WP2 – Triple Space knowledge representation

Lyndon J B Nixon

TripCom WP2
Innsbruck Meeting, 8 January 2008

Meeting agenda

- 1. WP2 Status
- 2. Task 2.6 - Triple Space distribution
  - Distribution in Triple Space (20 min, Berlin)
  - Metadata in Triple Space (20 min, Innsbruck)
- 3. Task 2.7 – OWL, WSMO and rules representation
  - Progress made to date
  - Plan to M24
- General discussion

WP2 Status: T2.6

- Aim of current task (to end March 08): implement the distribution („semantic clustering“ in contract) of Triple Space (data placement and recovery)
- We have proposals for the use of Distributed Hash Tables and RDF metadata in Triple Space distribution, being prototypically implemented
- Physical distribution not our concern (WAN built as Web and kernels address one another by IP, find one another by DNS). DHT use P2P overlay.

Distribution Manager

Implementation & Research Works

Lyndon J B Nixon

TripCom WP2
Innsbruck Meeting, 8 January 2008

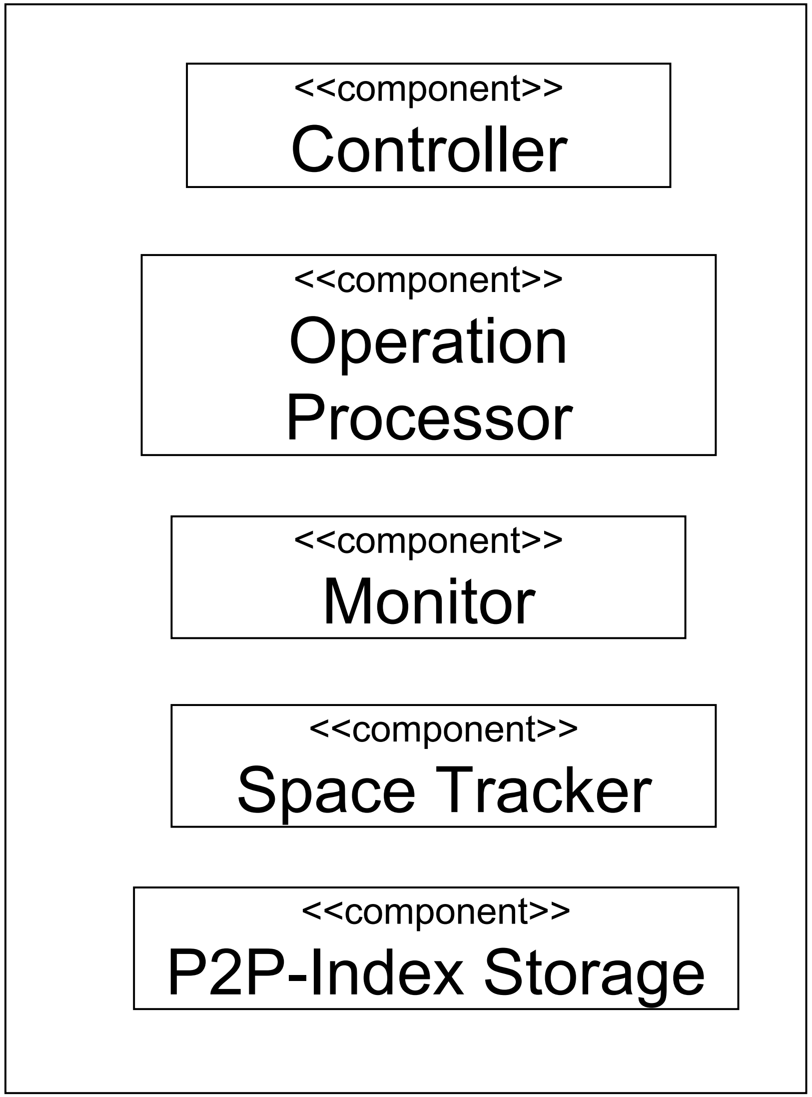
Outlines

- Implementation of Distribution Manager
- Research Works

Functionalities of Distribution Manager

1. Storage of indexes of Triples by Linda Out Operation (Optional indexing)
2. Forwarding of Linda rd operations which should be processed on other kernels and sending the answers to User client
3. Mapping a Linda rd operation without space URL to one or more Linda rd with space URL. Processing them and sending the answers to user client (Using the distributed index storage)

Components of Distribution Manager



The first α Prototype

- Functionality:
  - Distributed Index Storage System based on P-Grid
  - Storage and retrieval of indexes

    (s, p, o, space URL) with single template
  - Communication with mock objects as metadata manager, security manager
- Next Work:
  - Integration to the system bus (Javaspaces)
- Further Works:
  - Communication with the Metadata Manager
  - Communication with the Security Manager
  - And later Communication with the Query Processor

Implementation Challenges

- **Storage Layer** of P-Grid Implementation is not suitable for our distributed index storage. We need to communicate directly with the underlying database, and not to use the adapter classes of P-Grid implementation.
- Difficulties by keeping **tracks of started threads**, which are started for each Linda operations (rd or out operations).

Research Works

- Distributed Triple Space based on P-Grid Peer-to-Peer Overlay
- Efficient query evaluation and query answering for Distributed Triple Space (conjunctive and disjunctive queries)
- Usage of Metadata in combination with the DHT approach. Usage of semantic similarities and semantic hash functions.

Thank You!

Metadata - UIBK
Semantic Routing and Clustering

Reto Krummenacher (UIBK)
Lyndon J B Nixon

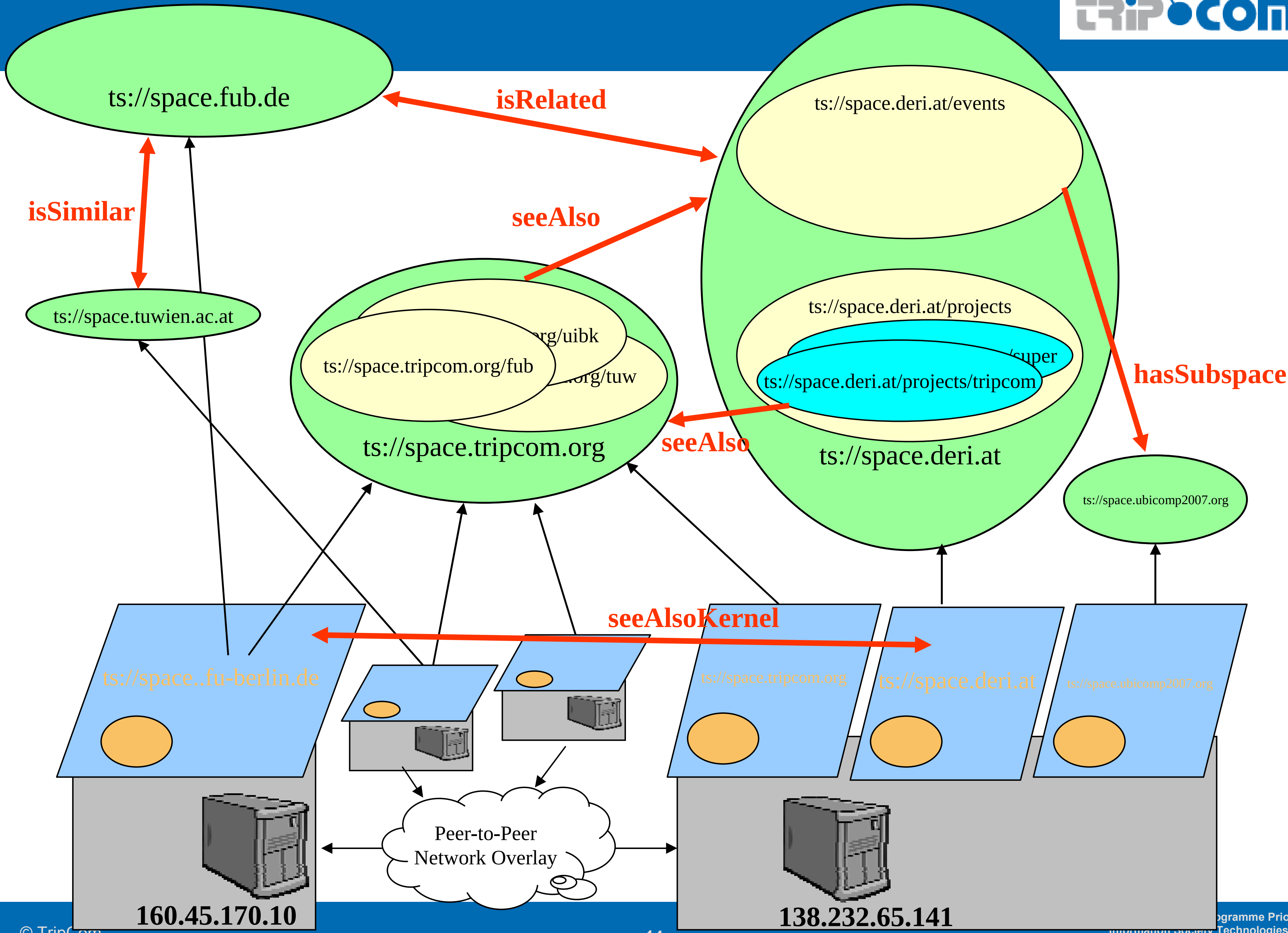
January 8, 2008

TripCom WP2

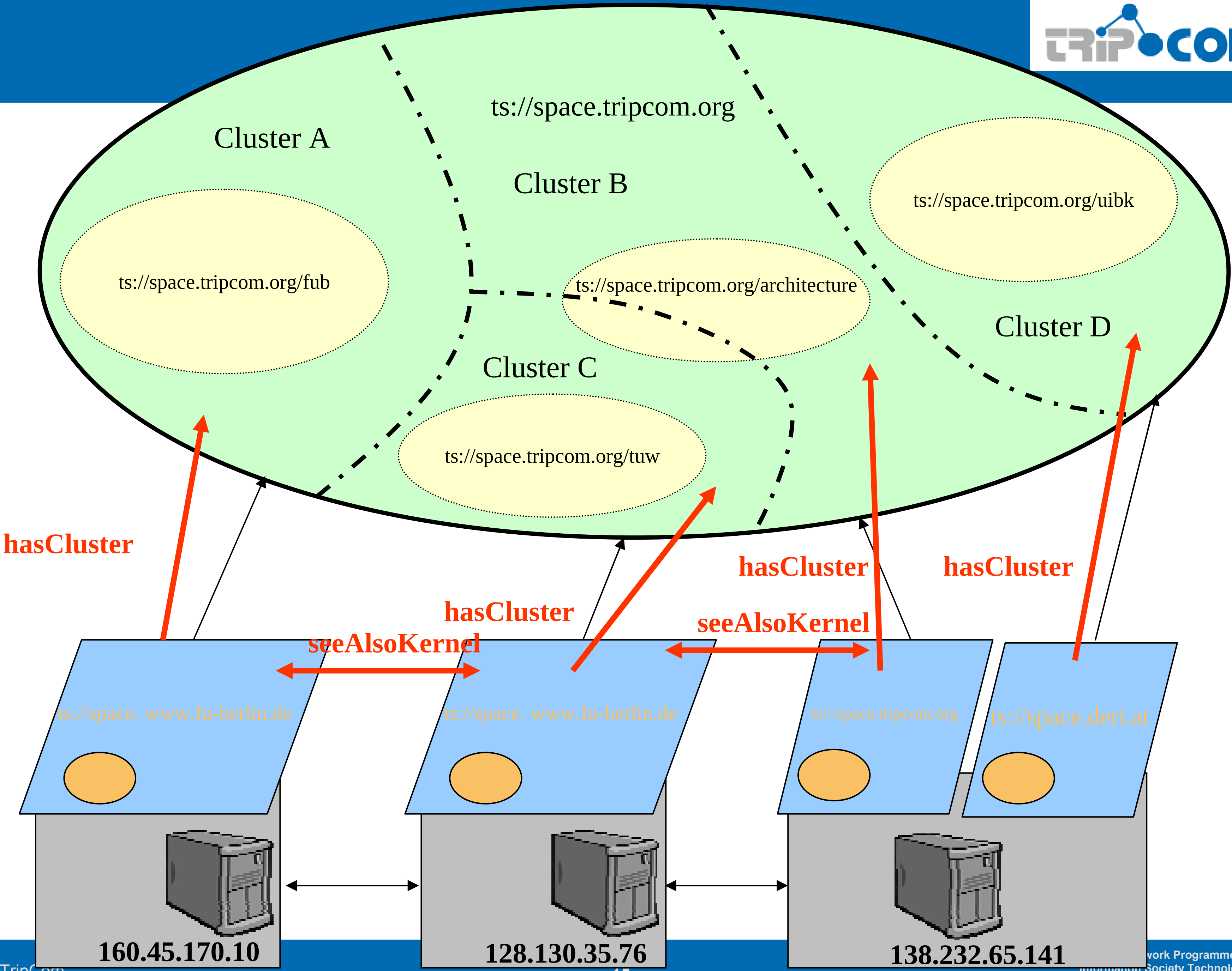
Innsbruck Meeting, 8 January 2008

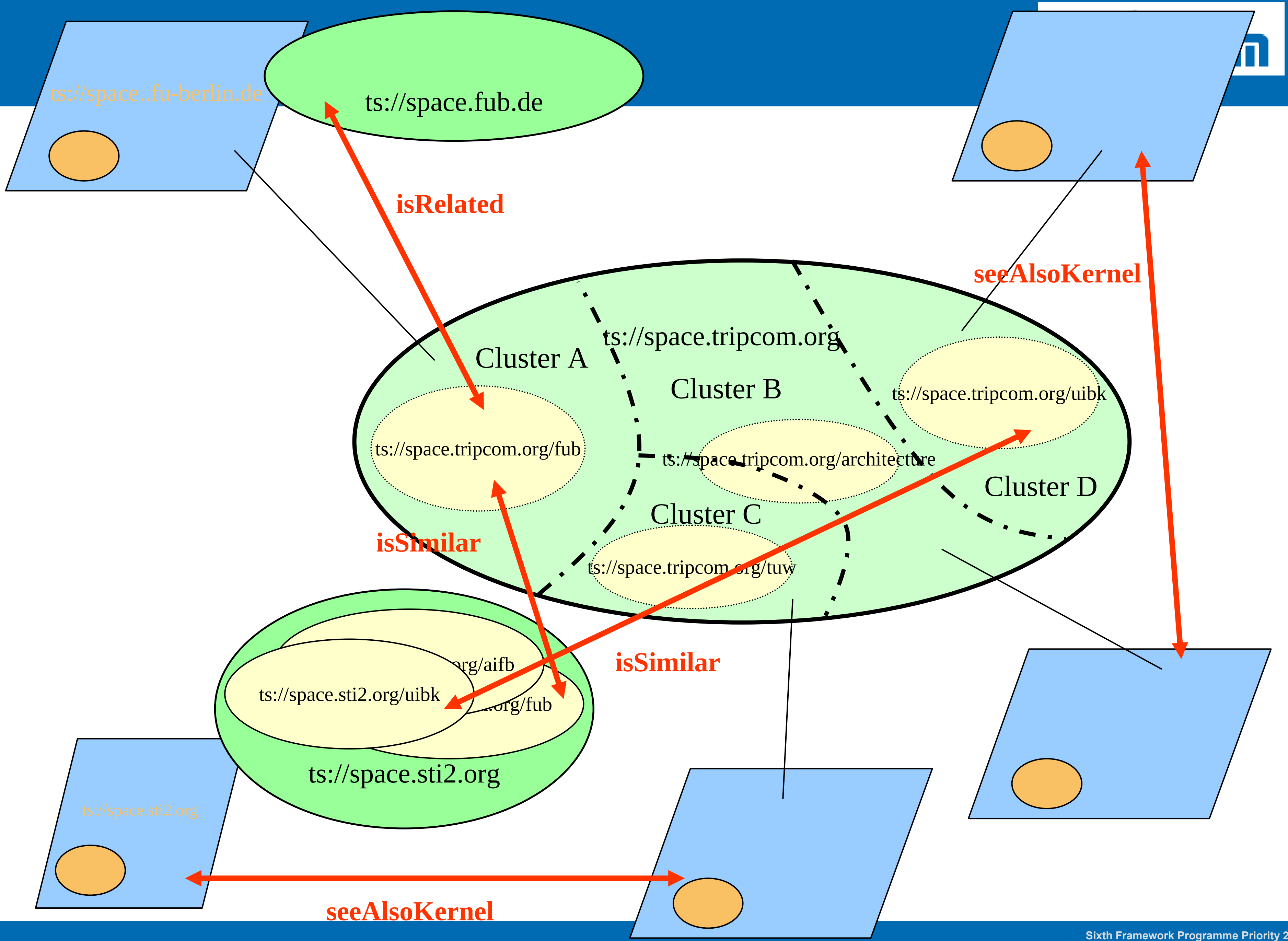
Semantic Routing

- Semantic Routing / Discovery by Metadata
  - Description of sources, not of data
    - Spaces are sources (logical entities)
    - Kernels are sources (physical entities)
  - Formal descriptions of relationships and links
    - Reasoning about search routes
    - More complex routing tables than e.g. in IP network
    - Transitive, symmetric, even reflexive links
    - Conditions on links and routes



ts://space.fub.de
isRelated
isSimilar
seeAlso
ts://space.tuwien.ac.at
rg/uibk
ts://space.tripcom.org/fub
org/tuw
ts://space.tripcom.org
ts://space.deri.at/events
ts://space.deri.at/projects
/super
ts://space.deri.at/projects/tripcom
seeAlso
ts://space.deri.at
hasSubspace
ts://space.ubicomp2007.org
seeAlsoKernel
ts://space..fu-berlin.de
ts://space.tripcom.org
ts://space.deri.at
ts://space.ubicomp2007.org
Peer-to-Peer
Network Overlay
160.45.170.10
138.232.65.141


ts://space.tripcom.org
Cluster A
Cluster B
Cluster C
Cluster D
ts://space.tripcom.org/fub
ts://space.tripcom.org/architecture
ts://space.tripcom.org/uibk
ts://space.tripcom.org/tuw
hasCluster
hasCluster
hasCluster
hasCluster
seeAlsoKernel
seeAlsoKernel
ts://space. www.fu-berlin.de
ts://space. www.fu-berlin.de
ts://space.tripcom.org
ts://space.deri.at
160.45.170.10
128.130.35.76
138.232.65.141


ts://space..fu-berlin.de
ts://space.fub.de
isRelated
seeAlsoKernel
ts://space.tripcom.org
Cluster A
Cluster B
Cluster C
Cluster D
ts://space.tripcom.org/uibk
ts://space.tripcom.org/fub
ts://space.tripcom.org/architecture
ts://space.tripcom.org/tuw
isSimilar
isSimilar
.org/aifb
ts://space.sti2.org/uibk
.org/fub
ts://space.sti2.org
ts://space.sti2.org
seeAlsoKernel

Description of sources

- Spaces are described by:
  - The relationship to other spaces
  - *Structural metadata*
- Clusters are described by:
  - Descriptions (i.e. metadata) about the published information.
    - What descriptions are necessary?
    - What type of metadata is necessary?
  - *Content metadata*

Description of sources

- Spaces are described by:
  - *Content metadata* too?!?
- ...or rather, the routes could be described by content metadata:
  - A cluster description is the minimal information necessary to understand the scope (nucleus, dimension) of a cluster. Similarly, a content-driven route would deliver the minimal description of the space a link points too!
  - ***The research on semantic clustering improves the results of semantic routing!***

How to discover a source?!

- Semantic routing:
  1. „pseudo-random“ walking from one space to another along the semantic routes
     - Where to find the „one space“?
     - Which routes to choose?
  2. Discovery of the first „one space“
     - Applying semantic cluster-like descriptions of spaces, i.e. the links to the spaces.
     - Semantic DHP...
  3. Chossing the right routes
     - Random along the known links
     - Matching the target of the query
       - Choose the most selective triple pattern (Bernstein et al., 2007)

Recap: Selectivity Estimation

Abraham Bernstein, Markus Stocker, and Christoph Kiefer:

SPARQL Query Optimization Using Selectivity Estimation.

6th ISWC, November 2007

OptARQ's Selectivity Estimation Approach

Selectivity of a triple pattern T (i.e., [ub:Student17 rdfs:label "Markus"])

$$sel(T) = sel(S) \times sel(P) \times sel(O)$$

Selectivity of the subject, predicate, and object

$$sel(S) = \frac{1}{|R|}, \quad sel(P) = \frac{|T_P|}{|T|}, \quad sel(O) = \begin{cases} \frac{h_c(P, O_c)}{|T_P|}, & \text{if } P\ bound; \\ \sum_{P_i \in P} \frac{h_c(P_i, O_c)}{T_{P_i}}, & otherwise. \end{cases}$$

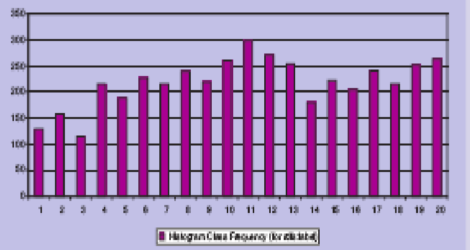
Example

R = number of resources = 500

$T_{rdfs:label}$ = number of triples matching rdfs:label = 500

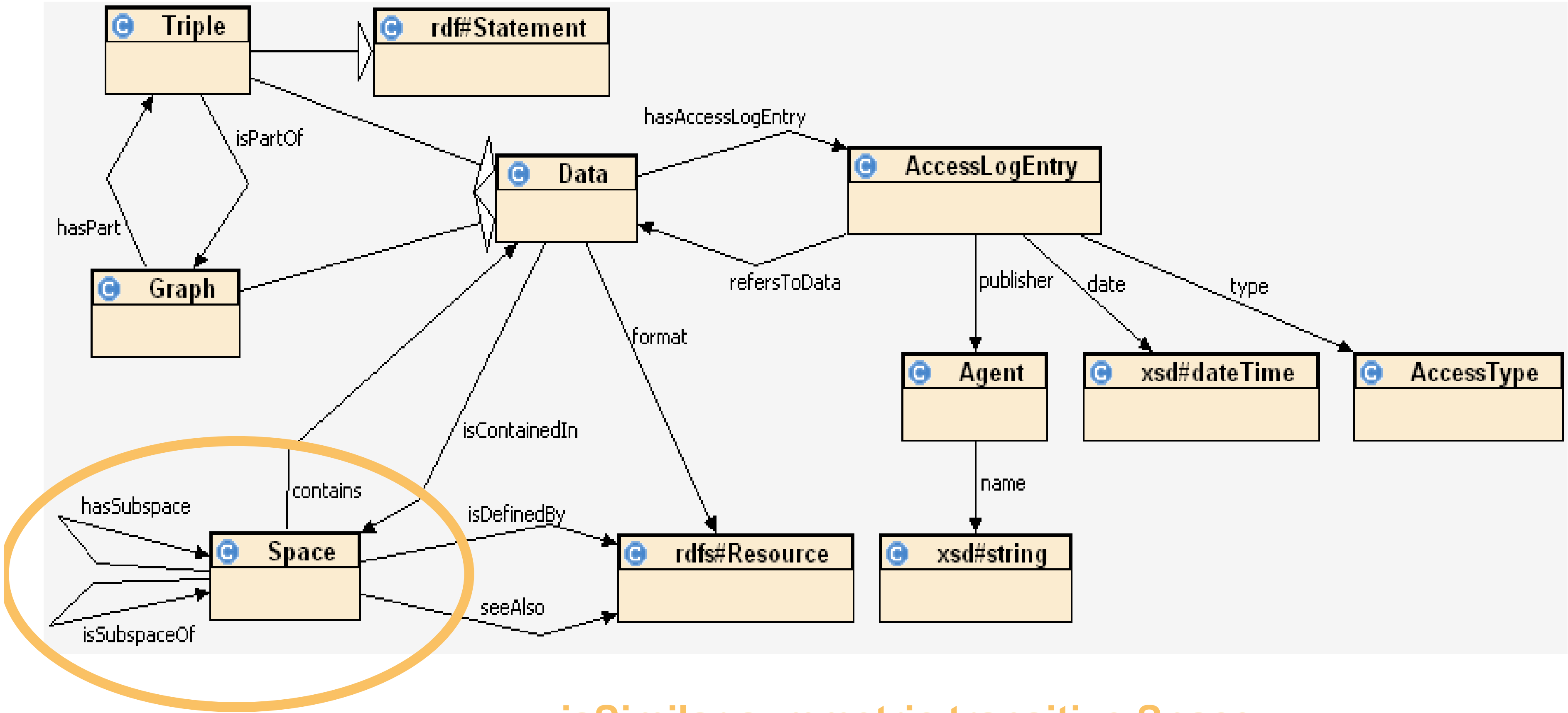
T = number of triples = 1000

$$sel([ub:Student17\ rdfs:label\ "Markus"]) = \frac{1}{500} \times \frac{500}{1000} \times \frac{300}{500} = 0.0006$$



The most selective triple pattern could indicate the description which leads to the most precise space, the new first „one space“!

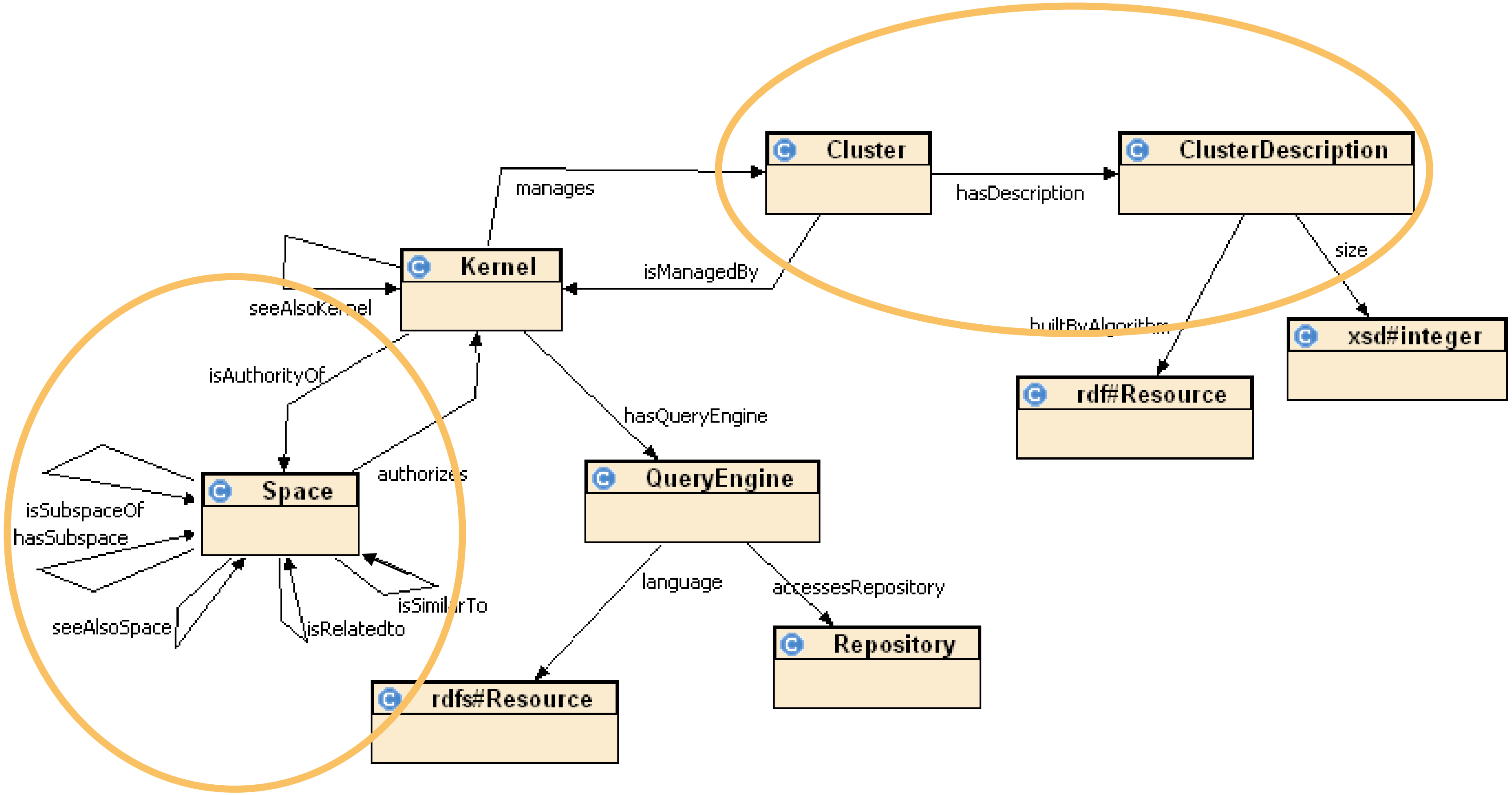
TS Ontology v1.0



- isSimilar symmetric transitive Space
- isRelated symmetric Space
- seeAlsoSpace Space

http://www.tripcom.org/ontologies/

TS Ontology v1.1



- Kernel links
- Clusters (descriptions)

Summary

- Metadata describes the sources of data
- Metadata describes the relationships of sources

- This allows for „physical routing tables“ through kernel relationships, and
- This allows as well for „logical routing tables“ through space relationships

- We do not have to treat every triple
  - Simplicity
  - Scalability
  - Preservation of semantics by „ignoring“ it

Expected Results in 2008

- Realization of Semantic Routing ideas...
  - Implementation
  - Evaluation
- Algorithms for the discovery of relevant spaces
  - Might be an alternative appraoch to DHT
  - Delivers the missing piece of semantic routing
- Semantic clustering approaches for self-organization
  - Selection of relevant metadata
  - Implementation of v1.x of TS ontology
- ...?

WP2 Status: T2.7

- Aim of current task (to end March 08): specify how other formalisms beyond RDF would be represented in Triple Space

- Syntax? Formalisms being considered: OWL and WSML both have RDF serializations
- Rules? One issue is the lack of stability in the specification of Semantic Web rules, we wait possibly on developments in RIF
- Semantics? Current work foresees triple-based representation of OWL Horst fragment and WSML Core fragment

WP2 Resources

WP 2: Triple Space Knowledge Representation						FUB	CEFR	LFUI	TUW	ONTO
					Target PM	22	0	15	6	18
		In charge	Start	End	Duration					
T2.1	RDF representation as tuples		1	6	6	6		1		[illegible]
T2.2	Triple space ontology		1	12	12	3		3.5		[illegible]
T2.3	Implementation RDF tuplespace		7	12	6	1		1		2.
T2.4	Specification semantic clustering		13	18	6	3.5	1	3	1	[illegible]
T2.5	Evaluation of T2.1 and T2.3 w.r.t. OWL and beyond		13	18	6	1		1		[illegible]
T2.6	Implementation semantic clustering		19	23	4	2		1	2	1.
T2.7	OWL and rules representation as tuples		19	24	5	3		4.5		
T2.8	Implementation self organization clustering		24	28	4	2.5		6	2	
T2.9	Evaluation of Implementation/Integration to WP6		29	36	7			1	1	[illegible]
					Sum	22	1	16	6	18

FUB 2.2 / 5

LFUI 1 / 5.5

TUW 0.2 / 2

ONTO 1.1 / 6.5

WP2 next steps (T2.6)

- Distribution system by DHT
  - Use of P-Grid P2P overlay without its storage layer – communication to be direct to underlying database
  - Indexing of (s,p,o,space URL) so lookup is based on single triple patterns
  - Selection of „most significant“ triple pattern(s), join of results from individual triple patterns
- Distribution system by metadata
  - Spaces are described in terms of relationship to other spaces as well as in terms of their content
  - First space selection by „most selective triple pattern“
  - Routing through relationships across spaces
  - Security policies on metadata operations to clarify

WP2 next steps (T2.7)

- Discussion of the knowledge representation paradigms and reasoning with:
  - OWL (DL)
  - WSML (LP)
- Theoretical foundation how to extend RDFS with WSML rules:
  - Embedding RDF in F-Logic
  - Embedding Extensional RDFS in First-Order Logic
- TODO:
  - Analyze the IRIS (WSML reasoner) to specify a subset of supported rules
  - Translate the data log rules to OWLIM rule language

End of Document