Semantic Web
Methoden, Werkzeuge und Anwendungen

Rudi Studer & Peter Haase

Institut AIFB, Universität Karlsruhe (TH) &
FZI Forschungszentrum Informatik &
Ontoprise GmbH

Universität Kassel
26. Januar 2006

Karlsruhe: Location for Semantic Technologies and Applications

Knowledge Management
B2B, EAI
Business Intelligence
Electronic Markets
eGovernment

Semantic Web Infrastructure
Ontology Management
Data, Web & Text Mining
Peer-to-Peer, Semantic Grid
Semantic Web Services

Agenda

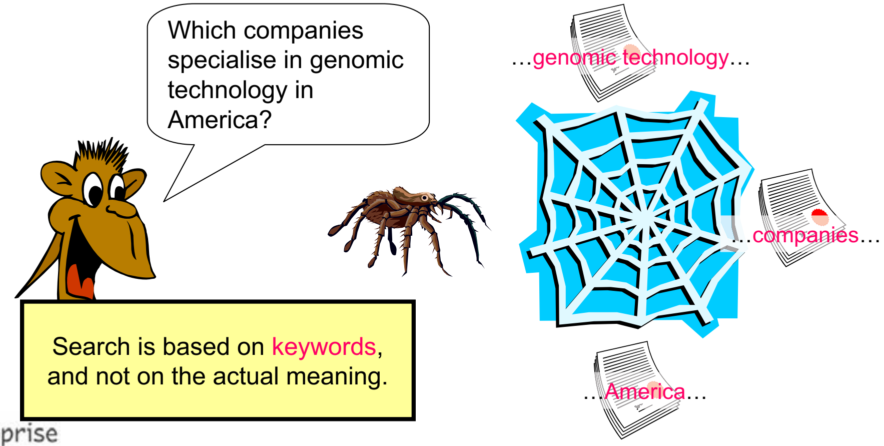
- Scenario: Semantic Search in a Digital Library

- Scalable Reasoning
- Mapping Ontologies
- Learning Ontologies
- Ontology Evolution

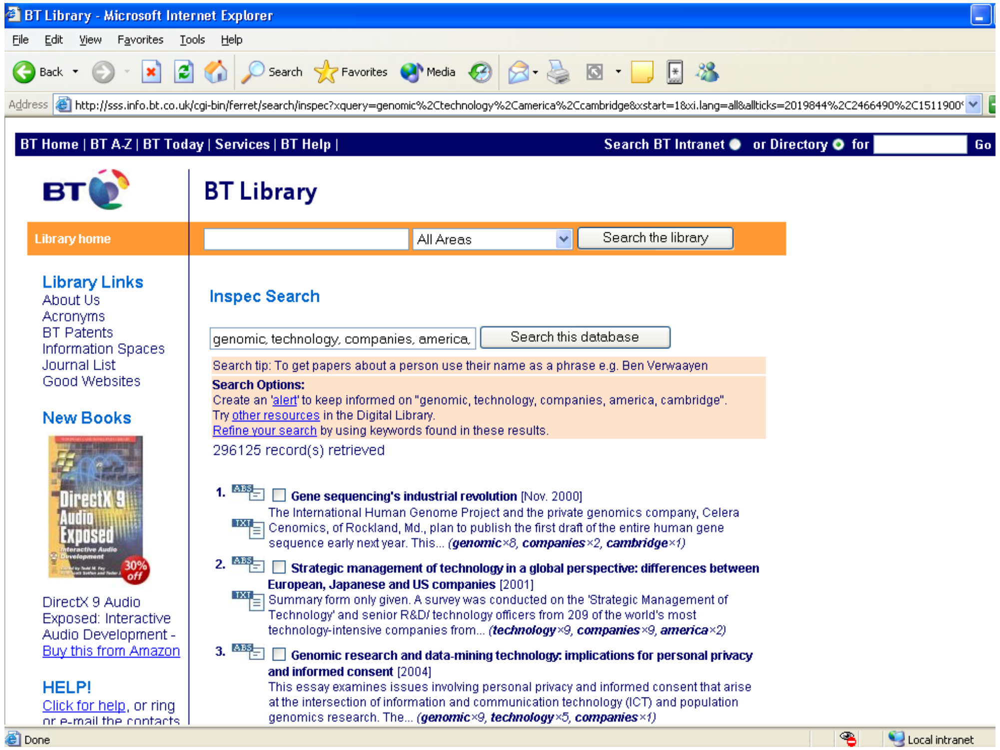
- Conclusion & Outlook

Problems with Search

What do the available information really mean?



Search is based on keywords, and not on the actual meaning.



BT Library

Library Links
About Us
Acronyms
BT Patents
Information Spaces
Journal List
Good Websites

New Books

DirectX 9 Audio Exposed: Interactive Audio Development - Buy this from Amazon

HELP!
Click for help, or ring or e-mail the contacts

Inspec Search

genomic, technology, companies, america,

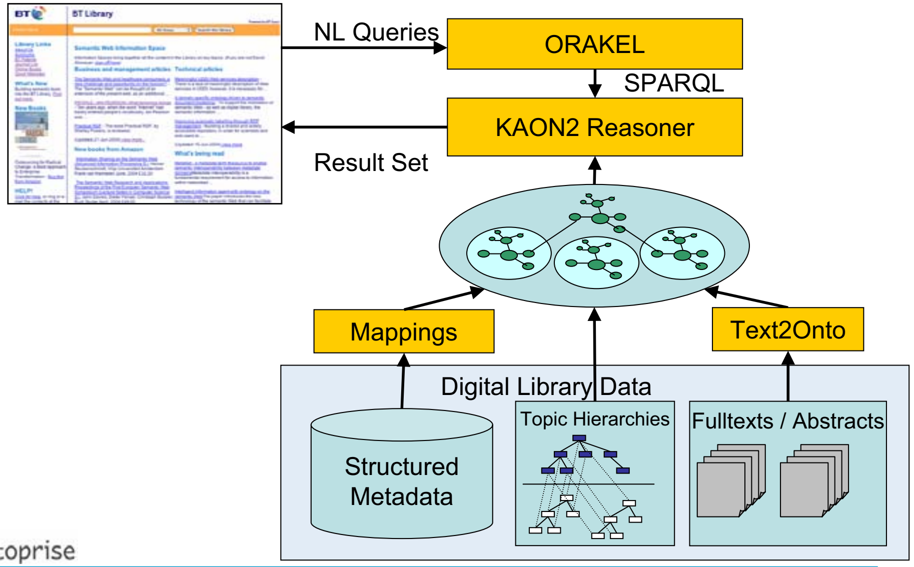
Search tip: To get papers about a person use their name as a phrase e.g. Ben Verwaayen

Search Options:
Create an 'alert' to keep informed on "genomic, technology, companies, america, cambridge".
Try other resources in the Digital Library.
Refine your search by using keywords found in these results.

296125 record(s) retrieved

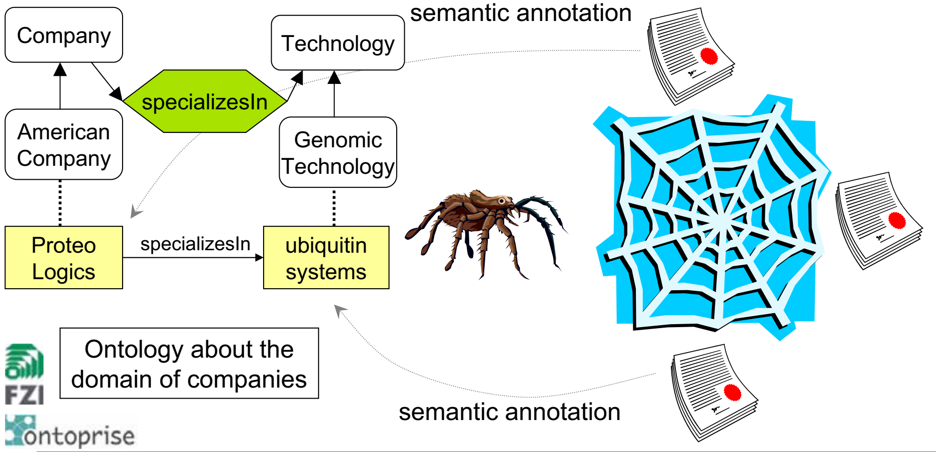
1. ☐ **Gene sequencing's industrial revolution** [Nov. 2000]
   The International Human Genome Project and the private genomics company, Celera Genomics, of Rockland, Md., plan to publish the first draft of the entire human gene sequence early next year. This... *(**genomic**×8, **companies**×2, **cambridge**×1)*
2. ☐ **Strategic management of technology in a global perspective: differences between European, Japanese and US companies** [2001]
   Summary form only given. A survey was conducted on the 'Strategic Management of Technology' and senior R&D/ technology officers from 209 of the world's most technology-intensive companies from... *(**technology**×9, **companies**×9, **america**×2)*
3. ☐ **Genomic research and data-mining technology: implications for personal privacy and informed consent** [2004]
   This essay examines issues involving personal privacy and informed consent that arise at the intersection of information and communication technology (ICT) and population genomics research. The... *(**genomic**×9, **technology**×5, **companies**×1)*

Conceptual Architecture



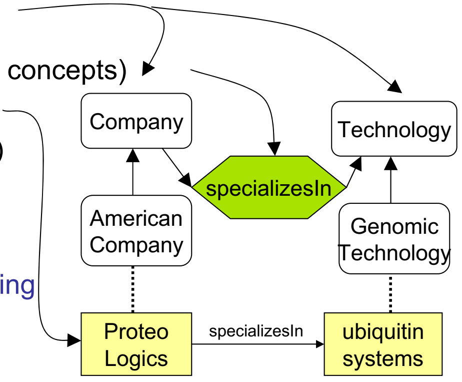
Semantic Annotation

Approach: Annotate information sources (documents) with semantic information



Web Ontology Language OWL

- Web Ontology Language (OWL) is a W3C Standard
  - three variants: OWL Lite, OWL DL, OWL Full
- Ontologies consist of:
  - concepts (=sets of objects)
  - roles (=connections between concepts)
  - individuals (=actual objects)
  - axioms (=truthful statements)
- Advantages of OWL:
  - precise semantics by grounding in description logics



Agenda

- Scenario: Semantic Search in a Digital Library

- Scalable Reasoning
- Mapping Ontologies
- Learning Ontologies
- Ontology Evolution

- Conclusion & Outlook

Practical Problem: Scalability

- Existing DL reasoners cannot answer queries over ontologies with many assertions

- Reasons:
  - reasoning in DL underlying OWL DL is NExpTime-complete
  - reasoning based on tableaux calculus
    - no specific query answering algorithms
    - difficult to identify facts relevant for the query

Comparison of Approaches

- General idea: use deductive database techniques for A-Box reasoning

Tableaux Calculi	Deductive Databases
• manage tuples **one-by-one** • to answer C(X), check whether C(a) holds for each a	• manage tuples **in sets** • very important!
• join optimizations are **difficult**	• join optimizations **supported** • core feature of relational databases
• difficult to be **goal-directed** • estimating relevant A-Box information is hard	• **magic sets** provide goal-directed search • selects only A-Box data relevant to the query

---

Deductive Database Techniques for Description Logics Reasoning

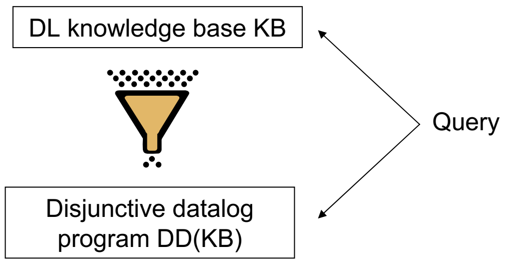
- Deductive databases can efficiently handle large data quantities
- Idea: apply techniques from the field of (disjunctive) deductive databases
  - join-order optimization
  - magic sets optimization

KB $\models \alpha$

if and only if

DD(KB) $\models \alpha$

for a ground fact α



DL knowledge base KB

Query

Disjunctive datalog program DD(KB)

OWL DL + Rules

- Extending OWL DL with rules is needed
- Query answering should be decidable (SWRL approach is undecidable)

- Chosen approach:
  - DL-safe rules:
  - restrict application of rules to individuals explicitly introduced in the ABox to achieve decidability
  - do not restrict component languages
  - …can be simply appended to the result of the reduction of description logics to disjunctive datalog

---

OWL Reasoner: KAON2

- Features
  - an API for programmatic management of OWL-DL ontologies,
  - a stand-alone server providing access to ontologies in a distributed manner,
  - an inference engine for answering queries (including support for SPARQL),
  - efficient access to instances via relational databases (available soon)

- Download (free for research purposes)
  - http://kaon2.semanticweb.org/

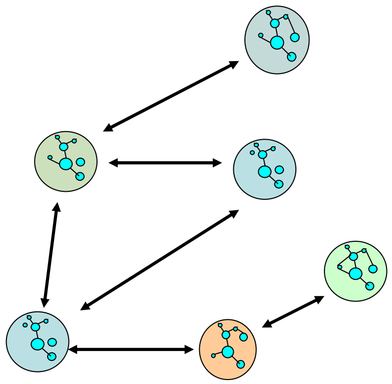
Agenda

- Scenario: Semantic Search in a Digital Library

- Scalable Reasoning
- Mapping Ontologies
- Learning Ontologies
- Ontology Evolution

- Conclusion & Outlook

Ontology Mappings

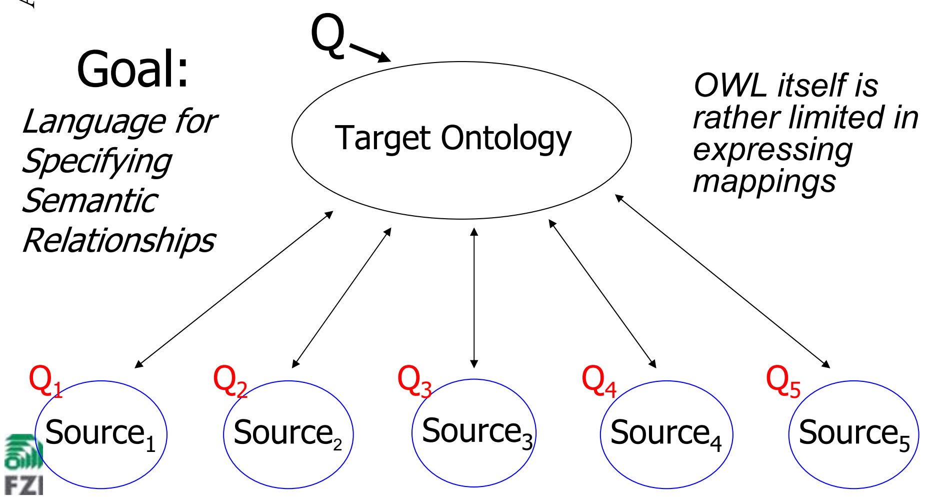


- Heterogeneous ontologies require mappings for interoperability

- Applications of mapping system:
  - Ontology Integration
  - Ontology Translation and Exchange

- Challenges:
  - Representation of and reasoning with mappings
  - Identification of mappings (alignment of ontologies)

Mapping Systems for Ontology Integration

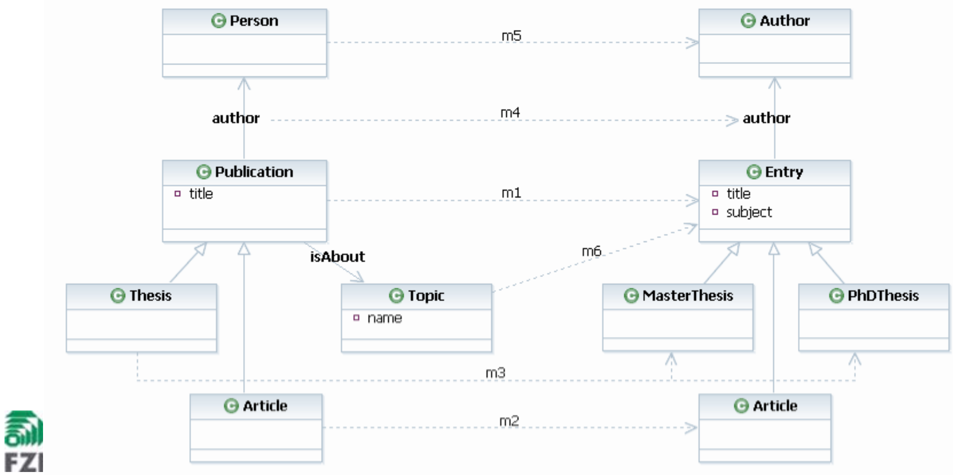


Goal:

Language for Specifying Semantic Relationships

OWL itself is rather limited in expressing mappings

Sample Mapping



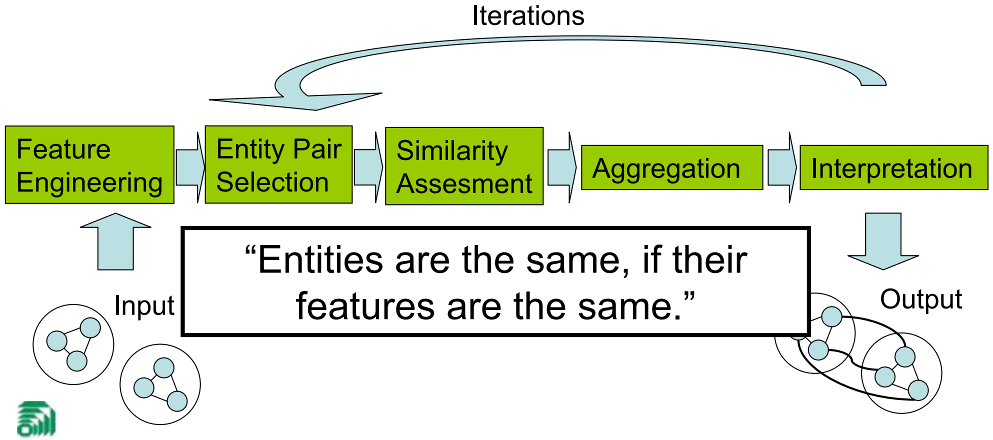
OWL DL Mapping System

- An OWL DL mapping system is a triple ($\mathcal{S}$, $\mathcal{T}$, $\mathcal{M}$), where
  - $\mathcal{S}$ is the source OWL DL ontology
  - $\mathcal{T}$ is the target OWL DL ontology
  - $\mathcal{M}$ is the mapping between $\mathcal{S}$ and $\mathcal{T}$
- Mapping: set of assertions
  - $q_S \sqsubseteq q_T$ (sound mapping)
  - $q_S \sqsupseteq q_T$ (complete mapping)
  - $q_S \equiv q_T$ (exact mapping)
  - where q_S and q_T are conjunctive queries over $\mathcal{S}$ and $\mathcal{T}$, respectively, with the same set of distinguished variables
- Semantics defined via translation into FOL, computing answers against $\mathcal{S} \cup \mathcal{T} \cup \mathcal{M}$

Decidability of Query Answering

- A mapping $q_S \sqsubseteq q_T$ is equivalent to an axiom $\forall \mathbf{x} : q_T(\mathbf{x},\mathbf{y}_T) \leftarrow q_S(\mathbf{x},\mathbf{y}_S)$
- Query answering undecidable with general implication mappings
- Decidable query answering:
  - Disallow non-distinguished variables in q_T to obtain safe rules:
    - $\forall \mathbf{x} : q_T(\mathbf{x}) \leftarrow q_S(\mathbf{x},\mathbf{y}_S)$
    - These rules directly correspond to SWRL rules
  - Require q_S to be DL-safe:
    - Each variable in a DL-atom must also occur in a non-DL atom (makes queries applicable only to explicitly introduced individuals)

Ontology Alignment Process



"Entities are the same, if their features are the same."

Similarities

	Feature	Similarity Measure
Concepts	label	String Similarity
	subclassOf	Set Similarity
	instances	Set Similarity
	…	
Relations		
Instances		

Machine learning can help to select and weight the features and measures.

From similarities to alignments:

$$sim(e,f)=\sum_k w_k sim_k(e,f)$$

$$align(e) = f \leftarrow sim(e,f)>t$$

FOAM

http://www.aifb.uni-karlsruhe.de/WBS/meh/foam

Framework for Ontology Alignment and Mapping

Fully or semi-automatic alignment of two or more ontologies

Agenda

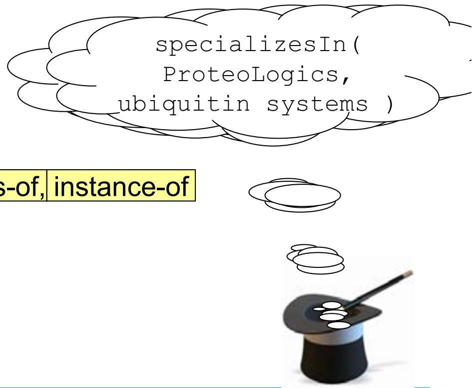
- Scenario: Semantic Search in a Digital Library

- Scalable Reasoning
- Mapping Ontologies
- Learning Ontologies
- Ontology Evolution

- Conclusion & Outlook

Ontology Learning

- **Extraction** of (domain) ontologies from natural language text
  - Natural Language Processing
  - Machine Learning
- Ontology Learning **tasks**
  - Concepts, instances
  - Taxonomic relations: subclass-of, instance-of
  - Relations
  - Relation instantiations
- Ontology Population



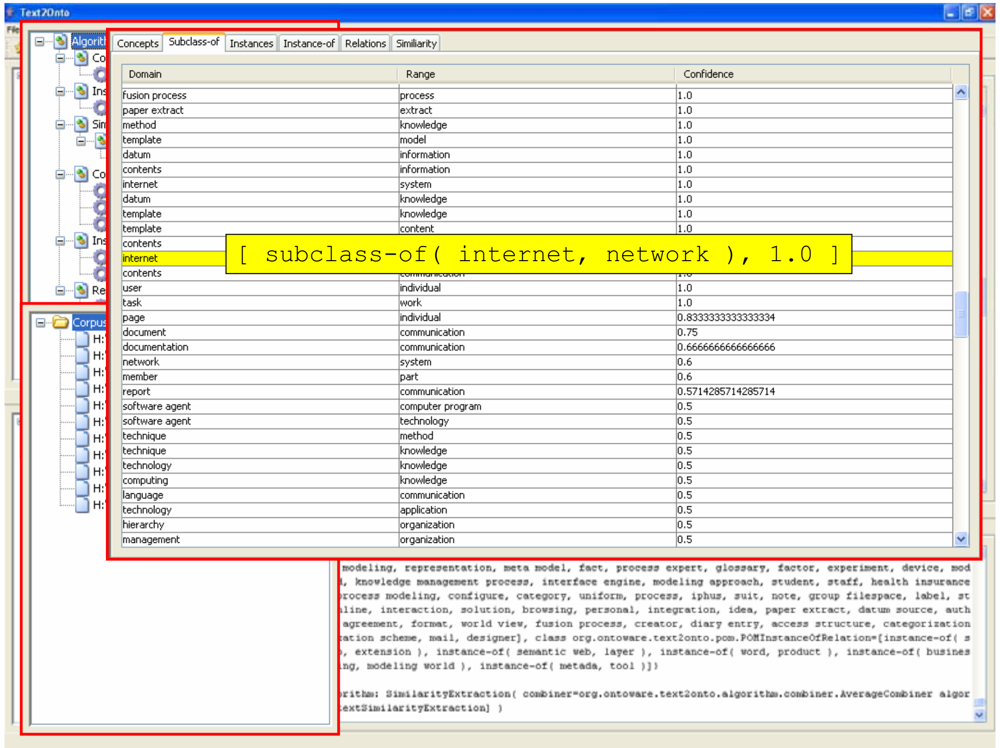
Ontology Learning - Challenges

- **Traceability**
  - Explanations, references
- Independence of a specific **ontology model**
  - User-defined consistency conditions
- Knowledge is **dynamic**
  - Support for ontology **maintenance**
  - Efficient updates of the ontology in case of changes to the corpus
- **Uncertainty** in knowledge acquisition
  - Ontology model supporting notions of confidence and relevance

---

Ontology Learning Tool: Text2Onto

- Support for semi-automatic **ontology extraction** from natural language text
- Support for **ontology maintenance** and data-driven **ontology evolution** by incremental ontology learning
- Model of Possible Ontologies (POM) based on confidence and relevance annotations
- Available at http://ontoware.org/projects/text2onto/



Domain	Range	Confidence
fusion process	process	1.0
paper extract	extract	1.0
method	knowledge	1.0
template	model	1.0
datum	information	1.0
contents	information	1.0
internet	system	1.0
datum	knowledge	1.0
template	knowledge	1.0
template	content	1.0
contents		
internet		
contents	communication	1.0
user	individual	1.0
task	work	1.0
page	individual	0.8333333333333334
document	communication	0.75
documentation	communication	0.6666666666666666
network	system	0.6
member	part	0.6
report	communication	0.5714285714285714
software agent	computer program	0.5
software agent	technology	0.5
technique	method	0.5
technique	knowledge	0.5
technology	knowledge	0.5
computing	knowledge	0.5
language	communication	0.5
technology	application	0.5
hierarchy	organization	0.5
management	organization	0.5

[subclass-of(internet, network), 1.0]

```
modeling, representation, meta model, fact, process expert, glossary, factor, experiment, device, mod
d, knowledge management process, interface engine, modeling approach, student, staff, health insurance
process modeling, configure, category, uniform, process, iphus, suit, note, group filespace, label, st
line, interaction, solution, browsing, personal, integration, idea, paper extract, datum source, auth
 agreement, format, world view, fusion process, creator, diary entry, access structure, categorization
ation scheme, mail, designer], class org.ontoware.text2onto.pom.POMInstanceOfRelation=[instance-of( s
), extension ), instance-of( semantic web, layer ), instance-of( word, product ), instance-of( busines
ing, modeling world ), instance-of( metada, tool )])

rithm: SimilarityExtraction( combiner=org.ontoware.text2onto.algorithm.combiner.AverageCombiner algor
extSimilarityExtraction] )
```

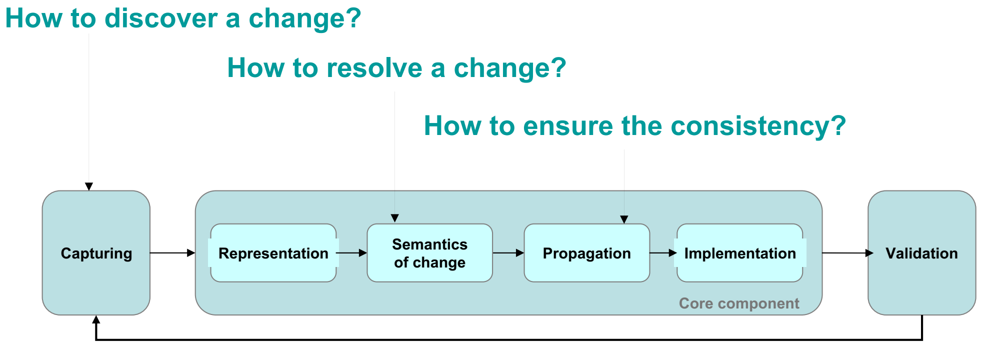
Agenda

- Scenario: Semantic Search in a Digital Library
- Scalable Reasoning
- Mapping Ontologies
- Learning Ontologies
- Ontology Evolution
- Conclusion & Outlook

Ontology Evolution - requirements

- **Functionality**
  - enable the handling of ontology **changes**
  - ensure the **consistency** of the underlying ontology and all dependent artefacts, e.g. instances
- **Guiding the user**
  - support the user to manage changes **more easily**
- **Refining the ontology**
  - offer advice to the user for **continual** ontology refinement
  - discover changes that lead to an **improved** ontology

Ontology Evolution - Process



How to discover a change?

How to resolve a change?

How to ensure the consistency?

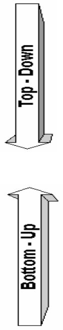
Capturing: Change Discovery

- Explicit request by the user
- Implicit request through learning

  - Structure-driven – exploits a set of heuristics to improve an ontology based on the analysis of the ontology structure

  - Data-driven - detects the changes based on the analysis of the ontology instances

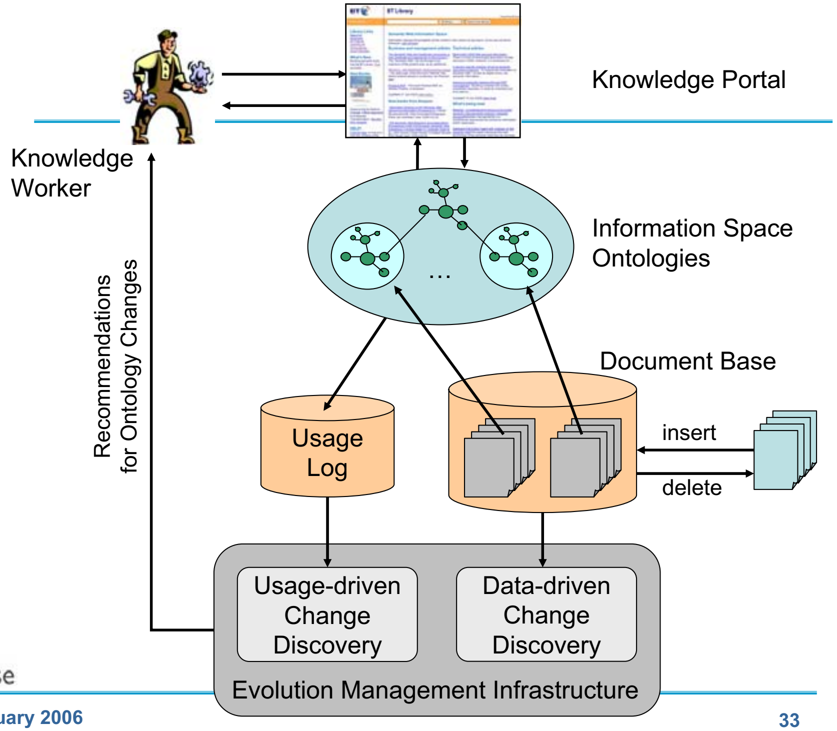
  - Usage-driven – takes into account the usage of the ontology

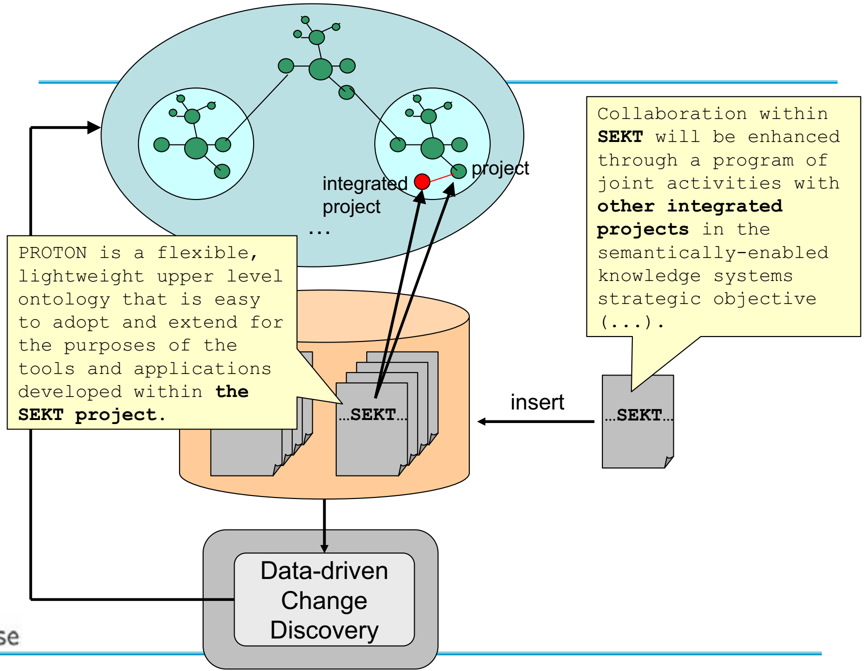


If no instance of a concept C use any of the properties defined for C, but only properties inherited from the parent concept, we can make an assumption that C is not necessary.

If all subconcepts have the same property, the property may be moved to the parent concept.

By tracking when an entity has last been retrieved by a query, it may be possible to discover that some entities are out of date.





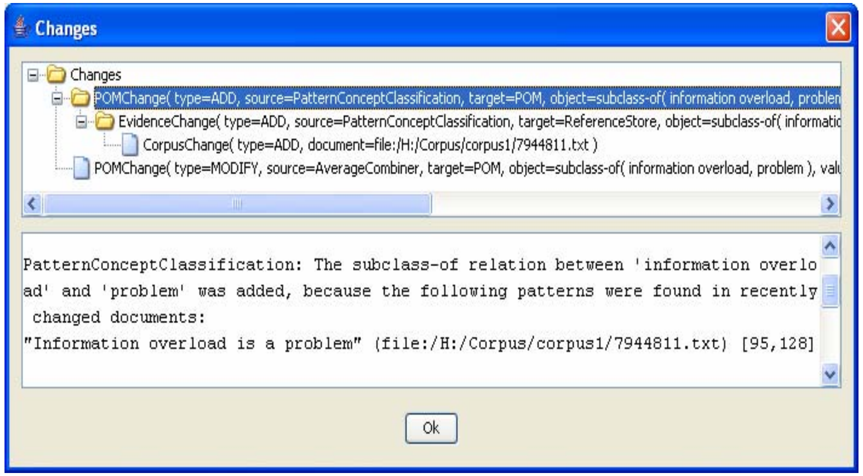
PROTON is a flexible, lightweight upper level ontology that is easy to adopt and extend for the purposes of the tools and applications developed within **the SEKT project.**

Collaboration within **SEKT** will be enhanced through a program of joint activities with **other integrated projects** in the semantically-enabled knowledge systems strategic objective (...).

Change Discovery in Text2Onto

- Data-driven Change Discovery
    - Deduction of ontology changes from changes to the data
- Incremental Ontology Learning
    - Update **evidence** for ontology elements based on observed **corpus changes**
    - Generate suggestions (and **explanations**) for **ontology changes** based on new evidence
- Ontology Change Strategies
    - How are different types of **ontology elements** affected by particular changes to the corpus?

Data-driven Change Discovery



```
Changes
  POMChange( type=ADD, source=PatternConceptClassification, target=POM, object=subclass-of( information overload, proble
    EvidenceChange( type=ADD, source=PatternConceptClassification, target=ReferenceStore, object=subclass-of( informati
      CorpusChange( type=ADD, document=file:/H:/Corpus/corpus1/7944811.txt )
  POMChange( type=MODIFY, source=AverageCombiner, target=POM, object=subclass-of( information overload, problem ), val
```

```
PatternConceptClassification: The subclass-of relation between 'information overlo
ad' and 'problem' was added, because the following patterns were found in recently
 changed documents:
"Information overload is a problem" (file:/H:/Corpus/corpus1/7944811.txt) [95,128]
```

Semantics of Change

- Consistency conditions
  - An ontology is consistent if it satisfies a given set if consistency conditions
  - Structural Consistency
    with respect to syntactic fragments
  - Logical Consistency
    (model-theoretic satisfiability)
  - User-defined Consistency
    outside of ontology model

SEKT Facts

- “Semantically Enabled Knowledge Technologies”

- EU IST Integrated Project (IP)
  - Start: Janurary 2004
  - Duration: 3 years
  - Budget: ~13 MEUR
  - Funding: ~10 MEUR
  - see http://www.sekt-project.com

- Part of ESSI Cluster
  - European Semantic Systems Initiative
  - SEKT, DIP (IP), Knowledge Web (NoE), ASG
  - see http://www.essi-cluster.org

Conclusion

- **Scalable reasoning**
  - Promising results for DL reasoning based on disjunctive deductive database techniques
  - Rule extensions to close paradigm gap

- **Ontology Mappings**
  - Methods for representation and identification of mappings
  - Query answering against heterogeneous ontologies

- **Ontology Learning**
  - POM to capture confidence and relevance in knowledge acquisition
  - Traceability: explanations, references

- **Ontology Evolution**
  - Support for ontology maintenance by data-driven change discovery
  - Semantics of change to ensure consistency

Future Work

- Networked ontology models
  - Including mappings, dependencies, modularization, …
  - Dynamics and change propagation
  - Global vs. local / partial consistency

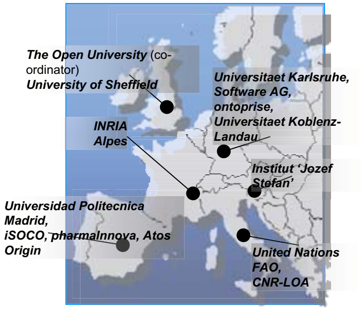
- Collaborative aspects
  - Distributed engineering
  - Argumentation and negotiation

- Context sensitivity
  - Representation of context
  - Combination of logic-based and probabilistic models
  - Reasoning with contexts

Lifecycle Support for Networked Ontologies

“Shaping the future infrastructures for semantic applications”

- EU IST Integrated Project
  - Start date: March 2006
  - Duration: 4 year project
  - Funding: € 10M (FP6)
  - http://www.neon-project.org/



- **Key outcomes from NeOn**
  - Open, scalable and service-centred **reference architecture**
  - The **NeOn toolkit** – a resource for engineering contextualized networked ontologies and semantic applications
  - Industry-strength documentation and reference material
  - Three **case studies** in two sectors: **pharmaceuticals** and **agriculture/fisheries**

References

- Ullrich Hustadt, Boris Motik, Ulrike Sattler: **Reducing SHIQ-Description Logic to Disjunctive Datalog Programs**. International Conference on Principles of Knowledge Representation and Reasoning, KR 2004
- Boris Motik, Ulrike Sattler, Rudi Studer: **Query Answering for OWL-DL with Rules**. International Semantic Web Conference 2004
- Marc Ehrig, York Sure: **Ontology Mapping - An Integrated Approach**. European Semantic Web Symposium, ESWS 2004
- Philipp Cimiano, Johanna Völker: **Text2Onto**. International Conference on Applications of Natural Language to Information Systems, NLDB 2005
- Peter Haase, Ljiljana Stojanovic: **Consistent Evolution of OWL Ontologies**. European Semantic Web Conference 2005
- Haase et al.: **A Framework for Handling Inconsistency in Changing Ontologies**, International Semantic Web Conference 2005
- York Sure, Rudi Studer: **Semantic Web Technologies for Digital Libraries,** *Library Management* 26 (4/5). April 2005.

Thank You!

For further information and relevant publications see
http://www.aifb.uni-karlsruhe.de/WBS

AIFB Portal enriched with OWL annotations, see
http://www.aifb.uni-karlsruhe.de/about.html