FRODO
A Framework for Distributed Organizational Memories
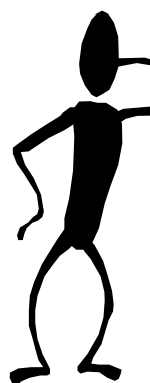
Andreas Dengel, Ansgar Bernardi
Andreas Abecker, Ludger van Elst
Markus Junker, Michael Sintek

21st SAB – Spring 2000

Agenda

- *Motivation: from KnowMore to FRODO*
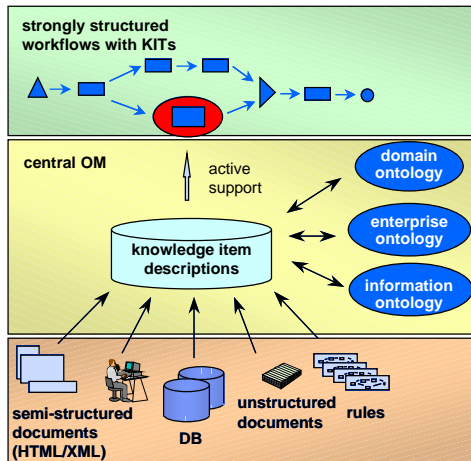
- *FRODO Research questions:*
  1. *Framework*
  2. *Ontologies*
  3. *DAU services for the OM*
  4. *Methodology for BPOKM*
  5. *Weak Workflow*

- *Work plan*

KnowMore: A Central OM with Proactive Information Supply

strongly structured workflows with KITs

central OM

active support

domain ontology

enterprise ontology

information ontology

knowledge item descriptions

semi-structured documents (HTML/XML)

DB

unstructured documents

rules

- strongly structured workflows with extensions for knowledge-intensive tasks
- active information supply relies on workflow context information
- information retrieval by reasoning over central ontologies
- access to various information sources via formal knowledge item descriptions

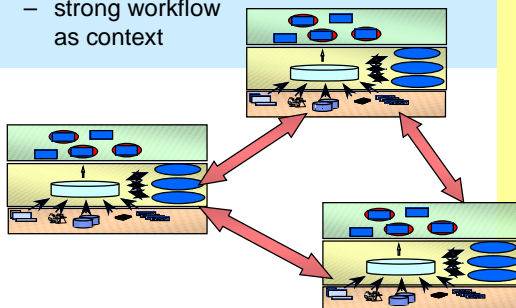Source: Knowledge Management Group

DFKI

Page 3

The Networked World Requires a Distributed Approach to Organizational Memories

KnowMore:
- distributed and heterogeneous knowledge sources
- knowledge-intensive, centralized inference
- global set of ontologies
- strong workflow as context

FRODO:
- independently introduced partial OMs with specialized ontologies
- distributed inference
- external knowledge sources with own ontologies
- weak workflow as context

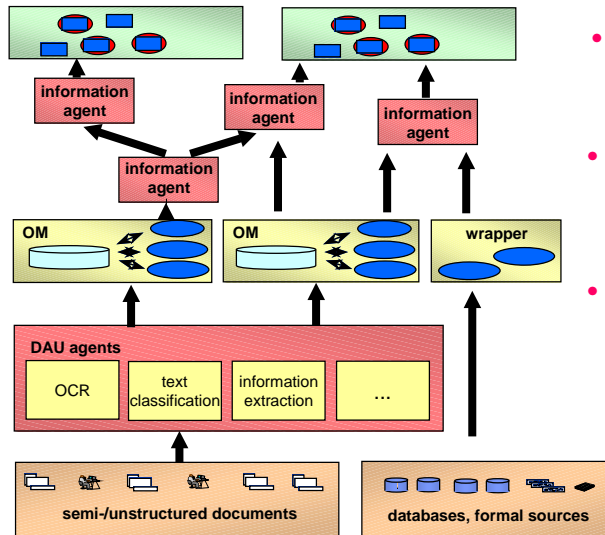*Scaling up leads to a new architectural approach.*

Source: Knowledge Management Group

DFKI

Page 4

The Decentralized OMs are Complemented by Agents for Information Input and Access

- decentralized OMs ask for cooperating information agents

- the agent paradigm facilitates pluggable components for information extraction and text analysis

- wrappers allow access to foreign IT systems, e.g., legacy databases, XML files, ...
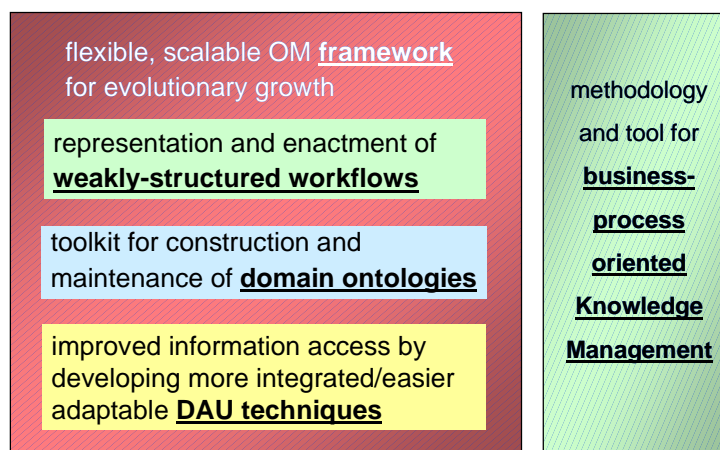
Source: Knowledge Management Group

Page 5

Practical Motivations Lead to Challenging Research Questions

flexible, scalable OM **framework** for evolutionary growth

representation and enactment of **weakly-structured workflows**

toolkit for construction and maintenance of **domain ontologies**

improved information access by developing more integrated/easier adaptable **DAU techniques**

methodology and tool for **business-process oriented Knowledge Management**

Source: Knowledge Management Group

Page 6

3

The OM Framework Aims at a Sound Integration of Heterogeneous Subsystems

FRAMEWORK APPROACH

- integration of
  - separately developed OM solutions and legacy (DB) systems
  - DAU components

- distributed problem solving

- based on
  - declarative knowledge representation
  - agents / speech acts / protocols
  - Internet-enabled: HTTP, XML, RDF, ...

The Framework Realization is Based on (Upcoming) Internet Standards

XML/RDF AND HTTP

- FIPA-like agents communicate with XML messages

- communication via HTTP

- involved XML technologies:
  - XSLT for message transformation and information extraction
  - RDF for representing meta-information
  - RDF Schema for representing ontologies
  - XML/RDF-based query and transformation languages for distributed inferences

- first prototypical realization of an RDF/RDF Schema query agent (based on an F-Logic implementation for RDF: SiLRI, Karlsruhe / Stanford)

- W3C membership under consideration

Agents Communicate Via XML Messages

```
<message type="message-type"
         sender="sender-url"
         receiver="receiver-url"
         additional-information >
   contents
</message>
```
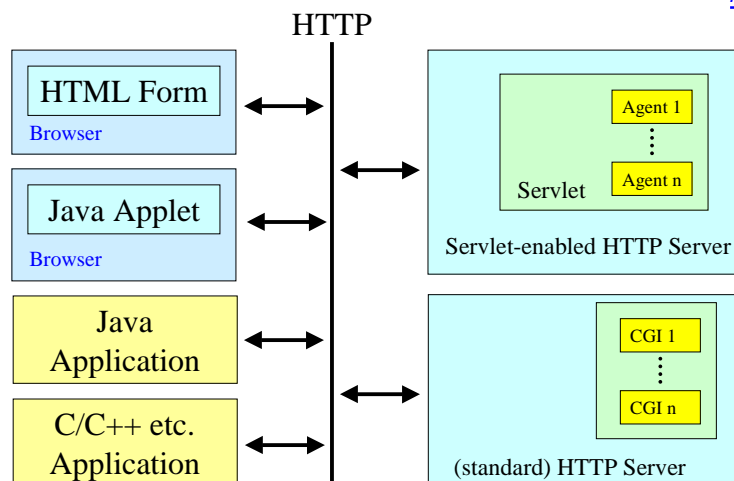
- *message types*: inform, request, agree, cancel, confirm, disconfirm, subscribe, ... (speech acts, FIPA)
- *additional information*: reply-with, in-reply-to, language, ontology, reply-by, protocol, ...
- *contents*: XML "forest"
- cooperation with the DAML initiative (DARPA, Jim Hendler) planned

---

Message Exchange is Based on Internet Techniques

HTTP

HTML Form — Browser ↔ Servlet-enabled HTTP Server (Servlet: Agent 1 ... Agent n)

Java Applet — Browser ↔

Java Application ↔ (standard) HTTP Server (CGI 1 ... CGI n)

C/C++ etc. Application ↔

Our Agent Platform Enables a Simple Integration of Foreign Software Components

CONSEQUENCES

- all foreign software components which are designed as CGI programs are immediately usable as agents (no explicit wrappers needed)

- access to agents simple for all software components that have access to a HTTP library (Java, C, C++, ...)

Higher level functionalities will be realized on this technical basis.
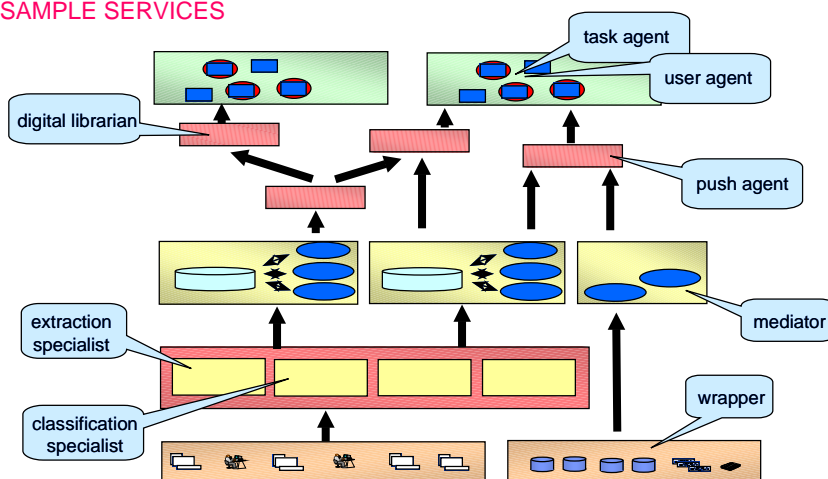
Source: Knowledge Management Group

Page 11

A *FRODO* System Instantiation Combines Heterogeneous Agents

SAMPLE SERVICES

Source: Knowledge Management Group

Page 12

FRODO Aims At an OM Middleware

- higher-level functionalities are realized by collaborative problem-solving, e.g.:
  - cooperative information retrieval
  - knowledge fusion from different sources
  - collaborative filtering by comparison of several user profiles
- collaborative problem solving will be enabled by:
  - specific speech acts & knowledge representation if required
  - protocols / transactions
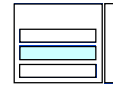 - identification of standard agents / agent communities
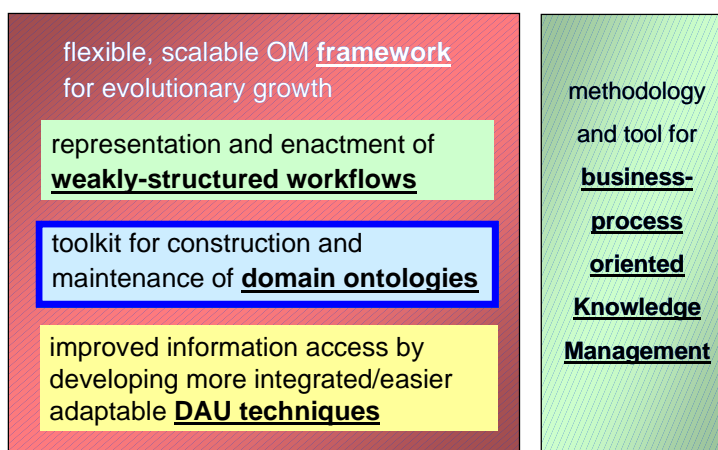
Ontologies are vital for communication and collaboration.

Research Topic 2: Domain Ontologies

flexible, scalable OM **framework** for evolutionary growth

representation and enactment of **weakly-structured workflows**

toolkit for construction and maintenance of **domain ontologies**

improved information access by developing more integrated/easier adaptable **DAU techniques**

methodology and tool for **business-process oriented Knowledge Management**

7

Acquisition and Maintenance of Domain Ontologies is Hard

BUILDING DOMAIN ONTOLOGIES

- in *KnowMore*, we sketched semi-automatic ontology acquisition exploiting *statistic* text analysis techniques

- in *FRODO*, we will couple
  - a widespread, state-of-the-art manual ontology modeling tool
  - *symbolic machine learning* for ontology acquisition
  - learning from user interaction
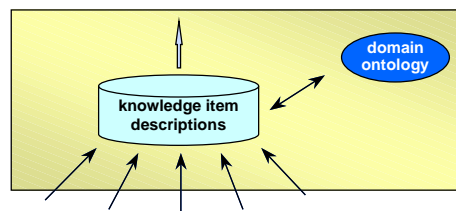
Ontology representation has already been investigated.

Source: Knowledge Management Group   Page 15

Use of RDF Schema Facilitates Ontology Sharing

- RDF / RDF Schema
  - future standard (ontologies and tools can be shared/reused)
  - extensible

- Tool support: Protégé-2000/RDF
  - close cooperation with Stanford Medical Informatics
  - open source
  - extensible
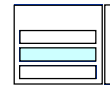 - other tools will soon be available

Manual ontology modeling will be supported by automatic knowledge acquisition from texts.
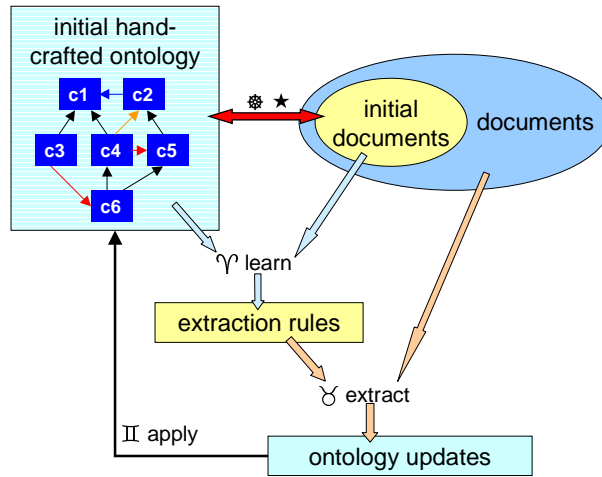
Source: Knowledge Management Group   Page 16

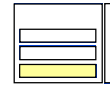## Information Extraction Will be Employed for Ontology Modeling Support

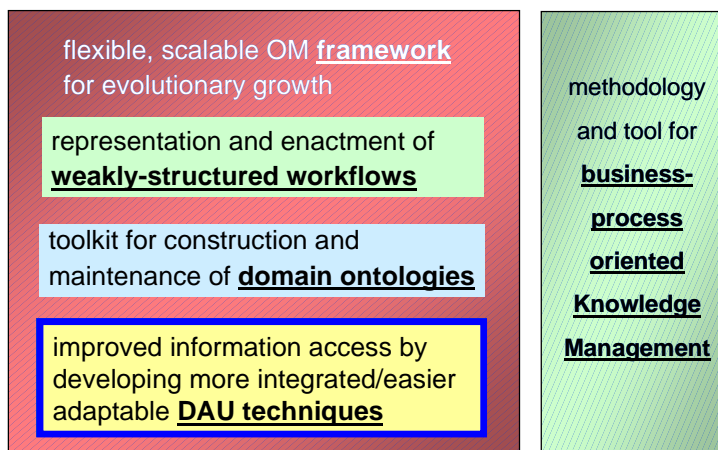INFORMATION EXTRACTION FOR LEARNING ONTOLOGIES

initial hand-crafted ontology

c1 c2
c3 c4 c5
c6

initial documents

documents

learn

extraction rules

extract

apply

ontology updates

Research Topic 3: DAU Techniques

flexible, scalable OM **framework** for evolutionary growth

representation and enactment of **weakly-structured workflows**

toolkit for construction and maintenance of **domain ontologies**

improved information access by developing more integrated/easier adaptable **DAU techniques**

methodology and tool for **business-process oriented Knowledge Management**

DAU Specialists Provide Essential Services in the OM Scenario
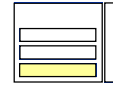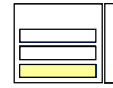
DAU SERVICES AND IMPROVEMENTS

- The overall OM goal is to satisfy users' situation-specific information needs

- DAU techniques:
  - *classify* documents with respect to formal ontologies,
  - *extract* data and information from texts

- Scientific topics include:
  - use of <u>background knowledge</u> from the overall OM scenario:
    - domain ontologies
    - (open) processes
    - user profiles
    - ...
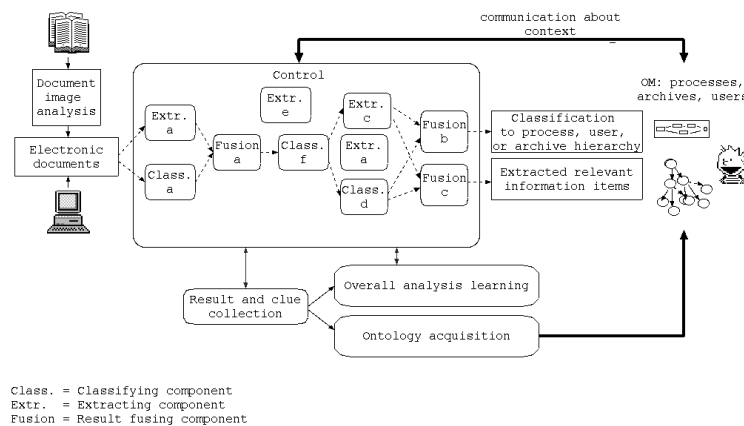 - processing of *structured* (HTML/XML) documents

---

We Will Identify Elementary Building Blocks and Control Mechanisms for DAU

FURTHER DAU RESEARCH

Class. = Classifying component
Extr. = Extracting component
Fusion = Result fusing component

Research Topics 4 & 5: Business Process Orientation

flexible, scalable OM **framework** for evolutionary growth

representation and enactment of **weakly-structured workflows**

toolkit for construction and maintenance of **domain ontologies**

improved information access by developing more integrated/easier adaptable **DAU techniques**

methodology and tool for **business-process oriented Knowledge Management**
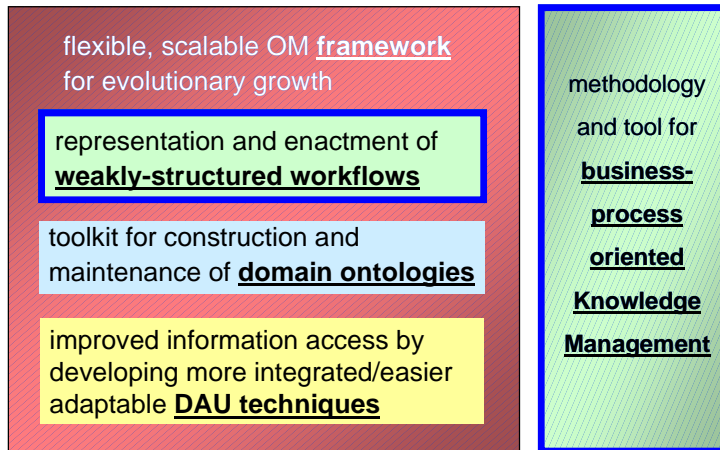
FRODO's Knowledge Management is Oriented Towards Business Processes

- business processes constitute the context for active situation-specific retrieval and task-specific archiving

- business process oriented knowledge management requires:
  - a systematic approach for introduction and evolution of OM solutions with appropriate modeling support
  - a suitable notion of processes beyond the limitations of strict workflow
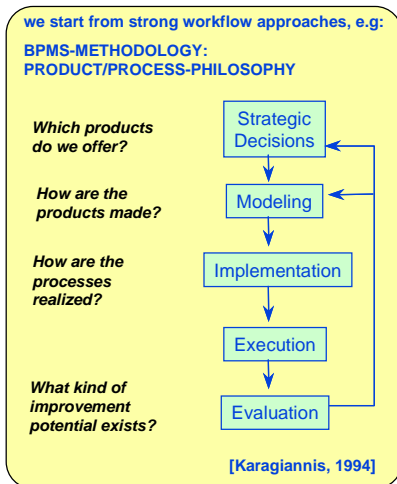
11

The Methodology Shall Result in a ‚Handbook for OM Introduction'

BUSINESS PROCESS ORIENTED KNOWLEDGE MANAGEMENT

we start from strong workflow approaches, e.g:

BPMS-METHODOLOGY: PRODUCT/PROCESS-PHILOSOPHY

Which products do we offer? → Strategic Decisions

How are the products made? → Modeling

How are the processes realized? → Implementation

→ Execution

What kind of improvement potential exists? → Evaluation

[Karagiannis, 1994]

Initial phase:

- OM-relevant entities and relationships in a company:
  who, where, why, what ...

- introduction guidelines:
  - identify suitable pilots
  - structure/evaluate knowledge sources
  - find communication streams
  - determine partial process models
  - ...

Iterative refinement:

- interleaved planning and execution of weakly-structured, interdependent activities and their knowledge needs

Source: Knowledge Management Group   DFKI   Page 23

The Methodology Will be Equipped With an Appropriate Modeling Tool

- modeling tasks:
  - business processes & information needs / flow
  - ontologies & knowledge item descriptions

- we will consider:
  - business process modeling (ARIS, ADONIS; UML)
  - machine learning for business process acquisition
  - integrated product and process modeling (bmb+f *Verbundprojekt* GiPP)
  - agent modeling
  - knowledge/ontological engineering (Protégé, KADS organizational model, ...)

The integrated modeling approach shall be based on standard techniques but must be extended for weak workflows.

Source: Knowledge Management Group   DFKI   Page 24

Weakly-Structured Workflow Will Serve as *FRODO*'s OM Task Level

- knowledge-intensive activities are seldom adequately represented by strictly formalized workflows

- weak workflow characteristics:
  - workflow control not completely predetermined
  - interleaved modeling/refinement and execution

- approach:
  - consider ad-hoc workflow, flexible workflow configuration
  - declarative formalism: logic-based, agent-based
  - open question: how can the *KnowMore* notion of context (based on strong workflow models) be transferred to the weak workflow scenario

---

Core Research Areas of *FRODO*

SUMMARY

- specialized speech acts, protocols, and standard agents for business-process oriented knowledge management

- distributed / networked knowledge representation

- domain ontology acquisition based on learning from texts and user interaction

- DAU techniques (classification / information extraction) for structured documents & with background knowledge

- declarative representation and enactment of weakly-structured workflows