Semantic Web Technologies II

SS 2008

02.07.2008

Semantic Web 2.0 III

Dr. Sudhir Agarwal
Dr. Stephan Grimm
Dr. Peter Haase
PD Dr. Pascal Hitzler
Denny Vrandečić

Übersicht

- Semantic Web
  - BT Digital Library
  - Project Halo
- Semantic Web 2.0? Web 3.0?
  - Semantic Blogging
  - OntoGame
  - Other examples

Semantic Web

Definition Semantic Web

"The **Semantic Web** provides a common framework that allows **data** to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework."

W3C, 2008

http://www.w3.org/2001/sw/

Semantic Web Technology Applications

- case studies in closed domains
- complex & comprehensive modeling
- team of knowledge engineers
- sophisticated reasoning
- unwieldy tools and obscure specifications

Begriffsklärung

- **Semantic Web**
  - Linked data
  - Oder Anwendungen von Linked data
  - Nutzt Semantic Web Technologies
- **Semantic Web Technologies**
  - W3C Standards und ihre Anwendung
  - Insbesondere RDF, OWL, SPARQL
- **Semantic Technologies**
  - Viel mehr! Computer arbeiten mit Bedeutung
  - Bsp: Cyc, LSI, Tagging, WordNet, Regeln, WolframAlpha, Goolge Squared, True Knowledge …

Semantic Web

Example: BT Digital Library

Scenario (BT Digital Library)

Bob works as technology analyst for British Telecom. His daily work includes research on new technological trends, market developments as well as the analysis of competitors.

*Bob's company maintains a **digital library** that gives access to a **repository of internal surveys and analysis documents**. The company also has a **license** with an **academic research database** which is accessed via a **separate interface**.*

*Depending on his work context, Bob uses the **topic hierarchies**, the **full-text search functionalities** or **metadata search facilities** provided by the two libraries to get access to the relevant data.*

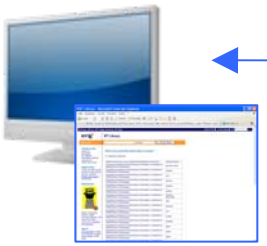
*However, Bob is often annoyed by the **differing topic hierarchies and metadata schemes** used by the two libraries as well as by a **cumbersome syntax for metadata queries**.*

- Heterogeneity of content
- Heterogeneity of search facilities
- Heterogeneity of data models (schemas)
- Interface design challenge

Why Ontology-Based Digital Libraries?

- Immediate support for unified structured queries against metadata and documents
- Easy integration of heterogeneous knowledge sources
- Easy integration with knowledge elicitation methods from unstructured content
- Mapping to natural language queries
- Generic, flexible and modular architecture

Conceptual Architecture



Knowledge Portal

Query Translation Component

Query Interface

Reasoning Engine

Query Answering Component

Built-In Library

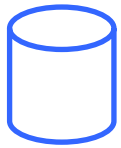
Full Text Indexing

Text Classification

Ontology Model and Knowledge Base

Internal Storage

Data Connectors



Ontology Learning

Document Metadata

Full-text documents

Ontology Model and Knowledge Base

- Ontology (PROTON top level ontology)
  - global conceptual model
  - aligned with established schemas (e.g. Dublin Core)
- Knowledge base of the digital library
  - actual bibliographical metadata, topic hierarchies, and full-text document content
  - data aligned with global ontology via mapping axioms

```
swrc:Book            rdfs:subClassOf    protont:Document
expl:document5127    rdf:type           swrc:InProceedings
expl:document5127    protont:title      "Digital Libraries"
```

- Query answering against knowledge base (SPARQL)

```
SELECT ?x WHERE {
    ?x rdf:type <http://proton.semanticweb.org/2005/04/protonu#Article> .
    ?x <http://proton.semanticweb.org/2005/04/protont#hasSubject> ?y .
    ?y rdfs:label ?z .
    match(?z,"Intellectual Capital")
}
```

Knowledge Portal

- Presentation layer for underlying content
- Interaction via standard interfaces
  - keyword-search, topic browsers etc.
- Interaction via natural language queries
  - converts natural language queries into SPARQL
- Translation step comprises
  - deep parsing of the questions
  - roughly, linguistic frames become query constraints
  - lexicon describes possible realizations of elements

" Who wrote books on 'digital libraries'?"
"Which journal articles were written by 'Tim Berners-Lee' (and for which journal)?"

Scenario Revisited

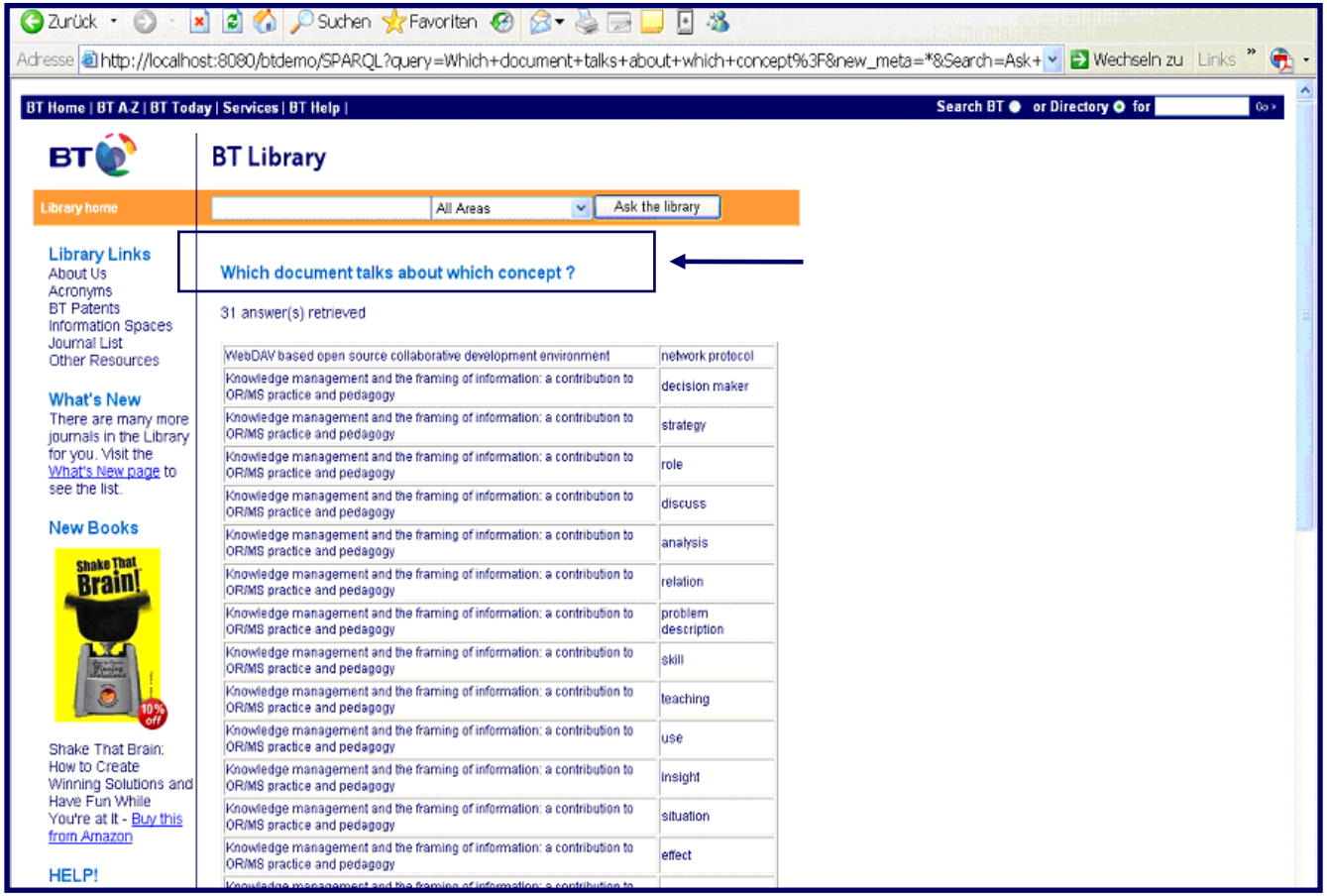
"Which journal articles were written by 'Tim Berners-Lee' for which journal?"

```
PREFIX protonu:<http://proton.semanticweb.org/2005/04/protonu#>
PREFIX protont:<http://proton.semanticweb.org/2005/04/protont#>

SELECT ?x ?z WHERE {
   ?x rdf:type protonu:Article .
   ?x protont:documentAuthor ?y .
   ?y rdfs:label ?ys .
   match(?ys, "Tim Berners Lee") .
   ?z rdf:type protonu:Journal .
   ?x protonu:publishedWithin ?z
}
```

"The Semantic Web"	"The Scientific American"
"WWW: Past, Present, and Future"	"IEEE Computer"
[...]	[...]

The BT Digital Library



Screenshot from BT Digital Library

Semantic Web

Example: Project Halo

Project Halo

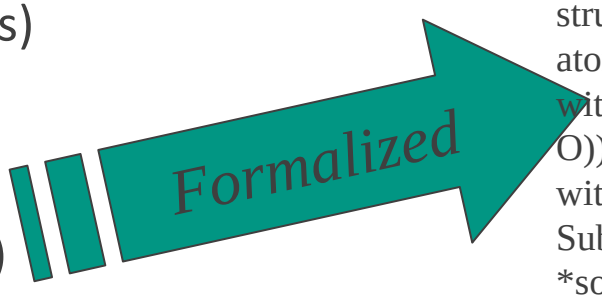
- "Building a digital Aristotle"
- A system that…
  - Encompasses much of the world's knowledge
  - Reasons over that knowledge
  - Answers novel scientific questions
  - Explains these answers
  - Is quite ambitous
- Multi-stage effort:
  - Start with a specific science (Chemistry)
  - Challenge with several teams
  - Answer AP-style questions
- Complete information at http://www.projecthalo.com/

Formalizing questions

- Example

  Which of the following compounds will produce a gas when HCl is added to the solid compound? HCl is a strong acid producing a yellow-green colored gas above the acid solution.

  - $Ba(OH)_2$ (s)
  - $CaCO_3$ (s)
  - $CuSO_4$ (s)
  - Na_3PO_4(s)
  - NaCl (s)

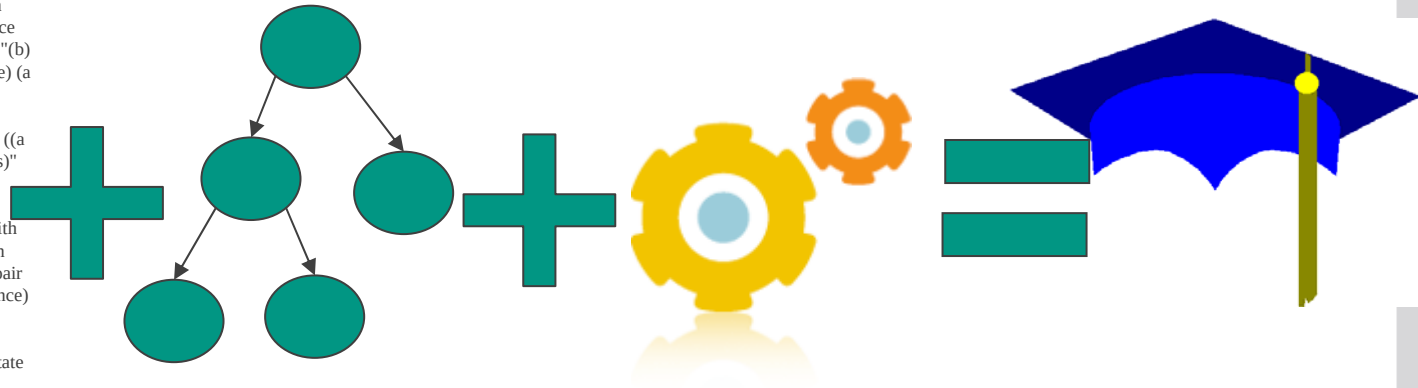


(every QF1 has (context ((:pair "(a) Ba(OH)2(s)" (a Reaction with (raw-material ((a HCl-Substance) (a Ba_OH_2-Substance with (state ((a State-Value with (value (*solid)))))))))) (:pair "(b) CaCO3(s)" (a Reaction with (raw-material ((a HCl-Substance) (a CaCO3-Substance with (state ((a State-Value with (value (*solid))))))))))) (:pair "(c) CuSO4(s)" (a Reaction with (raw-material ((a HCl-Substance) (a CuSO4-Substance with (state ((a State-Value with (value (*solid))))))))))) (:pair "(d) Na3PO4(s)" (a Reaction with (raw-material ((a HCl-Substance) (a Ionic-Compound-Substance with (state ((a State-Value with (value (*solid))))))) (has-basic-structural-unit ((a Ionic-Compound with (nested-atomic-chemical-formula ((a Chemical-Formula with (term ((:seq (:pair 3 Na) (:pair 1 P) (:pair 4 O)))))))))))))) (:pair "(e) NaCl(s)" (a Reaction with (raw-material ((a HCl-Substance) (a NaCl-Substance with (state ((a State-Value with (value *solid)))))))))))) (output ((forall (the context of Self) where (oneof2 (the result of (the2 of It)) where ((the value of (the state of It2)) = *gas)) (the1 of It) (comm [QF1-output-1] Self)))))

Background knowledge

- Formalizing questions is “just” question understanding
- Needs a huge amount of background knowledge = ontology
- And a reasoner to answer the question using the ontology

(every QF1 has (context ((:pair "(a) Ba(OH)2(s)" (a Reaction with (raw-material ((a HCl-Substance) (a Ba_OH_2-Substance with (state ((a State-Value with (value (*solid))))))))) (:pair "(b) CaCO3(s)" (a Reaction with (raw-material ((a HCl-Substance) (a CaO3-Substance with (state ((a State-Value with (value (*solid))))))))) (:pair "(c) CuSO4(s)" (a Reaction with (raw-material ((a HCl-Substance) (a CuSO4-Substance with (state ((a State-Value with (value (*solid))))))))) (:pair "(d) Na3PO4(s)" (a Reaction with (raw-material ((a HCl-Substance) (a Ionic-Compound-Substance with (state ((a State-Value with (value (*solid))))) (has-basic-structural-unit ((a Ionic-Compound with (nested-atomic-chemical-formula ((a Chemical-Formula with (term ((:seq (:pair 3 Na) (:pair 1 P) (:pair 4 O)))))))))))) (:pair "(e) NaCl(s)" (a Reaction with (raw-material ((a HCl-Substance) (a NaCl-Substance with (state ((a State-Value with (value *solid))))))))))) (output ((forall (the context of Self) where (oneof2 (the result of (the2 of It)) where ((the value of (the state of It2)) = *gas)) (the1 of It) (comm [QF1-output-1] Self)))))



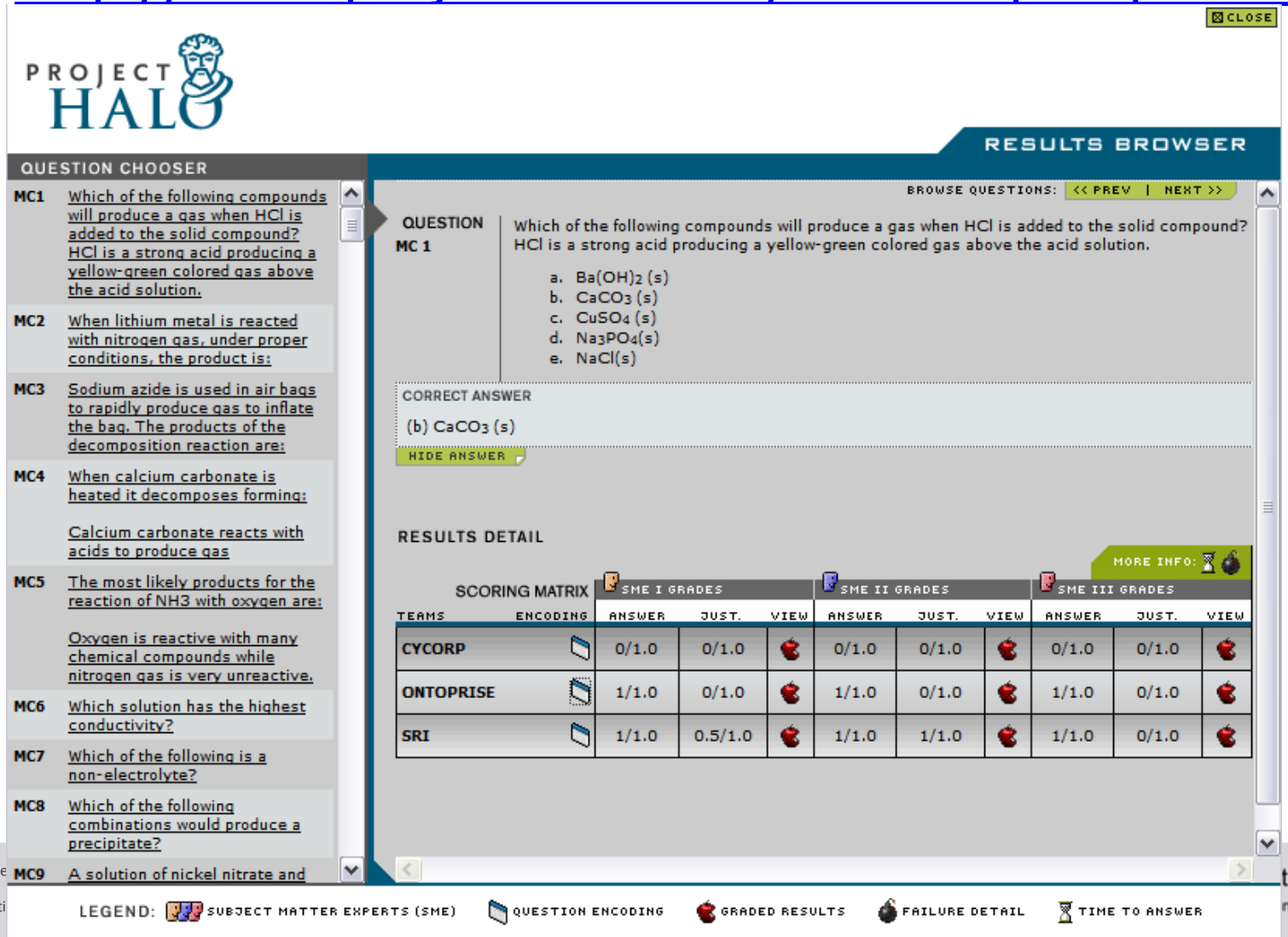
Evaluation

- Correctness
  - Was pretty high
- Justification
  - Considerably lower than correctness
- Speed
  - Was critical, but all systems faired well

- Results:
  - Human mean average in this test is AP-2.82
  - Project Halo scored an AP-3 – they would have passed!

Result browser

- http://www.projecthalo.com/halotempl.asp?cid=2



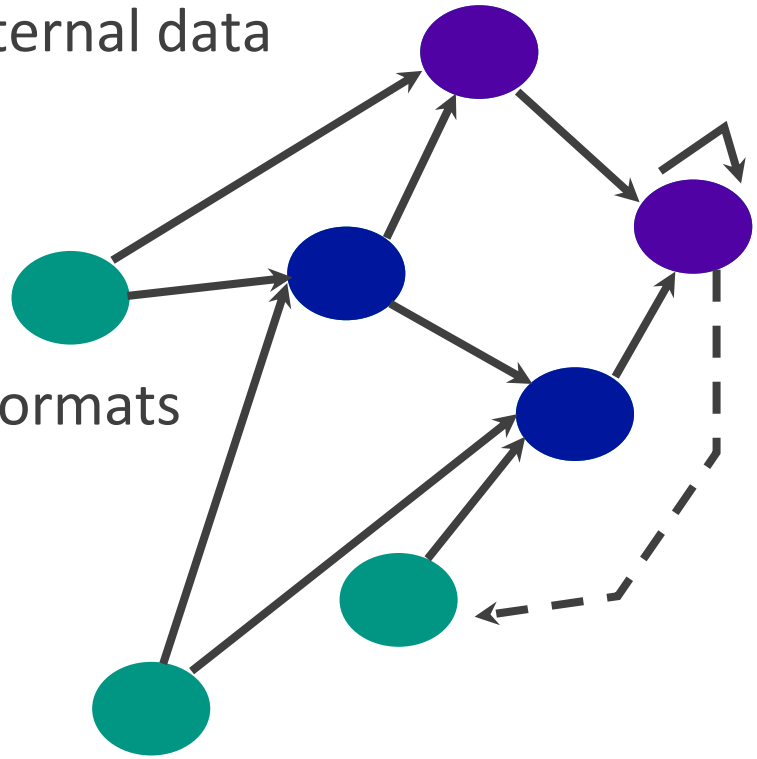
Semantic Web 2.0

Semantic Web of Data

- A World Wide Web (of data)
- Bottom-up, user-centred approach
- “A little semantics goes a long way”

Web Data Ecosystem

- Creation
  - expose application-internal data
  - tags
- Exchange
  - aggregators
  - other data, e.g microformats
  - data reconciliation
- Reuse
  - mashups



Semantic Web Application

- Nicht notwendig: RDF Backend
  - Implementierung ist irrelevant
  - Kann aber auch RDF-basiert sein – Joost
- Wichtig: Schnittstellen nach außen
  - RDF
  - SPARQL
  - Datenaustausch
- Nicht sein sondern schein

Semantic Web 2.0

Idea: Semantic Blogging

Chrissie's Blog

- loves to blog about movies she's seen for her friends
- "typical blogger"
  - blogging for 3 years
  - knows some HTML and CSS
  - can subscribe to RSS feeds
  - no knowledge of Semantic Web

Everything pink – Chrissies blog

Archive
June 2007
May 2007
April 2007
March 2007

About me
RSS-Feed

Blogroll
nutkidz
nakit-arts
Blog of the rings
Matrix Reblogged

Links
Gloria Cinema
Ecoshop
Legolas fanzine

Pirates of the Caribbean 3

June 21st, 2007

I just went with Till into the last part of the Pirates of the Caribbean, where our heroes (the adoringly cute Orlando Bloom and Keira Knightly reprise their roles) go to the end of the world to safe the one and only Captain Jack Sparrow (Johnny Depp! xOxOx!) from the claws of the Kraken. And guess what – Jack Sparrows daddy has a special appearance, played by old Rolling Stone Keith Richards! Weeeeha!

Best movie of the year, until know, without a question! Tons of fun, and colorful action.

no comments yet – post your comment - backtrack

Chrissie's blog workflow

- create new entry
  - enter title, write text, annotate with few tags like genre
- publish
  - entry saved in blog database
  - entry displayed on front page, archived, added to RSS feeds

Chrissie's Smoov blog workflow

- configure Smoov to show certain info
- create new entry
  - specify writing movie review, enter title, write text, identify movie, annotate with few tags
- publish
  - entry saved in blog database
  - Smoov processes entry
  - entry displayed on front page, archived, added to RSS feeds

Everything pink – Chrissies blog

Archive
June 2007
May 2007
April 2007
March 2007

About me
RSS-Feed

Blogroll
nutkidz
nakit-arts
Blog of the rings
Matrix Reblogged

Links
Gloria Cinema
Ecoshop
Legolas fanzine

Pirates of the Caribbean 3

June 21st, 2007

I just went with Till into the last part of the Pirates of the Caribbean, where our heroes (the adoringly cute Orlando Bloom and Keira Knightly reprise their roles) go to the end of the world to safe the one and only Captain Jack Sparrow (Johnny Depp! xOxOx!) from the claws of the Kraken. And guess what – Jack Sparrows daddy has a special appearance, played by old Rolling Stone Keith Richards! Weeeeha!

Best movie of the year, until know, without a question! Tons of fun, and colorful action.

Director George Verbinski
Running time 126 minutes
Starring Johnny Depp, Keira Knightley, Bill Nighy, Orlando Bloom, Geoffrey Rush
Info from Wikipedia

no comments yet – post your comment - backtrack

Everything pink – Chrissies blog

Archive
June 2007
May 2007
April 2007
March 2007

About me
RSS-Feed

Blogroll
nutkidz
nakit-arts
Blog of the rings
Matrix Reblogged

Links
Gloria Cinema
Ecoshop
Legolas fanzine

Pirates of the Caribbean 3

June 21st, 2007

I just went with Till into the last part of the Pirates of the Caribbean, where our heroes (the adoringly cute Orlando Bloom

See Pirates of the Caribbean 3
in the Gloria:
Today 16:00, 18:30, 21:00
Tomorrow 16:00, 18:30, 21:00
Reserve tickets now

old rolling Stone Keith Richards!
Weeeeha!

Best movie of the year, until know, without a question! Tons of fun, and colorful action.

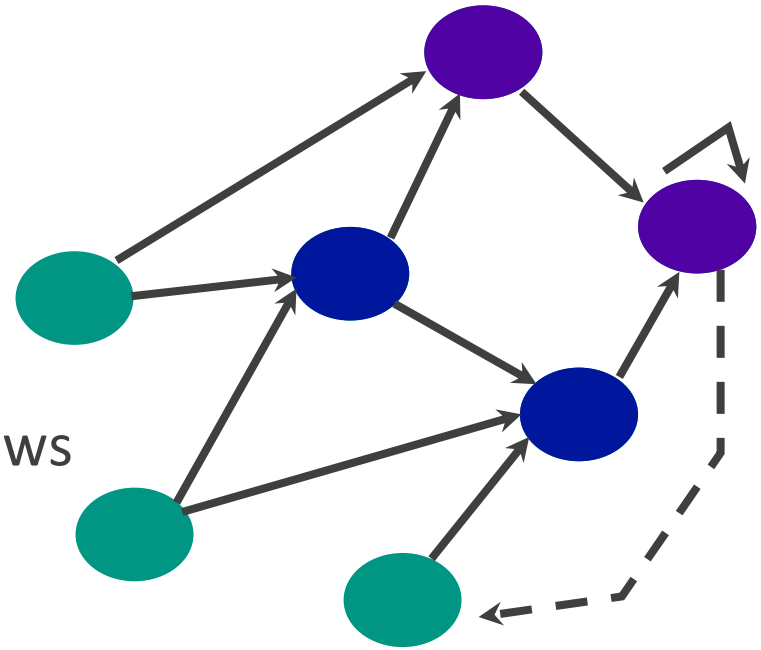
no comments yet – post your comment - backtrack

Director George Verbinski
Running time 126 minutes
Starring Johnny Depp, Keira Knightley, Bill Nighy, Orlando Bloom, Geoffrey Rush
Info from Wikipedia

See Pirates of the Caribbean 3
in the Gloria:
Today 16:00, 18:30, 21:00
Tomorrow 16:00, 18:30, 21:00
Reserve tickets now

Web Data Ecosystem

- Exchange
  - show aggregated reviews
- Reuse
  - spot trends
  - immediately produce meaningful movie recommendations



Web 2.0 or Semantic Web

blogging (use context)	
AJAX (dynamic display)	
interlinking of data (mashup)	large-scale interlinking of data
proprietary APIs	standardised APIs
niche vocabularies	comprehensive ontologies
screenscraping/RSS	well-defined data export
	data reconciliation

Web 2.0 or Semantic Web

blogging

AJAX

large-scale interlinking of data

standardised APIs

niche vocabularies

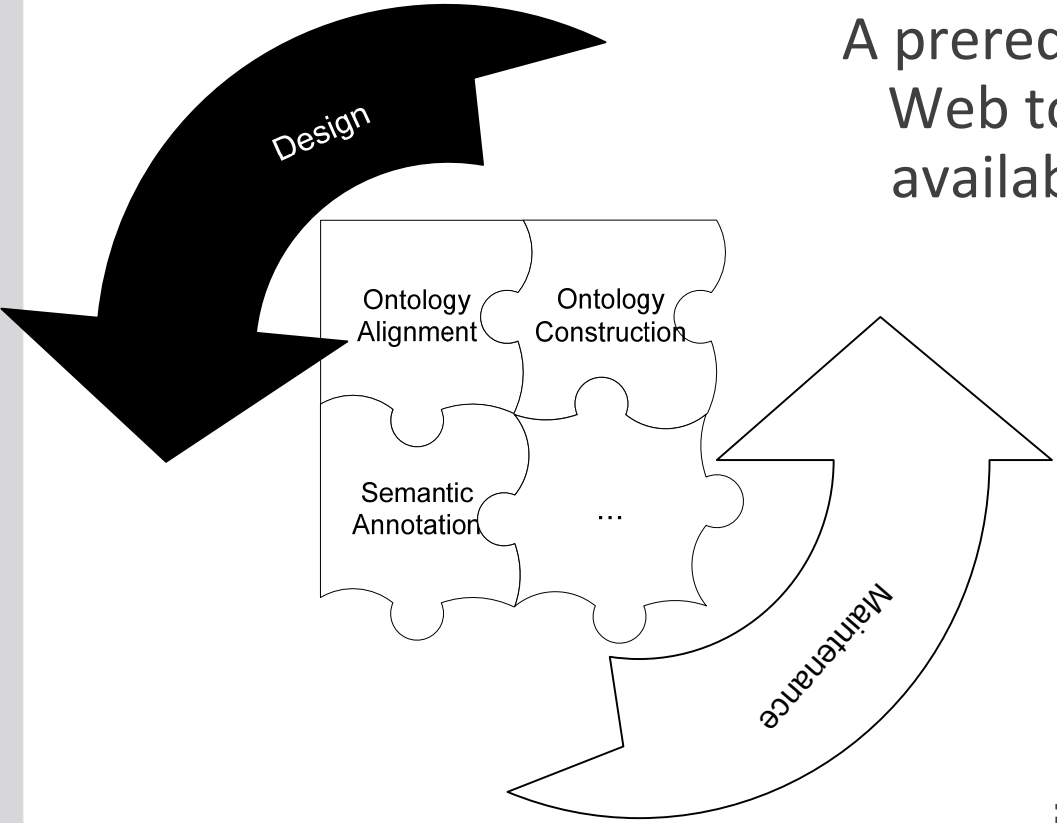
well-defined data export

data reconciliation

Semantic Web 2.0

Example: OntoGame

Lack of Human Contribution



A prerequisite for the Semantic Web to become a reality is the availability of **annotated data**.

Building the Semantic Web is not a one-time task, but a **continuous effort**.

Observation

There are tasks that are easy for humans but difficult for computers

Cf. Von Ahn

Not all the tasks on the Semantic Web can be automated.
Some at least partly require human intelligence.

The Million Dollar Question

How can we motivate people to dedicate their valuable time to build the Semantic Web?

- The EU can fund it
- We can hope for pure altruism
- **We can make it fun!**

A Huge Success: Luis von Ahn's Games with a Purpose

Some people are playing the game **more than 40 hours** per week.

The game collected more than 10 million consensual image labels.

The OntoGame Idea and Principles

Make people weave the Semantic Web by playing cool multi-player online games.

1. Fun and intellectual challenge
2. Consensus
3. Massive content generation

10 Challenges

1. Identifying suitable tasks in semantic content creation
2. Designing games
3. Designing a usable, attractive interface
4. Identifying suitable knowledge corpora
5. Preventing cheating
6. Defusing typical pitfalls of conceptual modeling
7. Distribution of labor
8. Fostering user participation
9. Deriving formal representations
10. Scalability and performance

OntoPronto:
Creating a Huge Domain Ontology

OntoTube: Annotating YouTube videos

Evaluation

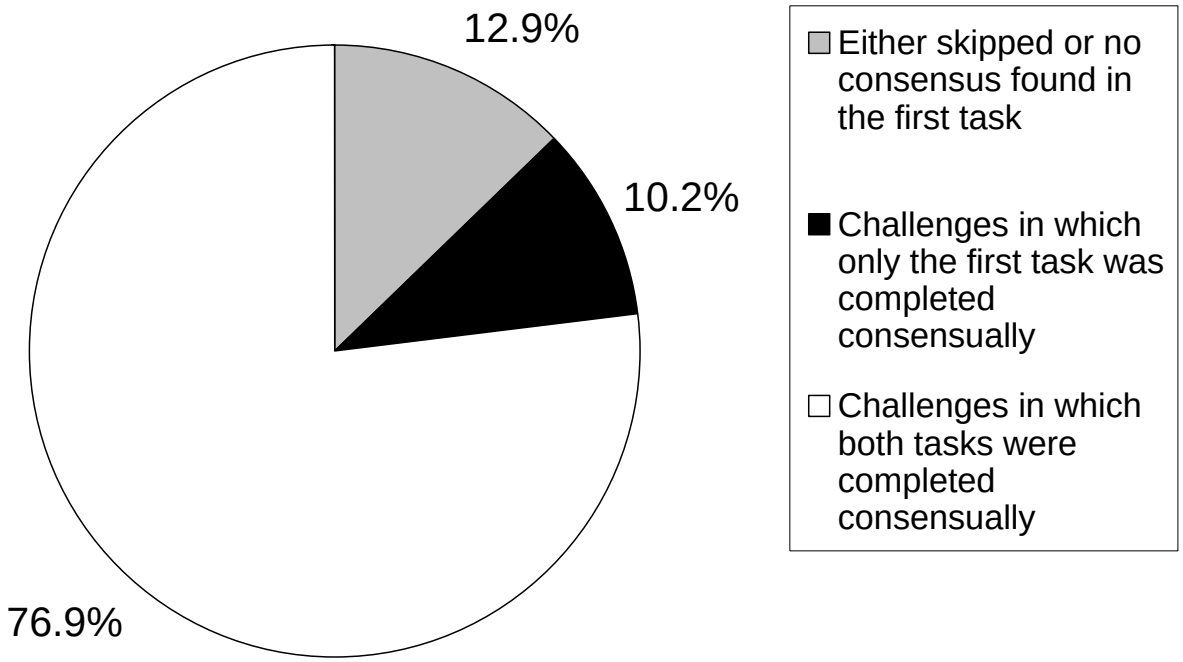
- Snapshot of RDF data 2 weeks after release
- 271 registered players within 2 weeks, 90% male
- More than 2500 games, 400 articles
- Is the game fun?
- Is the ontological data produced correct?

Katharina Siorpaes and Martin Hepp: Games with a Purpose for the Semantic Web.
IEEE Intelligent Systems, Vol. 23, No. 3, pp. 50-60, May/June 2008.

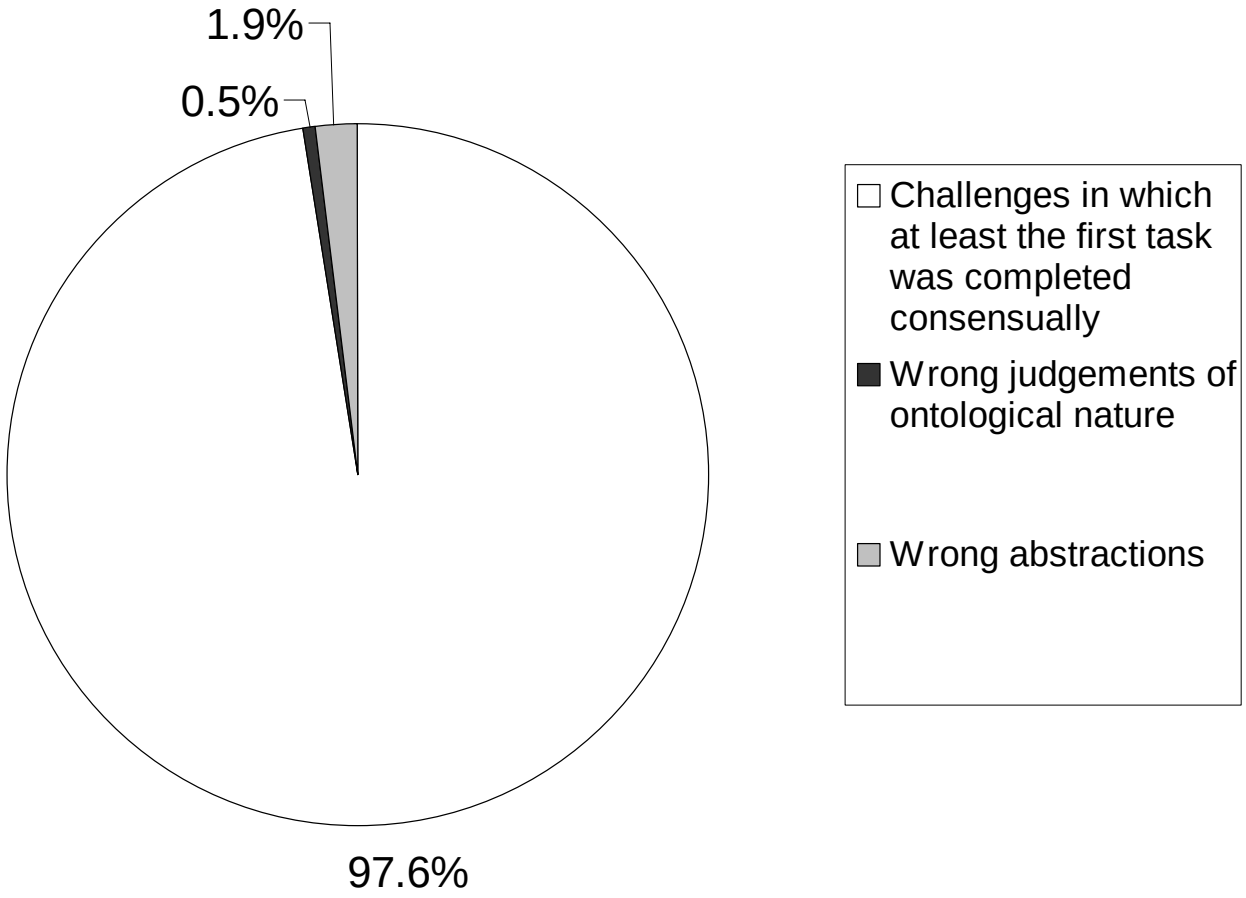
Fun Factor

- 35 players completed a survey
- Rules were understandable
- Entertaining (enthusiasts over-represented)
- Most rated the game „cool“ and entertaining
- Side effects: learning PROTON and Wikipedia
- Social component is weak (no communication, gender, age, nationality, etc.)
- „Constructive entertainment“

Consensus


12.9%
10.2%
76.9%
Either skipped or no consensus found in the first task
Challenges in which only the first task was completed consensually
Challenges in which both tasks were completed consensually

Correctness



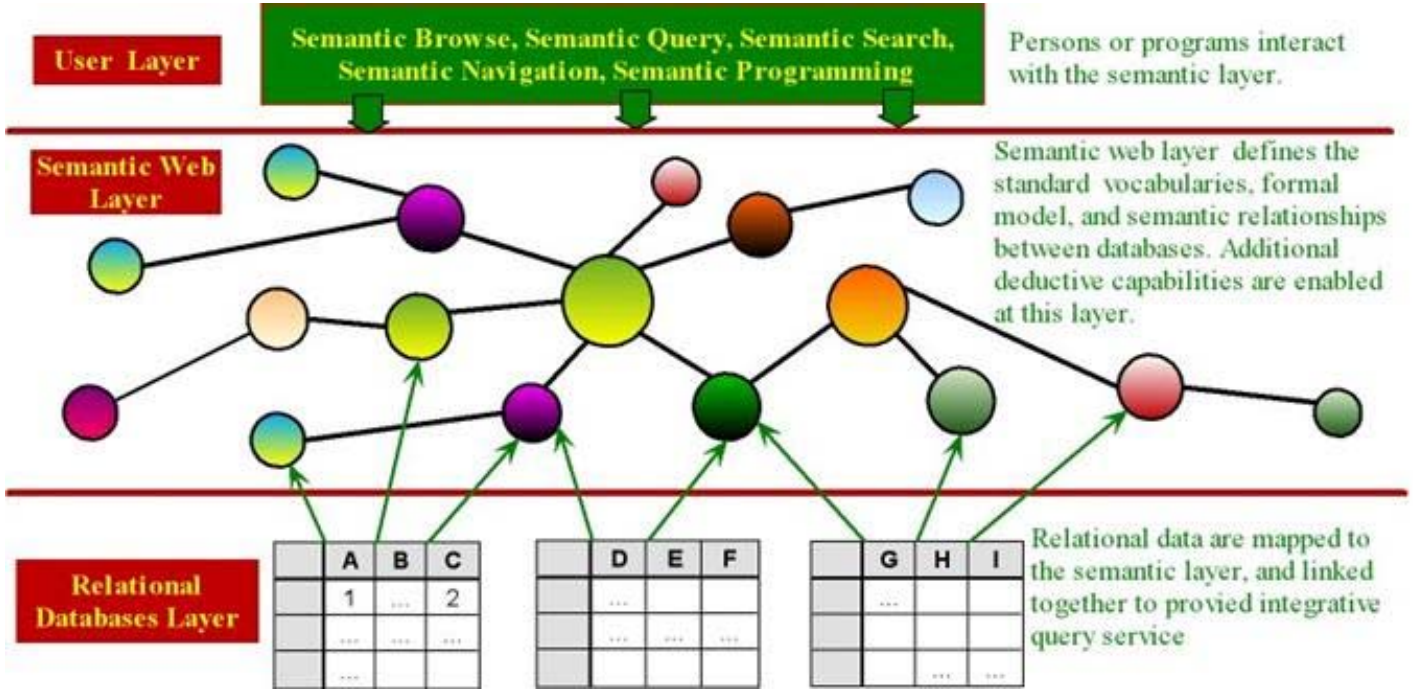
OntoGame

- Obviously, games for ontology construction, population, and alignment will not solve the knowledge acquisition bottleneck
- But they are a new way of combining human and computational intelligence and providing incentives
- Promising results
- Further incentives by increasing the human factor – e.g. „Meet your soulmate game“ ☺

Kurze Beispiele

Academy of Chinese Medicine Sciences

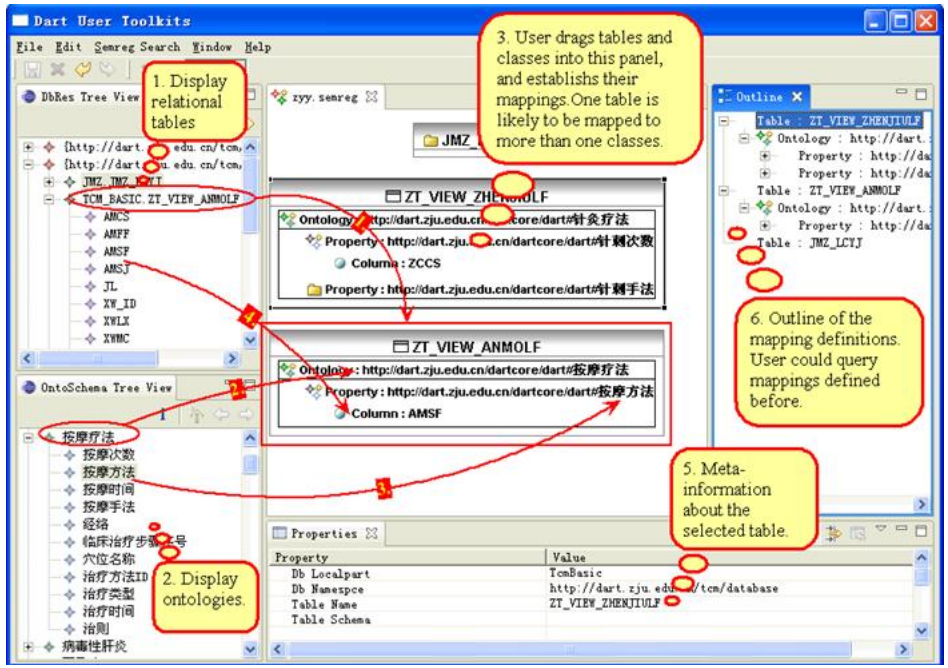
- Problem: Information über Traditionelle Chinesische Medizin (TCM) ist über **viele Institute und Datenbanken** in China verteilt. Die Datenstrukturen sind völlig unterschiedlich.
- Lösung: Integration der Datenbanken in einen verteilten Semantischen Grid, Suche und Dienste auf den integrierten Daten



Courtesy of Huajun Chen, Zhejiang University, CACSM (SWEO Case Study)

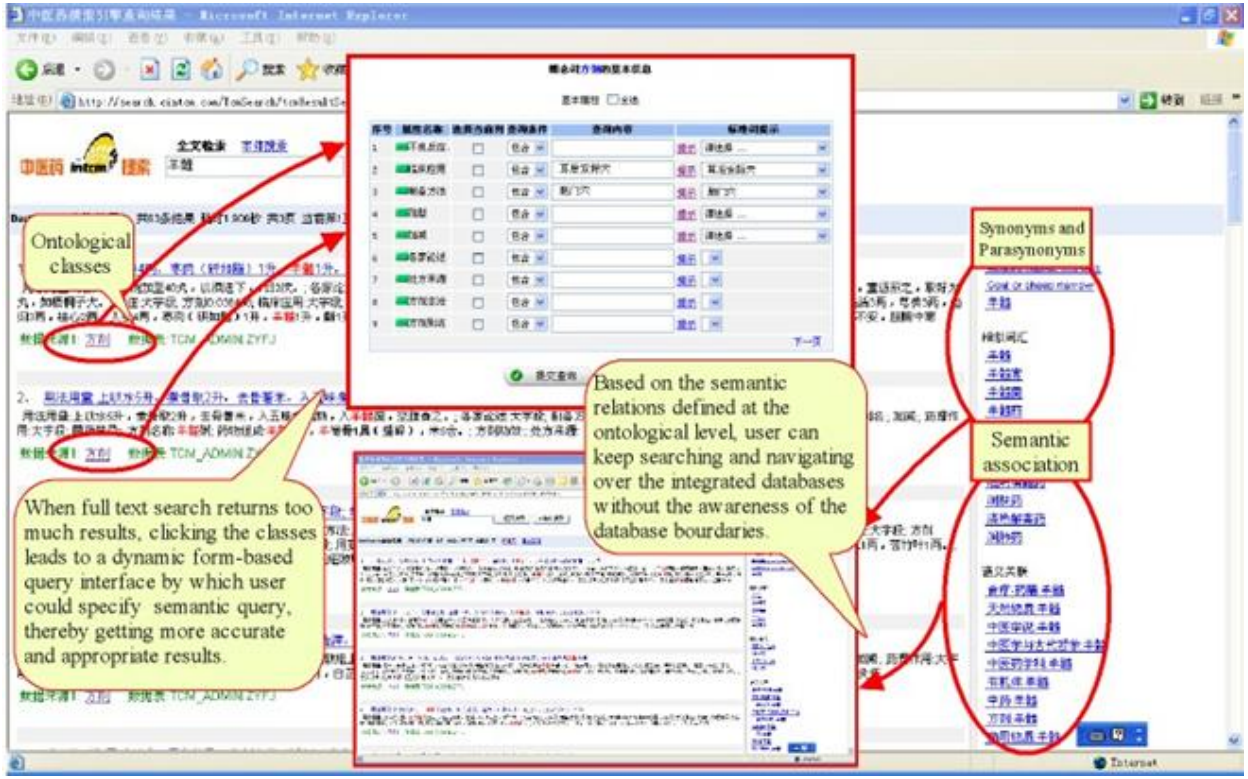
Academy of Chinese Medicine Sciences

- Datenintegration mit DartMapping
  - http://ccnt.zju.edu.cn/projects/dartmapping/
  - Tool zum Mapping der Relationalen Datenbanken in die Ontologie
- definiert Regeln, mit denen SPARQL queries in SQL umgewandelt werden
- Daten aus verschiedenen Datenbanken durch Ontologie vernetzt



Academy of Chinese Medicine Sciences

- http://search.cintcm.com/TcmSearch/tcmBasicSearch.luc
- online seit 2005
- Suchanfragen werden mit der Ontologie ausgedrückt



Vodafone live!

- Problem: Vodafone ist einer der größten Mobilfunkkonzerne der Welt. Verschiedene Anbieter liefern Content (Klingeltöne, Spiele, etc) die über die Platform *Vodafone live!* weltweit vertrieben wird. Die Daten verschiedenster Anbieter müssen integriert werden, sie sind komplex
- Lösung: Vodafone live! veröffentlichte ein RDF Vokabular, das die Content Provider einhalten. Die Metadaten werden in RDF geliefert und die Inhalte darüber in die Platform integriert
- Projekt:
  - Content ist primär Klingeltöne, Spiele, Wallpapers.
  - Metadaten über: kompatibles Endgerät, Content Ratings (Adult, Gaming, Violence, …), Gültigkeitsdauer des Angebots
  - Genaue Dokumentation, Validierung des XML über XML Schema
  - Einbindung der Provider wichtig

Courtesy of Kevin Smith, Vodafone Group R & D (SWEO Use Case)

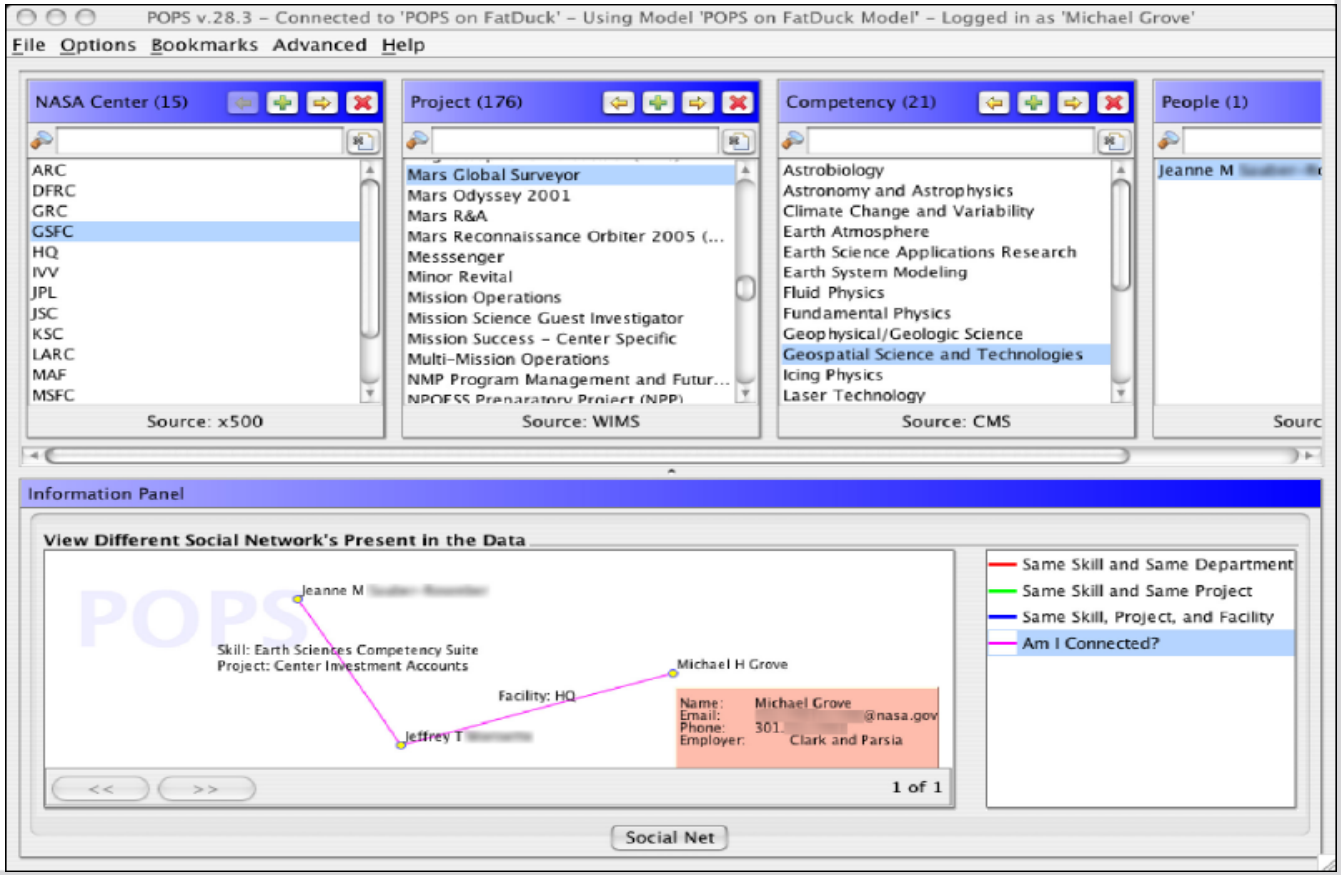
Vodafone live!

- Drei Vokabulare verwendet
  - Dublin Core für Metadaten
  - PRISM für Embargos und Gültigkeitsdaten
  - Internet Content Ratings Association (ICRA) Vokabular für Bewertung nach Gewalt, Erotik, oder Gambling
  - Wenige Erweiterungen
- Über das Portal kann nun der Benutzer sein Profil eingeben (Endgerät, Rating) und spezifische Inhalte finden
- Erhöhte den Umsatz signifikant, da durch die präzise Beschreibung die Inhalte besser gefunden wurden

Courtesy of Kevin Smith, Vodafone Group R & D (SWEO Use Case)

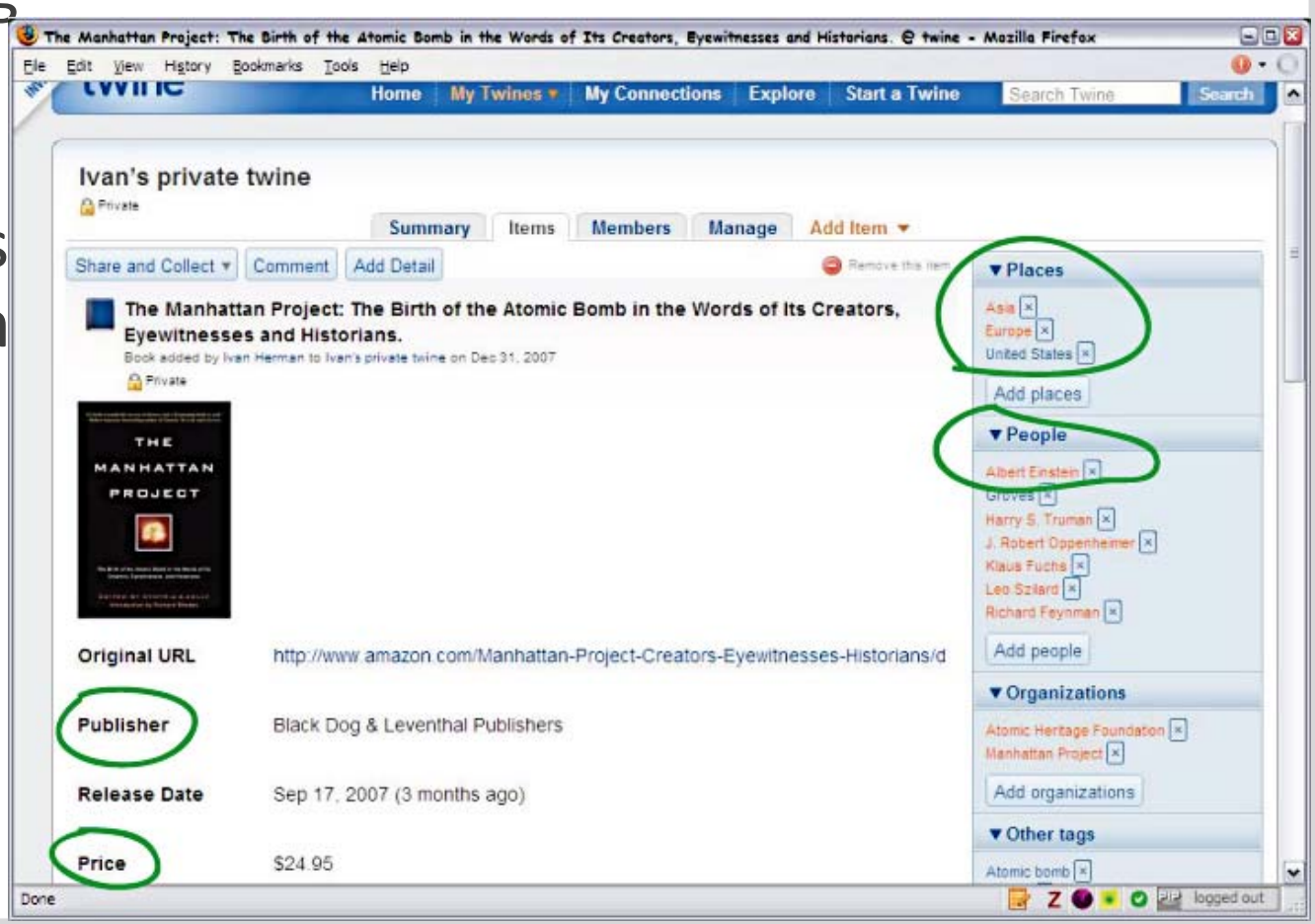
NASA Expert Finder

Expertise locater for nearly 70,000 NASA civil servants using RDF integration techniques over 6 or 7 geographically distributed databases, data sources, and web services…



Twine

- "Social Bookmarking on Steroids"
- Item relationships are based on ontologies
- Internals in RDF, will be available via APIs and SPARQL

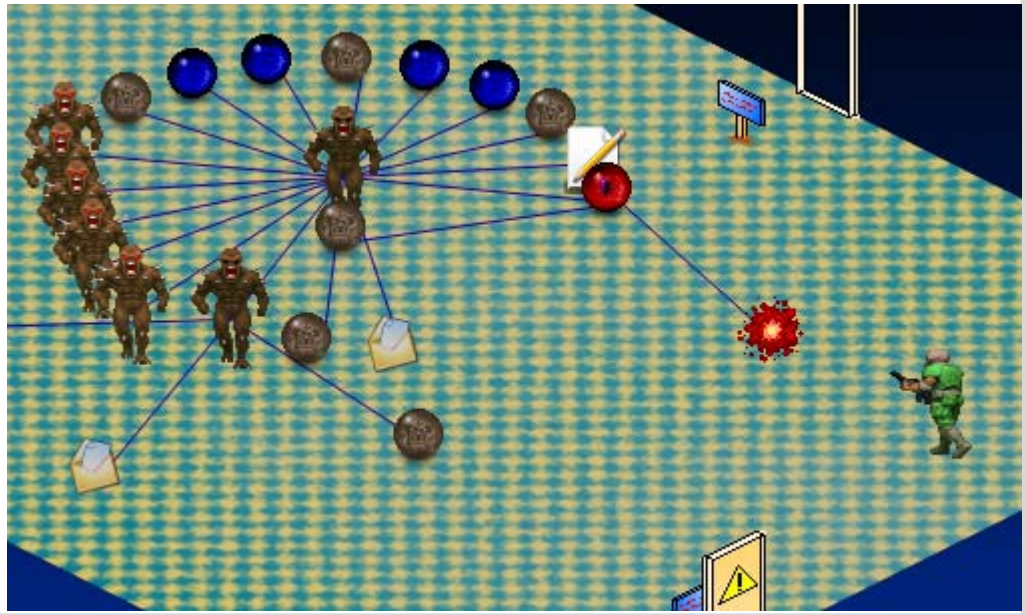


RDFRoom

AIFB

- www.dfki.uni-kl.de/~grimnes/2006/03/RDFRoom/
- RDFRoom is an isometric RDF viewer. It gives the user ways to view and manipulate his RDF data that might make him see the data in a brand new perspective.

"A lone soldier has been stranded in an alien world, filled with resources, literals and shifty anonymous nodes. Room upon room are filled with named graphs - can he find a way out?"



Dank

- Folien und Bilder von Katharina Siorpaes, Max Völkel, Anupriya Ankolekar, Leo Sauermann, Ivan Herman, Tim Berners-Lee und anderen

Current Semantic Web Layer Cake