Outline 2

Semi-Automatic Information and Knowledge Systems

•

Ontology Mapping & Alignment

Monika Lanzenberger



Introduction 3

- Ontology Mapping: for each ontological entity in the first ontology, we try to locate a corresponding entity in the second ontology, with the same or the closest semantics. It constitutes a fragment of more ambitious tasks such as the alignment of ontologies.
- Ontology Alignment: bringing two ontologies into mutual agreement, making them consistent and coherent with one and another. It may often include a transformation of the source ontologies removing the unnecessary information and integrating missing information.
- Whereas alignment merely identifies the relation between ontologies, mappings focus on the representation and the execution of the relations for a certain task.

•	n	tı	r	\neg		ш	L	~	М	1	\neg	'n	۱
•		ı		_	·	ı	и		ш	L	₩.	41	ı

- Similarity Layers
- General Alignment Process
- Application Scenarios
- Complexity & Evaluation
- Using Mappings

web

ML

Introduction: Similarity of Entities

1

- Entities are the same, if their features are the same or similar enough.
- Features represent a certain meaning
- Low similarity may not give evidence for alignments
- High similarity may give strong evidence for alignments
- Not every similarity is of equal importance



ML

Introduction: Relations Among Concepts

5

Introduction: Relations Among Instances

Equal

Similar

Different

Syntactically Equal

6

OWL Ontology Construct	Comparison Relationship	Description			
Concept	Equal	URI's equal.			
		Class member instances equal.			
	Syntactically equal	Labels are the same.			
	Similar	Superclasses are the same.			
		Subclasses are the same.			
		Data properties are the same.			
		Object properties are the same.			
		Similar low/high fraction of instances.			
	Broader than	Subclass superclass comparison.			
	Narrower than	Superclass subclass comparison.			
	Different	Class is different from all classes of the second ontology.			
CATT BOX	1	I			

sen **s** webs

Instances

Introduction: Relations Among Data Properties

ML

7

Introduction: Relations Among Object Properties

URI's equal.

Labels are the same.

Instances of the same concept.

Property members are the same.

Two instances linked via the same property to another instance.

Instance is different from all instances of the second ontology.

8

OWL Ontology Construct	Comparison Relationship	Description				
Data Properties	Equal	URI's equal.				
	Syntactically Equal	Labels are the same.				
	Similar	Data property domains are the same.				
		Data super properties are the same.				
		Data sub properties are the same.				
		Data properties members are the same.				
	Different	Data property is different from all data properties of the second ontology.				

Object Properties	Equal	URI's equal.
	Syntactically Equal	Labels are the same.
	Similar	Object property domains are the same.
		Object super properties are the same.
		Object sub properties are the same.
		Object properties members are the same.
	Different	Object property is different from all object properties of the second ontology.





- Introduction
- Similarity Layers
- General Alignment Process
- Application Scenarios
- Complexity & Evaluation
- Using Mappings

Selle

ML

9

Ontology Similarity

11

$$sim: \mathfrak{P}(E) \times \mathfrak{P}(E) \times O \times O \rightarrow [0,1]$$

Positiveness

$$\forall e, f \in \mathfrak{P}(E), O_1, O_2 \in O, sim(e, f, O_1, O_2) \geq 0$$

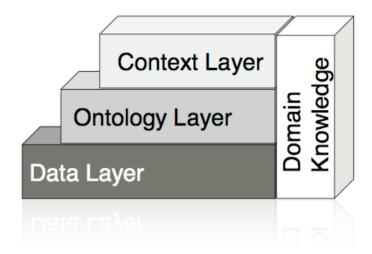
Maximality

$$\forall e, f, g \in \mathfrak{P}(E), O_1, O_2 \in O, sim(e, e, O_1, O_2) \ge sim(f, g, O_1, O_2)$$

Symmetry

$$\forall e, f \in \mathfrak{P}(E), O_1, O_2 \in O, sim(e, f, O_1, O_2) = sim(f, e, O_2, O_1)$$

[Ehrig 2007]



seme

[Ehrig 2007]

_ .

12

ML

$$sim: \mathfrak{P}(E) \times \mathfrak{P}(E) \times O \times O \rightarrow [0,1]$$

Two entity sets are identical

$$\forall e, f \in \mathfrak{P}(E), O_1, O_2 \in O, sim(e, f, O_1, O_2) = 1 \Leftrightarrow e = f$$

Two entity sets are similar / different to a certain degree

$$\forall e, f \in \mathfrak{P}(E), O_1, O_2 \in O, 0 < sim(e, f, O_1, O_2) < 1$$

Two entity sets are different and have no common characteristics

$$\forall e, f \in \mathfrak{P}(E), O_1, O_2 \in O, sim(e, f, O_1, O_2) = 0 \Leftrightarrow e \neq f$$

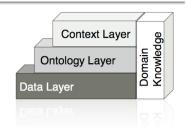
sen e

ML



[Ehrig 2007]

Data types such as integers, strings etc. are compared by operations such as relative distance and edit distance



Equal values:

$$sim_{equality}(v_1, v_2) := \begin{cases} 1 & \text{if } v_1 = v_2, \\ 0 & \text{otherwise.} \end{cases}$$

String similarity:

$$sim_{string}(v_1, v_2) := \max(0, \frac{\min(|v_1|, |v_2|) - ed(v_1, v_2)}{\min(|v_1|, |v_2|)})$$

Relative distance:

$$sim_{diff}(v_1, v_2) := 1 - \frac{|v_1 - v_2|}{maxdiff}$$

Selle

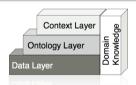
[Ehrig 2007]

ML

Ontology Layer

15

Similarity between sets of entities:



Single Linkage

$$sim_{single}(E, F) := \max_{(e,f)|e \in E, f \in F} (sim(e, f))$$

- Average Linkage
$$sim_{complete}(E,F) := \frac{\sum_{\forall (e,f)|e \in E, f \in F} sim(e,f)}{|E| \cdot |F|}$$

Multi Similarity

$$sim_{set}(E,F) := rac{\sum_{e \in E} \mathbf{e}}{|\sum_{e \in E} \mathbf{e}|} \cdot rac{\sum_{f \in F} \mathbf{f}}{|\sum_{f \in F} \mathbf{f}|}$$

with $\mathbf{e} = (sim(e, e_1), sim(e, e_2), ..., sim(e, f_1), sim(e, f_2), ...),$

[Ehrig 2007]

f analogously.

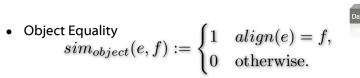


ML

web

Ontology Layer

Object similarity:



• Explicit Equality $sim_{explicit}(e,f) := \begin{cases} 1 & statement(e,"sameAs",f), \\ 0 & \text{otherwise}. \end{cases}$

Similarity between sets of entities:

Dice Coefficient

$$sim_{dice}(E,F) := \frac{2 \cdot |E \cap F|}{|E| + |F|}$$

• Jacquard Coefficient $sim_{jacquard}(E,F) := \frac{|E \cap F|}{|F| + |F|}$

Selle

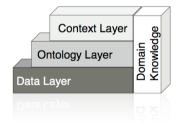
[Ehrig 2007]

ML

Ontology Layer

16

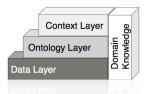
Label similarity:



 $sim_{label}(e, f) := sim_{string}(label(e), label(f))$

Taxonomic Similarity for Concepts: Extensional

$$sim_{taxonomic}(c_1, c_2) := \begin{cases} e^{-\alpha l} \cdot \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}} & \text{if } c_1 \neq c_2, \\ 1 & \text{otherwise.} \end{cases}$$



Extensional Concept Similarity:

$$sim_{extension}(c_1, c_2) := sim_{set}(\iota_C(c_1), \iota_C(c_2))$$

Domain and Range Similarity:

$$sim_{domran}(r_1, r_2) := 0.5 \cdot (sim_{object}(ran(r_1), ran(r_2))$$
$$+ sim_{object}(dom(r_1), dom(r_2)))$$

Concept Similarity of Instances:

$$sim_{parent}(i_1, i_2) := sim_{object}(c_1, c_2)$$

with $i_1 \in \iota_C(c_1), i_2 \in \iota_C(c_2)$

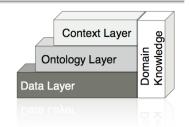
Selle

[Ehrig 2007]

Outline 19

- Introduction
- Similarity Layers
- **General Alignment Process**
- **Application Scenarios**
- **Complexity & Evaluation**
- Using Mappings

Frequency of usage of an entity or its characteristics in a certain context



$$sim_{use}(e, f) := sim_{diff}(Usage(e, con), Usage(f, con))$$

Example: Two books may be similar if their authors have many coauthored publications.

sem web

ML

ML

[Ehrig 2007]

ML

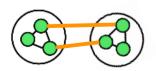
General Alignment Process

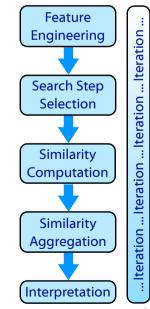
20





 $align: E \times O \times O \rightharpoonup E$, $\forall e \in E_{O_1} \ \exists f \in E_{O_2}, O_1 \in O, O_2 \in O :$ $align(e, O_1, O_2) = f$ $\vee align(e, O_1, O_2) = \bot$



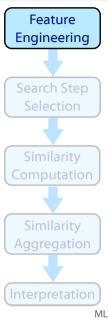




 $feat: O \times O \rightarrow \mathfrak{P}(F)$

Determine a list of features F:

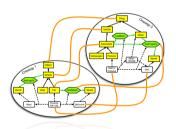
Extract characteristics of both ontologies, i.e. the features of their ontological entities (concepts C, relations R, and instances I) from intensional and extensional ontology definitions



serie webs

[Ehrig 2007]

Ontology Alignment Example: Step 1



The Car concept of ontology 1 is described by its label: Car, its superclass (subclassOf Vehicle), its concept sibling: boat, the direct property: hasSpeed, and its instance Porsche KA-123

The relation has Speed is described by the domain: Car and the range: Speed

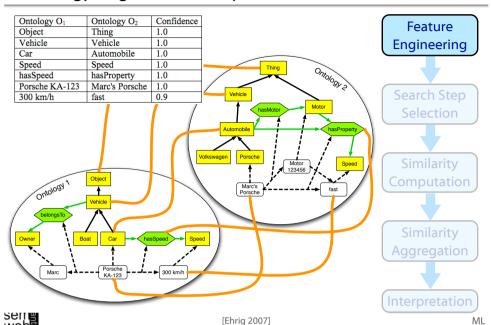
The instance Porsche KA-123 is characterized by the instantiated property instance: belongsTo: Marc and property instance: hasSpeed: 300 km/h

Feature
Engineering

Search Step
Selection

Similarity
Computation

Similarity
Aggregation

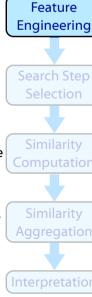


General Alignment Process: Step 1

24

Possible features:

- Identifiers: i.e. strings with dedicated formats, such as unified resource identifiers (URIs) or RDF labels.
- RDF/S Primitives: such as properties or subclass relations
- OWL Primitives: such as an entity being the sameAs another entity
- Derived Features: which constrain or extend simple RDFS primitives (e.g. most-specific-class-of-instance)
- Aggregated Features: i.e. aggregating more than one simple RDFS primitive, e.g. a sibling is every instance-of the parent-concept of an instance
- Domain Specific Features for instance, in an application where files are represented as instances and the relation hashcode-of-file is defined, we use this feature to compare representations of concrete files
- Ontology External Features: Any kind of information not directly encoded in the ontology, such as a bag-of-words from a document describing an instance



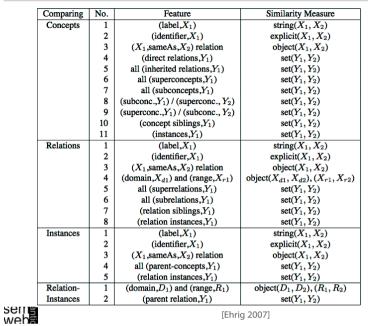


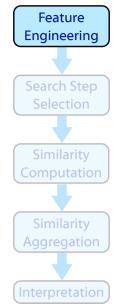
[Ehrig 2007]

ML

serie web

[Ehrig 2007]





Ontology Alignment Example: Step 2

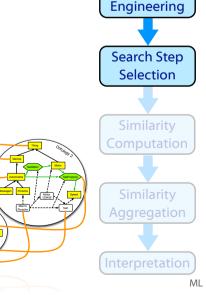
Compare all entities of the same type:

[Ehrig 2007]

55 candidate alignments:

- 42 concept pairs (6x7)
- 4 relation pairs (2x2)
- 9 instance pairs (3x3)

e.g., comparing
o₁: belongsTo with
o₂:hasProperty and with
o₂:hasMotor



$select: O \times O \to \mathfrak{P}(E \times E)$

Most common methods:

• compare all entities of O_1 with all entities of O_2 :

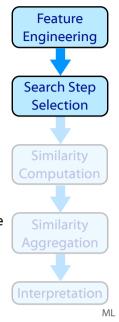
$$e, f \in E_1 \times E_2$$

any pair is treated as a candidate mapping

• or only compare entities of the same type

$$e, f \in (C_1 \times C_2) \cup (R_1 \times R_2) \cup (I_1 \times I_2)$$

 or use heuristics to lower the number of candidate mappings (e.g., applied in QOM) using strategies such as random or label, or change propagation



seme webs

ML

27

Feature

[Ehrig 2007]

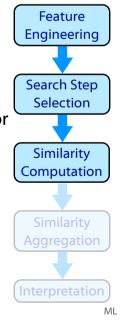
General Alignment Process: Step 3

 $sim: E \times E \times F \rightarrow [0,1]^k$

Based on the features F of the ontological entities we do the similarity computation for all pairs of candidates

Additional similarity measures exist

An example is described on the next slide



28



alignment: o₁:Car and o₂:Automobile

Computing the similarity of the candidate

Feature

Engineering

ſ	Comparing	No.	Feature	Similarity Measure			
ŀ	Concepts	1	$(label_1X_1)$	$string(X_1, X_2)$			
	Сопсери	2	(identifier X_1)	$\operatorname{explicit}(X_1, X_2)$			
			$(X_1, \text{sameAs}, X_2)$ relation	object (X_1, X_2)			
		3	(direct relations, Y_1)	$set(Y_1, Y_2)$			
		5	all (inherited relations, Y_1)	$\operatorname{set}(Y_1, Y_2)$			
		6	all (superconcepts, Y_1)	$set(Y_1, Y_2)$			
		7	all (subconcepts, Y_1)	$set(Y_1, Y_2)$			
		8	(subconc., Y_1) / (superconc., Y_2)	$\operatorname{set}(Y_1, Y_2)$			
		9	(superconc., Y_1) / (subconc., Y_2)	$ set(Y_1, Y_2) $			
		10	(concept siblings, Y_1)	$ set(Y_1, Y_2) $			
		11	(instances, Y_1)	$set(Y_1, Y_2)$			
	Relations 1 (label, X 2 (identifier,		$(label_{\cdot}X_1)$	$string(X_1, X_2)$			
			(identifier X_1)	explicit (X_1, X_2)			
			$(X_1, sameAs, X_2)$ relation	$\operatorname{object}(X_1, X_2)$			
		4	(domain, X_{d1}) and (range, X_{r1})	object $(X_{d1}, X_{d2}), (X_{r1}, X_{r2})$			
		5	all (superrelations, Y_1)	$set(Y_1, Y_2)$			
		6	all (subrelations, Y_1)	$set(Y_1, Y_2)$			
		7	(relation siblings, Y_1)	$\operatorname{set}(Y_1,Y_2)$			
		8	(relation instances, Y_1)	$\operatorname{set}(Y_1,Y_2)$			
	Instances		$(label, X_1)$	$string(X_1, X_2)$			
			(identifier, X_1)	$explicit(X_1, X_2)$			
			$(X_1, sameAs, X_2)$ relation	$\operatorname{object}(X_1, X_2)$			
		4	all (parent-concepts, Y_1)	$\operatorname{set}(Y_1,Y_2)$			
		5	(relation instances, Y_1)	$\operatorname{set}(Y_1,Y_2)$			
	Relation-	ation- 1 (domain, D_1) and (range, R_1)		object $(D_1, D_2), (R_1, R_2)$			
	Instances	2	(parent relation, Y_1)	$\operatorname{set}(Y_1,Y_2)$			
serre web			[Ehrig 2007]				

Feature **Engineering** Search Step Selection Similarity Computation

General Alignment Process: Step 4

31

Feature

Engineering

Search Step

Selection

ML

$agg: [0,1]^k \to [0,1]$

The individual similarity measures are weighted and combined

$$sim_{aggs}(e,f) := \frac{\sum_{k=1..n} w_k \cdot adj_k(sim_k(e,f))}{\sum_{k=1..n} w_k}$$

Some approaches for aggregation:

Averaging:

$$w_k = 1, adj_k(x) = id(x)$$

Linear Summation:

 w_k learned or manually assigned, $adj_k(x) = id(x)$

Linear Summation with negative evidence: wk: can have a negative value (e.g., superconcepts of the first entity have a high similarity with subconcepts of the second entity)

Sigmoid Function: emphasize high similarity and de-emphasize low similarity: $adj_k(x) = siq_k(x - 0.5),$

$$sig_k(x) = \frac{1}{1 + e^{-a_k x}}$$

Similarity Computation Similarity Aggregation

For every feature

 $sim_{label}(o_1:Car,o_2:Automobile) =$ $sim_{string}('Car', 'Automobile') = 0.0$

sim_{superconcept} (o₁:Car,o₂:Automobile) = $sim_{set}({o_1:Vehicle}, {o_2:Vehicle}) = 1.0$

Search Step Selection Similarity we compute a similarity. E.g., Computation

ML

32

Selle

[Ehrig 2007]

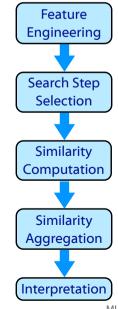
General Alignment Process: Step 5

 $inter: [0,1] \rightarrow \{alignment\}$

Aggregated similarity is compared to a threshold: every value above indicates an alignment

Determine the threshold:

- constant,
- $max(sim_{aqq}(e, f)) constant$,
- $max(sim_{aaa}(e, f))(1-p)$







[Ehrig 2007]

- Entities are similar if their position in the structure is similar
- Structure similarity is expressed through the similarity of the other entities in the structure
- Calculating the similarity for one entity pair, the similarity of the neighboring entity pairs are needed
- In a first round only basic comparison methods (e.g. string similarity) is applied (or pre-given alignments are used)
- In further rounds already computed pairs and use more sophisticated structural similarity measures
- When to stop the iteration:
 - 1. fixed number of iterations

Two ore more ontologies

Time resources are less critical

Finally high quality requirements

Human post-processing is required

- fixed time constraint
- 3. changes of alignments compared to a threshold

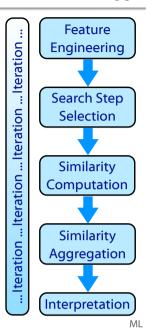
are combined into one target ontology

• By establishing alignments among entities

we identify equal entities which we can merge



[Ehrig 2007]



Introduction

- Similarity Layers
- General Alignment Process
- Application Scenarios
- Complexity & Evaluation
- Using Mappings

Selle

ML

Ontology Merging

35

Web Service Composition

36

- Agents or web service often use different representations of their domains resulting in different expressions on their goals, and their input or output
- Collaborate despite the heterogeneous representations
- Standard upper-level ontologies or ontology alignment
- Alignment needs to be fast, reliable, and correct
- Wrong results can lead to unjustified costs
- Sometimes user checks are possible
- Example: combine a booking service of an air carrier and a hotel reservation network

ML



ML

ML

- Users formulate a query in a specific query language based on a specific ontology
- Query is sent to a guery engine
- To access heterogeneous information sources the query needs to be re-written for the target ontologie(s)
- For the presentation of the answers the results have to be transformed back again
- Rewriting / Mapping should be fast and fully automatic
- Users may tolerate wrong results as long as the correct results are returned as well

Selle

[Ehrig 2007]

Outline 39

- Introduction
- Similarity Layers
- General Alignment Process
- Application Scenarios
- **Complexity & Evaluation**
- Using Mappings

- New information is inferred from distributed and heterogeneous ontologies
- Time constraints are not critical (for both, alignment and inference tools)
- Quality of the alignments is very important
- Alignment needs to be done automatically
- Wrong results may trigger additional wrong results in a cascading manner
- Detection of conflicting inconsistencies is required

[Ehrig 2007]

Many unsolved research issues

ML

Sem

Evaluation: Recall & Precision

40

Standard information retrieval metrics

Precision
$$p = \frac{\#correct_found_mapping}{\#found_mappings}$$

Recall $r = \frac{\#correct_found_mappings}{\#existing_mappings}$
F-Measure $f_1 = \frac{2pr}{p+r}$

- Compliance measure quality of identified alignments
- Performance measure quality of algorithm in terms of computational resources
- User-related measure overall subjective user satisfaction, e.g., how much user effort is needed
- Task-related measure quality of alignment for a certain use case or application scenario

sen e

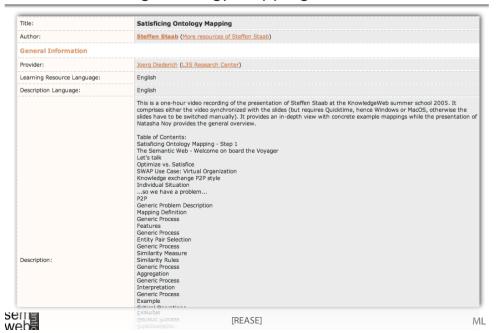
[Ehrig 2007]

Video Satisficing Ontology Mapping

43

ML

41



Ontology Alignment Evaluation Initiative + http://oaei.ontologymatching.org/ Apple FNR razr SimilarityFormeln The Ontolog...ent Source PSM OntologyMatching Recall and Precision CS652 SemWebLVAMaterial1 Ontology Alignment Evalua... **Ontology Alignment Evaluation Initiative** The increasing number of methods available for schema matching/ontology integration suggests the need to establish a consensu for evaluation of these methods. The Ontology Alignment Evaluation Innitiative is a coordinated international initiative to forge The goals of the Ontology Alignment Evaluation Innitiative are: · assessing strength and weakness of alignment/matching systems; comparing performance of techniques;
 increase communication among algorithm developer. improve evaluation techniques: most of all, helping improving the work on ontology alignment/matching through the controlled experimental evaluation of the techniques performances The means to achieve these goals are: · The organization of a yearly evaluation event; The publication of the tests and results of the event for further analysis. 2006 OAEI-2006 at the ISWC ontology matching workshop, Athens 2005 OAEI-2005 at the K-Cap workshop on integrating ontologies, Banff 2004 EON 2004 Ontology alignment contest, Hiroshima 2004 ParMIS 2004 I3CON, Gaithersburg Libs Library of competence tests and results Etiq On using Ontology Alignment Evaluation Initiative data Who Committee

Steffen Staab Video Slide - Reduction of Comparisons

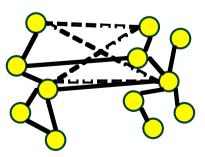
44

ML

- Random Selection
- Closest Label

Supp Sponsors

- Change Propagation
- Combination





ML

47

- $c = (feat + select + comp \cdot (\Sigma_k sim_k + agg) + inter) \cdot iter$
- NOM

$$c = O((n + n^2 + n^2 \cdot (\log^2(n) + 1) + n) \cdot 1)$$

= O(n^2 \cdot log^2(n))

PROMPT

$$c = O((n + n^2 + n^2 \cdot (1 + 0) + n) \cdot 1)$$

= O(n²)

QOM

$$c = O((n + n \cdot log(n) + n \cdot (1 + 1) + n) \cdot 1)$$

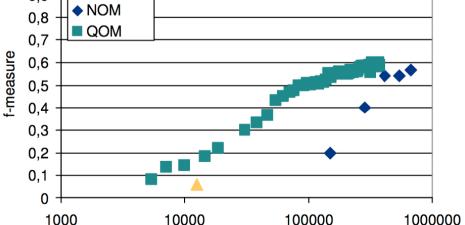
= $O(n \cdot log(n))$

Selle

[Slides Staab Video]

Steffen Staab Video Slide - Scenario 2

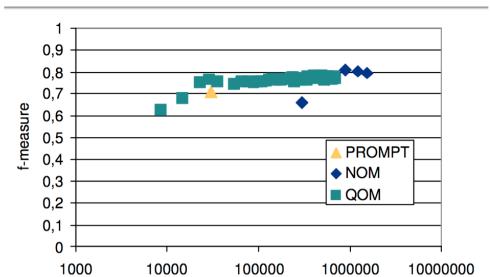
1 0,9 PROMPT • NOM



time (ms)

web

[Slides Staab Video] ML



Sell [Slides Staab Video]

time (ms)

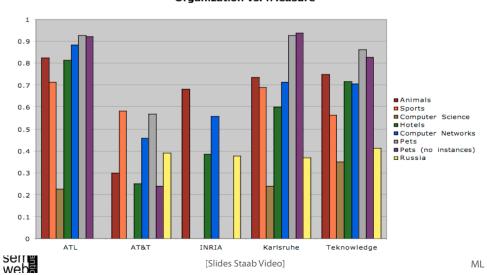
Steffen Staab Video Slide - Results

48

ML

Results for I3CON Ontology Alignment Experiment

Organization vs. fMeasure





Video Ontology Mapping and Alignment

Title: Author: Natasha Noy (More resources of Natasha Noy) **General Information** Joerg Diederich (L3S Research Center) Provider: Learning Resource Language: English Description Language: English This is a one-hour video recording of the presentation of Natasha Noy at the KnowledgeWeb summer school 2005. It comprises either the video synchronized with the slides (but requires Quicktime, hence Windows or MacQS, otherwise the slides have to be switched manually). It provide a high-level overview on ontology mapping while the presentation of Steffen Staab provides more details and example systems. Table of Contents: Ontology Mapping and Alignment Lots of Overlapping Ontologies on the Semantic Web Example Definitions of School Creating Correspondences Between Ontologies Semantic Integration Tasks Reasons for Mismatches Types of Mismatches Language-level Mismatches Ontology-level Mismatches Ontology-level Mismatches: Examples Some of the Differences Categories of Mappings Mapping Discovery: Information Sources Using a Common Reference Ontology Solve the problem before it arises Description: Using reference ontologies: Problems
Using Lexical Information Using Ontology Structure Using External Sources Using Prior Matches Mapping Composition Using Corpus of Matches

[REASE]

Outline 50

- Introduction
- Similarity Layers
- General Alignment Process
- Application Scenarios
- Complexity & Evaluation
- Using Mappings

sem webs

ML

Natasha Noy Video Slide - Using Mappings

52

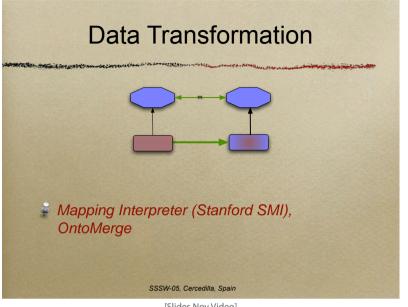
Using Mappings

- o Data transformation
- Query answering
- Reasoning with mappings
 - mapping composition (covered earlier)
- Generation of ontology extensions

SSSW-05, Cercedilla, Spain

51





webs

[Slides Noy Video]

ML

Data Transformation (II)

- Mapping interpreter (Stanford SMI)
 - Uses an instantiated ontology of mappings
 - mapping structure
 - Python rules
- OntoMerge
- Treats source ontologies with data and mapping axioms as a single ontology
- Uses a theorem prover to create new data

SSSW-05, Cercedilla, Spain

[Slides Noy Video]

ML

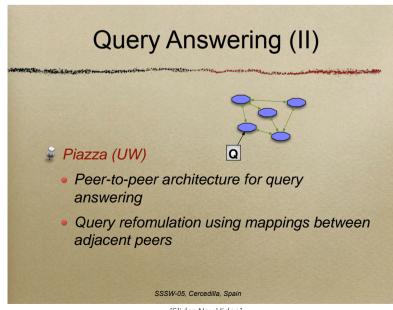
Natasha Noy Video Slide - Using Mappings

55

Natasha Noy Video Slide - Using Mappings

56

Query Answering Two settings one-to-one mappings global ontology SSSW-05, Cercedilla, Spain



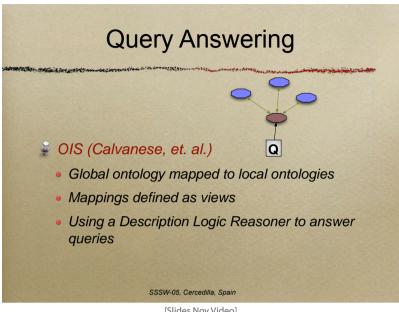


ML

Selle

[Slides Noy Video]

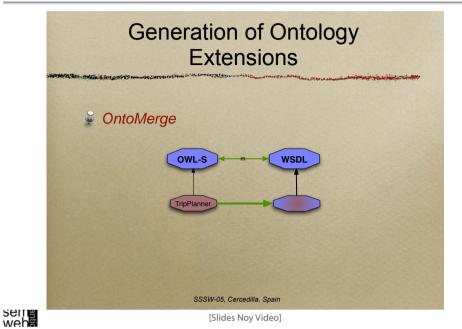
ML



webs

[Slides Noy Video]





References 60

[Ehrig 2007] Marc Ehrig: Ontology Alignment - Bridging the Semantic Gap, Dissertation, Springer, New York, 2007.

[Ehrig, et al. 2004] Marc Ehrig and York Sure: Ontology Mapping - An Integrated Approach, In: Proceedings of the First European Semantic Web Symposium. C. Bussler, J. Davis, D. Fensel, R. Studer, eds, 2004.

[REASE] REASE - Repository of EASE for Learning Units, http://rease.semanticweb.org, checked online 15.Jan.2008.

[Slides Noy Video] Natasha Noy: Ontology Mapping and Alignment, SSSW, 2005, pp. 39-45, http://dit.unitn.it/~accord/RelatedWork/Matching/Noy-MappingAlignment-SSSW-05.pdf, checked online 15.Jan.2008.

[Slides Staab Video] Steffen Staab: Satisficing Ontology Mapping, SSSW, 2005, http://rease.semanticweb.org/ubp/PUSH/search@srchDetailsLR?lrID=lr-leardiederich-1165406019525, checked online 15.Jan.2008.

ML