A Dataspase Odyssey: The iMeMex Personal DataSpace Management System

Lukas Blunschi, Jens Dittrich, Olivier Girard, Shant Karakashian, Marcos Salles

ETH Zurich, iMeMex.org
What is a DataSpace Management System?

- Vision paper by Mike Franklin, Alon Halevy and David Maier
  From Databases to Dataspaces: a New Abstraction for Information Management.
  SIGMOD Record, December 2005

- Idea: ALL data belonging to a particular person, task, or company gets assigned to a logical dataspace, no upfront investments for schemas.

- Dataspace independent from formats, devices, locations, and systems owning the data.

- What we need is a new kind of system able to manage dataspaces across devices and formats.
From Database Systems to DataSpace Systems

- DataSpace management systems can be seen as a hybrid of
  - a search engine,
  - a databases management system,
  - an information integration system,
  - and a data sharing system.

- iMeMex is one of the first DataSpace Management Systems

- iMeMex is designed as a highly extensible framework
  (Eclipse-like, OSGi service-oriented architecture)

- Current focus of iMeMex: personal dataspaces

- However, we are also targeting other applications including scientific dataspaces.
iMeMex Project: Integrated MeMex

- Project started 2 years ago
- Our goal: Build the first open source dataspace management system
- All Java, service-oriented (OSGi, cmp. Eclipse project)
- Plus platform specific extensions in C (currently for Windows and Mac)
- iMeMex is an extensible framework for dataspace management
- Supported by SNF (Swiss National Science Foundation)
- Demo at VLDB ‘05, papers at VLDB ‘06 and SIGIR PIM ‘06
- State of the system, milestones:
  - December ‘06: 0.42.0, RC1a developer version released/open source
  - Beg of Feb ‘07: 0.43.0, RC1b developer version
  - End of Feb ‘07: 0.43.x, RC2 developer version + RC1 AJAX client
  - End of May ‘07: 0.44.x, RC3 developer version + RC2 AJAX client
  - End of July ‘07: 0.45.x, RC1 end-user version
Literature on PIM and Dataspaces

- J. Dittrich, M. Salles, D. Kossmann, L. Blunschi
  iMeMex: Escapes from the Personal Information Jungle (Demo Paper)

- M. Franklin, A. Halevy, D. Maier
  From Databases to Dataspaces: A New Abstraction for Information Management

- J. Dittrich
  iMeMex: A Platform for Personal DataSpace Management

- J. Dittrich, M. Salles
  iDM: A Unified and Versatile Data Model for Personal Dataspace Management

- L. Blunschi, J. Dittrich, O. Girad, S. Karakashian, M. Salles
  A Dataspace Odyssey: The iMeMex Personal Dataspace Management System
  CIDR 2007 (Biennial Conference on Innovative Data Systems Research), Jan. 2007

- J. Dittrich, L. Blunschi, O. Girad, S. Karakashian, M. Salles, M. Färber
  From Personal Desktops to Personal Dataspaces: A Report on Building the iMeMex
  Personal Dataspace Management System
  BTW 2007 (Biennial German Database Conference), March 2007
Demo Screenshots
NOTE: we did not write any HTML/Javascript to generate this AJAX application. Our AJAX application was automatically generated from a Java program using GWT.
IMAP Mail Data Source

- inbox.misc2 email (1)
  - DDM Talk: Visual recognition: How...
  - Conference on Mobile Information
  - DDM Talk: Innovative Interaction P...
  - Die Welt zwischen 0 und 1: Invite...
  - Die Welt zwischen 0 und 1: Invite...
  - DDM Talk: Pictures of the Future: 5
  - DDM Talk: Wednesday June 28, P...
  - CFP for IEEE COMSWARE 2007
  - CFP-COMSWARE2007.pdf
  - DDM Talk: Thursday June 15, Prof...

- Fwd: IASTED PCDN 2007 Innsbruck
  - PDCN 2007 - Online CFP.pdf
  - 25th anniversary: The events and
  - DDM Talk: Wednesday June 21, P...

email messages
File System Data Source

file system data
ATOM Feed Data Source

ATOM news messages
Oracle Database Data Source

Database data source: either warehoused or mediated
Keyword Search

Search & Query: halevy

- inbox.misc2 email (1)
- testData (2)
- atom feed (3)
- rss feed (4)
- IMEMEXTEST database (5)

file://CIDR2007/Marcos' documents/papers/imemex.bib]bibtex[FHM05

file://CIDR2007/Marcos' documents/papers/imemex.bib]bibtex[DHMN05


file://CIDR2007/Marcos' documents/papers/imemex.bib]bibtex[DH05

A. Halevy and others. Crossing the Structure Chasm.
20 resource views found.
Search & Query Granularity: Resource Views

search & query granularity:
- fine granular resource views
- not files or documents
Here:
- entries from a BibTeX file
Search&Query Granularity: Resource Views

expanded view of a result
Navigate Inside Files

structure found inside a LaTeX file
Structural Search (NEXI-like)

structural queries beyond tcsh, similar to NEXI (Narrowed Extended XPath I)
Search History

search history
Keyword Prefix Search

Prefix "schwa" will also return results with "Schwarzenegger"
Popups for Graph Navigation

Popups allow navigation on subgraphs.
Upcoming Features

- Declarative Pay-as-you-go Information Integration Framework
- Data Sharing Framework using distributed iMeMex instances
- Cost-based optimization
- <your plugin here>
- Details:
  see our web-site: http://www.imemex.org