The Semantic Desktop
Supplementing a User‘s Memory

Andreas Dengel
Agenda

- Introduction to a typical knowledge worker and his problems in information management
- Some words about human perception and the role of ontologies
- Ontology creation and instantiation
- The Semantic Desktop
- Document Understanding via the Semantic Desktop
- Context-aware semantic services for the knowledge worker
Let me introduce Thomas to you, a typical but fictive knowledge worker

- Dr. Thomas Mustermann is head of Customer Relationship Management at the German Research Center for Artificial Intelligence (DFKI) in Kaiserslautern, Germany

- He is one of those guys who are tackled by Constant Multi-Tasking Craziness. Right now he works on about 15 tasks at the same time one of which is the organization of a training course on DFKI’s recent development: The Semantic Desktop

- When Andreas Dengel came back from APE 2010 held in Berlin, he told Thomas to invite Arnoud de Kemp (another knowledge worker) to the training course

- To fulfill his role and to complete all of his tasks, including the one he just received, Thomas has to acquire, organize, maintain, retrieve and use a whole bunch of information items

... but
Thomas’ problem, as for many knowledge workers, is the need to know more than he may remember.

On February 9th we will host a training course on the Semantic Desktop in Kaiserslautern and I was told to invite Arnoud.

On Thomas’ desktop, he stores around 12,000 files in about 2,300 folders in many applications and formats.
So what can we do in order to help knowledge workers like Thomas to do their job better?
When sending out the invitation, State-of-the-Art technologies may extract and store metadata about a document.

Example:

"Thomas sends invitation"

From: thomas.mustermann@dfki.de
Date: January 21, 2010 09:28:11
To: Arnoud de Kemp <dekemp@aka-verlag.com>
Subject: Semantic Desktop

Dear Arnoud:

Attached please find the agenda for our next training course on the Semantic Desktop. We would be very glad to welcome you.

If you have any question, don’t hesitate to contact me.

Best regards,
Thomas

Dr. Thomas Mustermann
Head of CRM
DFKI GmbH
Trippstadter Straße 122, 67663 Kaiserslautern, Germany
Phone: +49-631-20575-100
Email: Thomas.mustermann@dfki.de

... but is this enough?

© andreas.dengel@dfki.de - 2009 - Page 6
"We should no longer ask if a single information item we can avail of is sufficient.

We should rather ask how to establish a vivid, interactive Information Butler helping us to associate and to utilize the bits of knowledge captured within our directories."
How can we build an electronic memory like the MEMEX that helps us to remember?

The knowledge of humans is associative and perspective in which computers are lacking!

- Human beings may relate the content of a document into context because they have appropriate background knowledge.

- Understanding a message leads to concepts and ideas being part of mental models generated while reading.

- Computers may read the contents of documents but are not able to understand.
The limits of today's desktop information management lead to a cut between mental models and document contents.

Different applications manage different data

Emails is filed in Email folder

Attachments are stored in file folder

Sender of an Email is stored in an independent address repository

Related Websites are disregarded
From: Arnoud de Kemp <dekemp@aka-verlag.com>
Date: January 23, 2010 11:33:17
To: thomas.mustermann@dfki.de
Subject: Re: Semantic Desktop

Hi Thomas - its nice to hear from you.
Many thanks for sending the agenda. I am glad to accept and confirm my participation in the Semantic Desktop course on February 9.

Warm regards,
Arnoud

----------------------------------
Arnoud de Kemp
AKA GmbH
Postfach 103305
69023 Heidelberg
Germany
Email: dekemp@aka-verlag.com

A document is like a key, which while reading opens a system of links to other documents, to events, locations, persons, or tasks.

As part of a process, a document contains many textual relations to existing knowledge that is complemented by new facts and relationships captured in the document.

Thus, the unit of a message is variable and relative, depending on who reads it at what time and in which context.

“Arnoud confirms invitation“
Imaginations without terms are blind and terms without imaginations are empty*

- Our environment consists of items, facts and events that are „real“ and determine our lives („what is going on“)

- In order to express their thoughts, people use signs, symbols, or characters that may be understood by others („what I couch or explicate“)

- People reading texts put contents together and create their very individual imagination („what I mean“)

* I. Kant (1724 – 1804)
Hi Thomas -

it's nice to hear from you.

Many thanks for sending the agenda. I am glad to accept and confirm my participation in the Semantic Desktop course on February 9.

Warm regards,
Arnoud
Why don’t we make these mental models explicit so that people like Thomas and Arnoud may better remember all they have to do?
One approach is to look on the ideas of the Semantic Web that builds on predication and ontology to formally represent semantics.

A Theory of Ontology attempts to give answers to the question: \textbf{What is there?} 
(the Greek terms „ontos“ and „logos“ mean „to be“ and „word“)

Aristotle defined a system of ten categories, such as substance, quality, quantity, where, when, …

A Theory of Predication tries to answer the question: \textbf{What is it to say something about something?}

A subject is what a statement is about

A predicate is what a statement says about its subject

A common definition of an Ontology for Semantic Web researchers \textbf{is an explicit, formal specification of a conceptualization} \textless Tom Gruber, 1993 \textgreater
The enhanced Resource Description Framework (RDFS) provides the basis for describing meaning via ontologies.
An ontology provides a **shared vocabulary** to express facts about the world.

A fact is expressed as a Subject-Predicate-Object triple

Subjects, predicates, and objects are given as names for entities, also called resources or nodes

Entities *represent something*, *a person*, *an appointment*, *a website*, ...

Names are URIs, which are global in scope, always referring to the same entity in any RDF document in which they appear

The underlying structure of any knowledge can be viewed as a graph (of triples) consisting of nodes (subjects, objects) and labeled directed arcs (predicates) that link pairs of nodes.

```
<rdf:RDF>
  <rdf:Description rdf:about="http://www.aka-verlag.com/deKemp">
    <has-phone-no> 5082403</has-phone-no>
  </rdf:Description>
</rdf:RDF>
```
... but how can be provide a shared vocabulary that is understood by all users of any semantic technology?
The Semantic Desktop is an evolutionary approach towards the Semantic Web

A Semantic Desktop is a device in which an individual stores all her digital information like documents, multimedia and message interpreted as Semantic Web resources. Each resource is identified by a Uniform Resource Identifier (URI) and all data is accessible and queryable as RDF graph.

That way, 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 store, read and communicate via ontologies and Semantic Web protocols. The Semantic Desktop is an enlarged supplement to the user's memory.

Ontologies describe a particular vocabulary that can be used to describe aspects of real domains.

- The vocabulary may follow different “W-Dimensions” of knowledge (what, who, when, where, …)

- All workflow-relevant aspects of information can be described using a set of explicit categories

- The categories can be taken from other applications and formally represented using RDFS

Exemplary categories for describing the work context (in RDFS they are called **schemata**): Document Classes, Organizations, Groups, Persons, Events, Locations/Addresses, Appointments, Topics, Tasks.
Thomas already uses a Semantic Desktop on which he may create an application-independent **Personal Information Model**

Please note that the text in the attachment may itself contain relationships to already available knowledge
For this purpose, he uses the rich RDFS tool box allowing him to formally represent all aspects of information he needs.

- Schemata describe classes of objects in the work context by a fix pattern.
For this purpose, he uses the rich RDFS tool box allowing him to formally represent all aspects of information he needs.

Instances are exemplars or elements of a category having individuals pattern values.
For this purpose, he uses the rich RDFS tool box allowing him to formally represent all aspects of information he needs.

- Between the concepts of the ontology there are qualified relations called properties.
- Each instance has a is-a-relationship to its class, i.e. it complies the defined pattern.
For this purpose, he uses the rich RDFS tool box allowing him to formally represent all aspects of information he needs.
For this purpose, he uses the rich RDFS tool box allowing him to formally represent all aspects of information he needs.

URI assures the uniqueness of a resource.
Employing URIs, an application- and platform-independent unique representation for Thomas’ resources is created

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

All resources are identified by a URI (Uniform Resource Identifier)

http://www.aka-verlag.com for a Website
http://dfki.de/outlook/contact/0019E177 for a contact
imap://wmu@dfki.com/INBOX;UID=3 for a Email
http://dfki.de/outlook/contact/0019E177 for a file
file://Documents/Courses/Agenda#18 for a file
file://Documents/Courses/Invitations for a category
outlook://appointment/00000000ECD4B99358B9814B9DA for a calendar entry
The individual network of thoughts leads to a multi-dimensional and multi-perspective organization of content and for this reason necessarily to a “dematerialization” of traditional archiving concepts."
Over time Thomas’ continuous work has transformed the initial schema system into a comprehensive domain vocabulary. Because RDFS is a W3C-Standard, the vocabulary may be exchanged with Thomas’ colleagues and applications, i.e. it may be provided via a server.
Facts captured in new documents may be incorporated into Thomas’ PIMO using information extraction techniques.

Hi Thomas - it’s nice to hear from you. Many thanks for sending the agenda. I am glad to accept and confirm my participation in the Semantic Desktop course on February 9.

Warm regards,
Arnould

Arnoud de Kemp
AKA GmbH
Postfach 103305
69023 Heidelberg
Germany
Email: dekemp@aka-verlag.com
Thomas makes use of an ontology-based document understanding system helping him to extract relevant facts from all resources.

---

From: Arnoud de Kemp <dekemp@aka-verlag.com>
Date: January 23, 2010 11:33:17
To: thomas.mustermann@dfki.de
Subject: Re: Semantic Desktop

Hi Thomas -

It's nice to hear from you.

Many thanks for sending the agenda. I am glad to accept and confirm my participation in the Semantic Desktop course on February 9.

Warm regards,
Arnoud

Arnoud de Kemp
AKA GmbH
Postfach 10305
69023 Heidelberg
Germany
Email: dekemp@aka-verlag.com
Information extraction stepwise transforms the contents of documents into knowledge relating it to the existing ontology.

- **Segmentation**: Segmentation of the text into paragraphs, sentences, and words.
- **Indexing**: Description of the document using intrinsic text features.
- **Categorization**: Document classification.
- **Attribute Assignment**: Identification of potential attributes.
- **Unification**: Matching of attributes with known classes and instances.
- **Contextualization**: Extraction of properties and new instances.
Hi Thomas -

It's nice to hear from you. Many thanks for sending the agenda. I am glad to accept and confirm my participation in the Semantic Desktop course on February 9.

Warm regards,
Arnoud
Subsequently, potential attributes are recognized and it is checked whether they fit into the existing knowledge.

From: Arnoud de Kemp <dekemp@aka-verlag.com>
Date: January 23, 2010 11:33:17
To: thomas.mustermann@dfki.de
Subject: Re: Semantic Desktop

Hi Thomas,

It’s nice to hear from you.

Many thanks for sending the agenda. I am glad to accept and confirm my participation in the Semantic Desktop course on February 9.

Warm regards,
Arnoud

Arnoud de Kemp
AKA GmbH
Postal 103305
69023 Heidelberg
Germany
Email: dekemp@aka-verlag.com
Based on that new relations (properties) may be extracted and incorporated into the ontological context of the message.

From: Arnoud de Kemp  <dekemp@aka-verlag.com>
Date: January 23, 2010 11:33:17
To: thomas.mustermann@dfki.de
Subject: Re: Semantic Desktop

Hi Thomas -

its nice to hear from you.

Many thanks for sending the agenda, I am glad to accept and confirm my participation in the Semantic Desktop course on February 9.

Warm regards,
Arnoud

----------------------------------------------------------
Arnoud de Kemp
AKA GmbH
Postfach 103305
69023 Heidelberg
Germany
Email: dekemp@aka-verlag.com

The email content is related to the existing knowledge via semantic hyperlinks.
Let’s have a look for the real technology!
iDocument is a system for Ontology-based Document Understanding

- Documents are interpreted using ontologies as background knowledge.
- New documents are semantically annotated and connected to the knowledge base.
- New knowledge is created if contained in the document.

The Semantic Desktop acts as a personal “information butler”

The Semantic Desktop is a means to manage all personal information across application borders based on a shared vocabulary.

http://nepomuk.semanticdesktop.org


Semantic Desktop System Architecture

Semantic Applications

- PIMO Editor
- Personal Wiki
- Application Plugins
- Desktop Applications

web interfaces

Semantic Desktop Server

- Clustering
- Tagging
- Crawler
- User Context

Domain Ontologies

Sesame2 Repository
- Ont. Matching
- GUI invocation
- Desktop Search

Configuration Store
Service Store
Lucene Index

Aperture Crawlers
- Outlook
- Files
- e-mail

Outlook
filesystem
e-mail server
The social semantic desktop offers an ontology explorer for generating a PIMO

Classes, instances, and properties of the underlying domain
The Semantic Desktop acts as a personal “information butler”

The Semantic Desktop is a means to manage all personal information across application borders based on a shared vocabulary.

Relevant facts are extracted from all resources and proposed to the user for confirmation.

http://nepomuk.semanticdesktop.org


Symbols relationships, and facts may be explained

B. Adrian, B. Forcher, T. Roth-Berghofer, and A. Dengel, Explaining Ontology-Based Information Extraction in the NEPOMUK Semantic Desktop, Proceedings EXACT 2009, IJCAI Workshop on Explanation-Aware Computing, Pasadena, CA, (July 2009), accepted for publication
The Semantic Desktop acts as a personal “information butler”

The Semantic Desktop is a means to manage all personal information across application borders based on a shared vocabulary.

Relevant facts are extracted from all resources and proposed to the user for confirmation.

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

http://nepomuk.semanticdesktop.org


Context identification is a cornerstone of multi-context knowledge work support

- Relevant PIMO Concept(s)
- stimulus
- short-term
- medium-term
- long-term
- entailment spreading
- context switched
- potential context switch detected
- currently selected context

... what is the increased value of the Semantic Desktop
Browsing web sites creates an „information-push“ from the PIMO to the Semantic Desktop


http://dynaq.opendfki.de
A DropBox provides a service for the Semantic Desktop for filing and conceptualization.

Document and extracted meta data

Proposed concepts

Concepts as tags and connected folders
In addition we use a Semantic Wiki as part of a personal knowledge space.

A user may make use of our gnowsis semantic search

Sparql-Rules:

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

How to integrate Gutenberg's World?
Combing the PIMO with OCR technology allows to recognize entities in printed documents.

Contents of paper documents may be easily integrated into the Semantic Desktop

- Transformation of printed documents combined with OCR and Google book search
- Knowledge acquisition from the document image
  1. Detection of instances (visualization of hypotheses, acceptance via click)
  2. Presentation of existing and potentially new relations between instances
  3. Generation of new instances

B. Adrian, H. Maus, M. Kiesel, and A. Dengel,
Towards Ontology-based Information Extraction and Annotation of Paper Documents for Personalized Knowledge Acquisition
Workshop on Personal Knowledge Management PKM09 at the WM 09, Solothurn, Switzerland

© andreas.dengel@dfki.de - 2009 - Page 51
Using the Anoto-Pen new instances and facts may be easily collected via pen and paper.
We developed the **Semantic eInk** for pen and paper-based interaction with the Semantic Desktop

- With digital pen annotations on paper, content is semantically linked to existing contacts, topics, and events
  - Make annotations for document or text parts
  - Use ontological concepts

**Title**

**Pen and Paper-based Interaction with the Semantic Desktop**

**Author**

Marius Liwicki, Kingsa Schumacher and Andreas Dengel

German Research Center for AI (DFKI GmbH)
Knowledge Management Department, Kaiserslautern, Germany
{}@dfki.de

Nadir Weibel, Beat Signer, and Mosra C. Norrie
Institute of Information Systems, ETH Zurich, Switzerland
{}@iis.ee.ethz.ch

**Abstract**

In this paper we propose a system which recognizes and interprets the semantics of handwritten annotations on printed documents. The semantic information will be sent to the Semantic Desktop and onto Semantic Web-on-the-desktop components which support users in their information management. This allows a seamless integration of interactive paper into the individual knowledge work. The current implementation of the proposed system works with made notes and transforms them into digital format.

In workflows like reviewing, where the annotations have a meaning for the marked text, the problem of mapping the paper to the digital counterpart arises. A variety of approaches have been investigated to enable this kind of paper-driven digital services. They use cameras, Wearable Graphics Tablets, ultrasonic partitioning, RFID automatic barcode readers, or Aristo’s Digital Pen and Paper technology. The Aristo technology is particularly interesting because it is based on regular paper and the recording of the


8th IAPR International Workshop on Document Analysis Systems (DAS 08)

Let me summarize and conclude!
Summary

- The traditional Web has recently undergone an orthogonal shift into a Web of People/Web 2.0 focusing on collective intelligence, 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.

- Nowadays, a document is like a node in a network, a system of links to other documents, to events, locations, persons, or tasks.

- Trusted communities allow people to link with their colleagues (friends) and share information by making use of W3C Standards.

... towards the Social Semantic Desktop
Our strategy considers two major trends

Semantics Foundation

Semantic Web

Web 1.0

User Participation

Semantic Desktop

Web 2.0

Computer-Understandable Information

Community Relation


Thank you for your attention!

Address:
Prof. Dr. Andreas Dengel
DFKI GmbH
P.O. Box 2080
D-67608 Kaiserslautern
email: andreas.dengel@dfki.de
http://www.dfki.de/dengel