The Semantic Desktop:

The Intimate Supplement to Memory

- Recent Work-

Andreas Dengel

Agenda

- Some words about …
  - ... what is a semantic desktop
  - ... how to build a semantic desktop
  - ... how to integrate paper documents
  - ... more user observation
- Summary and next steps

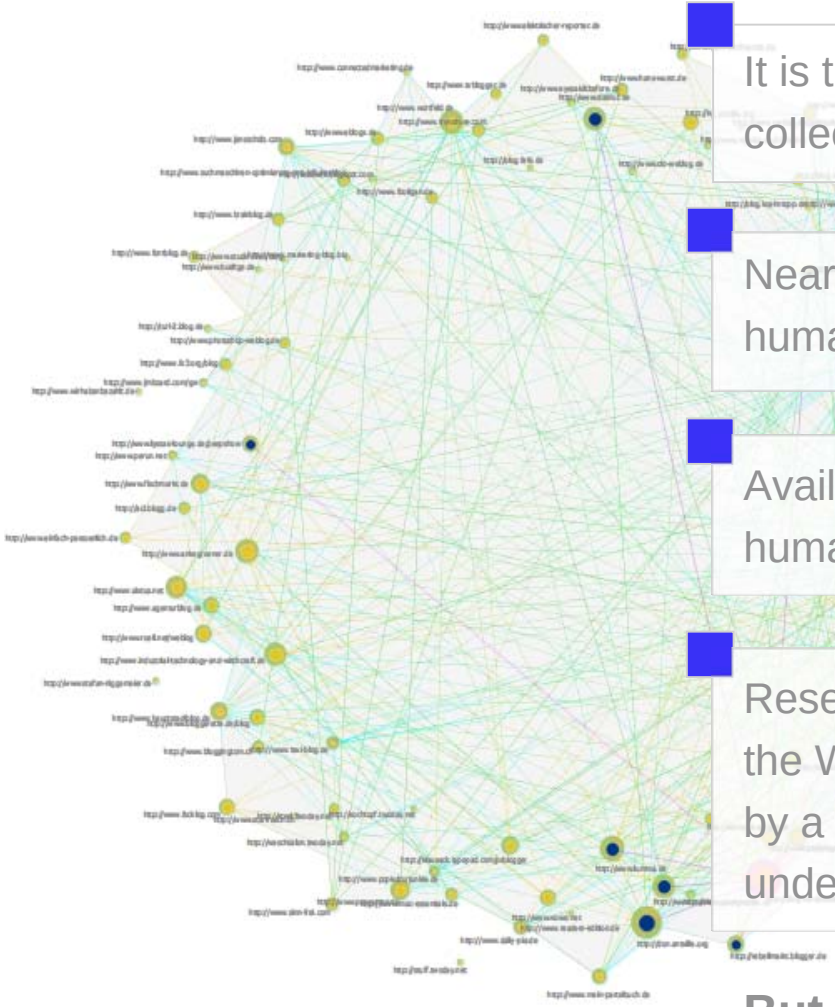
What is a Semantic Desktop?

Our definition of the Semantic Desktop

A Semantic Desktop is a device in which an individual stores all her digital information like documents, multimedia and messages. These are interpreted as Semantic Web resources, each is identified by a Uniform Resource Identifier (URI) and all data is accessible and queryable as RDF graph. Resources from the web can be stored and authored content can be shared with others. Ontologies allow the user to express personal mental models and form the semantic glue interconnecting information and systems. Applications respect this and store, read and communicate via ontologies and Semantic Web protocols. The Semantic Desktop is an enlarged supplement to the user's memory.

L. Sauermann, A. Bernardi and A. Dengel, *Overview and Outlook on the Semantic Desktop*, Proceedings International Semantic Web Conference, Galway, Ireland (Nov. 2005), pp. 1-19.

Starting Point



- It is the primary means of the information society to collect information of any kind
- Nearly everything we find in the Web is input by a human being through a computer
- Available documents are only fully understood by human beings
- Researchers attempt to transform the Web of links into the Web of meaning in which documents are described by a standardized vocabulary providing machine understandable semantics

But ...

... how can we build an appropriate vocabulary as a means to build an ontology to be shared with others?

Finding the right vocabulary reveals some problems

This also hold for documents

Depending on who reads a text (message),

... at what time

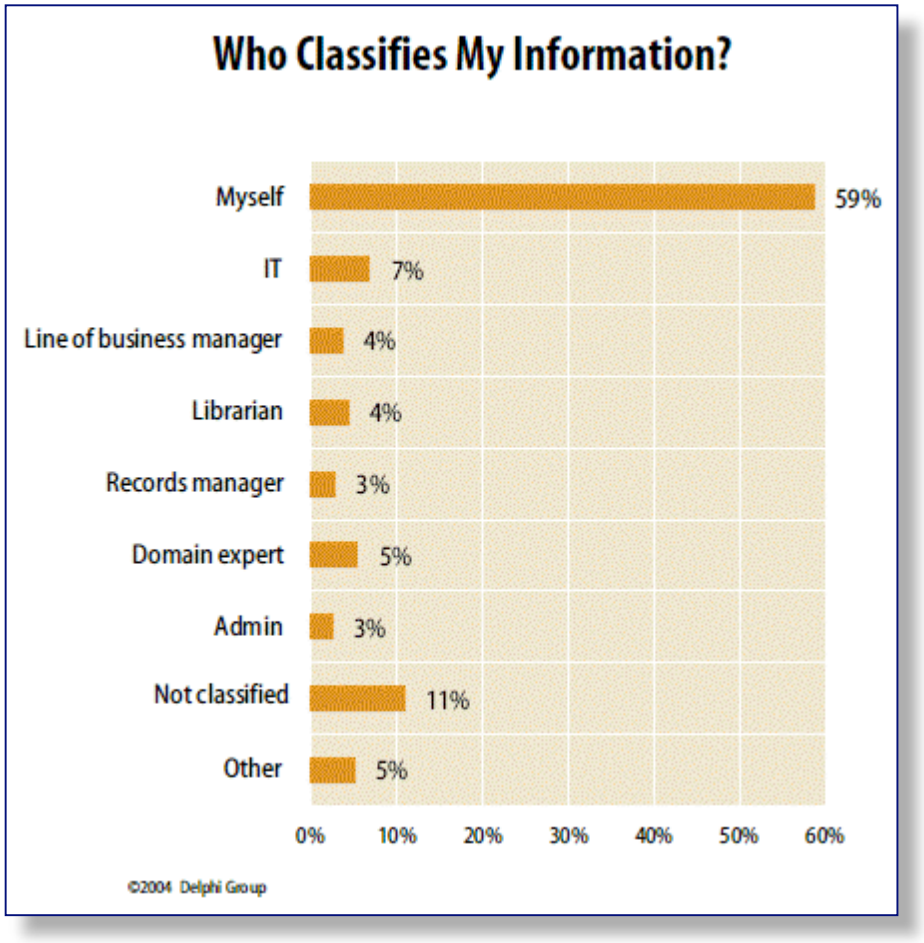
... in which aspects he/she is interested in,

... an which actual tasks he/she is working, and

... what expertise and experience is available,

it may be seen as valuable information, as a bootless statement or even as an annoyance

Daily practice requires different approaches



- Respective usability of a vocabulary strongly depend on the users document categorization results
- There is only few support in document categorization
- Users tend to prefer their own way to organizing their information with respect to their individual needs

Source: **Information Intelligence: Content Classification and the Enterprize Taxonomy Practice, Delphi Group Report, June 2004 (**http://www.delphigroup.com)

Some theses about knowledge evolution

- The bondage for formal organization of information inhibits creativity and limits the options of self-organization

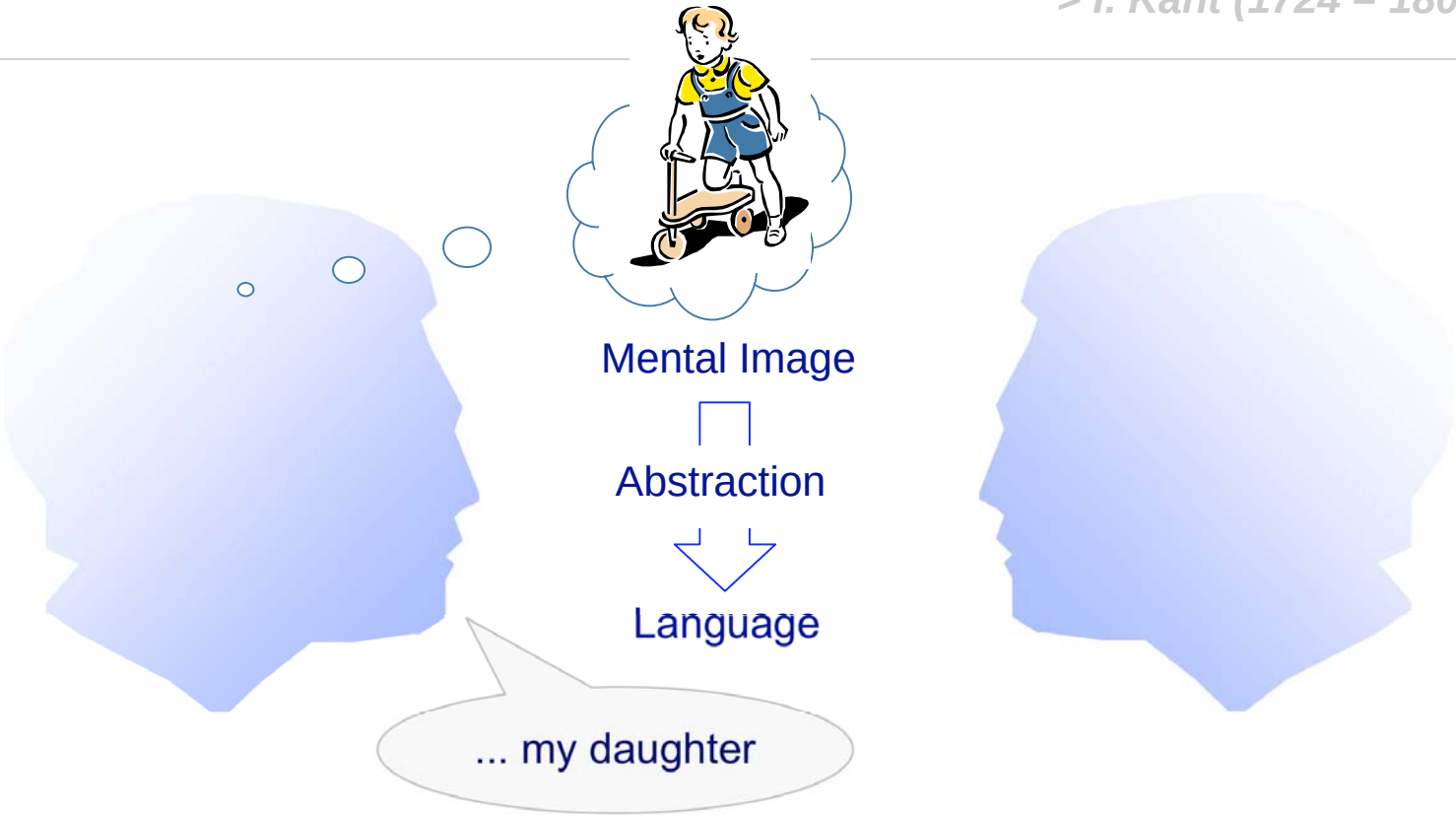
- A document is like a node in a net, a system of hyper-links to books, texts, pictures, etc.
  - Instead of a static objects, it is variable and relative depending on who reads it at what time and in which situation

- Individual trains of thoughts lead to multi-dimensional perspective organizations of contents and thus to a dematerialization of the classical archive
  - Structure and context of information and thus categorization are liable to an accelerating change

Subjectivity of terms ...

Imaginations without terms are blind, terms without imaginations are empty

> I. Kant (1724 – 1804)



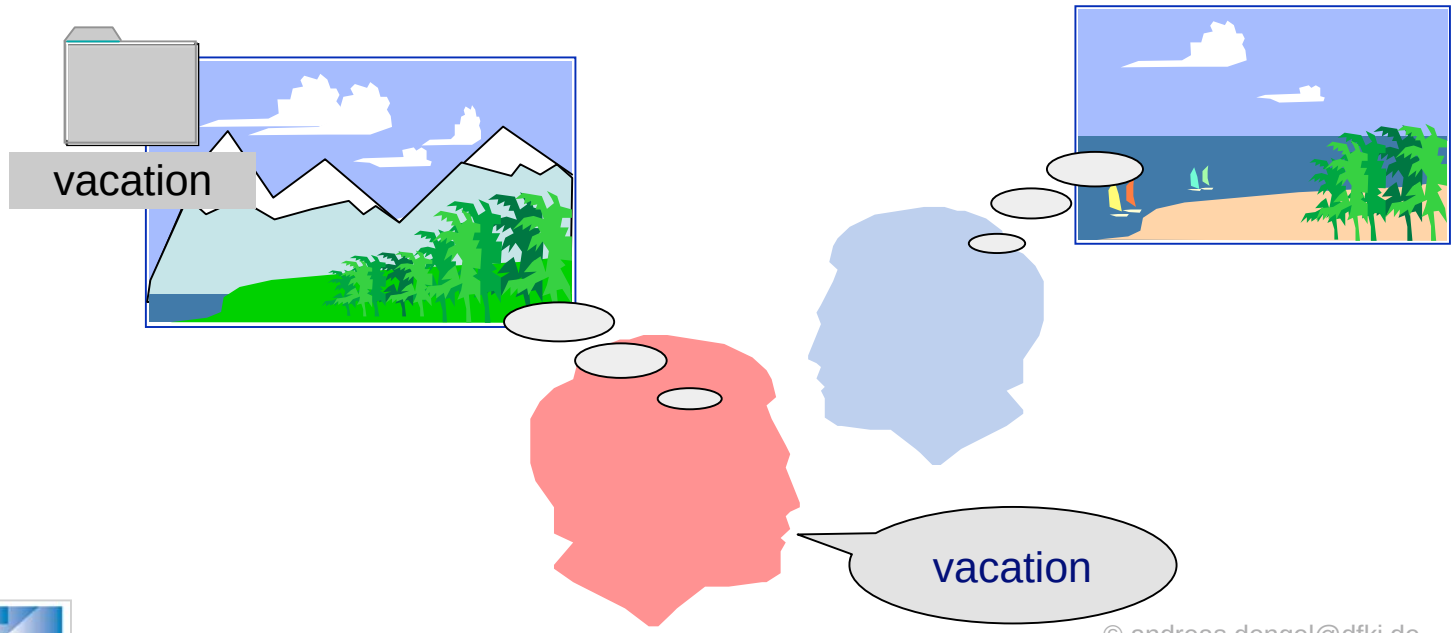
... is reflected within mental models

The perception of documents is subjective

In office environments people classify documents according to their preferences, i.e. they generate folders as categories and name them

- Resulting taxonomies correspond to subjective concepts of the world but …

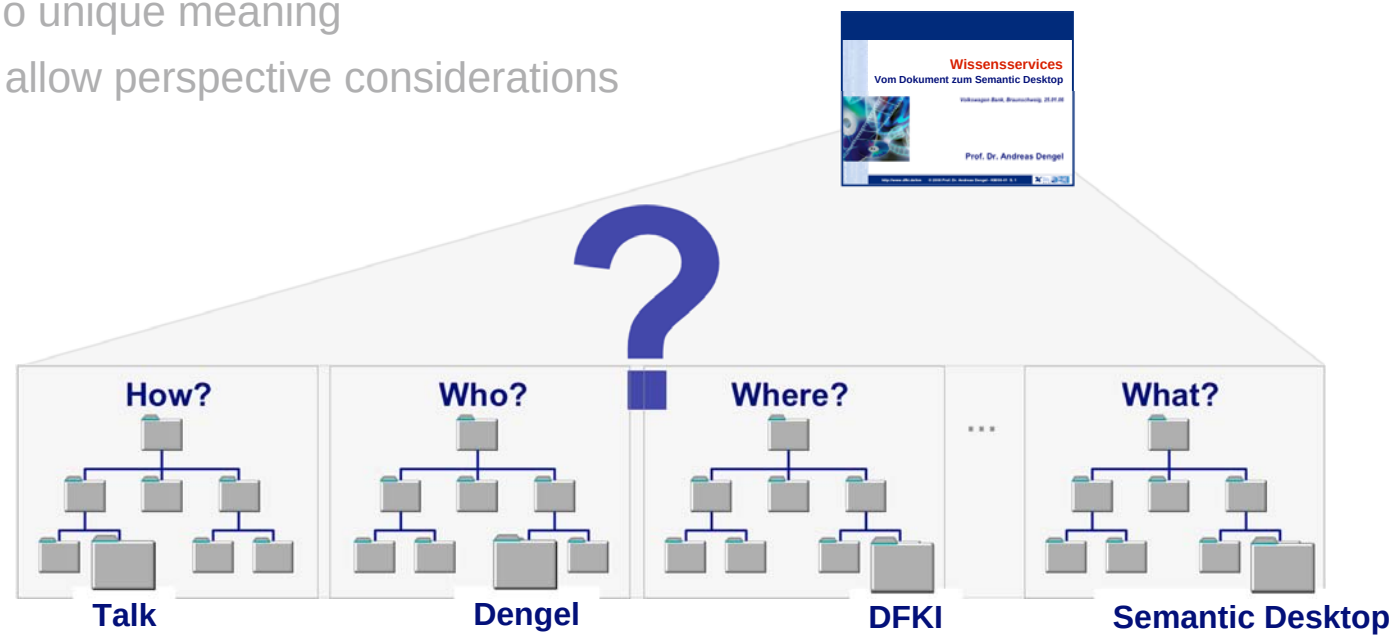
  … have no unique meaning



The perception of documents is subjective

In office environments people classify documents according to their preferences, i.e. they generate folders as categories and name them

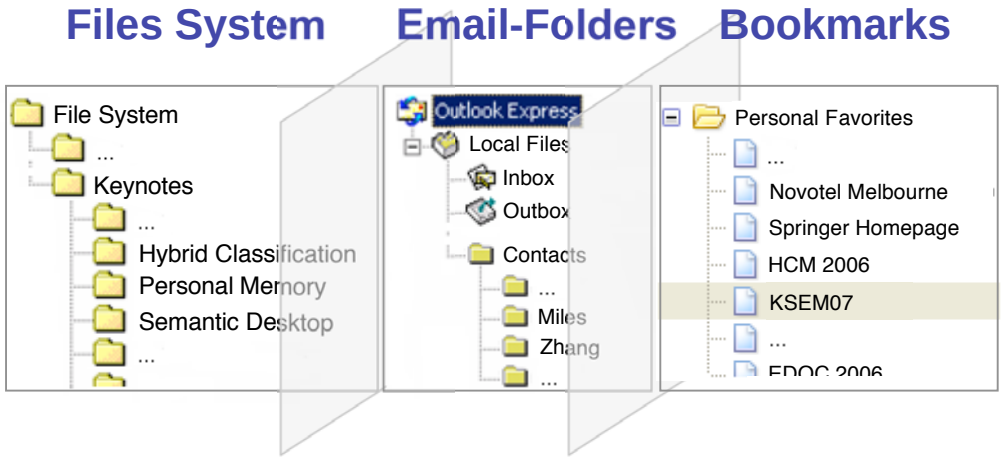
- Resulting taxonomies correspond to subjective concepts of the world but …
  - … have no unique meaning
  - … do not allow perspective considerations



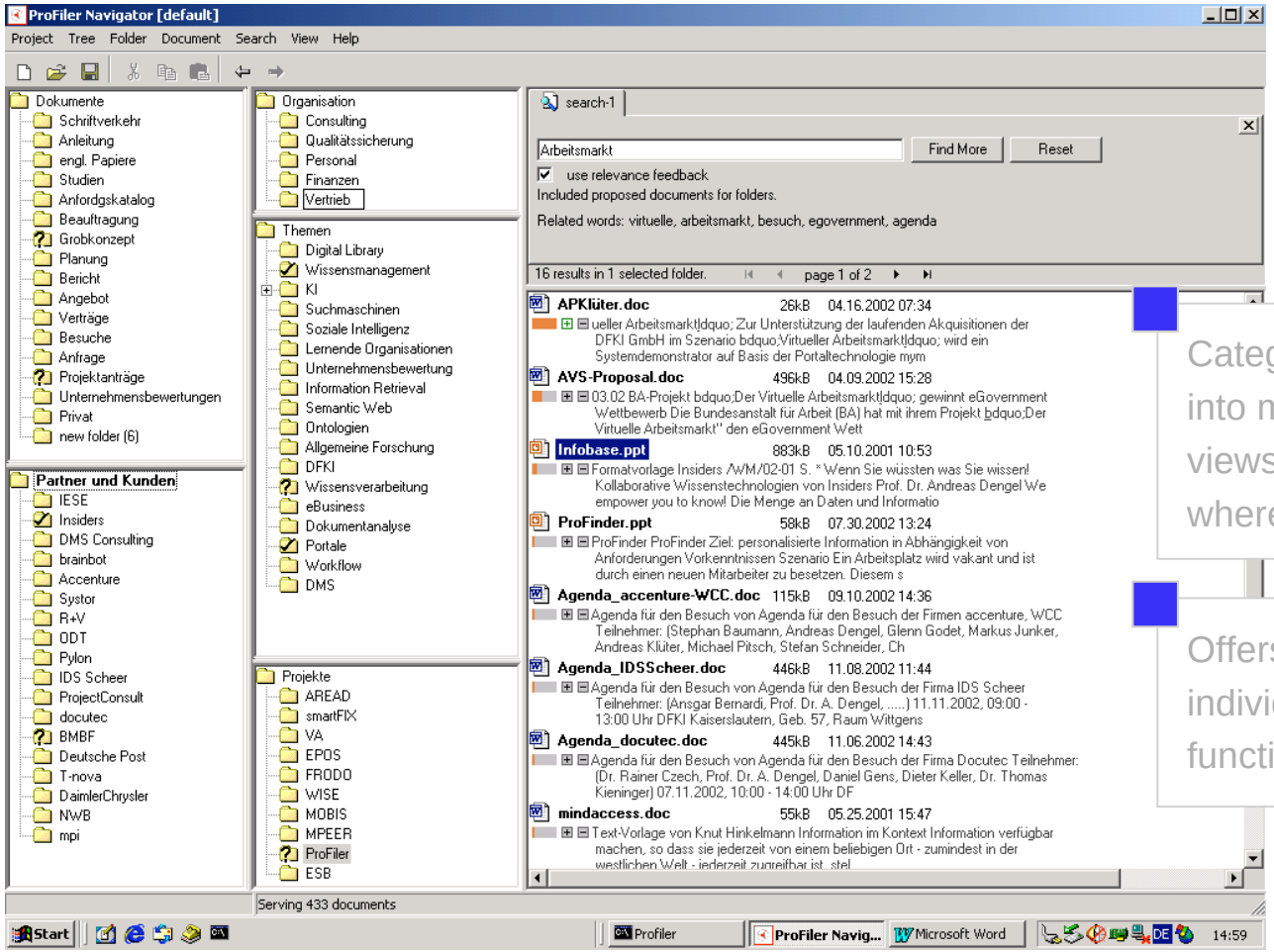
The perception of documents is subjective

In office environments people classify documents according to their preferences, i.e. they generate folders as categories and name them

- Resulting taxonomies correspond to subjective concepts of the world but …
  - … have no unique meaning
  - … do not allow perspective considerations
  - … are not integrative



Personal Memory



Categorizes documents into multi-perspective views (e.g. what, who, where, when, how)

Offers various types of individualized retrieval functions

A. Dengel, *Six Thousand Words about Multi-Perspective Personal Document Management*, Proceedings IEEE-EDM, Key Note Paper, Hong Kong, China (Oct. 2006), pp. 1-10.

There are obvious advantages in such an approach

- Contents (text) of information objects, whether we consider a taxonomy, a folder or a document, may be related or compared in the same way
- Communication between a user and her/his “Personal Memory” is driven by conceptualizations allowing to associate and to imagine in her/his own mental world
- In combination with the perspective directories, the user gets an excellent orientation and access point to her/his information space

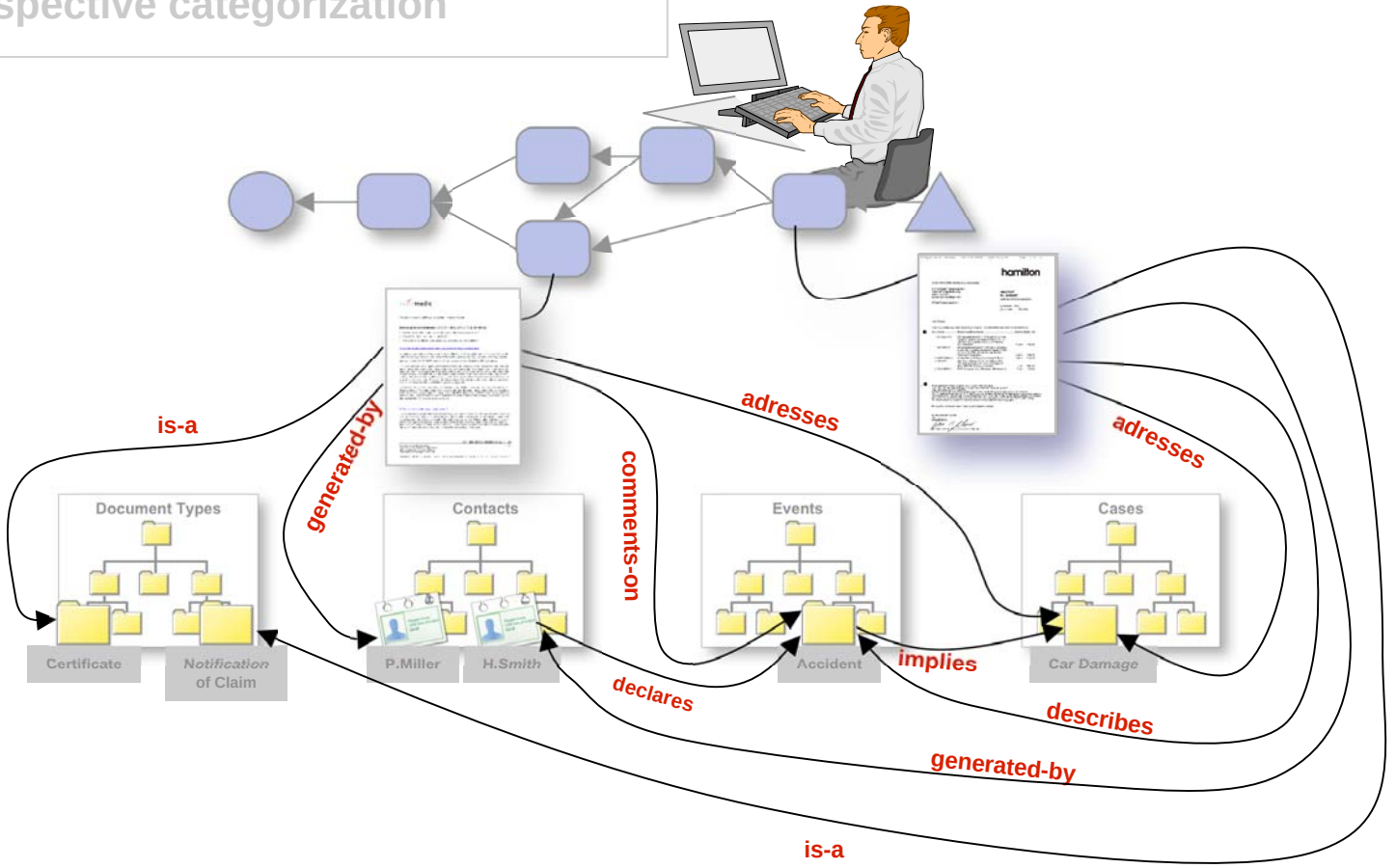
But ...

... how can we profit from the given implicit relationships among information items (resources)?

Generating semantic relations “on the fly”

Some predefined generic relations among folders generate semantics based on multi-perspective categorization

Example: Insurance-Workflow

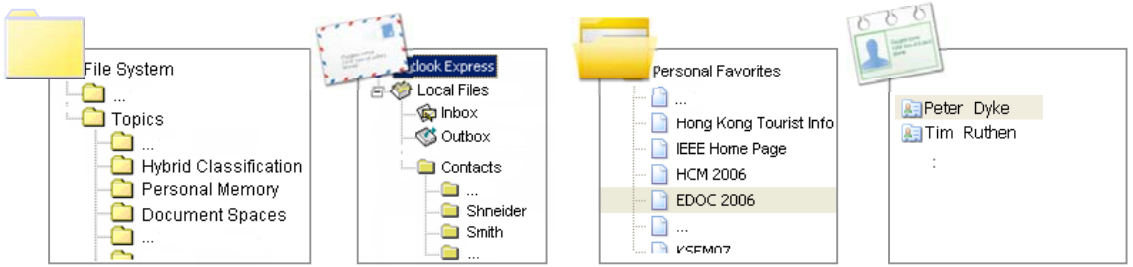


Bringing the Semantic Web to your desktop

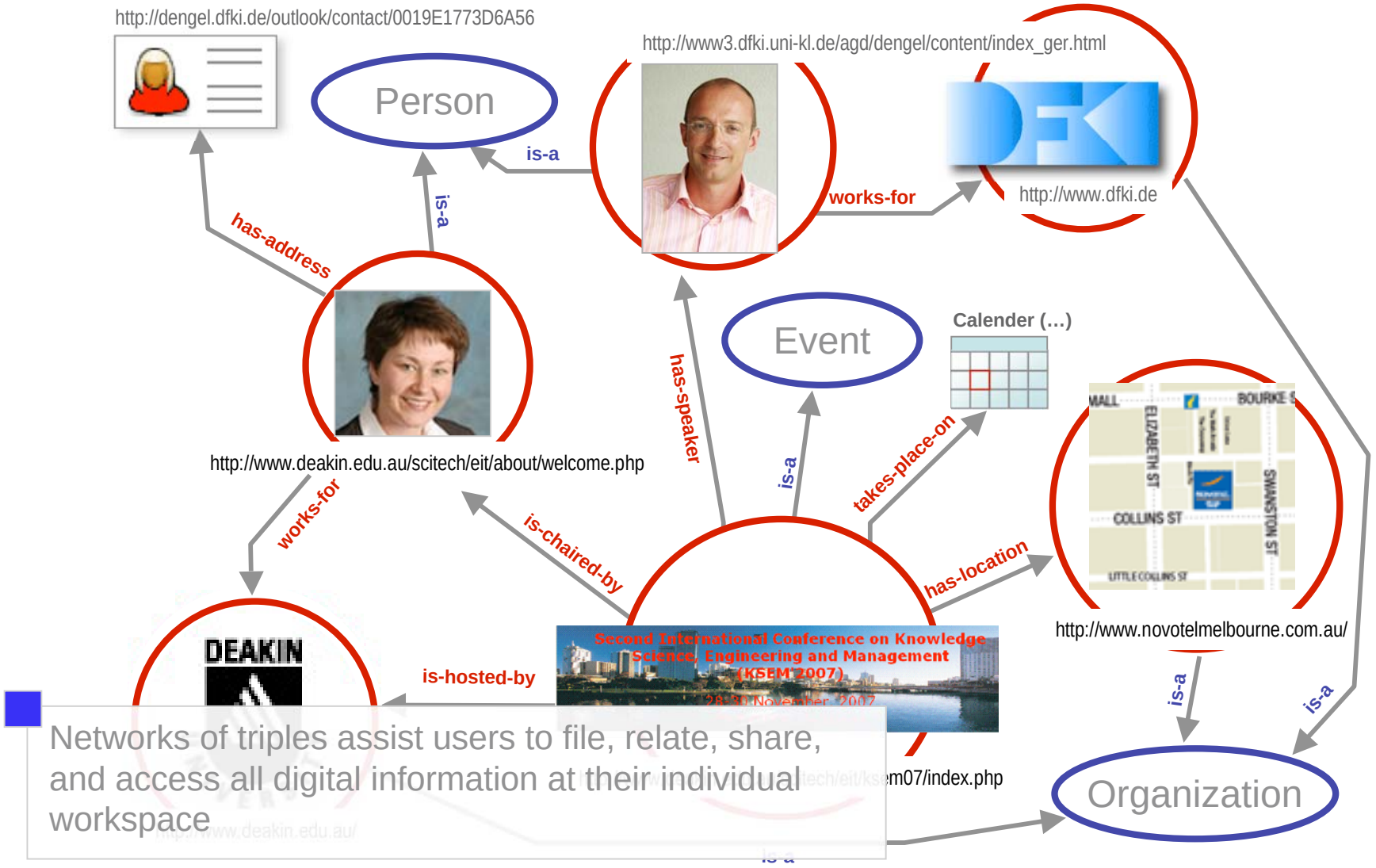
Each information item is a semantic web resource whether it is an email constituent (i.e. message, sender, recipient, attachment), an address (...), or a calendar event, ...

All resources are identified by a URI, such as

- imap://leo@gnowsis.com/INBOX/;UID=3 for an email
- file:///D:/EigeneDateien/Documents/Talks/Google.pdf for a file
- http://www3.dfki.uni-kl.de/agd/dengel/content/index_ger.html for a web site



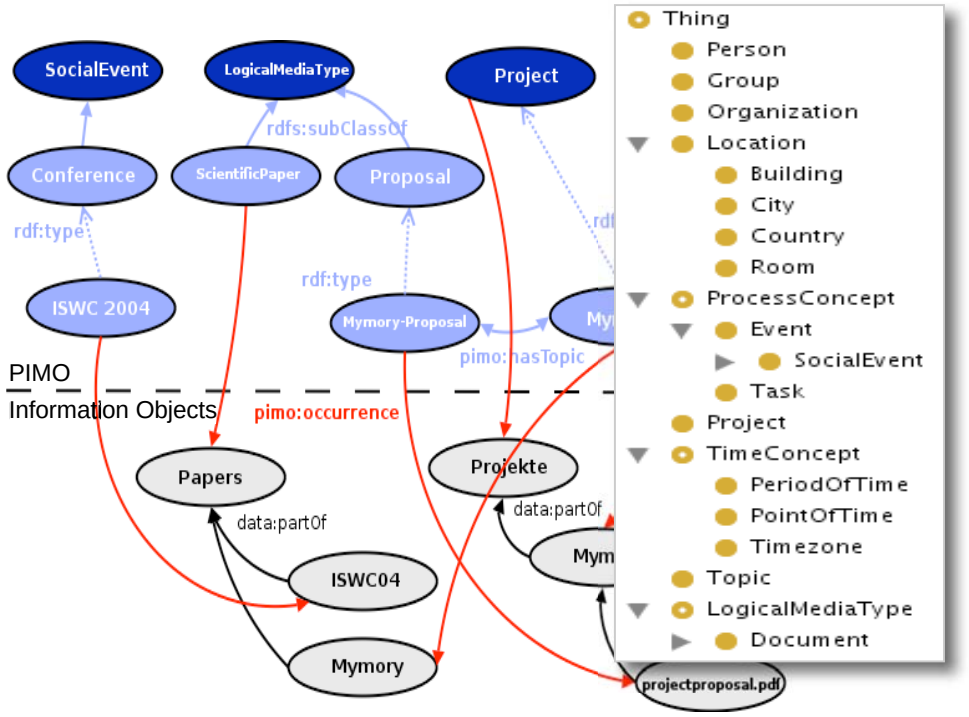
Buiding a Personal Information Model (PIMO)



Networks of triples assist users to file, relate, share, and access all digital information at their individual workspace

Personal Information Model (PIMO)

- PIMO captures all relevant information items across applications using RDF for the data and RDF/S ontologies for the semantics)



- The user may access shared categories, or create instances, classes, and properties (on the fly), and annotate the resources respectively

- Smushing identifies the synonymous resources and aggregates the data

  Promising approach! (see also http://esw.w3.org/topic/RdfSmushing)

L. Sauermann, L. v. Elst and A. Dengel, *PIMO - a Framework for Representing Personal Information Models*, Proceedings I-Media'07 and I-Semantics'07, Graz, Austria, J.UCS (Sep. 2007), pp. 270-277.

From data to PIMO

User Interface

read/write PIMO and resources

“Rebirth Engine”
- creating PIMO things from resources
- matching existing things
- …

Aperture.sf.net
- crawling framework
- extensible
- separated service

PIMO Store

RDF

Matching to Ontology

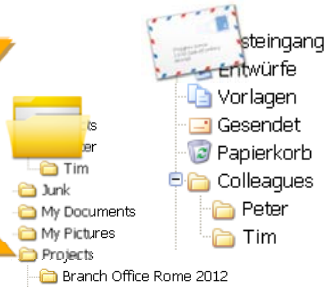
RDF Database

RDF

Adapter to RDF

Data

Person‘s files and e-mails



The first Step: The Semantic Desktop

As a workplace the Semantic Desktop is individually adaptable and service oriented as well as capable to independently compile and process knowledge

PIMO

It is a means to manage all personal information across application borders based on Semantic Web standards

By combining the PIMO with active user observation, the Semantic Desktops acts like an information assistant offering context-aware services

Relevant documents and best-practice solutions are offered using multi-dimensional perspectives with regards to tasks, contexts and processes

L. Sauermann, A. Bernardi and A. Dengel, *Overview and Outlook on the Semantic Desktop*, Proceedings International Semantic Web Conference, Galway, Ireland (Nov. 2005), pp. 1-19.

A. Dengel, *Knowledge Technologies for the Social Semantic Desktop*, in: Z. Zhang and J. Siekmann (Eds.): Proc. KSEM 2007,, LNAI 4798, Springer Publ. (Nov. 2007), pp. 2-9.

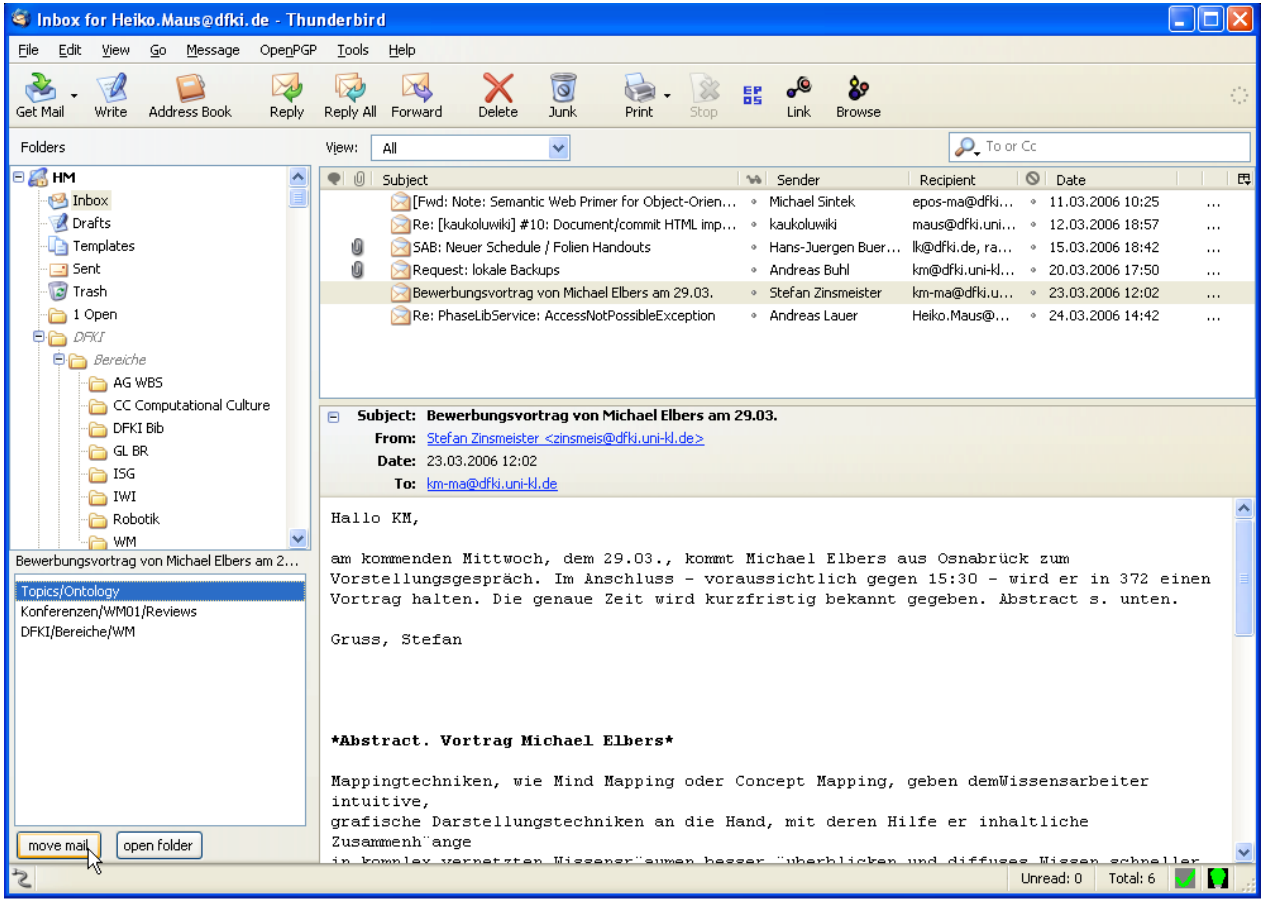
... what is the increased value of the Semantic Desktop

Context-aware services act like mental association

- Actions of knowledge workers are observed and collected in order to understand contextual behavior, i.e. browsing and clicking, reading and writing, task related actions
- Method is geared to human's short term memory, i.e. sensory signals are aggregated to more complex ones
- Observations lead to individualized context-aware services

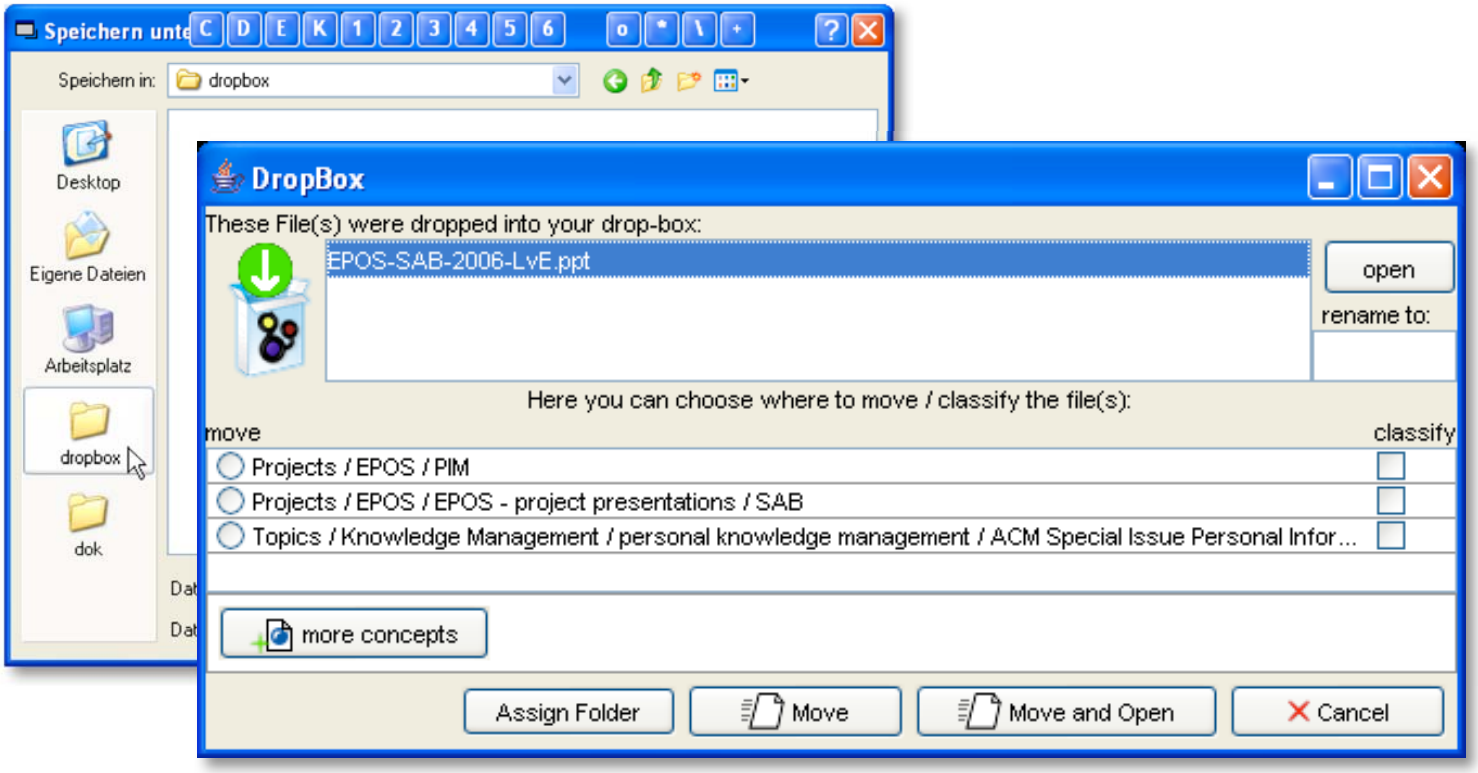
Sven Schwarz: *A Context Model for Personal Knowledge Management Applications*. In Modeling and Retrieval of Context, Second International Workshop, MRC 2005, Edinburgh, UK, July 31 - August 1, 2005

Context-aware services - ctd.



Adjustable concepts given in the PIMO classify incoming emails in order to allow for an automated multi-dimensional semantic filing

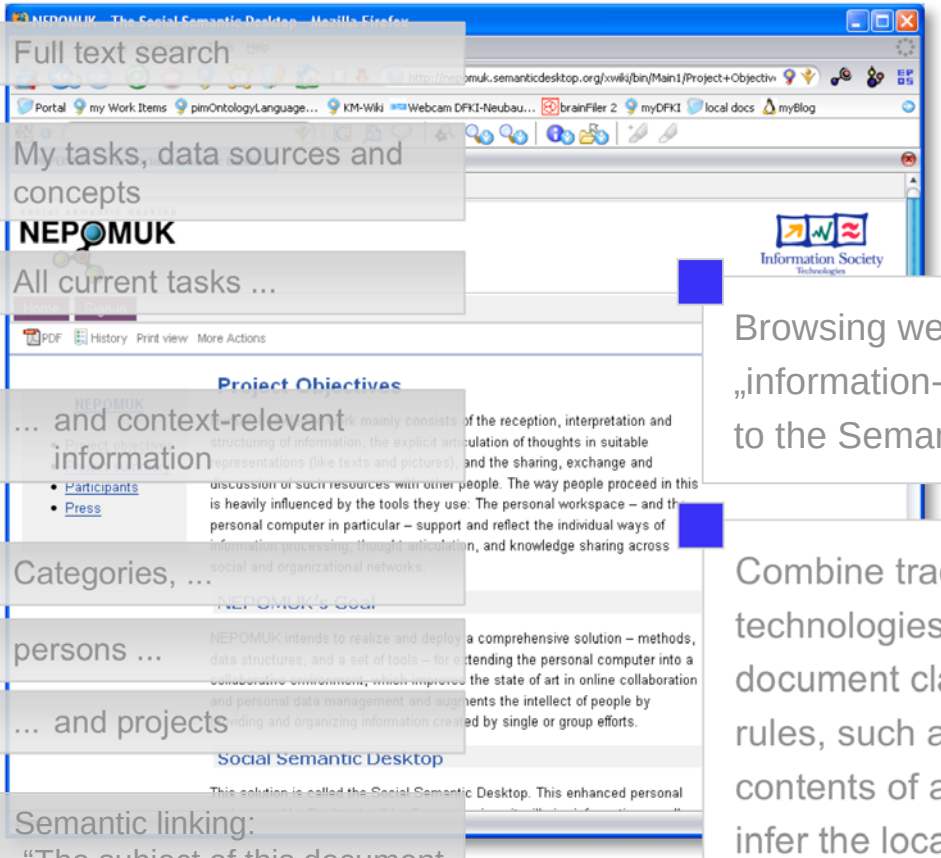
Context-aware services - ctd.



The „Save-as“ function offers recommendations for filing and the subsequent conceptualization of documents

L. Sauermann, A. Dengel, L. van Elst, A.Lauer, H. Maus and S. Schwarz, *Personalization in the EPOS Project*, Proceedings Proceedings of the Semantic Web Personalization Workshop at the ESWC 2006 Conference

Context-aware services - ctd.



Full text search

My tasks, data sources and concepts

All current tasks ...

... and context-relevant information

Categories, ...

persons ...

... and projects

Semantic linking:
"The subject of this document is the project NEPOMUK"

Browsing web sites creates an „information-push" from the PIMO to the Semantic Desktop

Combine traditional IR technologies, i.e. query expansion, document classification, ... with rules, such as "Extract the contents of an invitation e-mail and infer the location and attendees of the meeting"

H. Holz, H. Maus, A. Bernardi, O. Rostanin, *From Lightweight Proactive Information Delivery to Business Process Oriented Knowledge Management*, Journal of Universal Knowledge Management, JUKM, No. 5 (2005), pp. 101-127

Man Luo: *Semantic Meeting Annotation*. Diploma Thesis, (DFKI 2006).

Context-aware services - ctd.

DocuTag Semantic Tag Recommendations

Recommended Entities

Token: Entity: 0.5

across abap/4-programmi

Project

epos
nepomuk
aperture
gnowsis

Person

thomas roth-berghofer
malte kiesel
ansgar bernardi
dominik heim
sven schwarz
michael sintek
benjamin adrian

NEPOMUK

Leo Sauermann

Keywords

Gnowsis

NEPOMUK NEPOMUK aims at empowering individual knowledge workers to better exploit their personal information space and to maintain fruitful communication and exchange within social networks across organizational boundaries. NEPOMUK brings together researchers, industrial software developers, and representative industrial users, to develop a comprehensive solution for extending the personal desktop into a collaboration environment which supports both the personal information management and the sharing and exchange across social and organizational relations. This solution is called the Social Semantic Desktop. It will comprise a set of technical and methodological solutions for Supporting the knowledge life cycle, in particular the generation and exchange of personal thoughts via structured articulation in extended wiki-based semantic tools, goal-oriented organization by work process model integration, and sharing, exchange and alignment of metadata the management of all relevant information in the personal workspace via cross-media and cross-application linking and browsing of information items based on standard semantic web data structures, together with un-intrusive metadata generation support, knowledge communication within social networks and distributed search and storage to build, maintain, and employ interworkspace relations in large scale distributed scenarios. NEPOMUK realizes an open-source framework which allows integration of thirdparty components via pluggable adaptors. The core components and approaches will be actively promoted in open development communities in order to reach an early large-scale uptake and feedback. ...interfaces and data structures will be submitted to the appropriate ...committees like OASIS or W3C. The technical work packages ...tic desktop comprise tools for knowledge articulation and ...and data structures of the personal semantic web, and ...process support. The workpackages covering the ...the issues of distributed sear... ...and knowledge exchange. All te... ...tecture work package. Four... ...t, and test the solution in kn... ...stribution, organizational c... ...on dissemination and comm... ...ake. The project delivers a... ...semantic desktops, and...

Extracted Tokens and Symbols

insert new topic 0.5

across (-) activities (-) ansgar (-) ansgar bernardi (-) application (-) application-specific (-) architecture (-) articulation (-) benjamin (-) benjamin horak (-) branch-specific (-) collaboration (-) communication (-) community (-) components (-) comprise (-) consulting (-) cross-application (-) dfki (-) different (-) dissemination (-) distribution (-) employ (-) exchange (-) exploitation (-) generation (-) gnowsis (-) implementations (-) information (-) integration (-) interfaces (-) interworkspace (-) introducing (-) knowledge (-) large-scale (-) leo (-) leo sauermann (-) linking (-) maintain (-) management (-) metadata (-) microsoft (-) nepomuk (-) networks (-) office (-) open (-) open-source (-) organization (-) organizational (-) particular (-) process (-) project (-) promoted (-) reference (-) relations (-) relevant (-) representative

Kaukoluwiki

Menu

Main page
About
Recent Changes
WikiEtiquette
Find pages
Unused pages
Undefined pages
Page Index
Main
Projects
NEPOMUK
MyMory
People
Meetings
Documents

Your trail: NEPOMUK » Projects

NEPOMUK

Human knowledge work mainly consists of the reception, interpretation and structuring of information, the explicit articulation of thoughts in suitable representations (like texts and pictures), and the sharing, exchange and discussion of such resources with other people. The way people proceed in this is heavily influenced by the tools they use. The personal workspace - and the personal computer in particular – support and reflect the individual ways of information processing, thought articulation, and knowledge sharing across social and organizational networks.

Edit page Add Comment View page Page Info View RDF

tagthe.net

topicalizer

Wordnet

Web 2.0

Input:	Documents
Output:	Ranked list of 'Things' occurring in the document
Computation:	Choreography of Web 2.0 Services
Vocabularies:	RDFS, SKOS, PIMO language
RDF Store:	Gnowsis/Sesame2

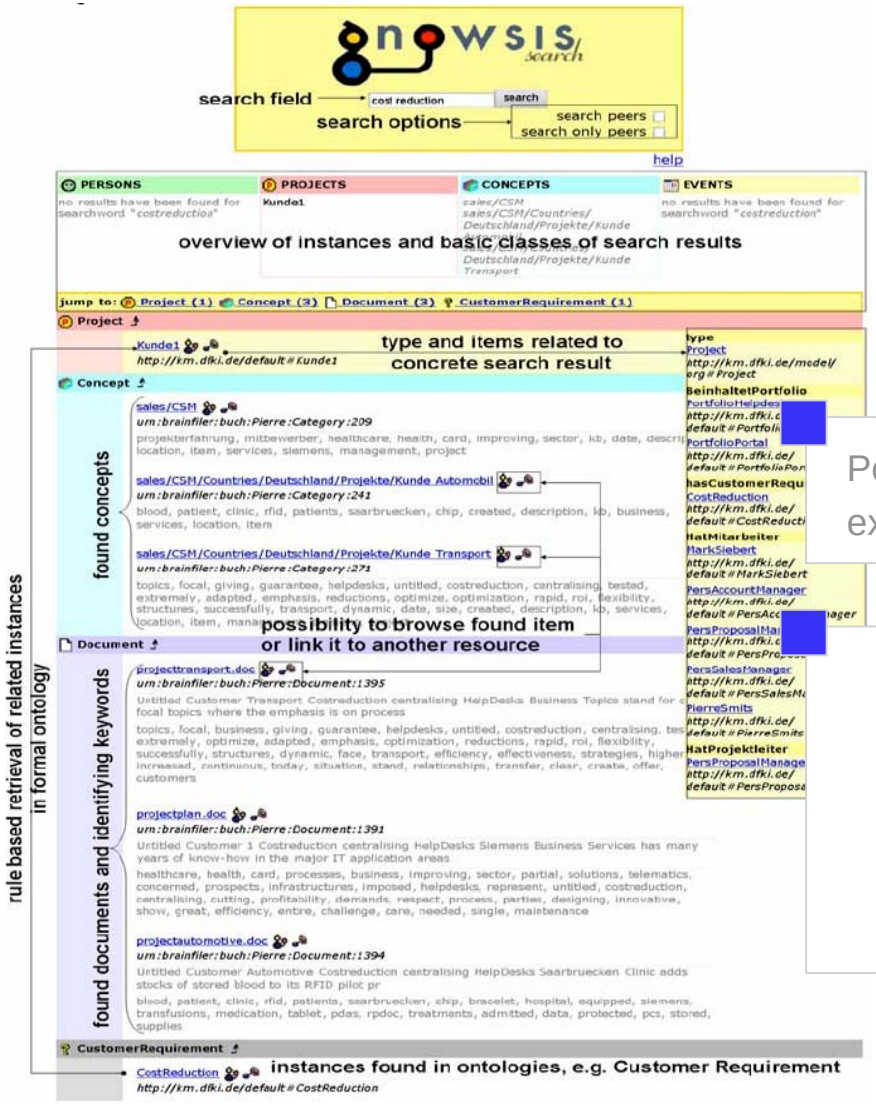
KM@DFKI Knowledge Management

Adaptive semantic search via *gnowsis*

- Exploit semantic links to search intelligent through all office information
- Context-aware and personalized to offer a better understanding of user needs
- Semantically described search results to meet the user's mental models

L. Sauermann, G.A. Grimnes, M. Kiesel, C. Fluit, D. Heim, D. Nadeem, B. Horak and A. Dengel, *Semantic Desktop 2.0: The Gnowsis Experience,* Proceedings 5th Int'l Semantic Web Conference, Athens, GA, USA LNCS 4273, Springer Publ. (Nov. 2006), pp. 887-900.

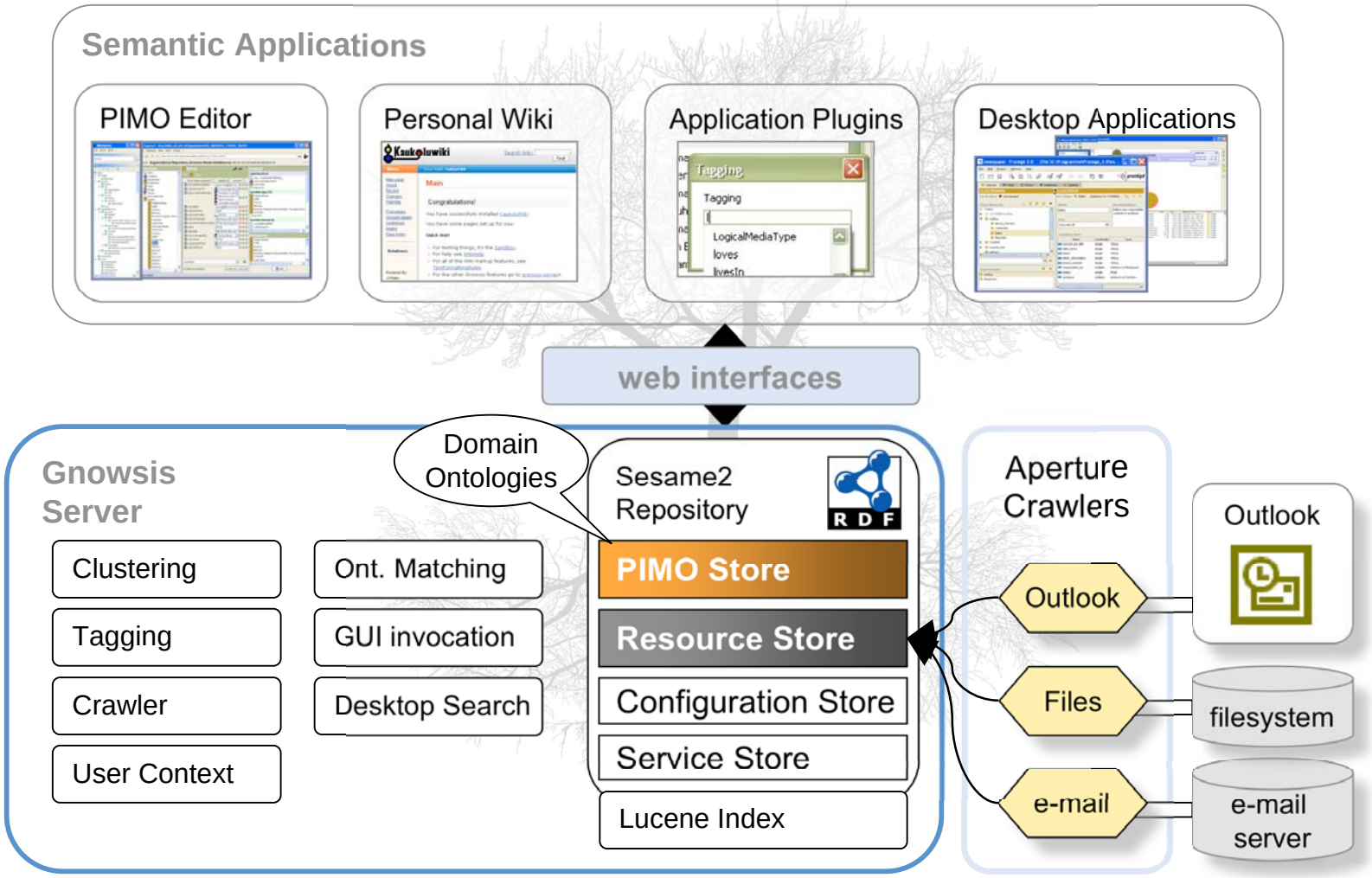
Adaptive semantic search via *gnowsis* - ctd.



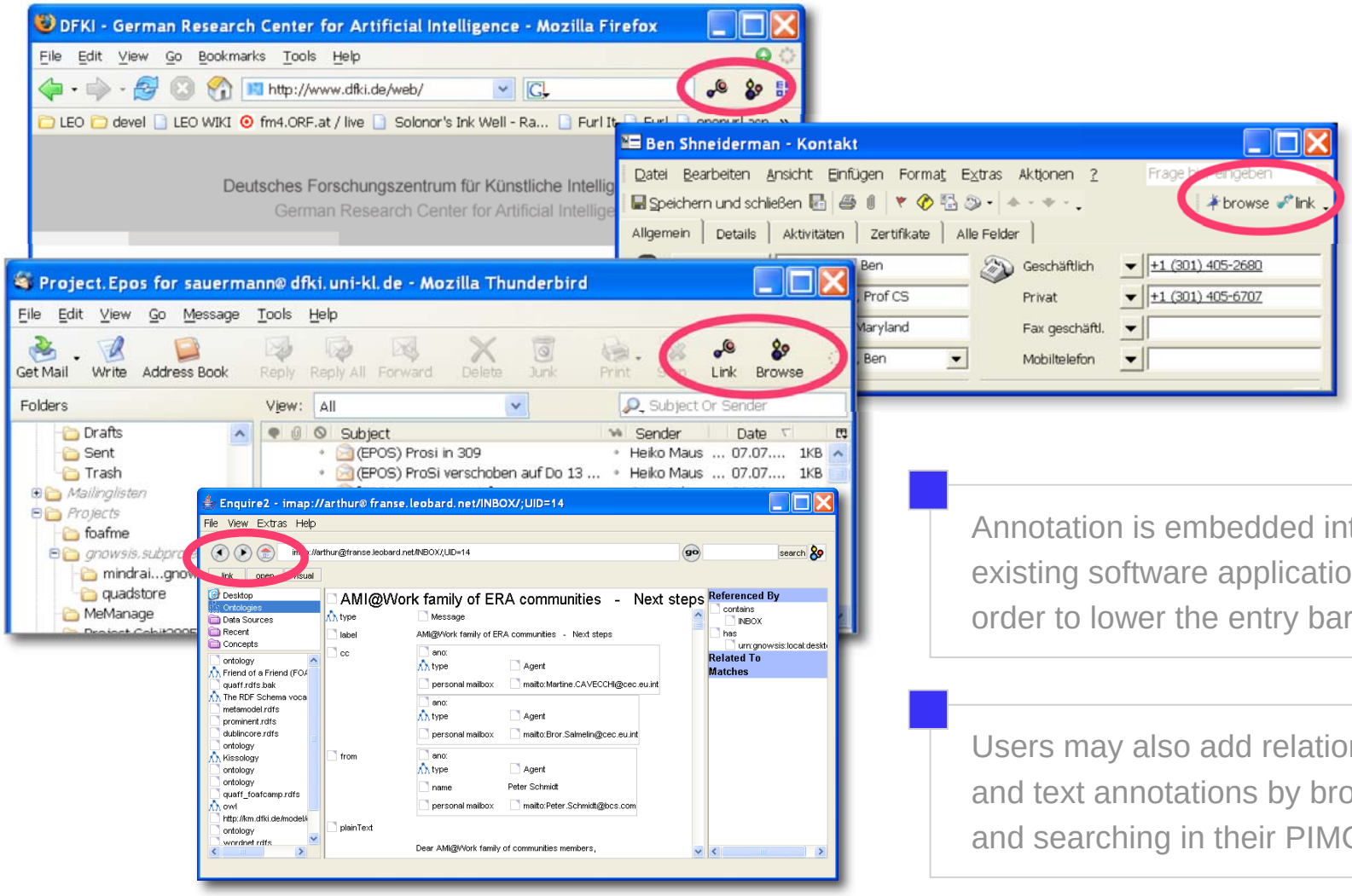
Personalized search in the PIMO via the expansion of customized SPARQL rules

```
# found a project? -> also show members
(?hit retrieve:item ?project),
(?project rdf:type org:Project) ->
   querySparql('CONSTRUCT   {
   ?project org:containsMember ?m.
   })
```

Semantic Desktop System Architecture



Manual annotation complements the automatic means

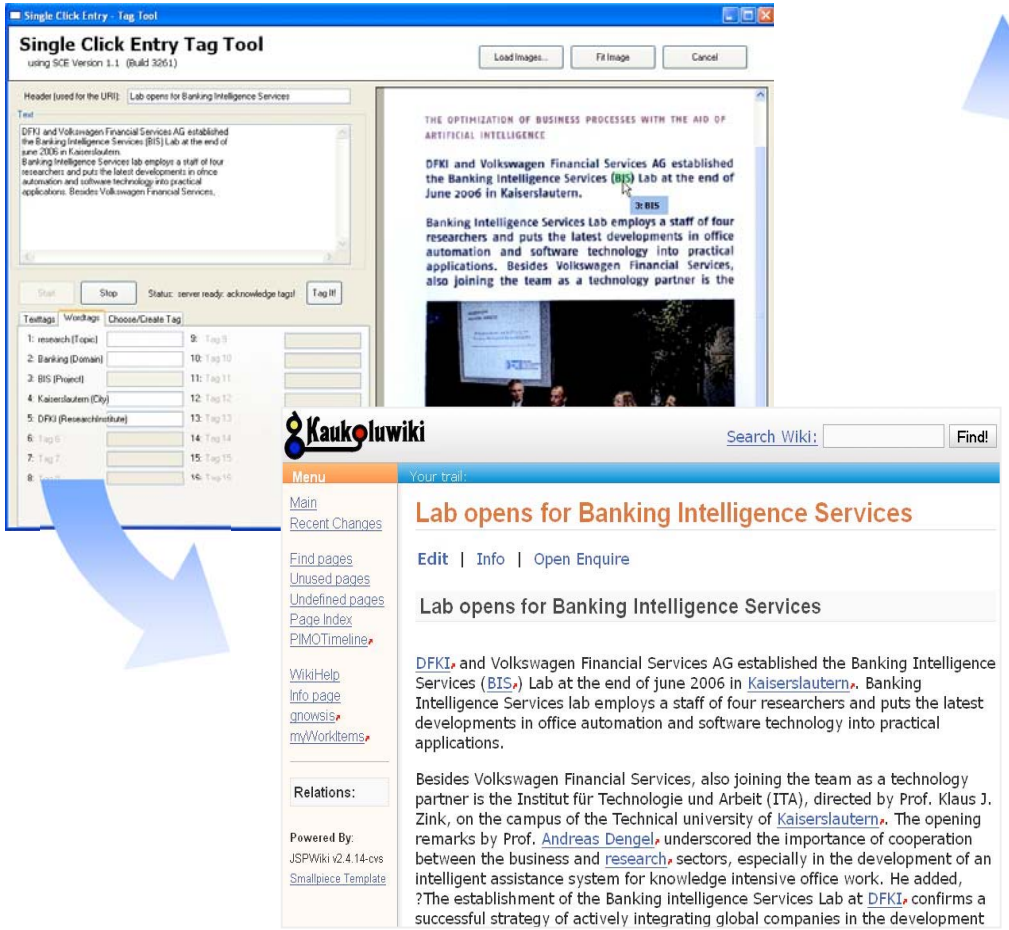


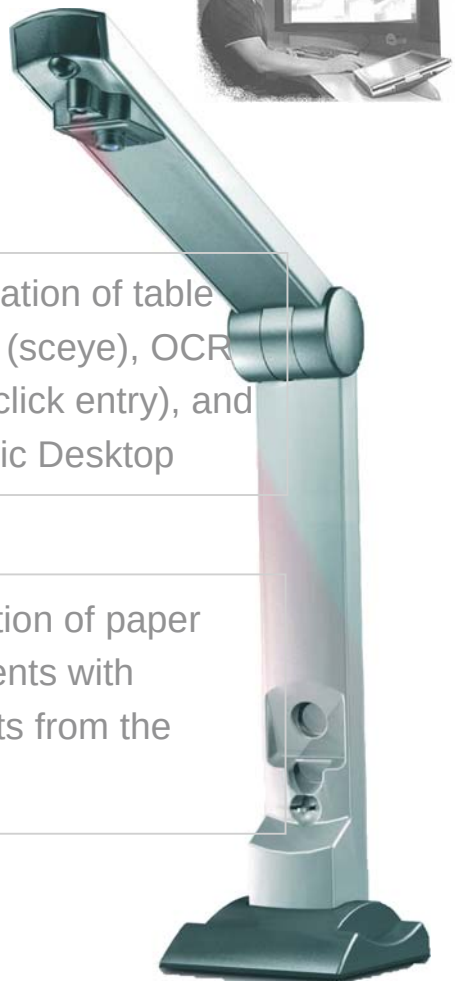
- Annotation is embedded into existing software applications in order to lower the entry barrier
- Users may also add relations and text annotations by browsing and searching in their PIMO

... and how can we bridge the gap to Gutenberg's world?

(find resources within documents)

Semantic Annotation of Paper-Based Information





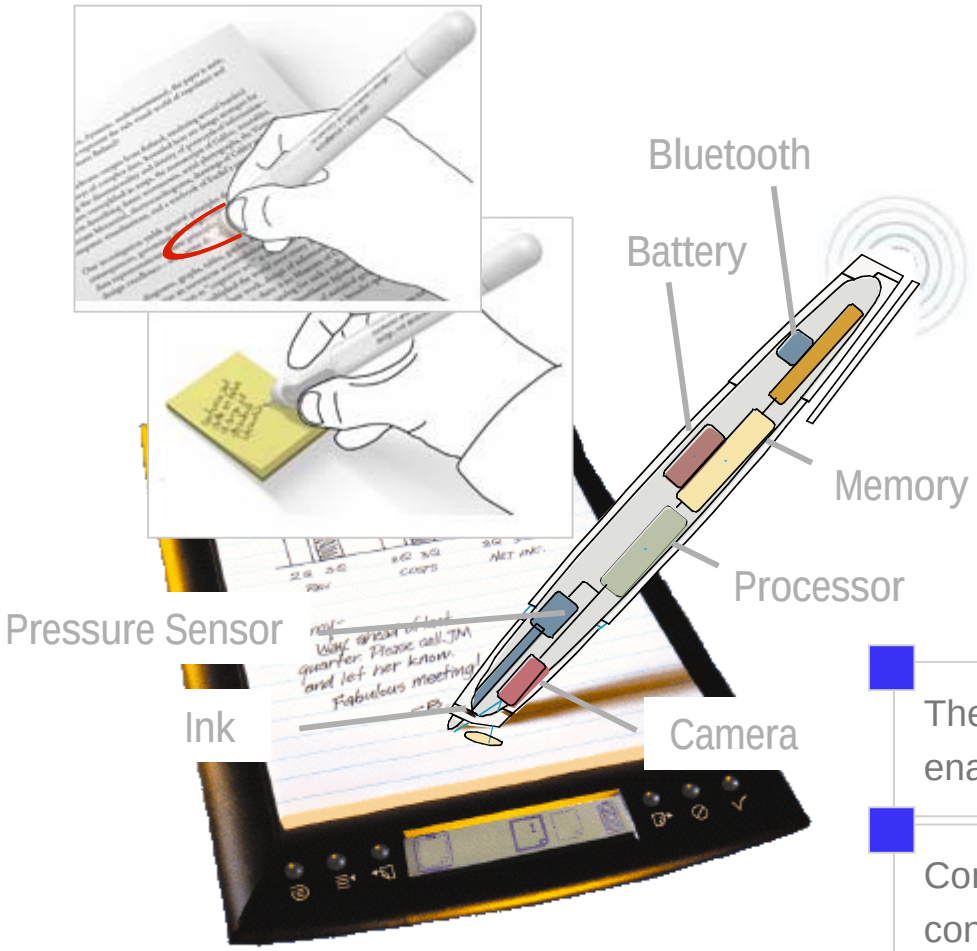
- Combination of table camera (sceye), OCR (single click entry), and Semantic Desktop
- Annotation of paper documents with concepts from the PIMO

H. Maus and A. Dengel, *Semantic Annotation of paper-based Information*, Proceedings CBDAR 2008, Curitiba, Brasil (Sep. 2007), pp. 158-160.

M. Kiesel and L. Sauermann, *Towards Semantic Desktop Wikis*, UPGRADE special issue on "The Semantic Web", Volume VI, pp. 30-34

In cooperation with: sceye Océ

Pen-based Knowledge Extraction

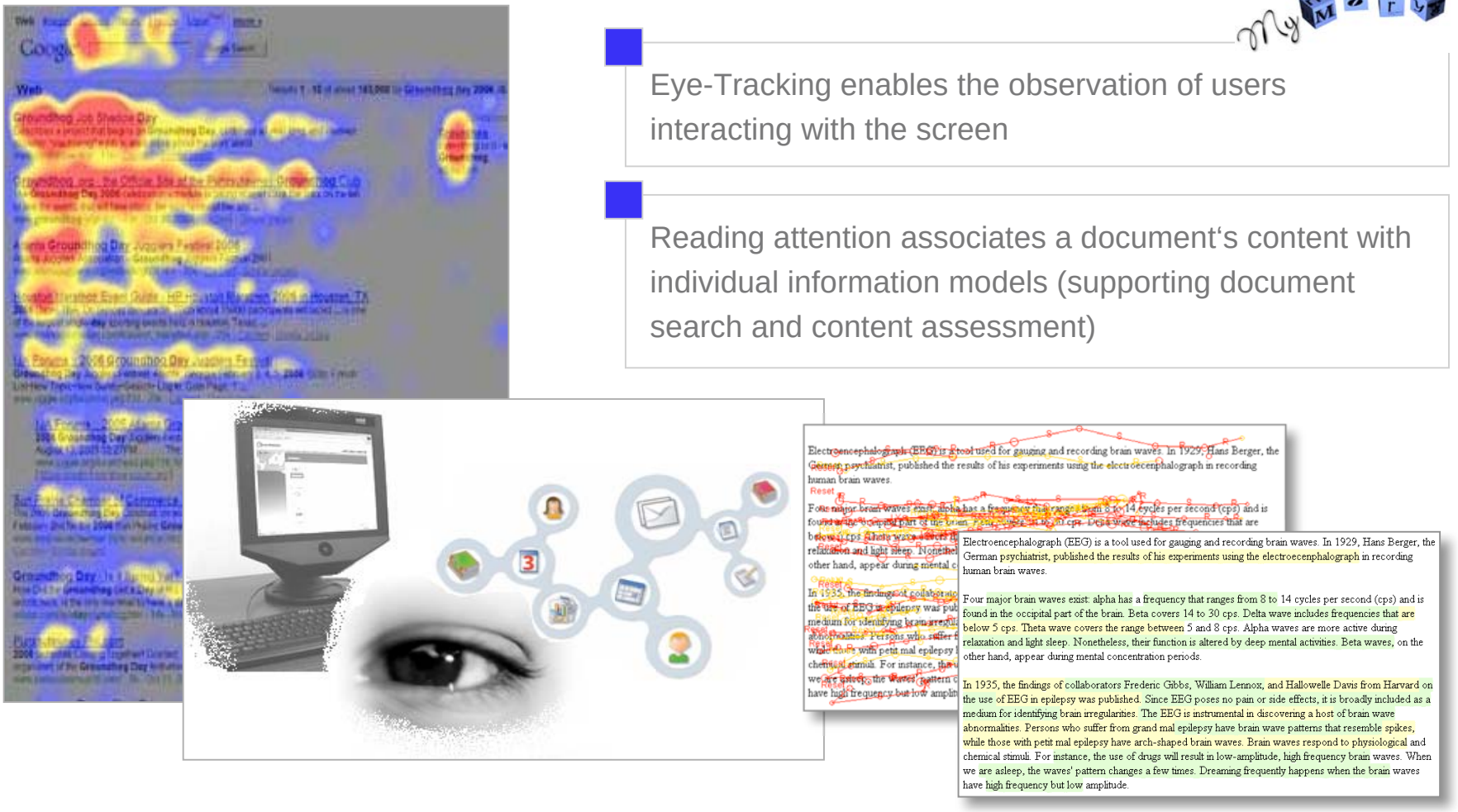


- The consideration of gestures and handwriting enables intuitive document processing
- Content is semantically linked to existing contacts, topics, events and other relevant documents

... but is there also a chance to learn more from user observation?

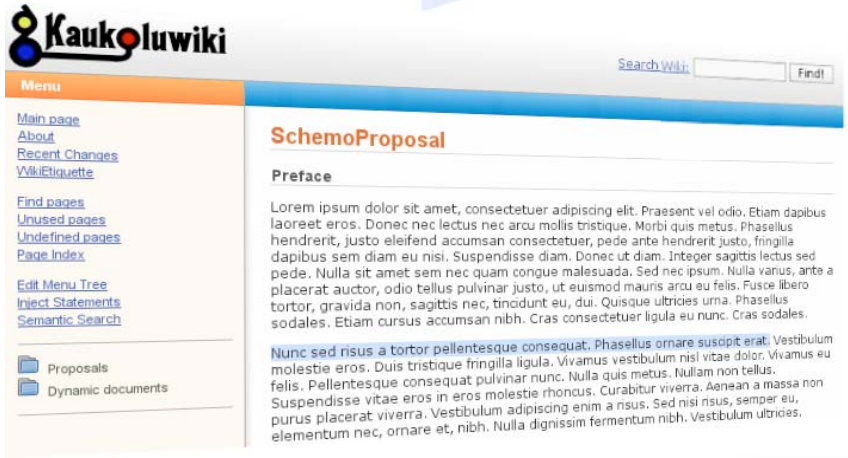
Attention-based Document Processing

- Eye-Tracking enables the observation of users interacting with the screen
- Reading attention associates a document's content with individual information models (supporting document search and content assessment)



G. Buscher, A. Dengel and L. van Elst, *High Level Eye Movement Measures for Relevance Assessments of Information Items*, submitted to CHI 2008, Florence, Italy (Apr. 2007).

Exploiting Attention Data within Semantic Wikis



- Keep track of attention and context information in your personal workspace
- Query and construct documents based on attention and context information

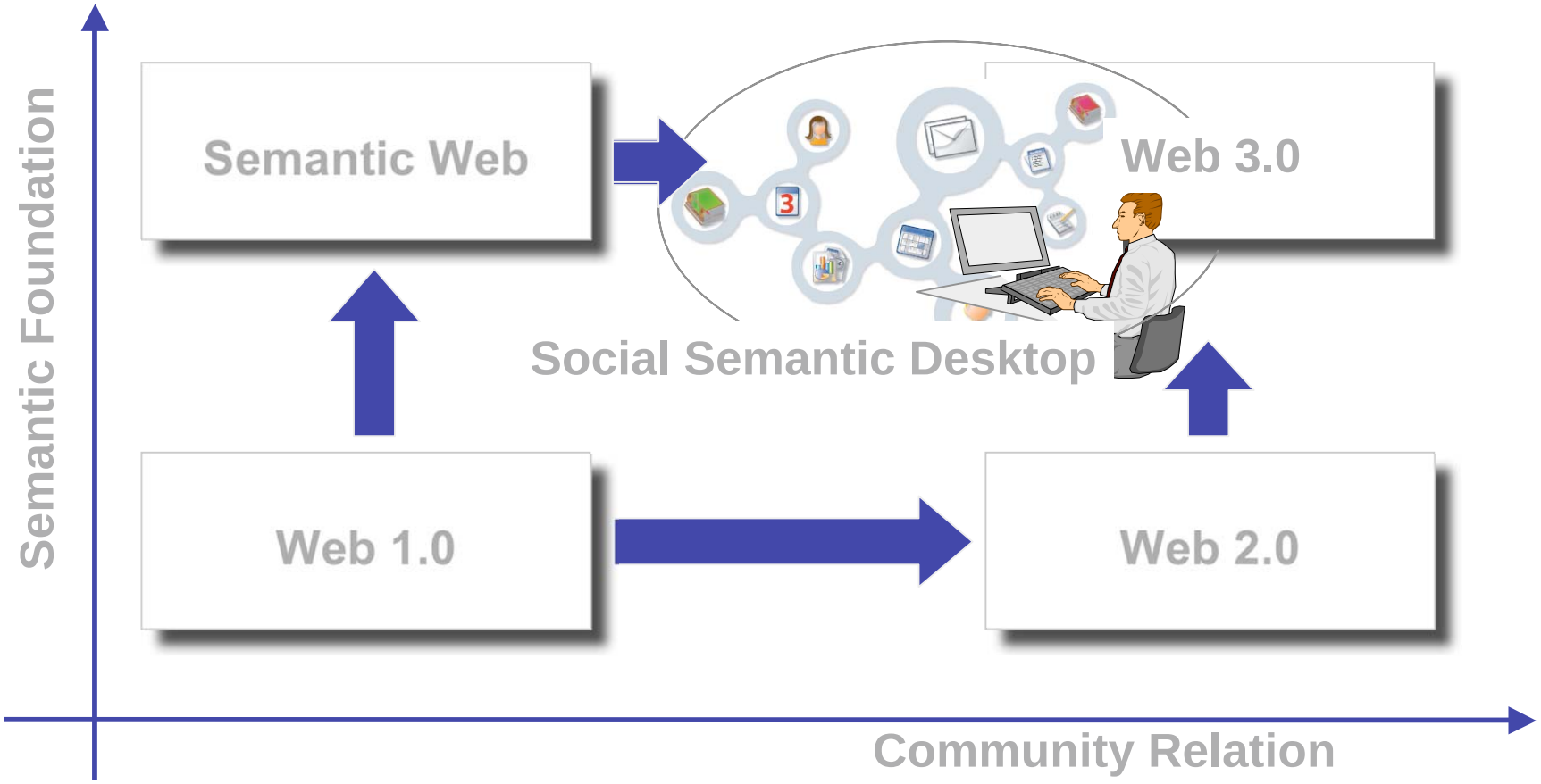
... and what are the chances of this approach?

Summary

- The traditional Web has recently undergone an orthogonal shift into a Web of People/Web 2.0
- Focus is set on folksonomies, collective intelligence, and the wisdom of trusted communities which influences office work as well
- The Semantic Desktop is a driving paradigm for desktop computing using Semantic Web standards but integrating native office applications and data
- The Web became part of our thinking and part of our workspace, and the documents we generate at our workspace become part of the Web
- Friend networks allow people to link with their friends and to traverse the network via these profiles, as well as to give comments, votes, and recommendations on their content published

... towards the Social Semantic Desktop

Our strategy considers two major trends



The next step: The Social Semantic Desktop

social semantic desktop
NEPOMUK

Semantic explicit knowledge becomes processable by computers

Social entails the demand of exchanging and interlinking knowledge from and among different workspaces

NEPOMUK realizes the basis for manifold exploitation

Individual exploitation concentrates on uptake, adaptation, and commercialization by tool adaptors & consultation services, i.e. spin-offs

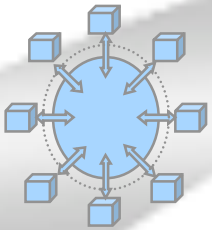
1 After first 18 months (initial project results are available)

social semantic desktop NEPOMUK

Exploit internally

- Internal exploitation
- Initial awareness via open source and scientific communities

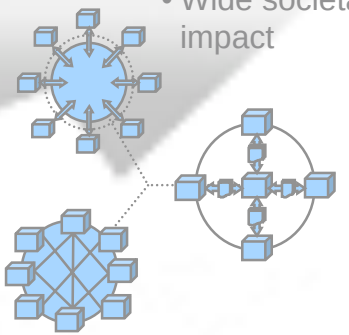
2 After 2-3 years



Exploit externally

- External application of project results by industrial partners
- Uptake in the scientific community
- Wide-scale awareness within open source community

3 After 3-5 years



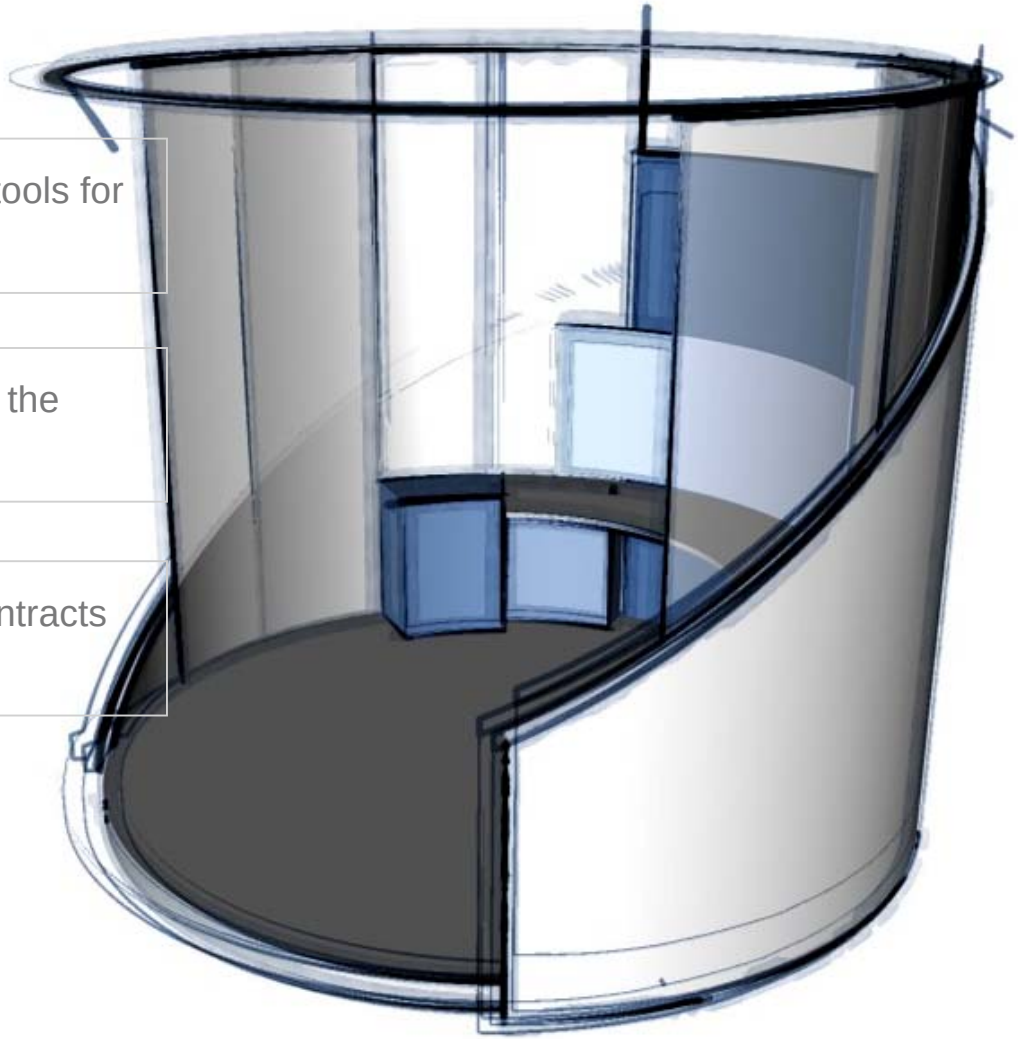
- Commercial exploitation by dedicated spin-offs
- Specific product development
- Wide societal impact

Collective exploitation will focus on establishing & maintenance of a vivid scientific society

- Pursuing standardization and platform & prototype development
- Collective support of standardization efforts (e.g. W3C working group)

Future office workspace as an integrating platform*

- Combine recent trends in physical tools for knowledge work support
- Show case in order to demonstrate the value of the semantic desktop
- Attract potential partners to give contracts to order to share DFKI's vision



* will be available end of October 2007

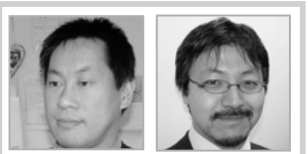
Thanks to my group!

Ass Profs & Post Docs

PhD Students

Software Engineers

Guests



Plus about 40 Master Students

Thanks to you for your attention!

Prof. Dr. Andreas Dengel
DFKI GmbH
P.O. Box 2080
D-67608 Kaiserslautern
email: andreas.dengel@dfki.de
http://www3.dfki.uni-kl.de/agd/dengel/content/index_eng.html