Text und Hypertext

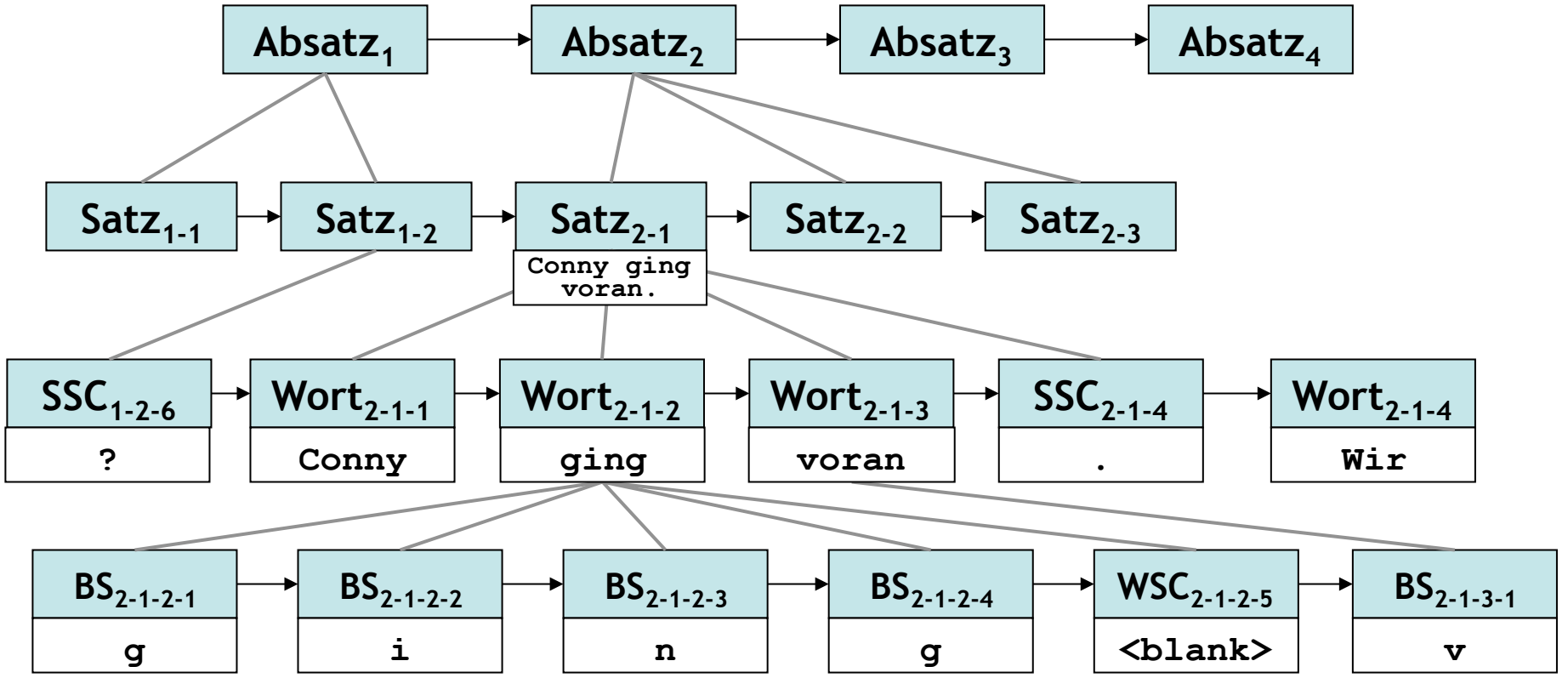
(und assoziative Verknüpfung)

Hans Uszkoreit

Text

- ☆ Text im weiteren Sinne ist eine Abfolge von alphabetischen Zeichen, unter Umständen angereichert um andere Objekte wie nicht-alphabetische Zeichen, Bilder oder Tabellen.
  - ○ plain text oder Klartext ist eine nur-Text Fassung von Text.
- ☆ Text im engeren Sinne ist eine Abfolge von natürlichsprachigen Sätzen, die in einem Sinnzusammenhang und einer intendierten Reihenfolge stehen. Der Text kann um andere Objekte wie nicht-alphabetische Zeichen, Bilder oder Tabellen angereichert sein, die inhaltlich in einem intendierten Zusammenhang mit den Sätzen stehen.

Lineare Abfolge



Strukturen

- implizite Struktur
  - inhaltliche Teile wie Beschreibungen, Definitionen, Anweisungen, etc.

- explizite Struktur
  - explizit für den menschlichen Leser
  - explizit für die Datenverarbeitung

  Erschließung der impliziten Struktur

mehrdimensionale Medien

- ☆ Tabellen nutzen die beiden Dimensionen, um funktionale Relationen zu kodieren
- ☆ Bilder nutzen die beiden Dimensionen, um Morphismen zu räumlich visueller Perzeption herzustellen (zwei oder dreidimensionalen visuellen Eindrücken)
- ☆ Filme sind Abfolgen von Bildern.

Unterschied zwischen dem Text und seiner medialen Präsentation

- ☆ je nach Darstellungsmedium muss der Text in Zeilen und Seiten (bzw. Fenster, Bildschirme, Frames etc.) unterteilt werden.

- ☆ Seitenzahlen
  - Seitenzahlen erfüllen zwei wichtige Funktionen
    - explizite Kodierung der Abfolge
    - Namen für Textteile als Ziel für Verweise

weitere Relationen

- ☆ Inhaltsverzeichnis (Abbildungsverzeichnis, Definitionsverzeichnis)

- ☆ Noten (Endnoten, Fußnoten)
- ☆ Verweise (interne, externe)
- ☆ Index (Sachindex, Personenindex, etc.)
- ☆ Referenzen (Literaturverzeichnis, Bibliographie, etc.)

- ☆ Verweise können auf Teile der Textstruktur zeigen oder auf Teile der Präsentationsstruktur (Seitenzahlen, Karteikarten)

Hypertext

- **Wikipedia de**
  Hypertext ist eine *multi-lineare* Organisation von Objekten, deren netzartige Struktur durch logische Verbindungen (so genannte Hyperlinks) zwischen Wissenseinheiten (Knoten, z. B. Texten oder Textteilen) hergestellt wird. Hypertext ist eine Anwendung des *Verweis-Knoten-Konzepts*.

- Hypertexte sind Texte, deren Textteile zusätzlich oder anstatt der linearen Abfolgerelation über andere Relationen miteinander verknüpft sind. Diese Verknüpfungen heißen Hyperlinks.

Hypertext

ca. 1500, Erasmus von Rotterdam: Seitenzahlen für Querverweise in Büchern

1932 Emmanuel Goldberg, Register in der Photoarchiven

1945, Vannevar Bush: Memex

1963, Doug Engelbart: NLS / Augment, Baumstruktur von Texten

1965, Ted Nelson: Xanadu

1975, Akscyn / McCracken (CMU): ZOG, später KMS (Knowledge Management System)

1976-1980, Allan Kay, Adele Goldberg, H.H. Ingalls (Xerox PARC): Objektorientierte Programmierung mit „Smalltalk“

1987, Bill Atkinson (Apple): HyperCard

1989, Tim Berners-Lee / Robert Cailleau (CERN): HTML / WWW

Intellectual Context

- ☆ a little history and some related projects

The Visionaries

Emmanuel Goldberg

New Methods of Photographic Indexing 8th International Congress of Photography, held in Dresden in 1931

http://www.sims.berkeley.edu/~buckland/goldberg.html

Vannevar Bush

As We May Think

The Atlantic Monthly – July 1945

http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm

The Pioneers

Ted Nelson

Hypertext, Xanadu

Doug Engelbart

the mouse and first hypertext
groupware applications

MyLifeBits

Gordon Bell

Project: MyLifeBits

Microsoft Bay Area
Research Center

topics

- ☆ processing and annotation of personal info-space
- ☆ document-related metadata, stand-off markup
- ☆ detection of instances of personal ontology
- ☆ markup and indexing
- ☆ associative interface for comprehension and production

Haystack

David Karger

MIT Computer Science and
Artificial Intelligence Lab

Individual Knowledge Access

A semantic web-like structure
for personal digital information

A Grand Challenge

Memories for Life

Outline of a Grand Challenge

"Managing information over a human lifetime"

Andrew Fitzgibbon

with Ehud Reiter

☆ http://www.nesc.ac.uk/esi/events/Grand_Challenges/proposals/Memories.pdf

example 1: personal digital memory

related projects

- MyLifeBits (Microsoft)
- Haystack (MIT)

and also relevant but with slightly different focus

- LifeRecorder (Bernt Schiele)
- LifeLog (DARPA) †
- LifeBlog (NOKIA)

The Teddy-Vision

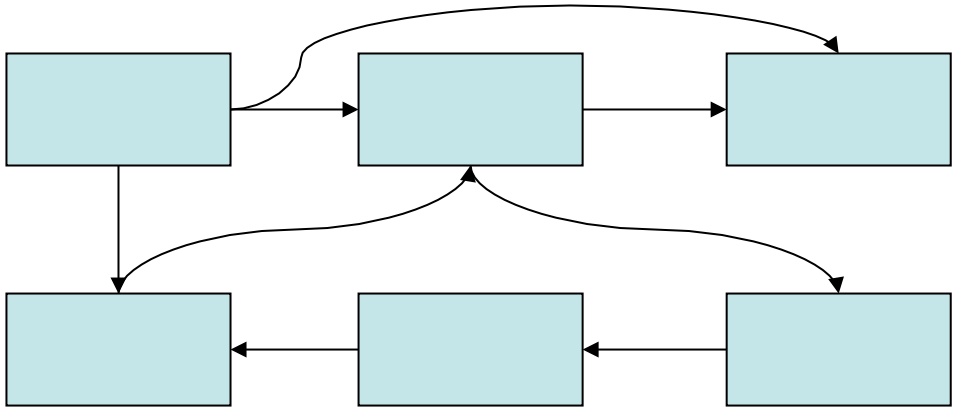
Don Norman: The Teddy

In: Norman 1992

"Turn Signals are the Facial Expressions of Automobiles"

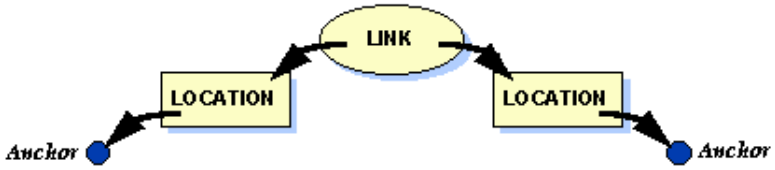
Hypertext

- ☆ Eine hypertextuelle Verknüpfung ermöglicht nichtlineare Abfolgen.
- ☆ z.B. Schleifen, Verzweigungen, Abkürzungen

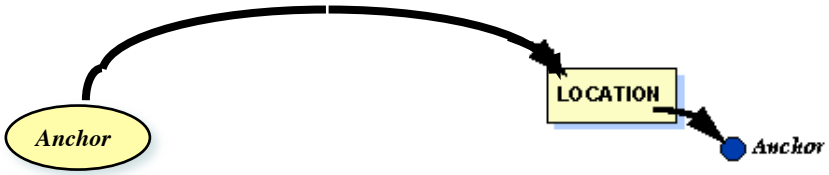


- ☆ hyperlinks

Simple Links



asymmetric HTML links



Linking what?

- Quelle (source):
  - part of a document (anchor): text string, image, part of image, etc.
- Ziel (target):
  - document - URL
  - URI (URL|URN)
  - part of a document (anchor)
- der Begriff Anker (anchor) wird sowohl für die Quelle als auch für das Ziel von asymmetrischen Hyperlinks verwendet

Functional Hyperlinks / Typed Hyperlinks

Functional Hyperlink ⟨person, homepage⟩
⟨company, homepage⟩
⟨section no., section⟩

Ambiguity ⟨person, homepage⟩
⟨person, email-address⟩

Relational Hyperlinks

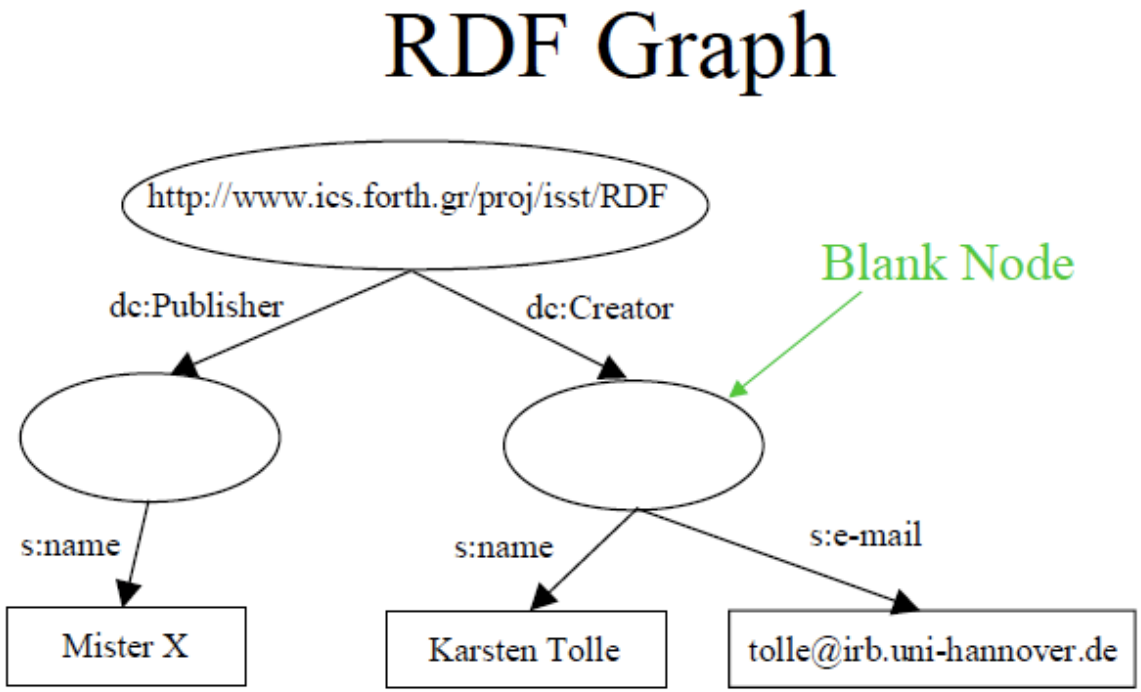
Relational Hyperlink {⟨person, homepage⟩
⟨person, email-address⟩}

Relational Labelled Hyperlink
{⟨person, „homepage“, homepage⟩
⟨person, „email“, email-address⟩}

Relational Typed Labelled Hyperlink

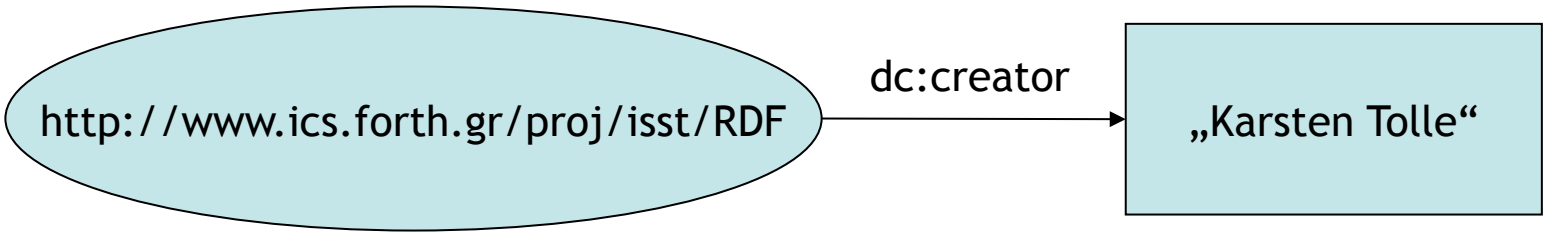
person: {⟨person, „homepage“, homepage⟩
⟨person, „email“, email-address⟩}

RDF Triples



RDF Triples

- ☆ triples der Form <objekt$_1$, relation, objekt$_2$>
- ☆ <subject, predicate, object> oder <s,p,o>



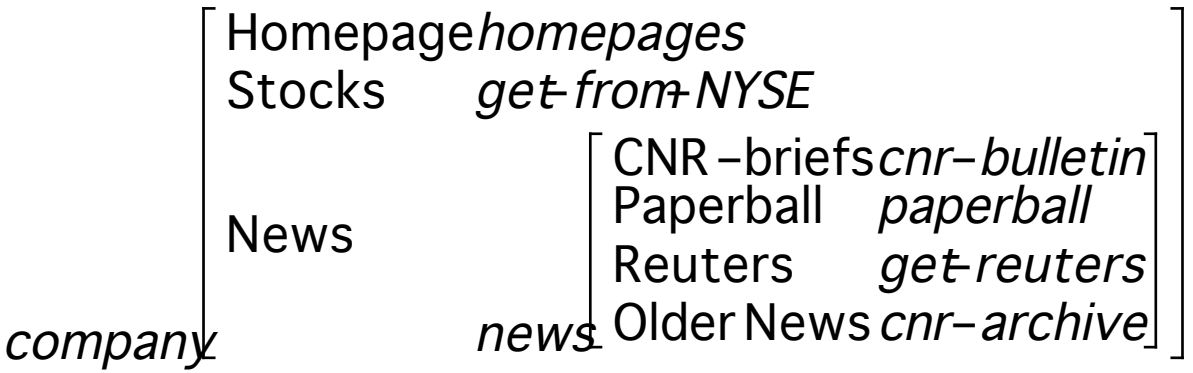
RELATIONAL ONTOLOGIES I (formalism)

$$type\begin{bmatrix} \text{FunRef}_1 & target_1 \\ \text{FunRef}_2 & target_2 \\ \vdots & \vdots \\ \text{FunRef}_n & target_n \end{bmatrix}$$

Targets may be

- ☆ **Short Pieces of Information (e.g., translation into Spanish)**
- ☆ **Regular Hyperlink (e.g., homepage)**
- ☆ **DB Access (e.g., lookup of account status)**
- ☆ **Start of a Process (e.g., start a credit check)**
- ☆ **Notification of a person (e.g., send query to expert)**
- ☆ **Search out of context (e.g., search in inter-, intra-. extranet)**

RELATIONAL ONTOLOGIES II (example)



Getypte Hyperlinks im UI

- ☆ Jedes bedeutungstragende Objekt hat eine Reihe von Hyperlinks, die das Objekt mit Zielen verbinden, je nach Typ des Links und Typ des Ziels können bei Auruf des Ziels unterschiedliche Dinge geschehen:
  - Anzeige im gleichen oder neuen Browserfenster
  - Anzeige auf dem gleichen oder auf einem anderen Bildschirm
  - Start eines Programms
  - Herstellen eine Verbindung
  - Schalten eines Geräts
  - usw.

- ☆ Die Auswahl aus den bestehenden Links kann über ein Menü, über Tasten, Gesten, Sprache geschehen.

INFORMATION MANAGEMENT - TASK

The task of modern information management

is to gather, maintain and supply

large volumes of digital information.

INFORMATION MANAGEMENT - COMPONENTS

- **acquisition** (gathering)
- **categorisation** (sorting w.r.t topics)
- **indexing** (by strings, words, terms, concepts)
- **summarisation** (condensing the information)
- **information extraction** (relevant data in text)
- **translation** (indicative translations)
- **delivering** (filtering, routing, push services)
- **presentation** (ranking, structuring, visualising)

The Ultimate Information Management

- ☆ The ultimate challenge for IM is the delivery of the right type of information exactly at the right time.
- ☆ How can IM know when the information is needed. It can't.
- ☆ If the needed information (decision basis) is linked with the decision trigger, it will be available at the right time.
- ☆ Therefore: use automatic hyperlinking for preparing the decision basis.

TOPICS OF KNOWLEDGE MANAGEMENT

- ☆ **Strategies for Knowledge Sharing**
- ☆ **Analysis und Evaluation of Corporate Knowledge**
- ☆ **Knowledge-Centered Workflow Management**
- ☆ **Knowledge-Centered Information Management**
- ☆ **Organisation and Exploitation of the Corporate Knowledge Base**
- ☆ **Learning Organisations**

CORPORATE KNOWLEDGE BASE

- ☆ **Combination of ”structured” and “unstructured” information in a distributed partitioned knowledge base**
- ☆ **Structuring of the knowledge base for different uses**
- ☆ **Transforming Information into Knowledge**

In order to become knowledge, information must be

- **immediately accessible**
- **grounded**
- **densely connected**
- **suited for inferences**

KNOWLEDGE SHARING

- ☆ One of the prime goals of KM is the sharing of knowledge

BUT

- ☆ How do we get people to offer knowledge
- ☆ How do we get people to consult the offered knowledge
- ☆ How do we get people to evaluate the consulted knowledge
- ☆ How can we learn from new decision situations

KNOWLEDGE SHARING – THE DILEMMA

- ☆ Sharing is beneficial for the company, not for the individual
- ☆ Therefore sharing needs to be rewarded
- ☆ Useless information is counterproductive
- ☆ Therefore rewarding needs evaluation
- ☆ Evaluation is an additional effort - does it require additional rewards?

Two Main Ideas

- ☆ augment human memory by a digital personal memory

- ☆ use associative memory as a guiding metaphor for large personal information repositories

Idea 1: Function (i.e. application)

- ☆ augment human memory by a digital personal memory
- ☆ enable the owner to
  - remember MUCH more than otherwise possible
  - combine ordered data storage, personal library and biography
  - recall the contents in new ways
  - keep contents in a processable way
  - process memorized contents
  - share memorized contents

Idea 2: Structure & Functionality (i.e., method)

- ☆ The WWW is in nature an associative information storage
- ☆ The role of hyperlinks
- ☆ Associative digital memories could be an important step between todays indexed digital contents and the semantic web

Differences between Human and Electronic Memories

Human Memory

- ☆ capacity of short term memory 7±2 chunks
  (Miller, G. 1956)

  decay in 200 ms

- ☆ capacity of long term memory 10^{8432} Bits (upper boundary)
  (Wang Y. & Liu D. 2003)

  forgetting over many years or never

Electronic Memory

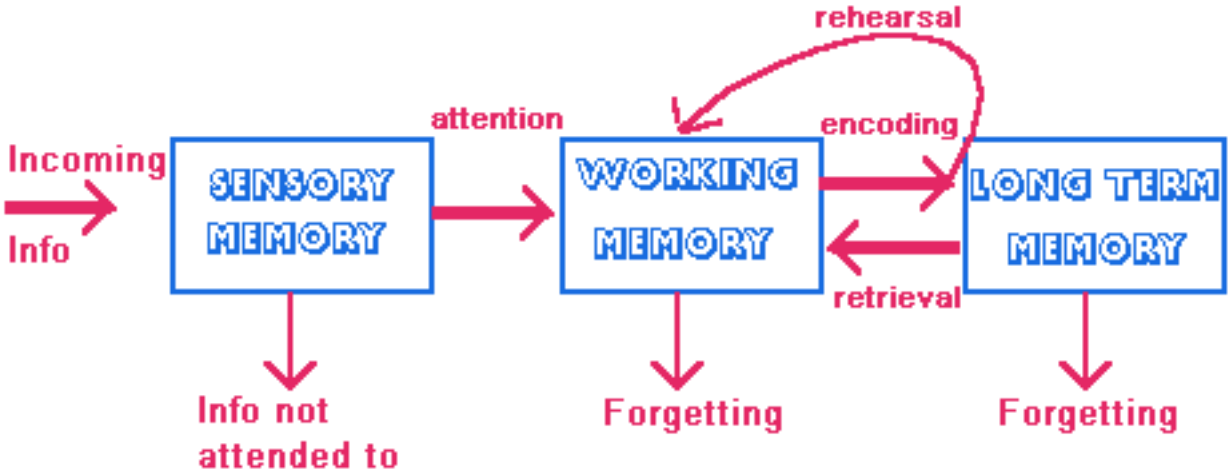
- ☆ short term just restricted in storage capacity

  no decay necessary/overwriting

- ☆ size of the WWW memory 2^{53} Bits

  forgetting through deletion

Human Memory



Personal Records

- ☆ Calendar entries
- ☆ Received and sent email messages
- ☆ Written documents
- ☆ Read documents
- ☆ Addresses
- ☆ Pictures, sounds, movies

Volume

- ☆ The amount of storage on a regular PC will soon be in the terabyte range.
- ☆ Already today, IR methods are applied on the desktop.

Motivation

- ☆ our individual long term memory is very limited
- ☆ the volume of digital information on our disks and other storage media explodes
- ☆ searching through our own files often is harder than a search on the web

Volume of One Decade

☆ calendar:	2000-12000 calender entries
☆ email:	20.000-100.000 messages
☆ addresses:	100-2000 addresses
☆ photographs:	1000-30000 pictures
☆ written papers, reports, reviews:	100-1000 documents
☆ music:	500-5000 titles
☆ talks:	50-500 slide sets
☆ read electronic papers:	200-2000 documents
☆ visited web-pages:	20.000-100.000 pages
☆ Web-pages mentioning the user	500-25.000 pages

Lifetime Experience Recording and Storage

- ☆ maximum resolution 720x1280 (HDTV 720P)
- ☆ 3 MBit/s through H264 compression
- ☆ 32 GB/d, 12 TB/y
- ☆ under 900 TB per life, taking into account some sleep

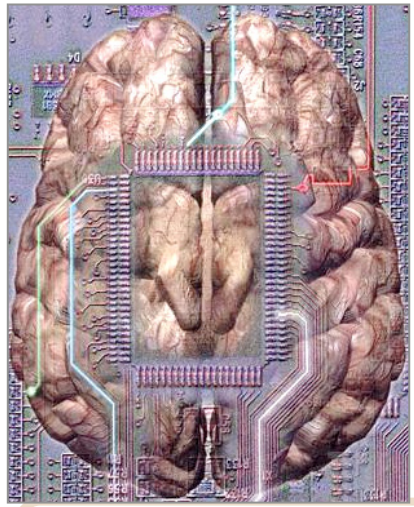
Special Issues

- ☆ Data -- Source, Privacy, Security, Secondary Data
- ☆ Evaluation -- Methodological Problems, Pragmatic Solution
- ☆ Life Experience Recording -- Idea, Objections, Privacy
- ★ Mutual Exploitation of Hard- and Software Technologies
  - structuring the memory
  - performant integration
  - performant access, presentation, display
  - application software

Components of PDAM

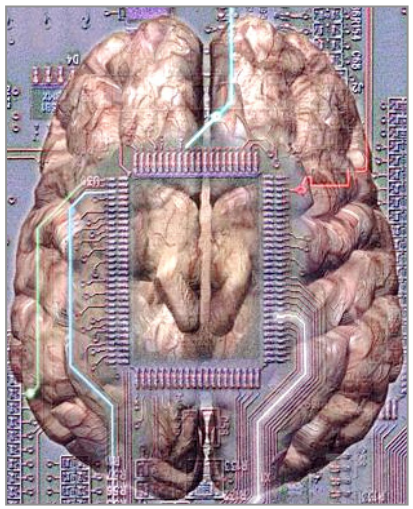
- ☆ Ontologies and Central Database
  - ○ Generic Personal Life Ontology
  - ○ Specialized Ontologies
  - ○ Ontology of Information Objects
  - ○ Entity and Record Databases

- ☆ The Memory Weaver
  - ○ Content Analyst
  - ○ Metadata Provider
  - ○ Integrator



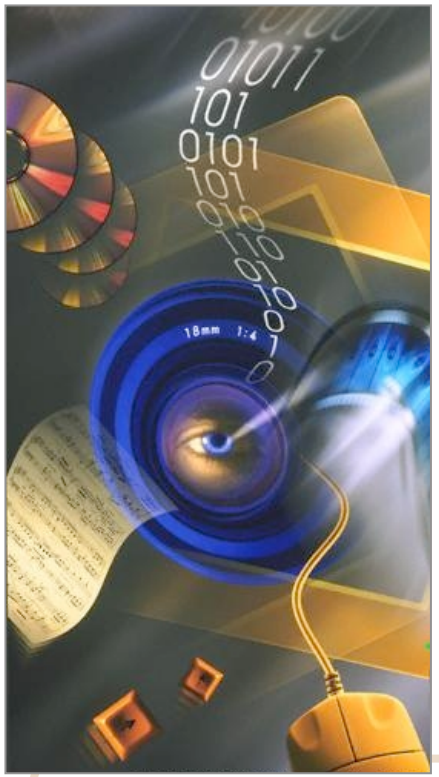
Augmented Cognition

- ☆ Different way of filling the memory
  - record things that you did not fully perceive
  - remember additional material
  - obtain metadata
- ☆ Different ways of keeping the memory
  - size
  - no absolute forgetting
- ☆ Different ways of using the memory
  - permanent analysis
  - embedding in applications



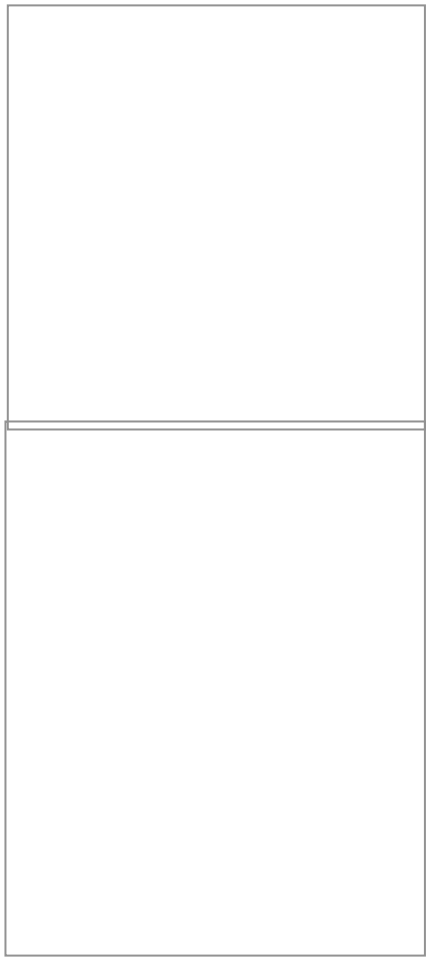
The Data Issue

- ☆ Because of privacy and confidentiality concerns, it is difficult to obtain realistic data of sufficient volume.
- ☆ Fabricating the data will be much too expensive, since a certain level of complexity and consistency needs to be guaranteed.
- ☆ Our project HyLaP will utilize personal data collected over a period of ten years.
- ☆ The data will have been filtered to remove confidential records. Some mail will be systematically anonymized.



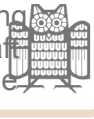
Recorded Life Experience

- ☆ Life experience recording will add masses of data to be included in the personal memory and strengthening
  the episodic component of memory.
- ☆ Research on such recordings are outside the scope of this project.
- ☆ However, we plan to cooperate with other centers to obtain data through life experience recordings.



Long Intensive Preparation

- ☆ First Concepts and Writeups (1998)
- ☆ Automatic Hyperlinking (TwentyOne, Hypercode, ExtraLink)
- ☆ Personal Memory Working Group (2002-2004)
- ☆ Course, workshop talks, two keynote lectures (1999-2004)
- ☆ Data Preparation (1999-2005)



Ontologies

- ☆ generic personal ontology
- ☆ special personal ontologies
- ☆ info-objects: doc-objects, DB-entries, print-objects, digital-experience, AV-objects

Generic Personal Upper Ontology (hand-crafted)

- ☆ persons (family, friends, co-workers, colleagues, customers, etc.)
- ☆ themes (hobbies, interests, work-related themes, etc.)
- ☆ external sources (web pages, books, information services, etc.)
- ☆ own works (papers, letters, ideas, graphics, photographs, etc.)
- ☆ communications (incoming mails, outgoing mails, meetings, etc.)
- ☆ locations (cities, offices, hotels, etc.)
- ☆ events (trips, meetings, conferences, etc.)
- ☆ functions (social roles, profession, jobs, offices in associations, etc.)
- ☆ timeline (mapping of all events onto multi-granular timescale)

Specialized Personal Ontologies

- "Trivial" cases of instance learning:
  - learn people from address book and mail files
  - learn meetings and trips from calendar
- More complex: topic learning
  - start with NEE
  - continue with term extraction
  - hypothesize relevant topics by occurrences in texts, filenames, email
- Great area for active learning
  - category of hypothesized topics can be solicited from user
  - or just checked with user

Ontology of Information Objects

- ☆ Ontology of document and DB record types
- ☆ calendar entries, emails, address book entries, reports, papers, bibliographic references, photographs, mp3 titles, etc.
- ☆ knows about recipients of emails
- ☆ knows about relationship between calendar entries and locations
- ☆ knows about literature lists in publications, co-authors, etc.

Types of Information Objects

Email

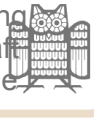
Calendar
Addresses
References

Documents
- texts
- movies
- pictures
- music

life recordings
- IM chat logs
- IMAV chat logs
- conversations
- phone calls
- meetings
- perspectives

Ontologically Structured Association DB

- ☆ Typed hyperlinks can be used to present associations.
- ☆ For many relations (roles, attributes) introfuced in concepts and instantiated in instances hyperlinks can employed.
- ☆ The choice of relations to be presented as typed hyperlinks depends on user model, preferences, situational context.



Associative Search

New wireless voice technology introduced
Posted at 5:09 PM PT, Feb 8, 1999

By Stephen Lawson, InfoWorld Electric

NTT Labs on Monday brought Dick Tracy into the enterprise, introducing a wireless voice and data system that can use a wrist radio at the Demo 99 conference.

AirWave technology, demonstrated for the first time in the United States at this week's conference in Indian Wells, Calif., is based on a wireless PBX. Small, handheld phones -- and a wrist radio that looks like an oversized watch -- can be used to make voice calls and exchange data around a building or campus. The handheld phones can be switched to a public cellular mode to become conventional cell phones.

Internet. There is no airtime charge for AirWave communications in the building or campus.

AirWave systems are scheduled to be available through distribution partners by the end of this year, priced as low as $400 per user.

Associative Search

New wireless voice technology introduced
Posted at 5:09 PM PT, Feb 8, 1999

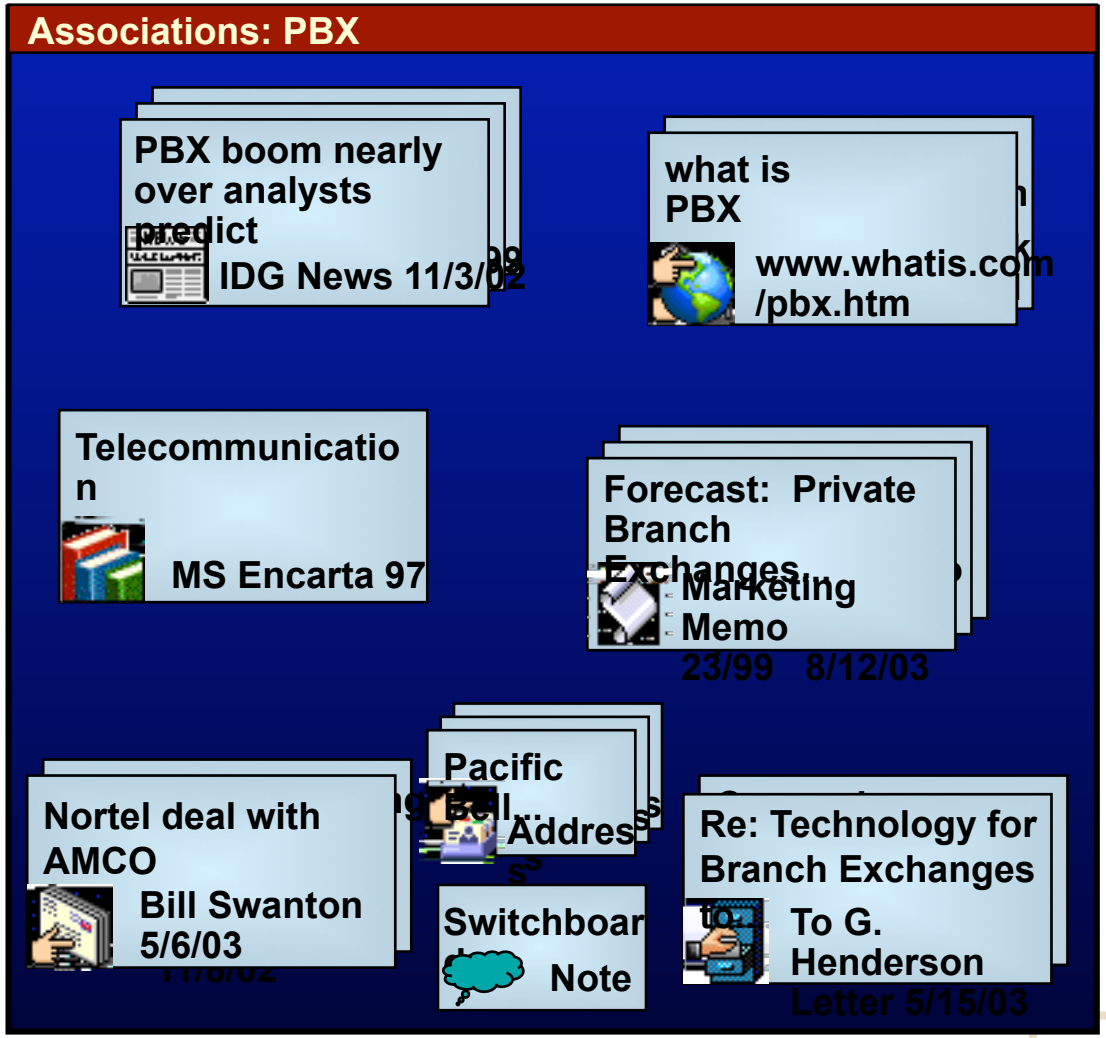
By Stephen Lawson, InfoWorld Electric

NTT Labs on Monday brought Dick Tracy into the enterprise, introducing a wireless voice and data system that can use a wrist radio at the Demo 99 conference.

AirWave technology, demonstrated for the first time in the United States at this week's conference in Indian Wells, Calif., is based on a wireless **PBX**. Small, handheld phones -- and a wrist radio that looks like an oversized watch -- can be used to make voice calls and exchange data around a building or campus. The handheld phones can be switched to a public cellular mode to become conventional cell phones.

Internet. There is no airtime charge for AirWave communications in the building or campus. AirWave systems are scheduled to be available through distribution partners by the end of this year, priced as low as $400 per user.

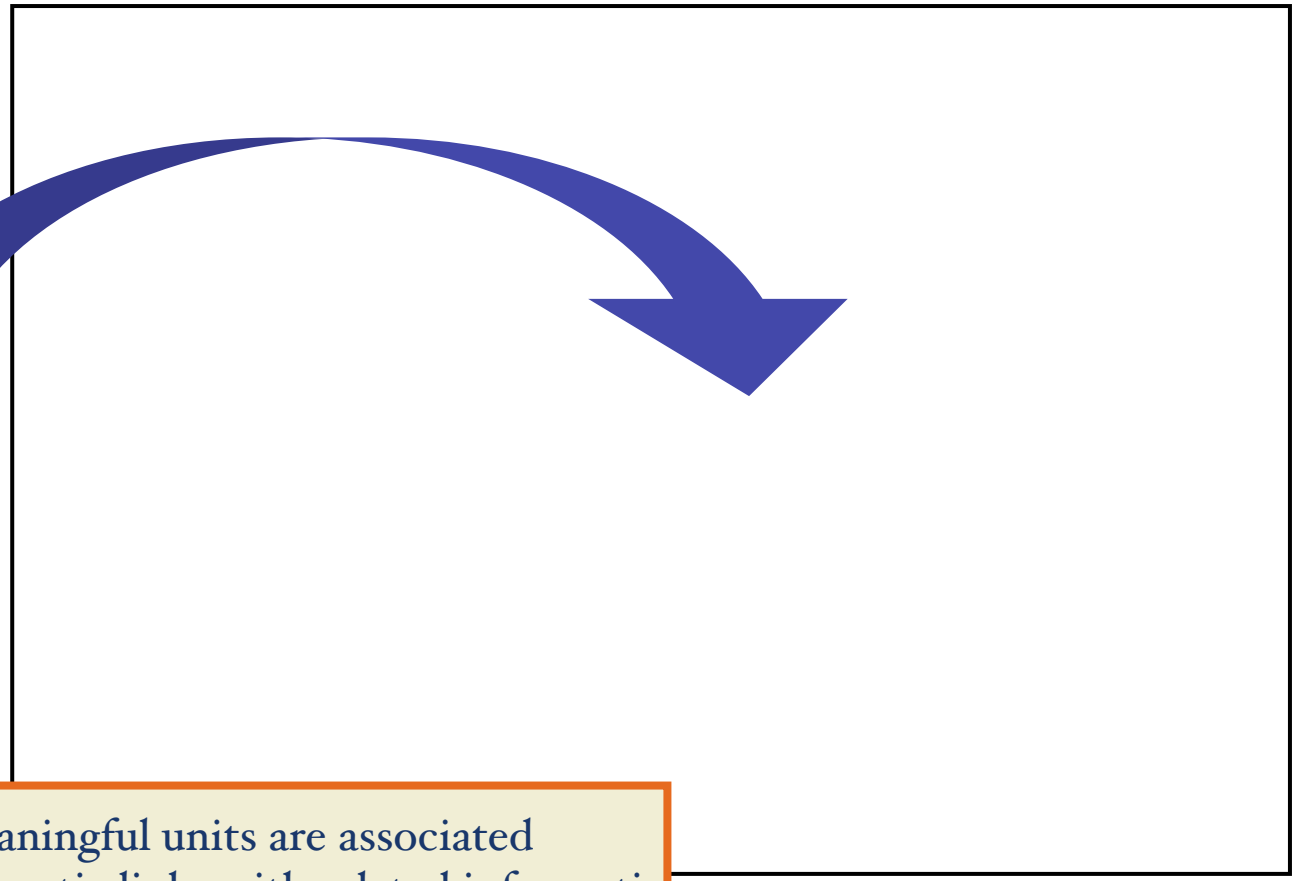
ASSOCIATIVE SEARCH



Densely Associated Content

videte nunc quo adfectent iter apertius quam antea. nam superiore parte legis quem ad modum Pompeium oppugnarent, a me indicati sunt; nunc iam se ipsi indicabunt. iubent venire agros **Attalensium** atque Olympenorum quos populo Romano Servili, fortissimi viri victoria adiunxit, deinde a in Macedonia regios qui pa T. Flaminini, partim L. Paul Persen vicit vi parti sunt, deinde ag optimum et fructuosissi Corinthium qui L. Mu imperio ac felicitat vectigalia populi Romani adiunctus est, post autem agros in Hispania apud novam duorum Scipionum eximia virtute possessos; tum vero ipsam veterem Carthaginem vendunt quam P. Africanus nudatam tectis ac moenibus sive ad notandam Carthaginiensium calamitatem, sive ad testificandam nostram victoriam, sive oblata aliqua religione ad aeternam hominum memoriam consecravit.

Attalia (Attaleia, Antalya)
Coastal city in Pamphylia
Map
Other References
History
Comments
My Comments



All meaningful units are associated via semantic links with related information distributed all over the digital global knowledge base.