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?
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.
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 Enterprise 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.
Imaginations without terms are blind, terms without imaginations are empty

> I. Kant (1724 – 1804)

... is reflected within mental models
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

Files System

Email-Folders

Bookmarks

© andreas.dengel@dfki.de - 2007
Personal Memory

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

Offers various types of individualized retrieval functions

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

- is-a
- generated-by
- comments-on
- declares
- implies
- describes
- addresses

Certificate
Notification of Claim
Accident
Car Damage
Certificate
P.Miller
H.Smith

implies

© andreas.dengel@dfki.de - 2007
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
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](http://esw.w3.org/topic/RdfSmushing))


© andreas.dengel@dfki.de - 2007
From data to PIMO

User Interface
- Person's files and e-mails

Data
- Person's files and e-mails

```
© andreas.dengel@dfki.de  -  2007
```

"Rebirth Engine"
- creating PIMO things from resources
- matching existing things
- ...

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

read/write PIMO and resources
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.

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.


© andreas.dengel@dfki.de - 2007
... 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.
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.

Context-aware services - ctd.

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”


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
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

Adaptive semantic search via *gnowsis* - ctd.

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

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


In cooperation with: © andreas.dengel@dfki.de - 2007
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).

Exploiting Attention Data within Semantic Wikis

Query and construct documents based on attention and context information

Keep track of attention and context information in your personal workspace

© andreas.dengel@dfki.de - 2007
... 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:

- **Semantic Foundation**
  - Semantic Web
  - Web 1.0
  - Web 2.0

- **Community Relation**
  - Social Semantic Desktop
  - Web 3.0
The next step: The Social Semantic Desktop

**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

After 1-2 years
- Initial exploitation
- Initial awareness via open source and scientific communities

Exploit internally

After first 18 months (initial project results are available)
- External application of project results by industrial partners
- Uptake in the scientific community
- Wide-scale awareness within open source community

Exploit externally

After 2-3 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