Introduction to the Ontology Alignment Evaluation 2005

Jérôme Euzenat
INRIA Rhône-Alpes
euzenat@inrialpes.fr

Heiner Stuckenschmidt
vrije Universiteit
heiner@cs.vu.nl

Mikalai Yatsevich
Università degli Studi di Trento
Dipartimento di Informatica e Telecomunicazioni
yatskevi@unitn.it

Outline

Goal

- Improving the performances of the ontology matching field ...

... through the comparison of algorithms ...

... on various sets of tests:

Created an "Ontology Alignment Evaluation Initiative" for permanence.

Schedule

Preparation June 1st-July 1st;

Execution July 1st-September 1st;

Evaluation August 15th + September 1st-7th;

Rules

- From a set of couple of ontologies (in OWL)...

... Use one automatic system...

... With the same set of parameters...

... To output an alignment (in the ontology alignment format).

- All general purpose resources authorized.

Validation

- Participants provide their systems. . .

. . . and their parameters;

- Organizers test this:

Participants

Name	System	Benchmarks	Directory	Anatomy	Validated
U. Karlsruhe	FOAM	√	√		
U. Montréal/INRIA	OLA	√	√		√
IRST Trento	CtxMatch 2	√	√		
U. Southampton	CMS	√	√	√	
Southeast U. Nanjin	Falcon	√	√	√	√
UC. Dublin	?	√	√		
CNR/Pisa	OMAP	√	√		

General comments

- More participants than last year
- No americans?
- We took into account most of the remarks of last year;
- The time devoted to this experiment (3 month during summer) was too short
- The final results are not yet on the web!

Test set

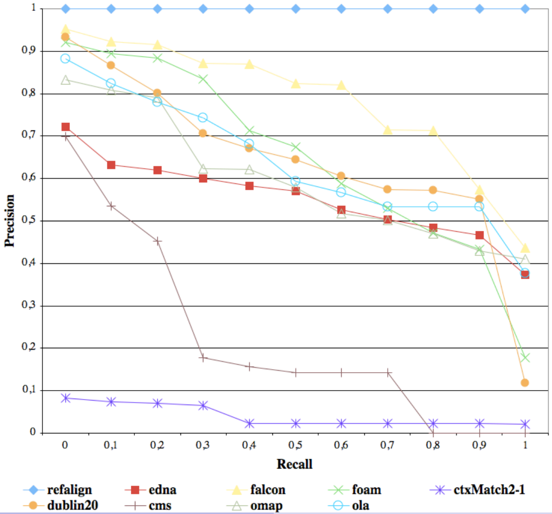
- improving on last year EON test bench (improved by circular references);
- based on a bibliography ontology in OWL-DL in RDF/XML featuring reference to outer ontologies;
- containing 33 named classes, 24 object properties, 40 data properties, 56 named individuals and 20 anonymous individuals;
- reference 1-1 alignments with "=" relation and 1. confidence;
- three group of tests: simple (4), systematic (46), and real-life (4);
- systematically altered by combining the transformations on names, comments, instances, classes, properties.

Results (precision and recall)

algo	edna	falcon	foam	ctxmatch	dublin	cms	omap	ola
1xx	0.96	1.00	0.98	0.10	1.00	0.74	0.96	1.00
2xx	0.41	0.90	0.89	0.08	0.94	0.81	0.31	0.80
3xx	0.47	0.93	0.92	0.08	0.67	0.93	0.93	0.50
means	0.45	0.91	0.90	0.08	0.92	0.81	0.35	0.80
1xx	1.00	1.00	0.65	0.34	0.99	0.20	1.00	1.00
2xx	0.56	0.89	0.69	0.23	0.71	0.18	0.68	0.73
3xx	0.82	0.83	0.69	0.22	0.60	0.18	0.65	0.48
means	0.61	0.89	0.69	0.24	0.72	0.18	0.70	0.74

In green, the validated results In red the very good performances.

Results (precision/recall profiles)



Comments

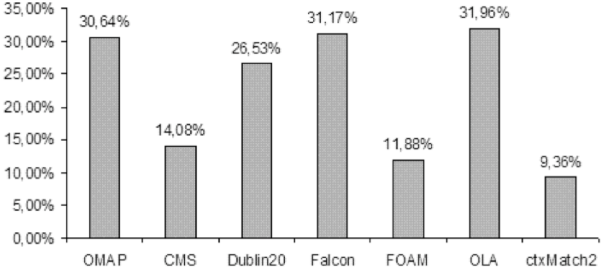
- Syntactic quality of results improved;
- Global quality of results improved (the tests were more difficult);
- The results of the best system improved (even with wider challenge);
- Last year, il was possible to win with a system mainly accounting on one feature of ontology; this year the winners combined features intelligently;
- We really need some semantically correct evaluation methods;
- First test with results validated!

Goal

Test ontology matching systems on web directories which can be treated as shallow ontologies:

- Real world matching task composed from parts of Google, Yahoo and Looksmart taxonomies
- Some reference mappings available
- Only true positives: allows to measure recall but not precision

Results


35,00%
30,00%
25,00%
20,00%
15,00%
10,00%
5,00%
0,00%
30,64%
14,08%
26,53%
31,17%
11,88%
31,96%
9,36%
OMAP
CMS
Dublin20
Falcon
FOAM
OLA
ctxMatch2

Comments

- The matching task is hard
  - The best systems found 30% of mappings or have recall about 30%
  - All the systems together found 60% of mappings
- The matching task is discriminant
  - 44% of the mappings found by any of the matching systems was found by only one system

Future work

- Extend the dataset in order to add true negatives
- Allows to measure precision
- Apply the similar dataset construction techniques to expressive ontologies (such as in medical ontologies task)
- Opens possibility to have quantitive analysis of results in medical ontologies case

Data set

The GALEN Anatomy model:

- Developed by the University of Manchester
- About 10.000 concepts and 30 relations about human anatomy
- Available as a (Protege) OWL file ($\approx$ 4MB)

The Foundational Model of Anatomy FMA

- Developed at University of Washington
- About 60.000 terms and 50 relations for talking about human anatomy and development
- Available as database plugin to Protege and as an exported RDF schema files ($>$ 40MB !) further translated in OWL.

Comments

- This has been a challenge for organizers;
- This was a challenge for tools (mostly for OWL managers);
- We are currently not able to clearly evaluate the results
- Beside partial attempts, there is no reference set of mappings for these ontologies;
- Plans: finding a set of core mappings automatically that can help to check results on these mappings.

Lesson learned

- More and more tools, and more robust;
- We can measure quality increase;
- It is difficult to find "real world" test sets;
- Real world test sets are difficult to evaluate;
- Validation of results is difficult, but not impossible.

Future plan

- Need for new real world cases;
- Need for new evaluation metrics;
- Need for more real world alteration of the benchmarks;
- We would like to accept continuous submissions;

Conclusion

- We will now listen to the participants;
- We will have a general discussion them;
- Thank you all!

http://oaei.inrialpes.fr/2005

AIFB

Framework for Ontology Alignment and Mapping

Marc Ehrig and York Sure
Institute AIFB, University of Karlsruhe
Integrating Ontologies Workshop
Banff, Canada
October 2, 2005

Alignment Definition

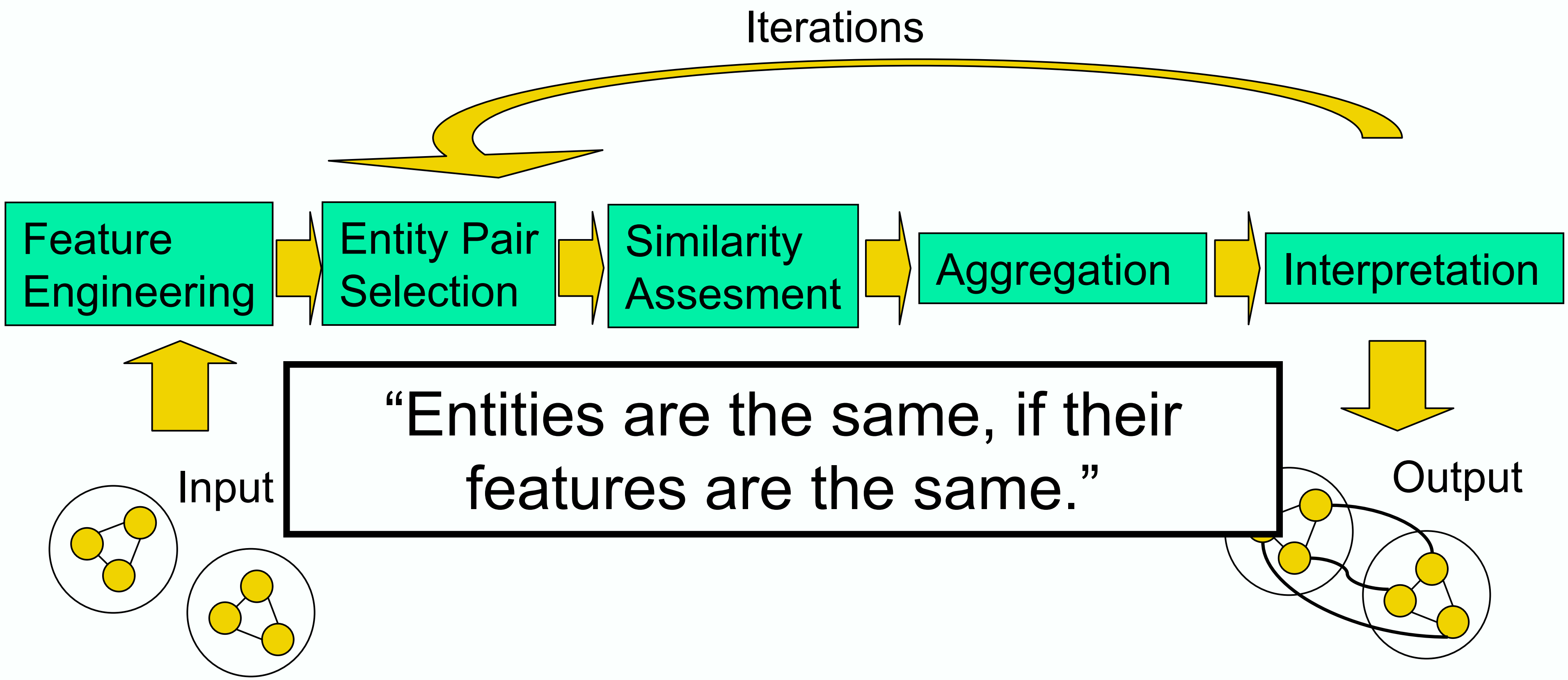
- Given two ontologies O_1 and O_2, aligning one ontology with another means that for each entity (concept C, relation R, or instance I) in ontology O_1, we try to find a corresponding entity, which has the same intended meaning, in ontology O_2.

- $align(e_{1i}) = e_{2j}$

Use Cases

- Query and Answer Translation
- Ontology Merging
- Ontology Evolution / Versioning
- Data Integration
- Reasoning
- …

Process



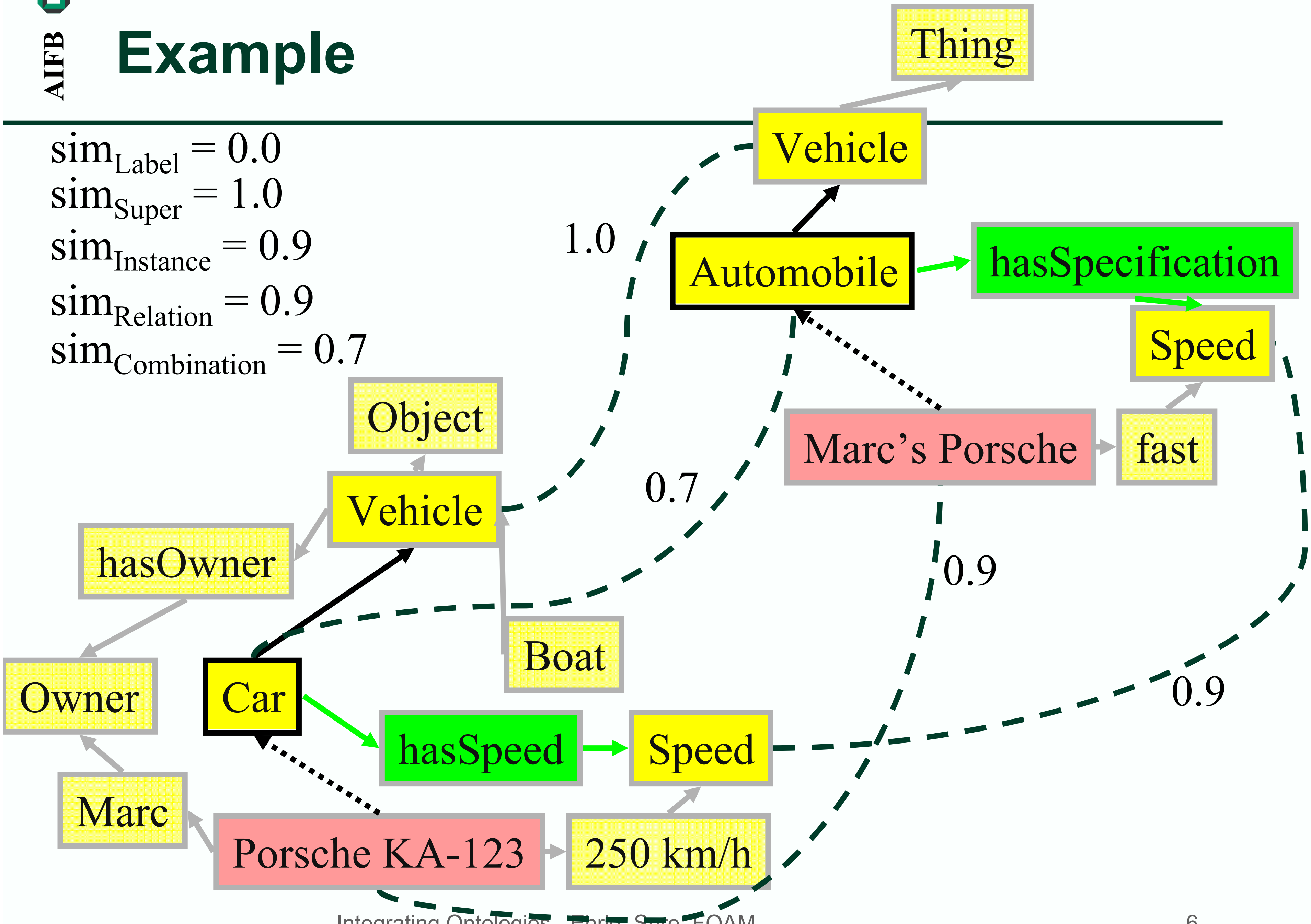
"Entities are the same, if their features are the same."

Similarity Rules

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

Example

$sim_{Label} = 0.0$
$sim_{Super} = 1.0$
$sim_{Instance} = 0.9$
$sim_{Relation} = 0.9$
$sim_{Combination} = 0.7$



AIFB

Different Aspects of Ontology Alignment

- Quality
- Time Constraints
- User Interaction
- Task-based Alignment
- Complex Alignments
- Ontology Merging
- Mapping Maintenance and Evolution
- …

Strengths

- Labels or identifiers are important and help to align most of the entities.
- The structure helps to identify alignments, if the labels are not expressive.
- A more expressive ontology results in better alignments; an argument in favor of ontologies compared to simple classification structures.
- The generally learnt weights have shown very good results.

Weaknesses

- The approach cannot deal with consequently changed labels. Especially translations, synonyms, or other conventions make it difficult to identify alignments.
- The system is bound to OWL-DL or lesser ontologies.

- Problems with the representation format of alignments.

General Comments

- improvement compared to last year
- middle to upper ranking
- valuable results

Open Source Tool

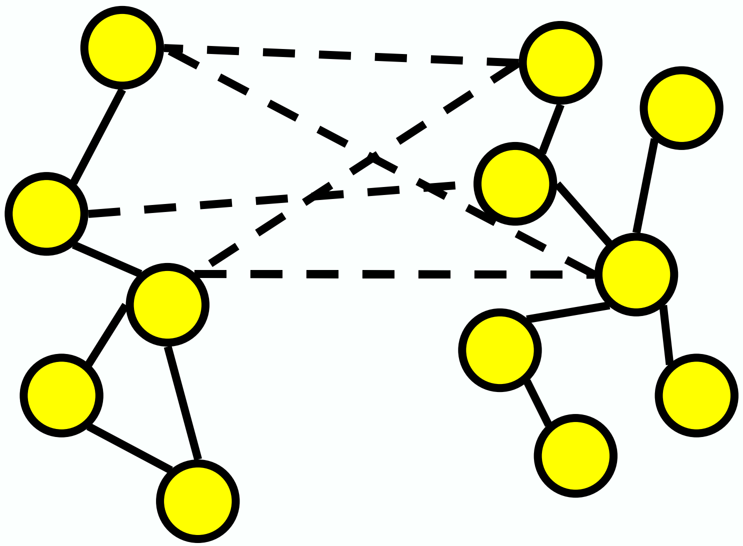
http://www.aifb.uni-karlsruhe.de/WBS/meh/foam

foam

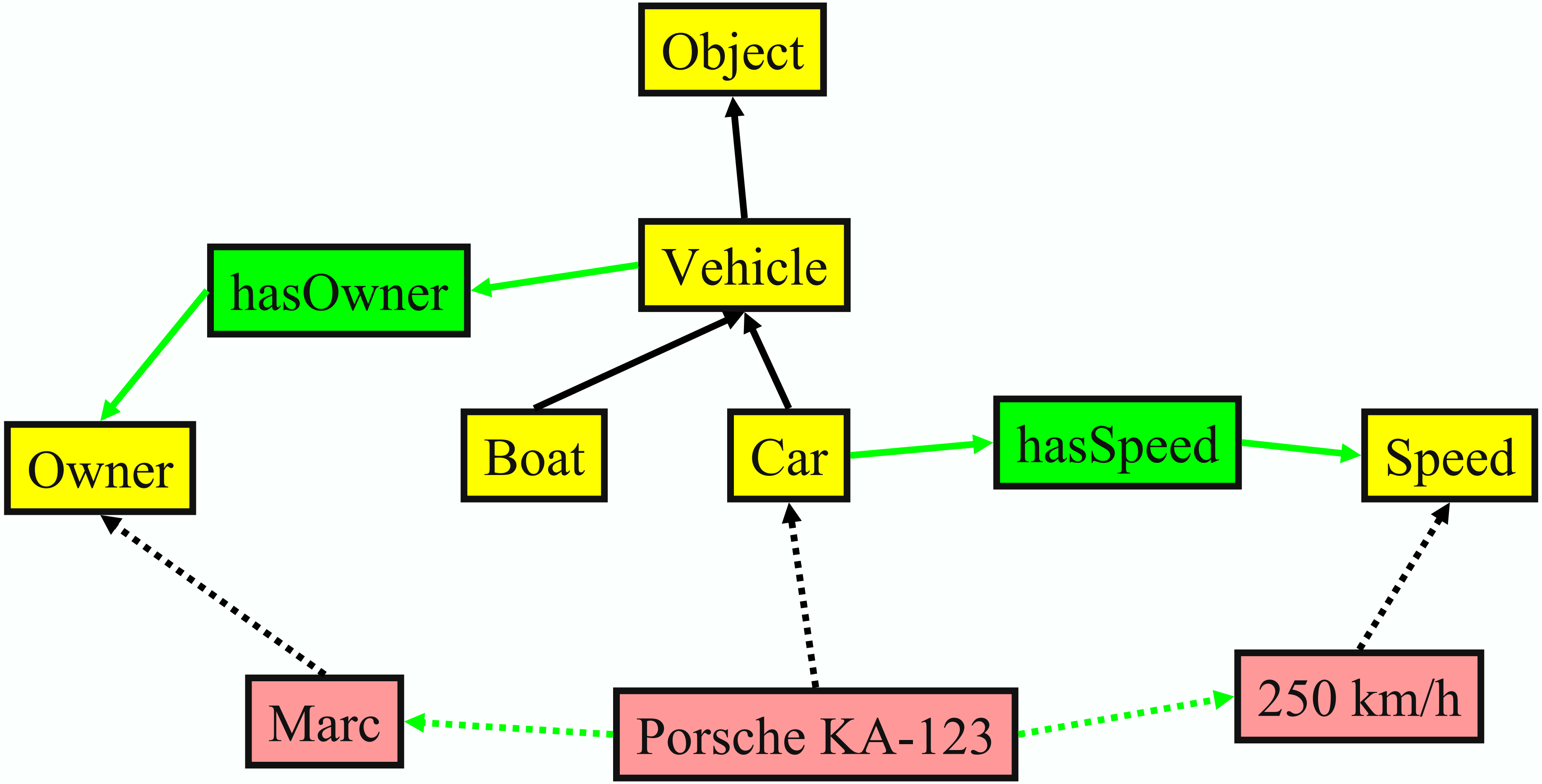
Framework for Ontology Alignment and Mapping

Candidates

- Complete
- Random selection
- Closest label
- Change propagation



Ontology Example



Similarity Measure

- String similarity

$$sim_{String}(s_1, s_2) = \max(0, \frac{\min(|s_1|, |s_2|) - ed(s_1, s_2)}{\min(|s_1|, |s_2|)})$$

- Object Similarity
- Set similarity

Combination

- How are the individual similarity measures combined?

- Linearly
- Weighted
- Special Function

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

Interpretation

- From similarities to alignments

- Threshold
- $align(e_{1j}) = e_{2j} \leftarrow sim(e_{1j}, e_{2j}) > t$

#	Name	Prec.	Rec.	F-measure	Time
101	Reference alignment	1.0	1.0	1.0	2.96
102	Irrelevant ontology	-	-	-	207.14
103	Language generalization	1.0	1.0	1.0	180.95
104	Language restriction	1.0	1.0	1.0	177.63
201	No names	0.90	0.65	0.75	175.99
202	No names, no comments	0.85	0.57	0.68	176.59
203	No comments	1.0	1.0	1.0	174.21
204	Naming conventions	0.96	0.93	0.94	185.09
205	Synonyms	0.80	0.67	0.73	174.46
206	Translation	0.93	0.76	0.84	172.15
207		0.95	0.78	0.86	167.89
208		0.96	0.87	0.92	164.20
209		0.81	0.57	0.67	168.63
210		0.92	0.67	0.77	164.31
221	No specialization	1.0	1.0	1.0	172.92
222	Flattened hierarchy	1.0	1.0	1.0	127.63
223	Expanded hierarchy	0.99	1.0	0.99	142.70
224	No instance	1.0	0.99	0.99	42.09
225	No restrictions	1.0	1.0	1.0	171.13
228	No properties	1.0	1.0	1.0	112.60

CROSI Mapping System

Advanced Knowledge Technologies IRC,
University of Southampton, UK

Hewlett Packard Laboratories, Bristol, UK

Bo Hu, Yannis Kalfoglou, Dave Reynolds, Nigel Shadbolt
{bh, y.kalfoglou, nrs}@ecs.soton.ac.uk, dave.reynolds@hp.com

Structure

- CROSI project
- Semantic Intensity Spectrum
- CMS – (CROSI Mapping System)
- OAEI’05 Results
- Issues

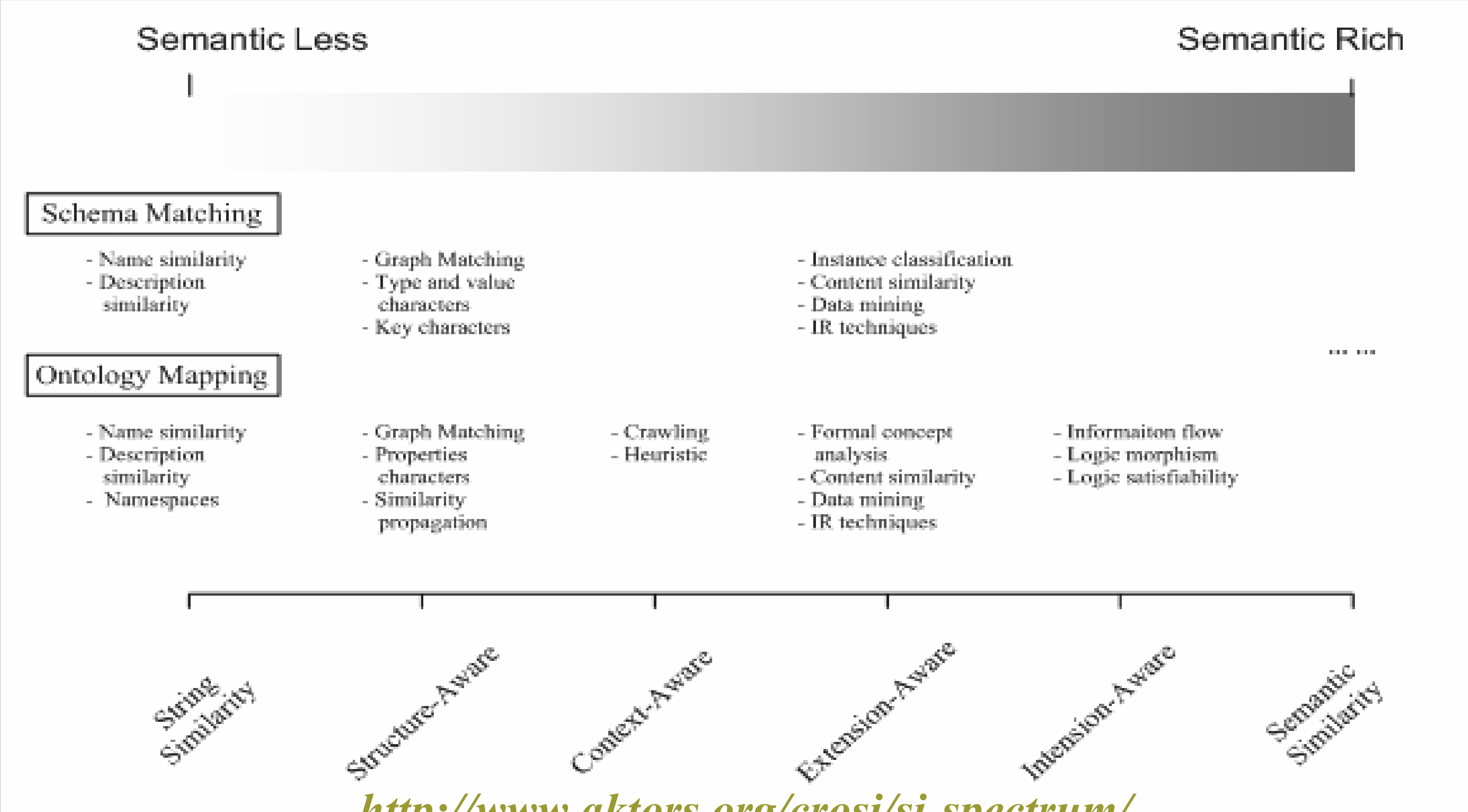
CROSI Project

- **C**apturing, **R**epresenting, and **O**perationalising **S**emantic **I**ntegration
- One year project (Oct'04—Oct'05) funded by Hewlett Packard Labs @Bristol, UK
- Deliverables
  - a survey of the state-of-the-art of semantic integration techniques and systems
  - a method for classifying semantic integration technology
  - a principled architecture for deploying ontology mapping systems
  - an ontology mapping system (CMS)
- more on: *http://www.aktors.org/crosi/*

Semantic Intensity Spectrum

- A classification technique for *both* database schema matching and ontology alignment approaches

Semantic Intensity Spectrum

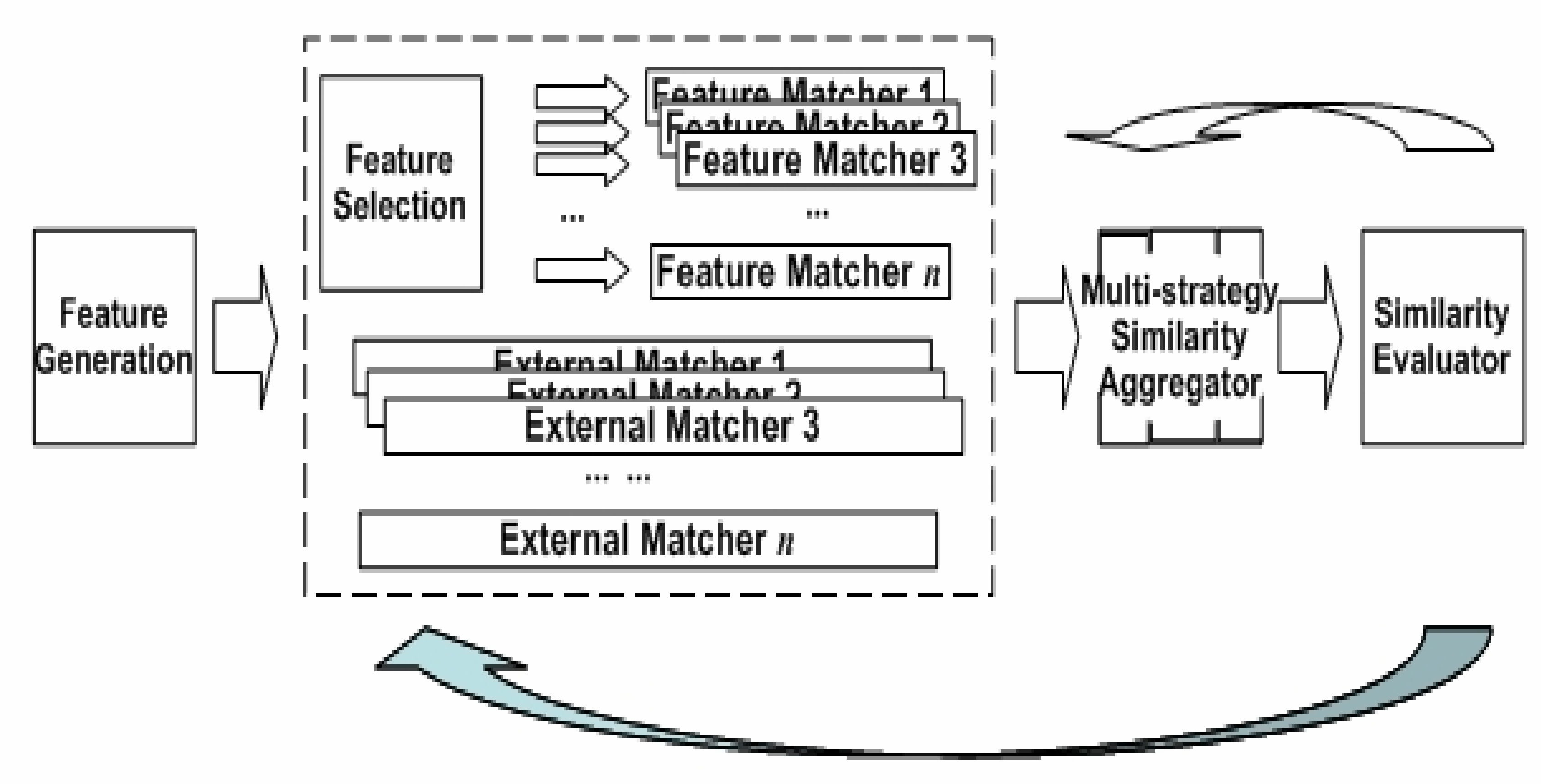


http://www.aktors.org/crosi/si-spectrum/

Semantic Intensity Spectrum

- A classification technique for *both* database schema matching and ontology alignment approaches
- A designer's aid for navigating through different mapping approaches with emphasis on the use of semantics

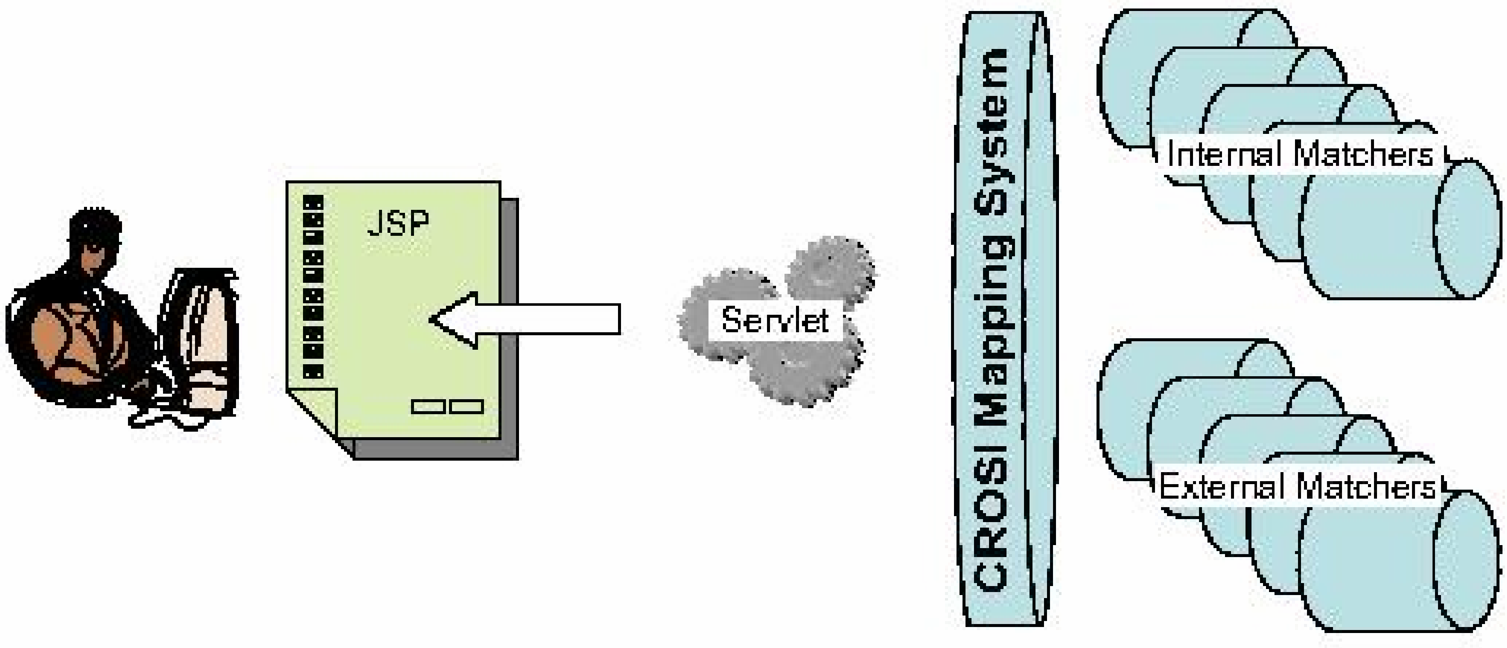
A principled architecture



CMS: design commitments

- Avoid *reinventing the wheel*
  - Use existing packages to enhance internal matchers
  - Use existing mapping/alignment systems as external matchers
- Semantically enriched matchers based on the definition of concepts
  - Propagate similarity along concept hierarchies
  - Refine concept similarity by taking into account the names, domains and ranges of declared properties
  - Compute similarity using WordNet hierarchies

CMS: deployment choices



OAEI'05 Results
benchmark cases

- When appropriate, we combined four internal matchers:
  - *CanoName:* concept and property names
  - *StructurePlus*: concept definitions
  - *WNDisSim:* WordNet based semantic distance
  - *HierarchyDisSim:* concept hierarchy-based semantic distance
- Only results wrt. concepts are saved

OAEI'05 Results
benchmark cases cnt'd.

- Weaknesses:
  - Does not perform well when both concept and property names are replaced by random strings
    - *should random strings be consider as a valid mapping?*
  - WordNet *synsets* need context information otherwise unrealistic mappings are proposed
    - e.g. MotionPictures ⇔ LectureNotes
- Strengths:
  - Penalisation of concept similarities is based on:
    - the subtle differences between property domains and ranges
    - their respective locations in the concept hierarchies

OAEI'05 Results
web directories case

- Had to deal with only concept names and single-line hierarchies
  - We found some cyclic definitions
- Canonical name and hierarchical distance matchers were used
- CMS performed well in the majority cases

OAEI’05 Results
medical ontologies case

- Size was the major obstacle for parsing and processing the FMA ontology
  - Naming conventions
  - Different focus of FMA and OpenGalen
- 100% mapping is identified between a few concepts
- More mappings can be identified when aligning concepts against individuals is allowed
  - *should a broader sense of mapping be adopted?*

Issues

- Difficult to validate results of real-life cases
- Numeric similarity scoring is confusing in the context of ontology mapping
- A pragmatic use-case to make the most out of ontology mapping/alignment technology

Falcon-AO: Aligning Ontologies with Falcon

NingSheng Jian, Wei Hu, Gong Cheng, Yuzhong Qu

Department of Computer Science and Engineering
Southeast University, P. R. China

Outline

- What is Falcon-AO
- How does Falcon-AO work
- Results
- Conclusion
- Outlook

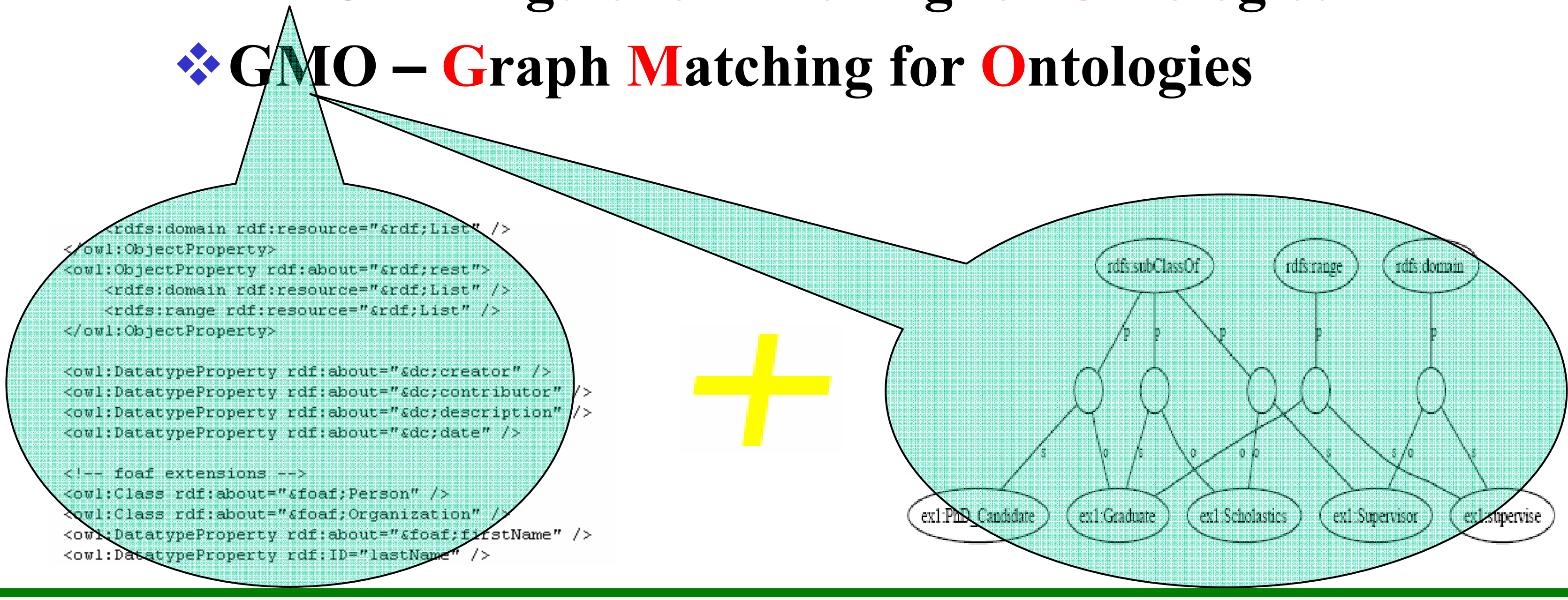
What is Falcon

Finding, Aligning, Learning ontologies, and ultimately for Capturing knowledge by an ONtology-driven approach.

A suit of methods and tools for the Semantic Web applications

What is Falcon-AO

- **Aligning Ontologies with Falcon**
- **An integration of two matchers**
  - **LMO – Linguistic Matching for Ontologies**
  - **GMO – Graph Matching for Ontologies**



LMO – Linguistic Matching for Ontologies

- **String Similarity (SS)**
  - **Edit Distance of Local Names**
- **Document Similarity (DS)**
  - **Entity ⇒ Virtual Document (VD)**

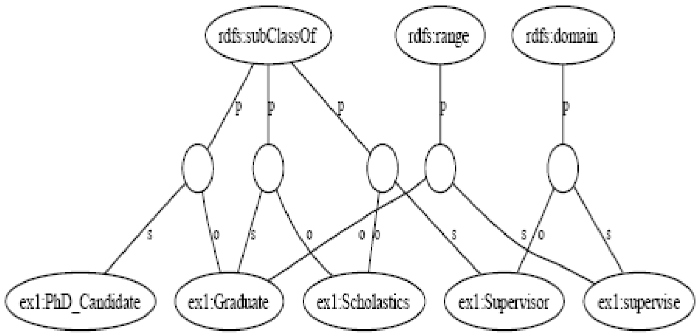
Local Name Label Comment *Neighbors'VD …*

Weighted *Bag of terms*

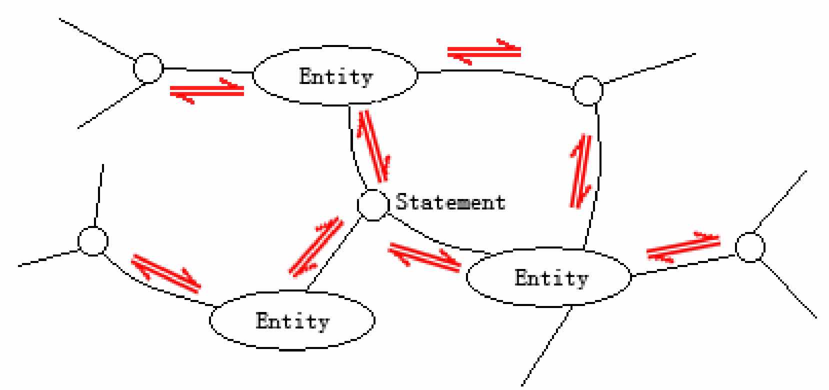
  - **TermWeighting = TF * IDF**
  - **Cosine Similarity in VSM (Vector Space Model)**
- ***LinguisticSimilarity = 0.2 * SS + 0.8 * DS***

GMO – Graph Matching for Ontologies

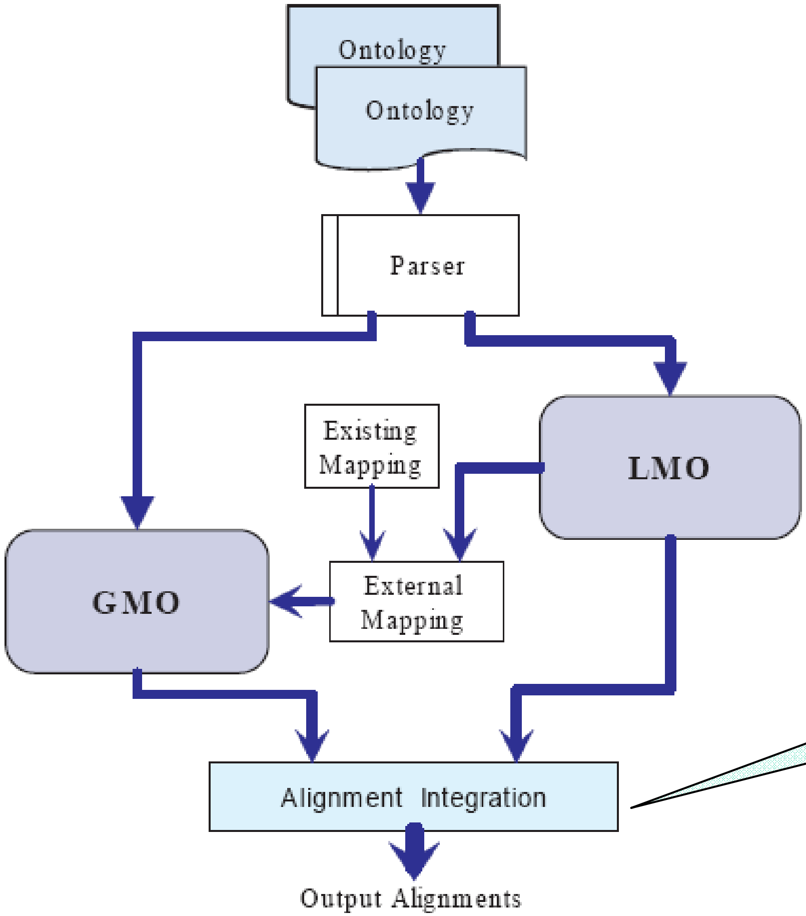
- **Bipartite Graph Model**



- **Similarity Accumulation**



Architecture of Falcon-AO



Linguistic Comparability
&
Structural Comparability

Results

Test Cases	Average Precision	Average Recall	Average F-Measure	Average Run Time
#101 - #104	**1.00**	**1.00**	**1.00**	**5s**
#201 - #210	**0.96**	**0.95**	**0.95**	**63s**
#221 - #247	**0.99**	**1.00**	**0.99**	**4s**
#248 - #266	**0.71**	**0.60**	**0.63**	**60s**
#301 - #304	**0.93**	**0.81**	**0.86**	**20s**

Low ***High***

Linguistic Comparability

Conclusion

- Both linguistic and structural
- Quickly with high linguistic comparability
- External mapping as input
- No lexical database

Outlook

- Very large ontologies
- Lexical database
- Many-to-many mapping
- Measurement of comparability

Thank you.

OLA in the OAEI 2005

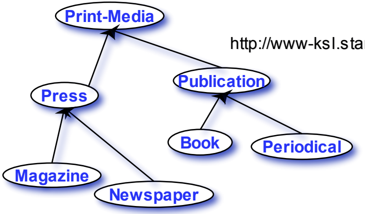
Jérôme Euzenat[2] & Philippe Guégan[1] & Petko Valtchev[1]

[1]Université de Montréal
Montréal (Qc), Canada
Petko.Valtchev@umontreal.ca

[2]INRIA Rhône-Alpes
Monbonnot, France
Jerome.Euzenat@inrialpes.fr

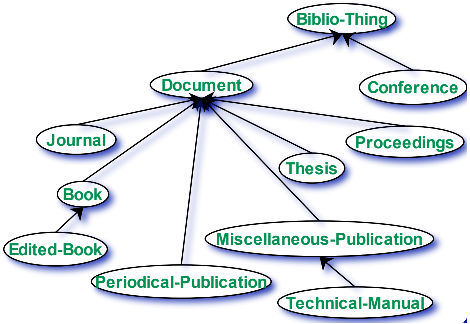
WOI 2005, October 2nd, Banff

Alignment: Problem

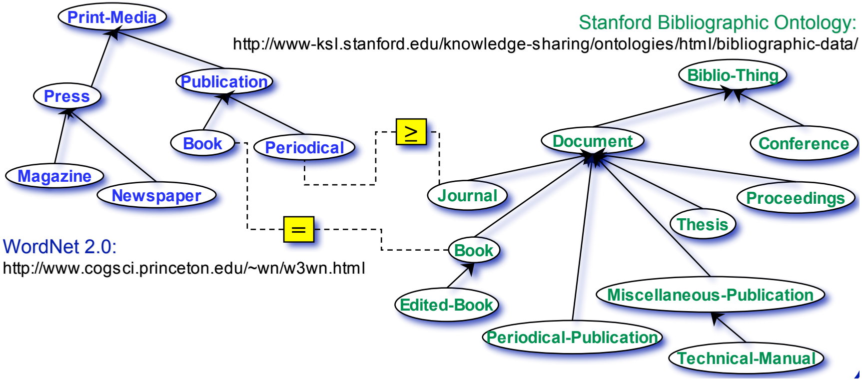


WordNet 2.0:
http://www.cogsci.princeton.edu/~wn/w3wn.html

Stanford Bibliographic Ontology:
http://www-ksl.stanford.edu/knowledge-sharing/ontologies/html/bibliographic-data/



Alignment: Goal



Outline

1. The problem
2. The system
3. Adaptations
4. Results

Ola/VisOn: Objectives

Collaborative project U. of Montreal/INRIA, on top of OWL API and Align API featuring:

- **parsing** and **visualization** of OWL-Lite and OWL-DL ontologies,
- **computation of similarities** between entities from two ontologies,
- **extraction of alignments** from a pair of ontologies,
- **manual construction of alignments** on two ontologies,
- **automated completion** of an existing (partial) alignment,
- **visualization** of alignments,
- **comparison** of alignments.

Principles of the alignment in OLA

All-encompassing comparison. All the available knowledge taken into account when aligning (neighborhood).

Highest automation level. Although (semi-)manual, the could benefit from automated "draft" alignment generation.

Category-dependent comparison. Entities divided into categories: Only same-category entity pairs compared. Similarity functions tailored acording to the category.

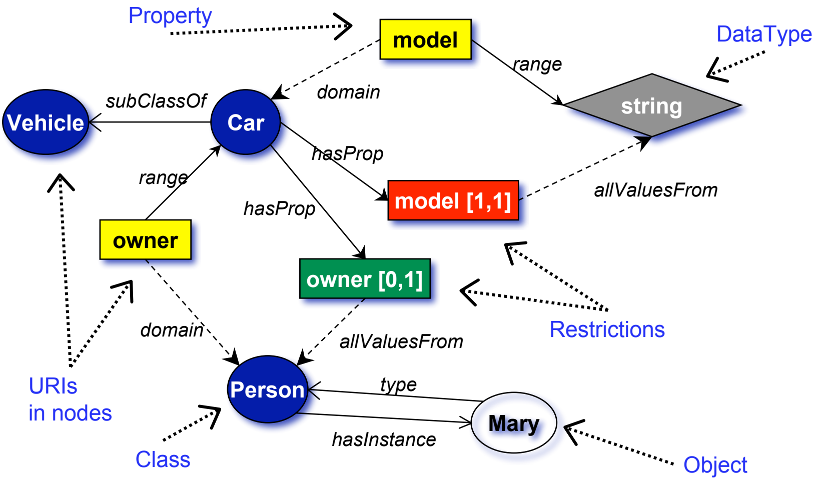
Comparability of similarity results. Useful properties of the similarity functions:

- *normalization*,
- *positiveness*,
- *maximalness*, etc.

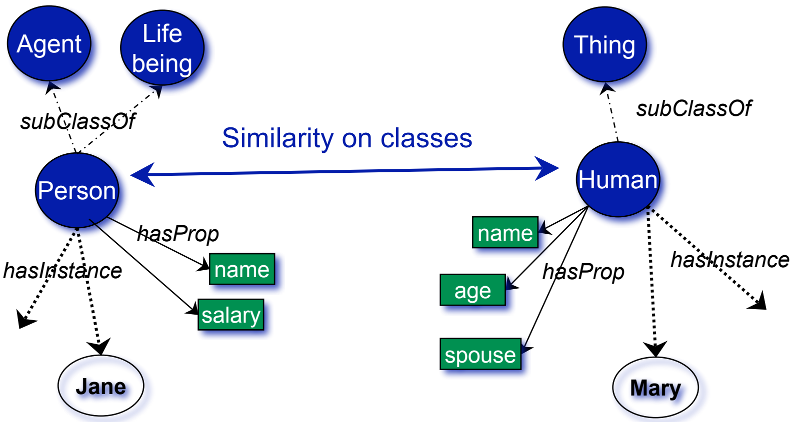
The process

1. **Parse and transform**: ontologies mapped into OL-Graphs, *labeled digraph* structures where *semantically related* entities are directly (physically) *connected*,
2. **Compare**: similarity for all same-category *cross-ontology* entity pairs extracted,
3. **Rank and align**: entity pairs ranked category-wise and best candidates made to alignment cells.

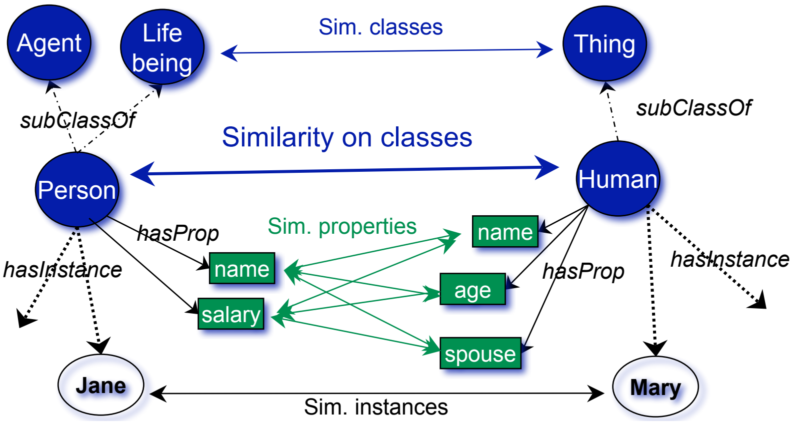
OL-Graph: Structure



OL-Graph: Comparing nodes



OL-Graph: Comparing nodes



Similarity model: Structure

Similarity factors for a category X from the OL-Graph:

- similarities of **identifying** or **describing** terms,
- similarities of the **neighbor node** pairs (same category, cross-graph), related by the **same relationship**,
- similarities of further category-spécific local features,

Similarity model: Generic function

Similarity function = **weighted linear** combination of **collection** similarities.

For a category X and its set of relationships $\mathcal{N}(X)$, the similarity measure $Sim_X : X^2 \to [0,1]$ is:

Definition

$$Sim_X(x, x') = \sum_{\mathcal{F} \in \mathcal{N}(X)} \pi_{\mathcal{F}}^{X} MSim_Y(\mathcal{F}(x), \mathcal{F}(x'))$$

where weights $\pi_{\mathcal{F}}^{X}$ sum to a unit:

$$\sum_{\mathcal{F} \in \mathcal{N}(X)} \pi_{\mathcal{F}}^{X} = 1.$$

Similarity model: Concrete functions

Similarity factors for all node categories [2]:

Funct.	Node	Factor	Measure
Sim_O	$o \in O$	$\lambda(o)$	sim_L
		$a \in A$, $(o,a) \in \mathcal{A}$	$MSim_A$
Sim_A	$a \in A$	$r \in R$, $(a,r) \in \mathcal{R}$	Sim_R
		$o \in O$, $(a,o) \in \mathcal{U}$	$MSim_O$
		$v \in V$, $(a,v) \in \mathcal{U}$	$MSim_V$
Sim_V	$v \in V$	value literal	type dependent
Sim_C	$c \in C$	$\lambda(c)$	sim_L
		$p \in P$, $(c,p) \in \mathcal{A}$	$MSim_P$
		$c' \in C$, $(c,c') \in \mathcal{S}$	$MSim_C$
sim_D	$d \in D$	$\lambda(r)$	XML-Schema
Sim_R	$r \in R$	$\lambda(r)$	sim_L
		$c \in C$, $(r,\texttt{domain},c) \in \mathcal{R}$	$MSim_C$
		$c \in C$, $(r,\texttt{range},c) \in \mathcal{R}$	$MSim_C$
		$d \in D$, $(r,\texttt{range},d) \in \mathcal{R}$	Sim_D
		$r' \in R$, $(r,r') \in \mathcal{S}$	$MSim_R$
Sim_P	$p \in P$	$r \in R$, $(p,r') \in \mathcal{S}$	Sim_R
		$c \in C$, $(p,\texttt{all},c) \in \mathcal{R}$	$MSim_C$
		$n \in \{0,1,\infty\}$, $(p,\texttt{card},n) \in \mathcal{R}$	equality

Similarity: Computation

- **Circularity** = fixed-point computation (SoE) [1],
- **Non-linear**: multiple choices unforseeable beforehand,
- However **a solution always exists** - process converges [3],

- Computed by an **iterative approximation process**:

  Step 0 initialize similarity by text comparison:
  - *string-based* or
  - *lexical* (WordNet).

  Step n+1 re-compute similarity from **Step n** values.

  Stop value increase drops below a *threshold*.

Adaptation: Weight computation

- a real **headache**: weights *a priori* fixed for entire benchmarks,
- compare entities on actual description space (EON):

Definition

$$Sim_X^+(x, x') = \frac{Sim_X(x, x')}{\sum_{\mathcal{F} \in \mathcal{N}^+(x,x')} \pi_{\mathcal{F}}^X}$$

where $\mathcal{N}^+(x, x')$ is the set of all relationships $\mathcal{F}$ s.t. $\mathcal{F}(x) \cup \mathcal{F}(x') \neq \emptyset$.

Adaptation: Weight computation (count'ed)

- WordNet similarity made simpler and more purposeful,
- adapt weights to refect the "**importance**" of a factor within the ontologies:
  - *intuition*: the bigger the number of links, the higher the respective weight,
  - $\pi_{\mathcal{F}}^{X}$ follows the average number of $\mathcal{F}$ links per X entity:

Definition

$$\pi_{\mathcal{F}}^{X} = \frac{\sum_{x \in X(\mathcal{O}_1 \cup \mathcal{O}_2)} |\mathcal{F}(x)|}{|X(\mathcal{O}_1 \cup \mathcal{O}_2)|}$$

OAEI Results

Test group	Prec.	Rec.
1XX	1.00	1.00
2XX	0.80	0.73
3XX	0.50	0.48
H-means	0.80	0.74

3XX definitely require further reflection.

Lessons learned

- Significant **improvement** w.r.t. EON: more than 30% precision gain
- **Well** on the **second** test category
- **But**: "real-world" ontololgies way too hard
- Where to look for **further** improvement:
  - (light weight) **inference** to complete the entity descriptions,
  - **category borders** could be made similarity-permeable.

Questions?

http://www.iro.umontreal.ca/∼owlola/OAEI.html

References

Gilles Bisson.
Learning in FOL with similarity measure.
In *Proc. 10th American Association for Artificial Intelligence conference, San-Jose (CA US)*, pages 82–87, 1992.

Jérôme Euzenat and Petko Valtchev.
Similarity-based ontology alignment in OWL-lite.
In *Proc. 15th ECAI*, pages 333–337, Valencia (ES), 2004.

Petko Valtchev.
Construction automatique de taxonomies pour l'aide à la représentation de connaissances par objets.
Thèse d'informatique, Université Grenoble 1, 1999.