. The "Tag Widgets" label points to the "Bikes" tag in the "Filter by Tag" section.

Phlat Browser [Cutr+06]

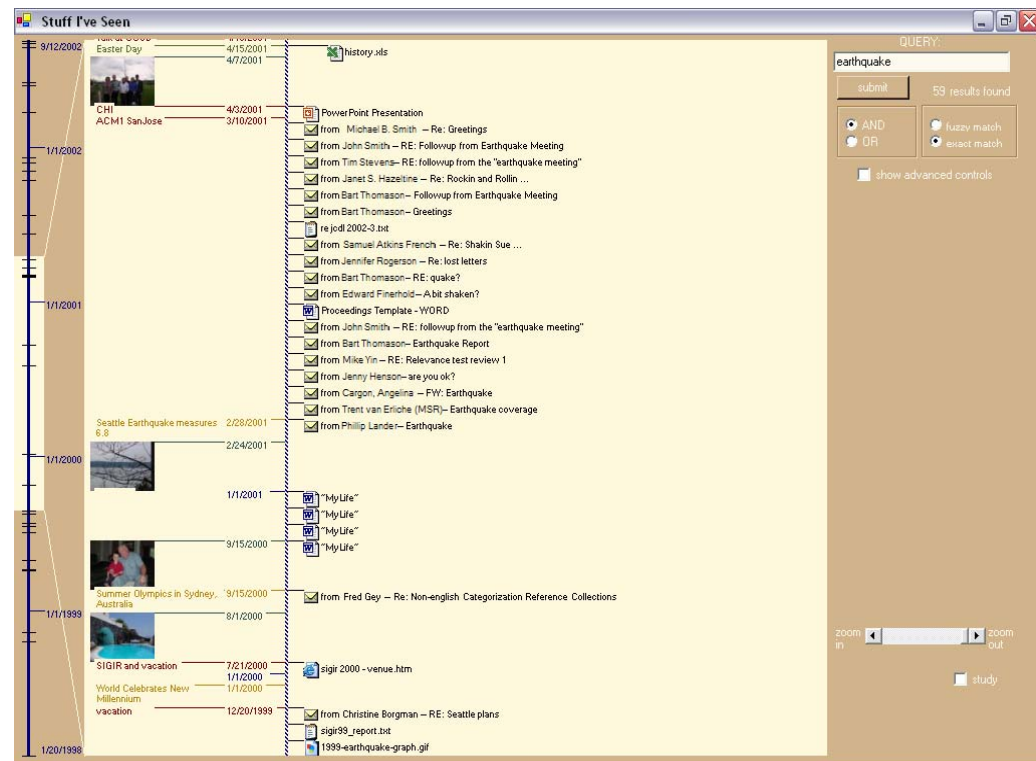
UIs für PIM-Suche

- Anforderungen an neuartige UIs für PIM-Suche*
 - Kombination von text-basierter Suche und Browsing
 - Filterung, Sortierung entlang mehrerer Metadatenfacetten
 - Extensiver Einsatz von Kontextinformationen
 - Überblicks- und Detailansichten
 - Sinnvolle Gruppierungs- und Kategorisierungsfunktionen
 - Vereinheitlichung von Suche und anderen PIM-Aktivitäten
 - Möglichkeit, unterschiedliche UIs zu verwenden

* [Barr95], [Duma+01], [Berg+03], [Cutr+06b], [Marc+06]

UIs für PIM-Suche – Kontextvisualisierung

- Zeit als wichtigster Organisationsfaktor in PIM
 - Rolle des episodischen Gedächtnis
 - Wichtige öffentliche und personenbezogene Ereignisse
 - SIS [Duma+03]
 - Stuff I've Seen
 - Memory Landmark Interface
 - Vorgänger von MS Desktop Search



UIs für PIM-Suche – Kontextvisualisierung

- UIs für kontextbasierte Suche: Ortskontext [Toya+03]

- WWMX Browser UI

- Geovisualisierung via Weltkarte

- Constraint Panels

- Map

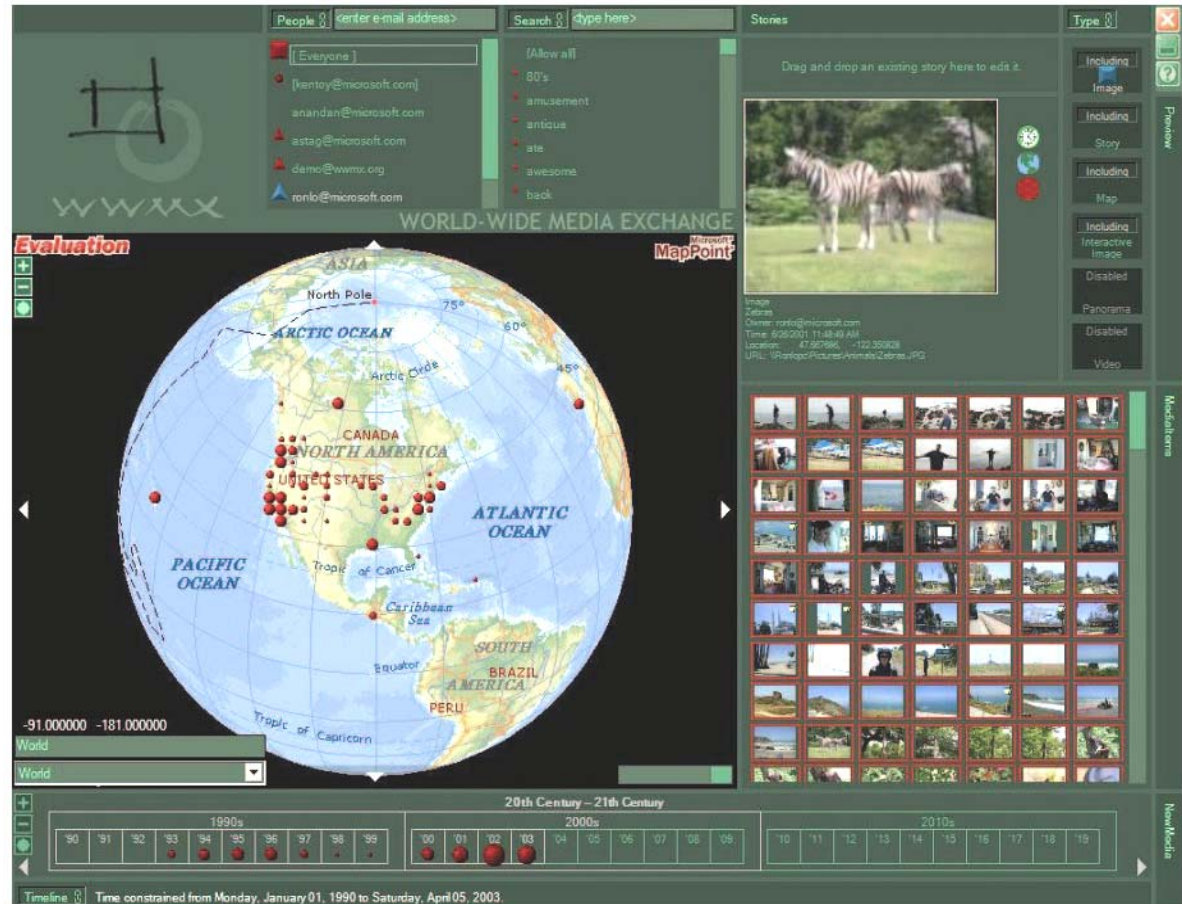
- Timeline

- People

- Keywords

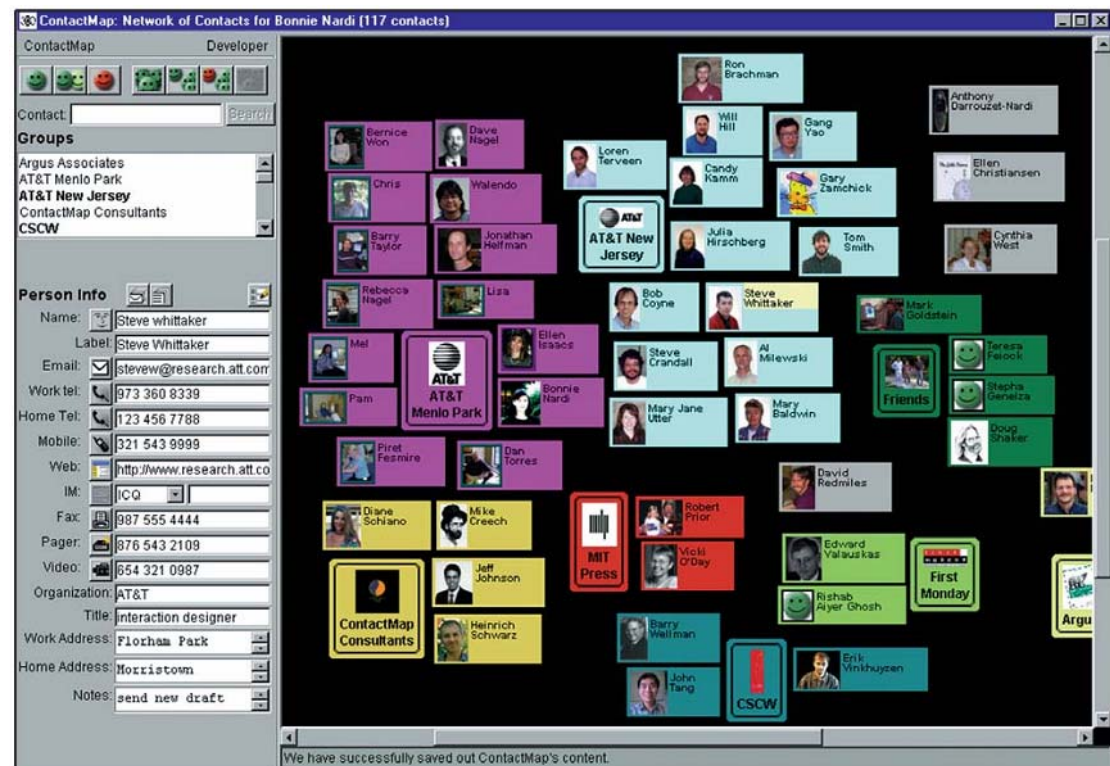
- Media types

- <http://wmx.org/>



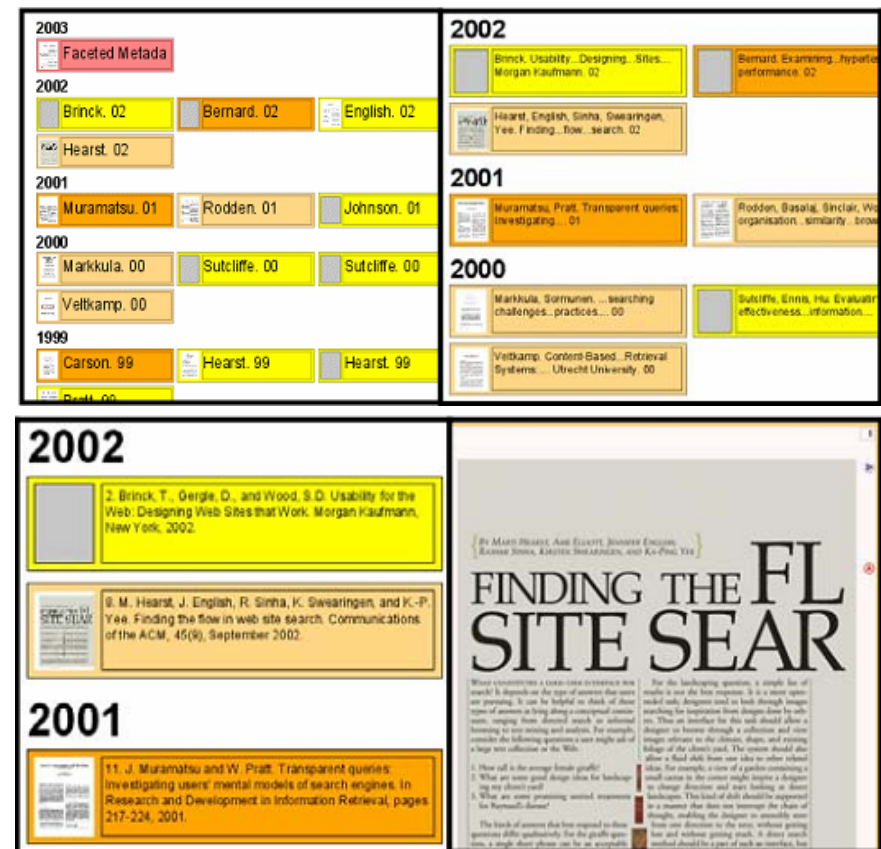
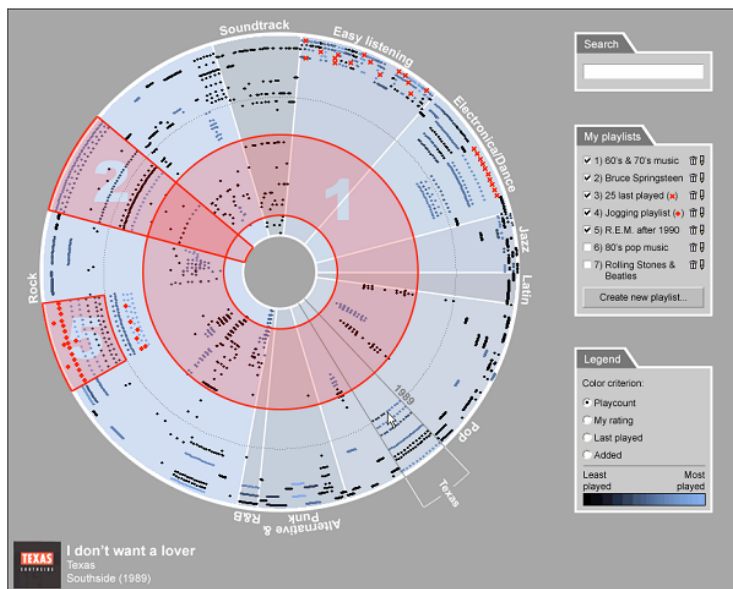
UIs für PIM-Suche – Kontextvisualisierung

- UIs für kontextbasierte Suche: Personen- und Projektkontext
 - Z.B. ContactMap [Whit+04], DeepaMehta [Rich+04]
 - Darstellung des den Nutzer umgebenden sozialen Netzwerks
 - Dateien, Emails, Bookmarks assoziiert mit (Gruppen von) Personen
- Kommunikationsfunktionen



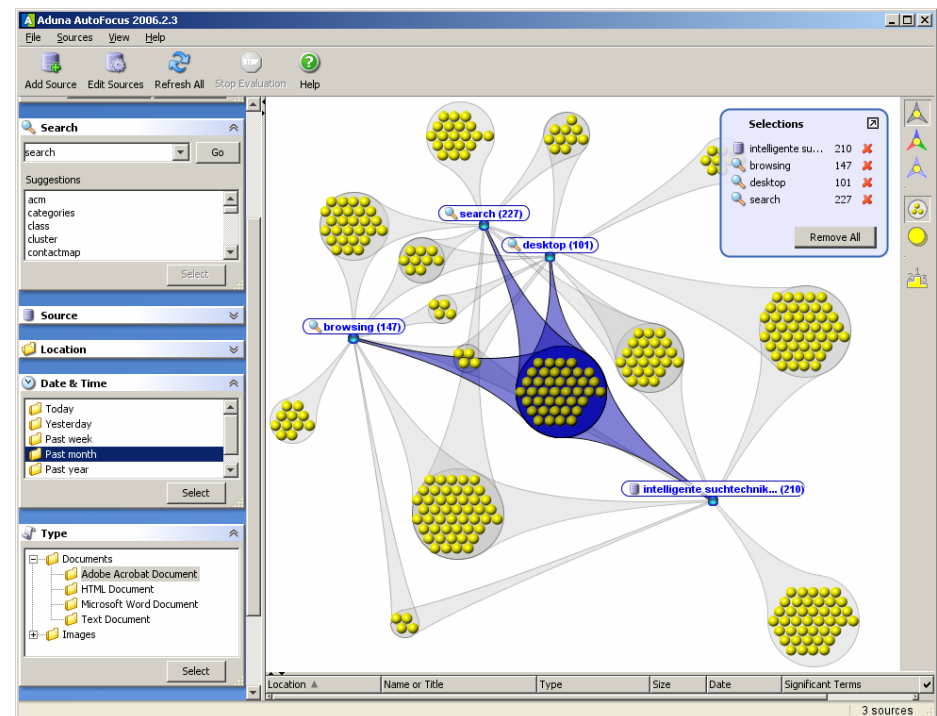
UIs für PIM Suche – Überblick und Detail

- Überblicke- und Detailansichten
 - Visual Information Seeking Mantra von Shneiderman [Shne+96]
 - “Overview first, zoom and filter, then detail-on-demand“
 - „Disc Visualization“ für Musikbibliotheken [Torr+04]
 - Semantic Zooming [Bier+04]



UIs für PIM Suche – Kategorisierung

- Gruppierungen und Kategorisierungsfunktionen
 - Sinnvolle Organisation von Suchergebnissen ausschlaggebend
 - Anzeige von Ergebnissen in ihrem semantischen Kontext [Hear06]
 - Schnelle Entscheidungshilfe bei mehrdeutigen Queries [Chen+00]
 - Automatisches Clustering vs. feste Kategoriensysteme
 - ClusterMaps
 - z.B. In AdunaFocus →
 - basiert auf Sesame
 - Clusterbildung anhand mehrerer Suchanfragen

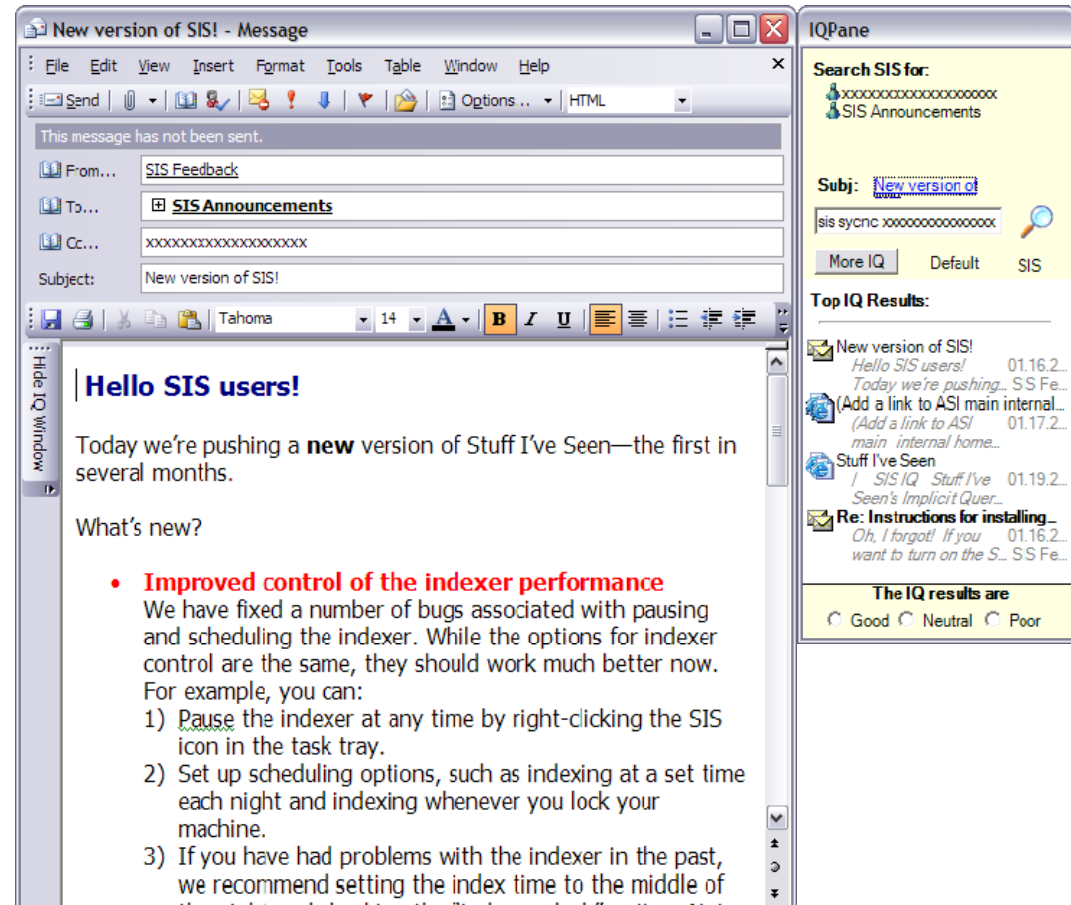


UIs für PIM-Suche – Fazit

- Benutzerschnittstellen für PIM-Suche – Fazit
 - Aktuelle UIs für Desktop Suche unzureichend
 - Neuartige Visualisierungen in Einzelprototypen vorhanden
 - Kontextvisualisierung
 - Multidimensionales Metadatenbrowsing
 - Semantisches Zooming
 - Gruppierungs- und Kategorisierungsfunktionen
 - Integrationsbedarf in existierende Desktop-Lösungen sowie miteinander!
 - Problem der „UI Fragmentation“
 - Bedarf an personenspezifischen UIs
 - Angepasst an Präferenzen, Fähigkeiten, etc.

Aktuelle Forschungsfragen

- Proaktive Suche: „Finding without Searching“
 - Automatische Suche anhand aktueller Aufgabe
 - Implicit Query
 - SIS [Duma+04]
 - Analyse von Emails
 - Extraktion von Metadaten
 - Betreff
 - Text
 - Sender, etc.
 - Suche anhand Aufgabenkontext



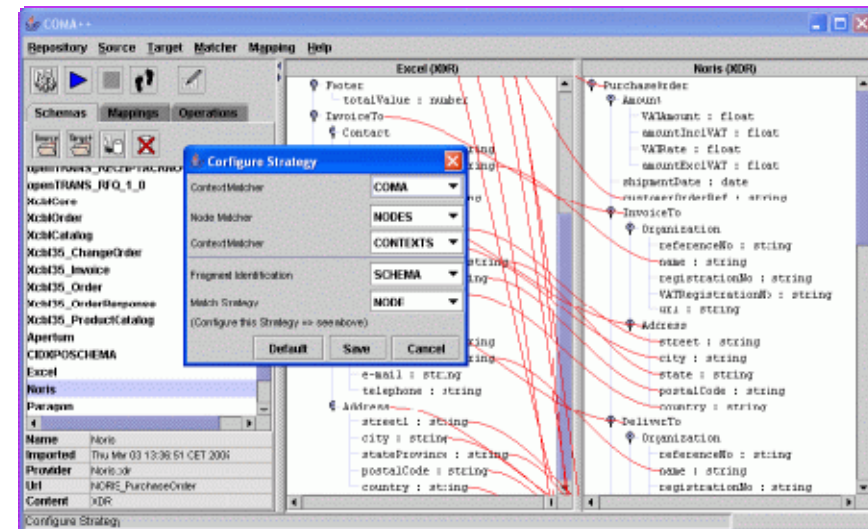
Aktuelle Forschungsfragen

- Wechselwirkung zwischen PIM- und Web-Suche
 - Semantic Desktops verwalten wertvolle Kontextinformationen
 - Diese können bei der Suche im (Semantic) Web ausgenutzt werden [Chir+06]
 - Z.B. Suche nach „Madonna mit Kind“
 - Kunstliebhaber?
 - Popmusik-Fan?
- Vorverarbeitung von Suchanfragen auf dem Desktop
 - „Query expansion“ auf der Basis des persönlichen Kontexts
- Kein Bedarf an serverseitiger Kontextverarbeitung
 - Keine Privacy Probleme



Aktuelle Forschungsfragen

- Semantische Integration – Ontology Matching
 - Anwendungsübergreifende Suche bzw. Reasoning fordert Integration der Datenschemata bzw. Ontologien
 - (Vollständig) manuelle Abbildung nicht skalierbar
 - Ontology Matching
 - Shared Ontologies als gemeinsame Grundlage
 - Mapping Heuristiken
 - Machine-Learning-Verfahren



COMA++ [Aumu+05]

- Integration von Anwendungen bzw. User Interfaces!

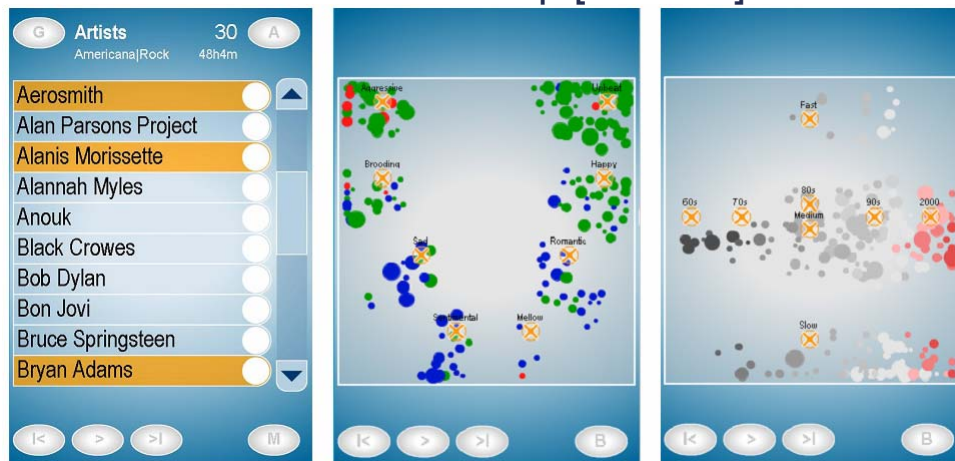
Aktuelle Forschungsfragen

- PIM: Hauptfokus mobiler Endgeräte
 - Personal Digital Assistant = Personal Information Manager
 - Mobile Suche
 - Vermeidung von Texteingabe
 - Überblick- und Detail-Techniken!
 - Relevanz von Orts- und Gerätekontext
- Ergebnisvisualisierung, Interaktionen

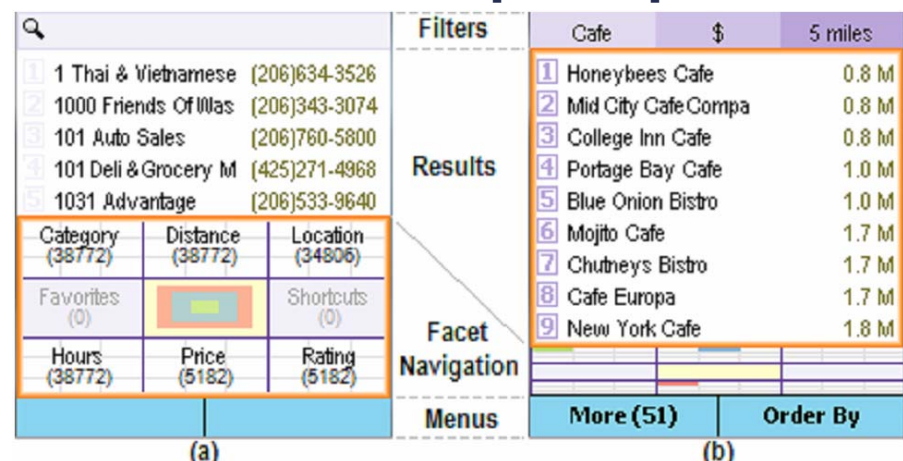


Apple iPhone 2007

Artist Map [Guli+05]



FaThumb [Karl+06]



Weitere Herausforderungen

- Sicherheitsaspekte in PIM
 - Umfassendere Suchmöglichkeiten auf dem Desktop
 - Neue Zusammenhänge durch Data Integration
 - Verschmelzung von Web- und Desktop-Suche
 - Persönliche Informationen u.U. auf mehreren Endgeräten verteilt
 - Sicherheitsrichtlinien wichtiger als je zuvor
- Performanzprobleme in PIM
 - Erhöhter Rechenaufwand für
 - Modellierung, Indizierung, Inferenzen, Suche
 - Visualisierung
 - Beschränkte Rechenleistung insb. auf mobilen Endgeräten

Zusammenfassung

- Personal Information Management
- Besonderheiten der PIM-Suche
 - Rolle von Assoziationen bzw. Kontextinformationen
- Semantic Desktops
 - Semantic Desktop Systeme
 - Semantic Web-Technologien für die Modellierung von PI
- Neuartige UIs für PIM-Suche
- Aktuelle Forschungsfragen und Herausforderungen
 - Ontology Matching und Reasoning
 - Mobile PIM-Suche
 - Sicherheit

Zusammenfassung

Vielen Dank für Ihre Aufmerksamkeit!

Literatur

- [Aumu+05] David Aumueller, Hong Hai Do, Sabine Massmann, Erhard Rahm: ***Schema and ontology matching with COMA++***. SIGMOD Conference 2005: 906-908
- [Barr 95] Deborah K. Barreau. ***Context as a Factor in Personal Information Management Systems***. JASIS 46(5): 327-339, 1995
- [Berg+03] Ofer Bergman, Ruth Beyth-Marom, and Rafi Nachmias. ***The user-subjective approach to personal information management systems***. JASIST, 54(9):872-878, 2003.x
- [Bern+04] B. J. Jansen and A. Spink. ***How are we searching the World Wide Web? A comparison of nine search engine transaction logs***. Information Processing and Management 42 (2006) 248–263
- [Chen+00] Hao Chen and Susan T. Dumais. ***Bringing order to the Web: automatically categorizing search results***. CHI 2000
- [Chir+05] Paul Alexandru Chirita, Rita Gavrioloaie, Stefania Ghita, Wolfgang Nejdl, and Raluca Paiu. ***Activity Based Metadata for Semantic Desktop Search***. ESWC 2005, LNCS 3532, pp. 439–454, 2005.
- [Chir+06] Paul Alexandru Chirita, Claudiu S. Firan, and Wolfgang Nejdl. ***Summarizing Local Context to Personalize Global Web Search***. CIKM06, November 5–11, 2006, Arlington, Virginia, USA.
- [Cutr+06] Edward Cutrell, Daniel C. Robbins, Susan T. Dumais, and Raman Sarin. ***Fast, flexible filtering with phlat***. In CHI 2006, pages 261-270, 2006.
- [Cutr+06b] Edward Cutrell, Susan T. Dumais, and Jaime Teevan. ***Searching to eliminate personal information management***. Commun. ACM, 49(1):58-64, 2006.

Literatur

- [Cutr+06c] Edward Cutrell. ***Search user interfaces for pim.*** In Personal Information Management - A SIGIR 2006 Workshop, 2006.
- [Dach+06] Raimund Dachsel, Markus Weiland: ***TimeZoom: a flexible detail and context timeline.*** CHI Extended Abstracts 2006: 682-687
- [Deck+04] Stefan Decker, Martin R. Frank: The Networked Semantic Desktop. WWW Workshop on Application Design, Development and Implementation Issues in the Semantic Web 2004
- [Dey01]
- [Duma+01] Susan Dumais, Edward Cutrell, Hao Chen. ***Optimizing Search by Showing Results In Context.*** SIGCHI'01, March 31-April 4, 2001, Seattle, WA, USA.
- [Duma+03] Susan T. Dumais, Edward Cutrell, Jonathan J. Cadiz, Gavin Jancke, Raman Sarin, and Daniel C. Robbins. ***Stuff I've seen: a system for personal information retrieval and re-use.*** In SIGIR, pages 72-79, 2003.
- [Duma+04] Susan T. Dumais, Edward Cutrell, Raman Sarin, and Eric Horvitz. ***Implicit queries (iq) for contextualized search.*** In SIGIR, page 594, 2004.
- [Free+96] Eric Freeman, David Gelernter: ***Lifestreams: A Storage Model for Personal Data.*** SIGMOD Record 25(1): 80-86 (1996)
- [Gemm+06] Jim Gemmell, Gordon Bell, Roger Lueder. ***MyLifeBits: A Personal Database for Everything.*** Commun. ACM 2006 49(4), p88-95
- [Guli+05] Rob van Gulik, Fabio Vignoli: ***Visual Playlist Generation on the Artist Map.*** ISMIR 2005: 520-523

Literatur

- [Hear 06] Marti A. Hearst. ***Clustering versus faceted categories for information exploration.*** Commun. ACM. 2006 49(4), p59-61
- [Karg+05] David R. Karger, Karun Bakshi, David Huynh, Dennis Quan, Vineet Sinha: ***Haystack: A General-Purpose Information Management Tool for End Users Based on Semistructured Data.*** CIDR 2005: 13-26
- [Lans 88] Lansdale, M. ***The psychology of personal information management.*** Applied Ergonomics 19, 1 (1988), 55–66.
- [Marc 06] Gary Marchionini. ***Exploratory search: from finding to understanding.*** Commun. ACM, 49(4):41-46, 2006.
- [Page+98] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. ***The PageRank citation ranking: Bringing order to the web.*** Technical report, Stanford University, 1998.
- [Rich+03] Jörg Richter, Max Völkel, and Heiko Haller: ***DeepaMehta – A Semantic Desktop.***
- [Ring+03] Meredith Ringel, Edward Cutrell, Susan T. Dumais, Eric Horvitz: ***Milestones in Time: The Value of Landmarks in Retrieving Information from Personal Stores.*** INTERACT 2003
- [Rose+04] Daniel E. Rose and Danny Levinson. ***Understanding user goals in web search.*** In WWW2004, pages 13-19, 2004.
- [Shne+] Ben Shneiderman, ***The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations.*** In Proceedings of the IEEE Symposium on Visual Languages, pages 336-343, Washington. IEEE Computer Society Press, 1996.

Literatur

- [Schr+06] Monica M. C. Schraefel, Max Wilson, Alistair Russell, and Daniel A. Smith. ***mSPACE: improving information access to multimedia domains with multimodal exploratory search.*** Commun. ACM, 49(4):47-49, 2006.
- [Teev+04] Jaime Teevan, Christine Alvarado, Mark S. Ackerman and David R. Karger. ***The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search.*** CHI 2004, p415-422, Austria, 2004
- [Teev+06] Jaime Teevan, William Jones, and Benjamin B. Benderson. ***Personal Information Management.*** Commun. ACM, 49(1):41-43, 2006.
- [Torr+04] Marc Torrens, Patrick Hertzog, and Josepg Lluís Arcos. ***Visualizing and Exploring Personal Music Libraries.*** 2004
- [Toya+03] Kentaro Toyama, Ron Logan, Asta Roseway, and P. Anandan. ***Geographic Location Tags on Digital Images.*** ACM MM'03, Berkeley, USA, p156-166
- [Yee+03] Ka-Ping Yee, Kirsten Swearingen, Kevin Li, and Marti A. Hearst. ***Faceted metadata for image search and browsing.*** CHI2003, p401-408
- [Whit+04] Steve Whittaker, Quentin Jones, Bonnie A. Nardi, Mike Creech, Loren G. Terveen, Ellen Isaacs, John Hainsworth. ***ContactMap: Organizing communication in a social desktop.*** ACM Trans. Comput.-Hum. Interact. 11(4), p445-471, 2004