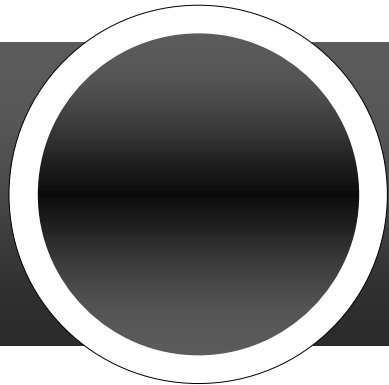


Seminar: Business Intelligence –
Teil II: Data Mining & Knowledge Discovery



Web Knowledge Management

Anne Jannasch

13.02.2004

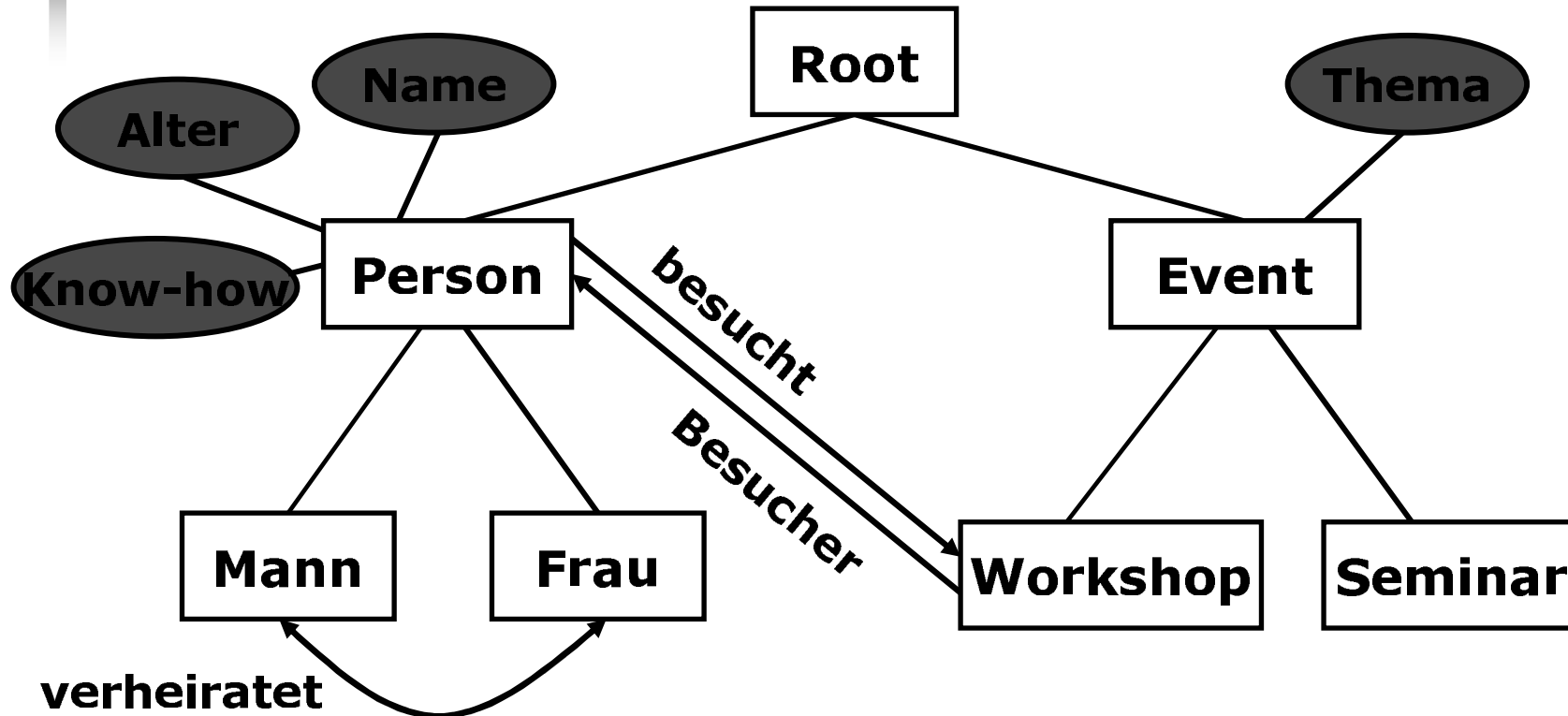
Überblick

- **Ontologie-Begriff**
- **Ontologiesprachen**
 - RDF und RDFS
 - OIL
 - F-Logic
- **Ontologiewerkzeuge**
- **Ontologie-basierte Systeme**
 - Ontobroker
 - On-To-Knowledge
- **Methoden zur Extraktion und Verwendung von taxonomischen Beziehungen**

Ontologie-Begriff

- Definition: „Eine Ontologie ist eine Spezifikation einer Konzeptualisierung.“ (Tom Gruber)
- Charakteristika einer Ontologie:
 - Konzepte
 - Beziehungen
 - Eigenschaften
 - Regeln
- Organisation einer Ontologie:
→ Taxonomien
- Wissensbasis = Ontologie + Instanzen

Beispiel für eine Ontologie



- Regeln:
 - „Verheiratet“ ist symmetrisch.
 - „Besucht“ und „Besucher“ sind invers.
 - Wenn eine Person einen Workshop besucht, eignet sie sich Wissen über ein Thema an.

Überblick

- ✓ Ontologie-Begriff
- **Ontologiesprachen**
 - RDF und RDFS
 - OIL
 - F-Logic
- Ontologiewerkzeuge
- Ontologie-basierte Systeme
 - Ontobroker
 - On-To-Knowledge
- Methoden zur Extraktion und Verwendung von taxonomischen Beziehungen

RDF und RDFS

- RDF steht für Resource Description Framework
- Datenmodell zur Repräsentation von Metadaten:
 - Ressourcen
 - Alles, was durch RDF beschrieben werden kann.
 - Eigenschaften
 - Charakteristiken, Attribute oder Relationen einer Ressource.
 - Aussagen
 - Bestehen aus Subjekt, Prädikat und Objekt.

RDF und RDFS

- Beispiel: Die Ressource <http://www.w3.org/Home/Lassila> hat den Autor Ora Lassila.

| | |
|------------------------|---|
| Subjekt (Ressource) | http://www.w3.org/Home/Lassila |
| Prädikat (Eigenschaft) | Autor |
| Objekt (Wert) | „Ora Lassila“ |

- In XML-Syntax:

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:s="http://description.org/schema/">
  <rdf:Description about="http://www.w3.org/Home/Lassila">
    <s:Creator>Ora Lassila</s:Creator>
  </rdf:Description>
</rdf:RDF>
```

RDF und RDFS

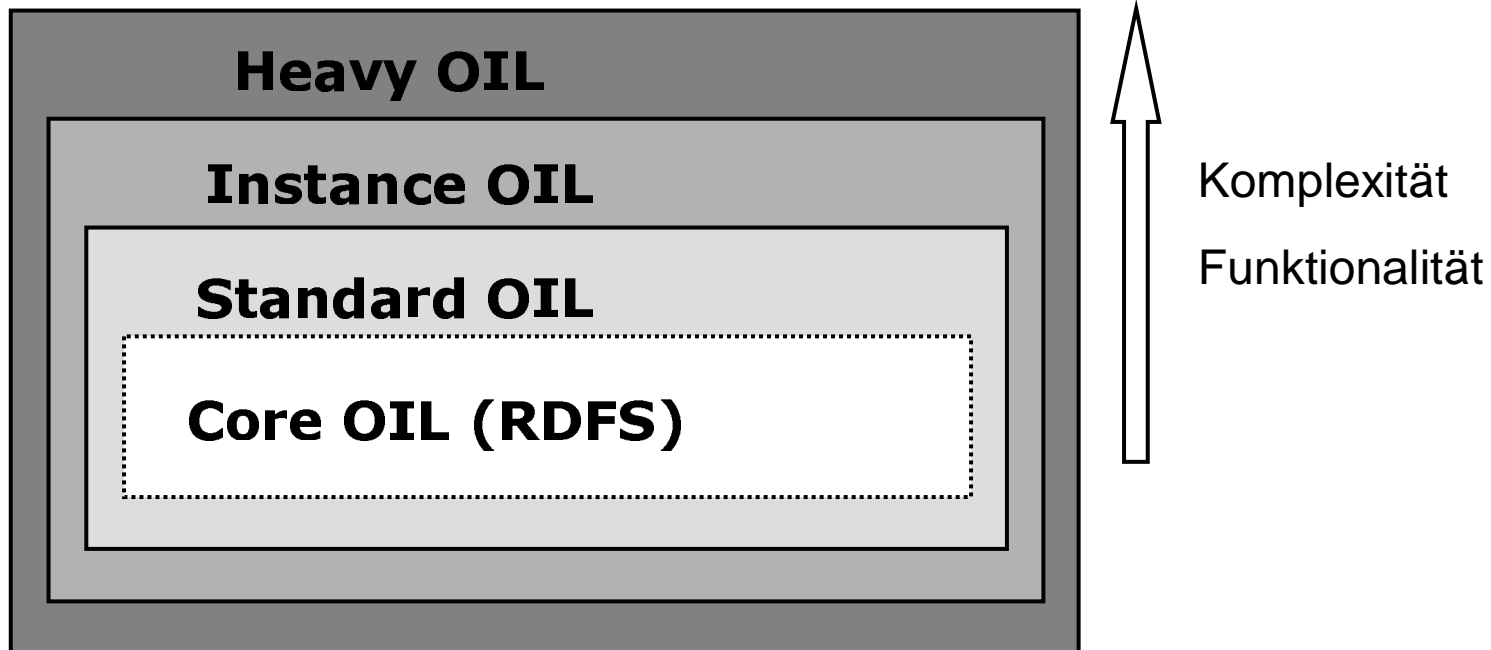
- RDFS steht für RDF-Schema.
 - Erweiterung von RDF
 - Formale Definition von Klassen und Eigenschaften
 - Vererbungshierarchien
 - Einschränkungen
- RDFS bietet Beschreibung von Schemata, die mit Ressourcen instanziiert werden können.
 - Semantische Interpretation von Ressourcen

OIL

- OIL steht für Ontology Inference Layer
- OIL führt drei Sprachfamilien zusammen:
 - Frame-basierte Systeme
 - Frames (Klassen)
 - Slots (Attribute)
 - Deskriptive Logik
 - Konzepte
 - Rollen
 - Inferenzen
 - Web-Sprachen
 - RDF und RDFS

OIL

- OIL ist in Schichten eingeteilt:



F-Logic

- Kombination der Repräsentationsformalismen Frames und Prädikatenlogik
- Definition von Inferenzregeln
- Modellierungsprinzipien:

| F-Logic | Bedeutung |
|----------------|--|
| $C1 :: C2$ | C1 ist Unterklasse von C2. |
| $O : C$ | O ist eine Instanz von C. |
| $C1[A=>>C2]$ | Für die Klasse C1 ist ein mehrwertiges Attribut A definiert, dessen Werte eine Instanz der Klasse C2 sind. |
| $->$ | Impliziert (wird zur Modellierung von Regeln verwendet). |

Überblick

- ✓ **Ontologie-Begriff**
- ✓ **Ontologiesprachen**
 - RDF und RDFS
 - OIL
 - F-Logic

→ **Ontologiewerkzeuge**

- **Ontologie-basierte Systeme**
 - Ontobroker
 - On-To-Knowledge
- **Methoden zur Extraktion und Verwendung von taxonomischen Beziehungen**

Ontologiewerkzeuge

- Hoher Arbeitsaufwand beim Erstellen von Ontologien
- Ontologiewerkzeuge helfen bei der Entwicklung und Verwendung von Ontologien
 - Lokal installierbar
 - Web-basierend

Protégé-2000

- Protégé-2000
 - Ontologie-Editor der Universität Stanford
 - Kostenlos im Internet erhältlich
 - In Java geschrieben
 - Erweiterbar durch Plugins

Protégé-2000

- Wissensmodell
 - Frame-basiert
 - Protégé-2000 Ontologien bestehen aus:
 - Klassen: Konzepte aus dem zu modellierenden Weltausschnitt
 - Slots: Attribute der Klassen
 - Facetten: Eigenschaften von Slots
 - Axiome: Zusätzliche Einschränkungen
- Speicherarten der erstellten Wissensbasis
 - Standard Text File
 - RDF
 - JDBC

Protégé-2000

The screenshot displays the Protégé-2000 interface with a project named 'newspaper'. The main window shows a class hierarchy on the left and a detailed view of the 'Editor' class on the right. The 'Slots' tab is selected in the top menu, and a 'salary' slot configuration dialog is open in the foreground.

Class Hierarchy (Left Panel):

- THING^A
 - SYSTEM-CLASS^A
 - Author^A
 - News_Service
 - Columnist^M
 - Editor^M
 - Reporter^M
 - Content^A
 - Advertisement^A
 - Article
 - Layout_info^A
 - Billing_Chart
 - Content_Layout
 - Prototype_Newspaper
 - Rectangle
 - Section
 - Library
 - Newspaper
 - Organization
 - Person
 - Employee^A
 - Columnist^M
 - Editor^M
 - Reporter^M
 - Salesperson
 - Manager

Editor Class Properties (Right Panel):

- Name:** Editor
- Documentation:** Editors are responsible for the content of sections.
- Constraints:** editor-employees-salary-c
- Role:** Concrete
- Template Slots Table:**

| Name | Type | Cardinality | Other Facets |
|----------|----------|-------------|-------------------|
| sections | Instance | multiple | classes={Section} |
| respons | | | |
| name | | | |
| date_hir | | | |
| salary | | | |
| current_ | | | |
| phone_r | | | |
| other_in | | | |

Salary Slot Configuration Dialog (Foreground):

- Name:** salary
- Documentation:** periodic wage paid to someone for work
- Value Type:** Float
- Cardinality:**
 - required
 - multiple
 - at least: []
 - at most: 1
- Default:** []
- Inverse Slot:** []
- Minimum:** 1.000
- Maximum:** 15.000

Überblick

- ✓ Ontologie-Begriff
- ✓ Ontologiesprachen
 - RDF und RDFS
 - OIL
 - F-Logic
- ✓ Ontologiewerkzeuge
- **Ontologie-basierte Systeme**
 - Ontobroker
 - On-To-Knowledge
- Methoden zur Extraktion und Verwendung von taxonomischen Beziehungen

Ontologie-basierte Systeme

- Wissensmanagement ist kritischer Erfolgsfaktor für Unternehmen
- Dokument-Management-Systeme mit Schwächen

→ Ontologie-basierte Systeme

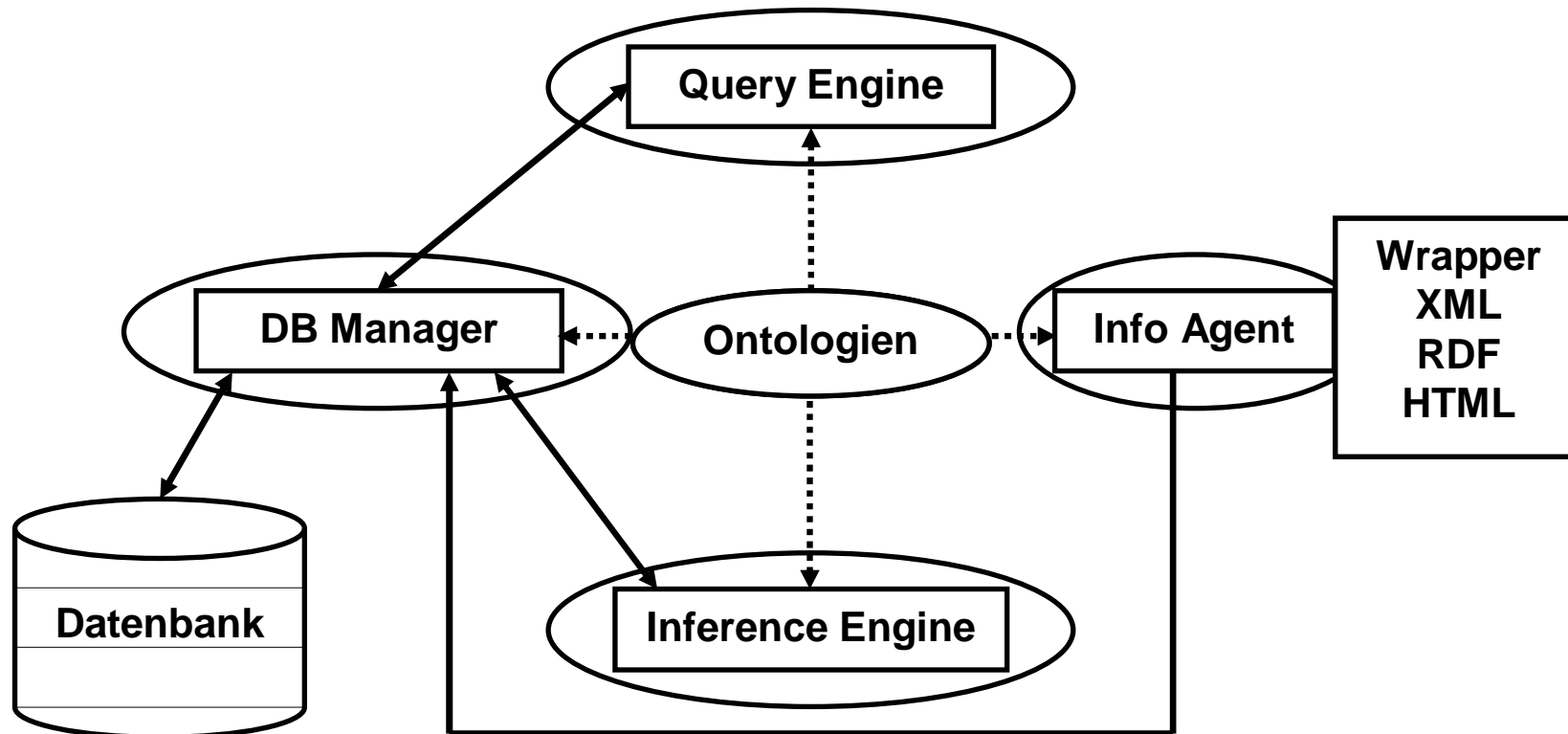
- Verbinden Wissen mit Hilfe von Ontologien
- Leiten neues Wissen ab

Ontobroker

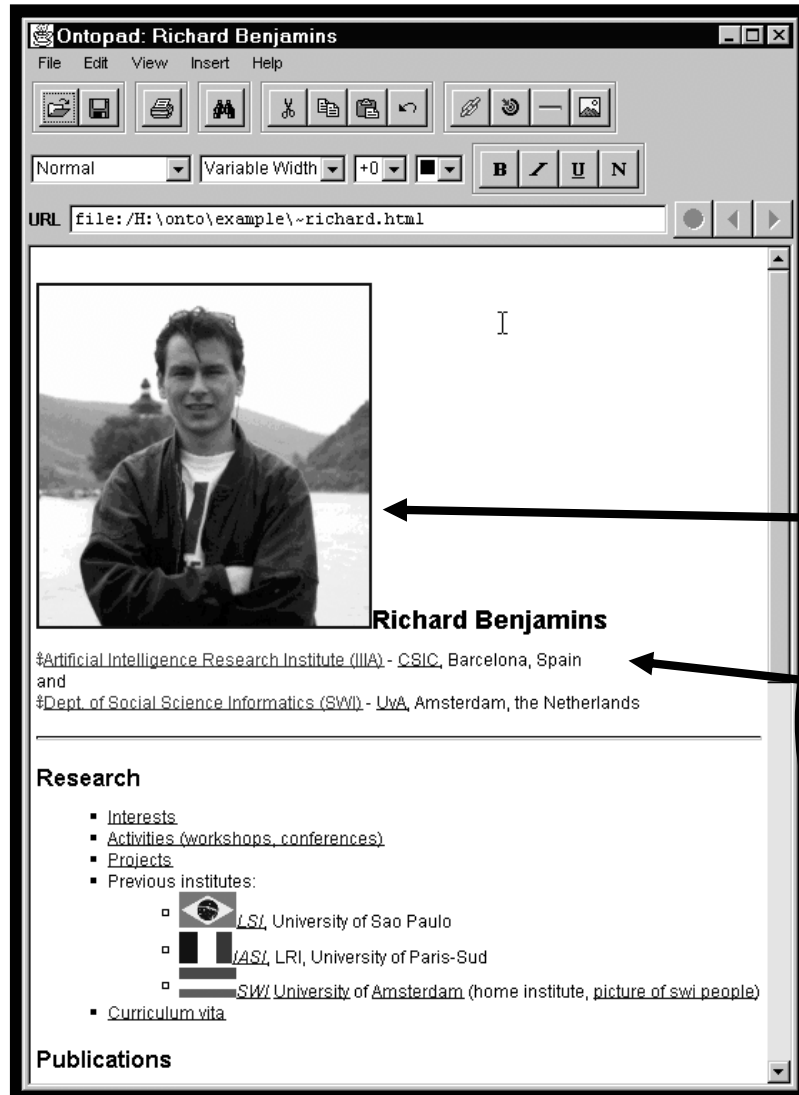
- Firma Ontoprise in Karlsruhe
- Ontologie-basiertes Wissensmanagement-System
- Zugriff auf verschiedenartige Quellen
- Ableiten von neuem Wissen aus vorhandenen Fakten

Ontobroker

- Aufbau von Ontobroker



Ontobroker



```

<HTML>
  <HEAD>
    <TITLE>
      Richard Benjamins
    </TITLE>
    <A onto="page:Researcher"> </A>
  </HEAD>
  <H1>
    <A HREF="pictures/id_rich.gif">
      <IMG align="middle"
        SRC="pictures/richard.gif"></A>
    <A onto="page[photo=href]"
      href="http://www.iiia.csic.es/~richard/
        pictures/richard.gif" ></A>
    <A onto="page[firstName=body]"
      Richard</A>
    <A onto="page[lastName=body]"
      Benjamins </A>
  </H1> <P>
    <A onto="page[affiliation=body]"
      href="#card">
      Artificial Intelligence Research Institute
      (IIIA)</A> -
    <A HREF="http://www.csic.es/">CSIC</a> Barcelona,
    Spain <BR>
    and <BR>
    ...
  
```

Ontobroker

- Funktionsweise von Ontobroker
 - Eigenständige Java-Anwendung
 - Server

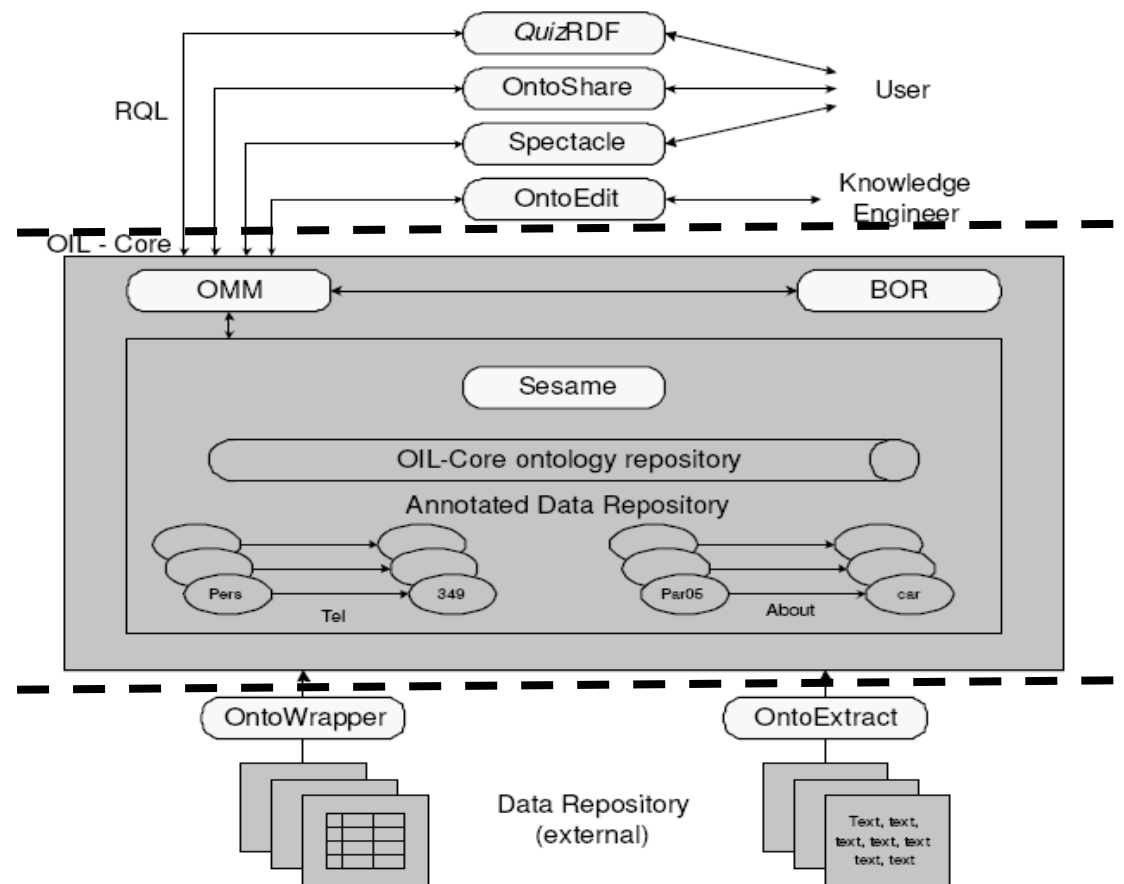
 - Anwendungen
 - (KA)2 Initiative
 - Integration nur bei annotierten Seiten
 - Verschiedene „Ontogruppen“
 - Kein Ersatz für Suchmaschinen
- Schwerpunkt liegt auf Integration in Anwendungen einzelner Gruppen

On-To-Knowledge

- Europäisches Projekt, 2000 – 2002
- Entwicklungsziele:
 - Werkzeugumgebung
 - Ontologiesprache OIL
 - Methodologie
- Zentrales Ergebnis:
 - Ontologie-basierte Werkzeug-Umgebung für Wissensmanagement
 - Unterstützung von Firmen-Netzwerken und Internet

On-To-Knowledge

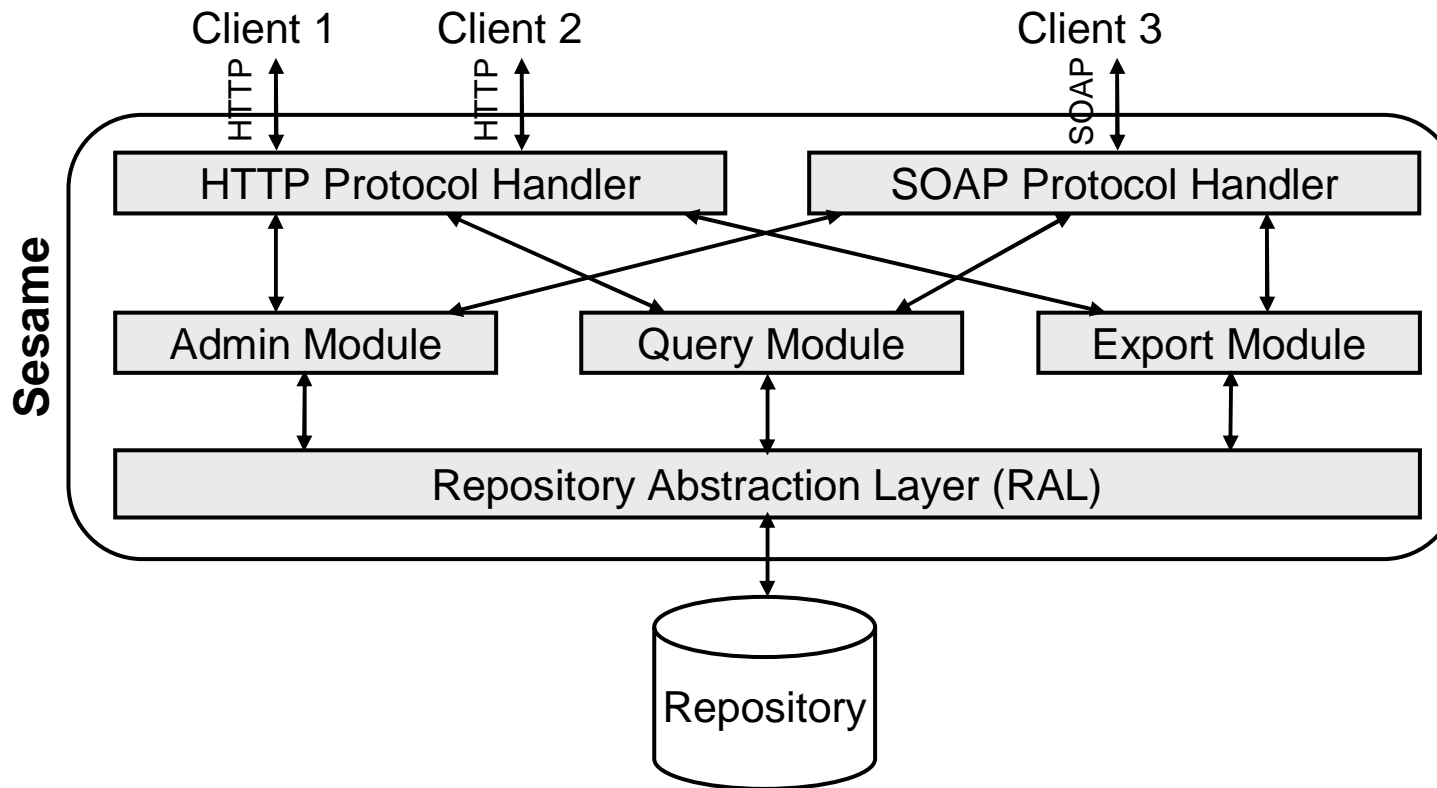
- Aufbau



On-To-Knowledge

- Sesame

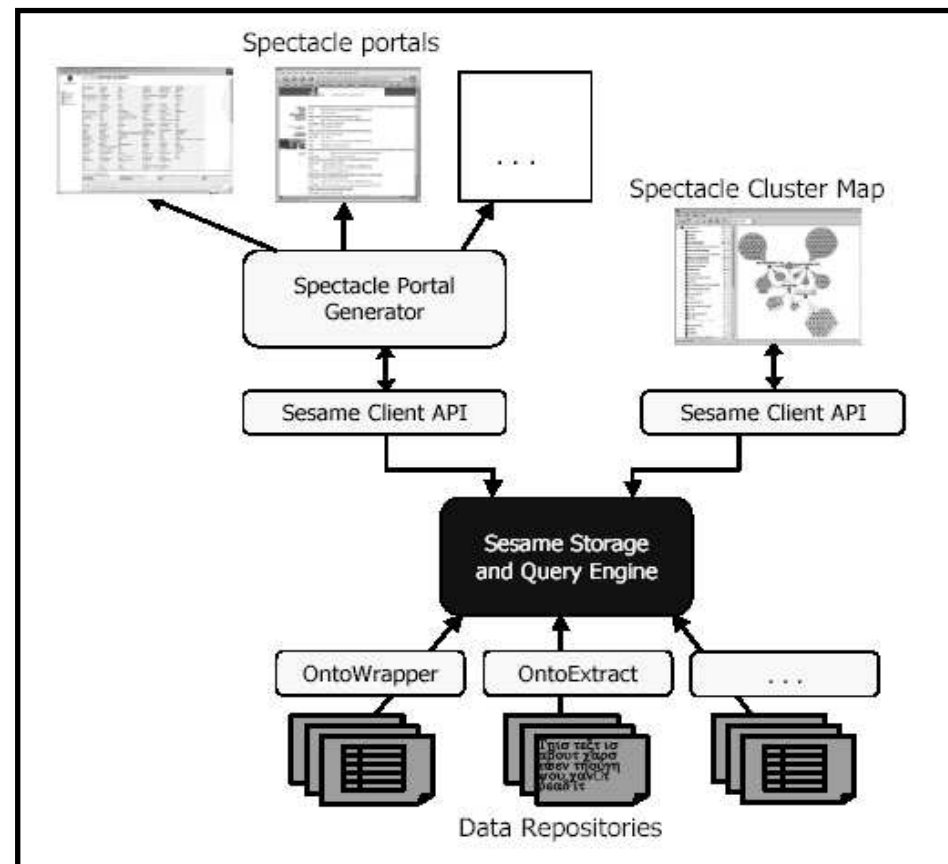
→ Persistentes Speichern, Archivierung und Abfrage von RDF und RDFS-Daten



On-To-Knowledge

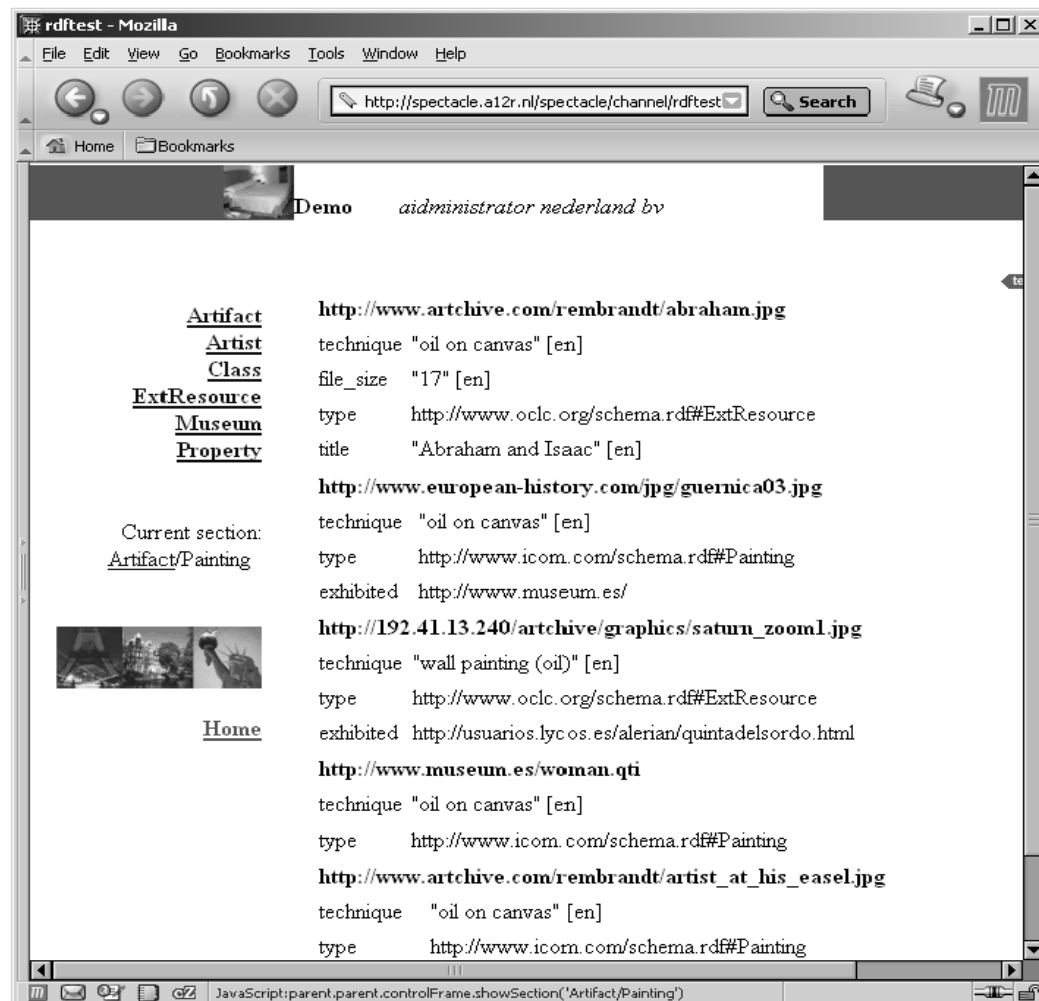
- Spectacle

- Plattform mit maßgeschneiderter Informationsdarstellung
- Erzeugen von Informationsrepräsentationen
- Formatiert und gegliedert
- „Navigieren“



On-To-Knowledge

- Spectacle
 - RDF-Explorer



On-To-Knowledge

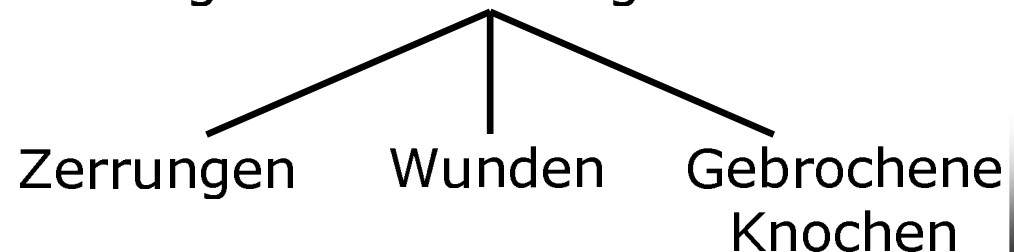
- **OntoShare**
 - Gemeinsamer Zugriff einer Arbeitsgruppe
 - Interessen als Anwenderprofil
 - Speichern, Wiedergewinnen und Zusammenfassen von Informationen
 - Informieren anderer Anwender
- **Funktionalitäten**
 - E-Mail Benachrichtigung
 - Suchen im gemeinsamen Speicher
 - Personalisierte Information

Überblick

- ✓ Ontologie-Begriff
- ✓ Ontologiesprachen
 - RDF und RDFS
 - OIL
 - F-Logic
- ✓ Ontologiewerkzeuge
- ✓ Ontologie-basierte Systeme
 - Ontobroker
 - On-To-Knowledge
- **Methoden zur Extraktion und Verwendung von taxonomischen Beziehungen**

Extraktion taxonomischer Beziehungen

- Syntaktische Analyse
 - Suche nach Instanzen von lexikalisch-syntaktischen Mustern
- Beispiel:
 - Muster: ...NP {,NP} * {,} oder andere NP ...
 - Satz: „Zerrungen, Wunden, gebrochene Knochen oder andere Verletzungen kommen häufig vor.“
 - Taxonomische Beziehungen: Verletzungen



Extraktion taxonomischer Beziehungen

- Statistisch basierte Analyse
 - Erstellen von taxonomischen Beziehungen
 - Folgern der Semantik einer neuen Klasse
 - Klasse in Beziehung zu vorhandenen Klassen setzen
 - Distributionale Daten
 - Keine Muster-Erstellung, sondern Automatisierung
 - Automatische Klassifikationsmethoden

- Verfahren der k-nächsten Nachbarn
 - Neues Objekt in Klasse mit meisten „k-nächsten Nachbarn“
 - Neue Klasse in taxonomische Struktur
 - Hyponym jener Klasse mit größter Anzahl von Hyponymen mit nächsten Nachbarn

Verwendung taxonomischer Beziehungen

- Kombination der Informationen über taxonomische Struktur mit statistischen Daten über Konzepte
- Tree Descending Algorithmus
 - Semantik jeder Klasse spiegelt wichtigste semantische Eigenschaften ihrer Hyponyme wieder
 - Abstieg von Wurzel zu Blättern
 - Neues Objekt in Klasse mit größter Anzahl ähnlicher Objekte
- Tree Ascending Algorithmus
 - Kombination taxonomischer und distributionaler Ähnlichkeit zwischen nächsten Nachbarn

Zusammenfassung

- ✓ **Ontologie-Begriff**
→ „Spezifikation einer Konzeptualisierung“
- ✓ **Ontologiesprachen**
→ Formale Darstellung von Ontologien
- ✓ **Ontologiewerkzeuge**
→ Entwicklung und Bearbeitung von Ontologien
- ✓ **Ontologie-basierte Systeme**
→ Verbinden und leiten neues Wissen mit Ontologien ab
- ✓ **Methoden zur Extraktion und Verwendung von taxonomischen Relationen**
→ Automatisierung des Ontologie-Erzeugungsprozesses



**Vielen Dank
für
Ihre Aufmerksamkeit !**