ASMOV Results for OAEI 2009

Yves Reginald Jean-Mary
E. Patrick Shironoshita
Mansur R. Kabuka

INFOTECHSoft
INNOVATIVE HEALTHCARE SOLUTIONS

Automated Semantic Mapping of Ontologies with Verification (ASMOV)

- Algorithm:
  - Iterative similarity measurements
  - Semantic verification
  - User interaction

- Applications:
  - Information Integration
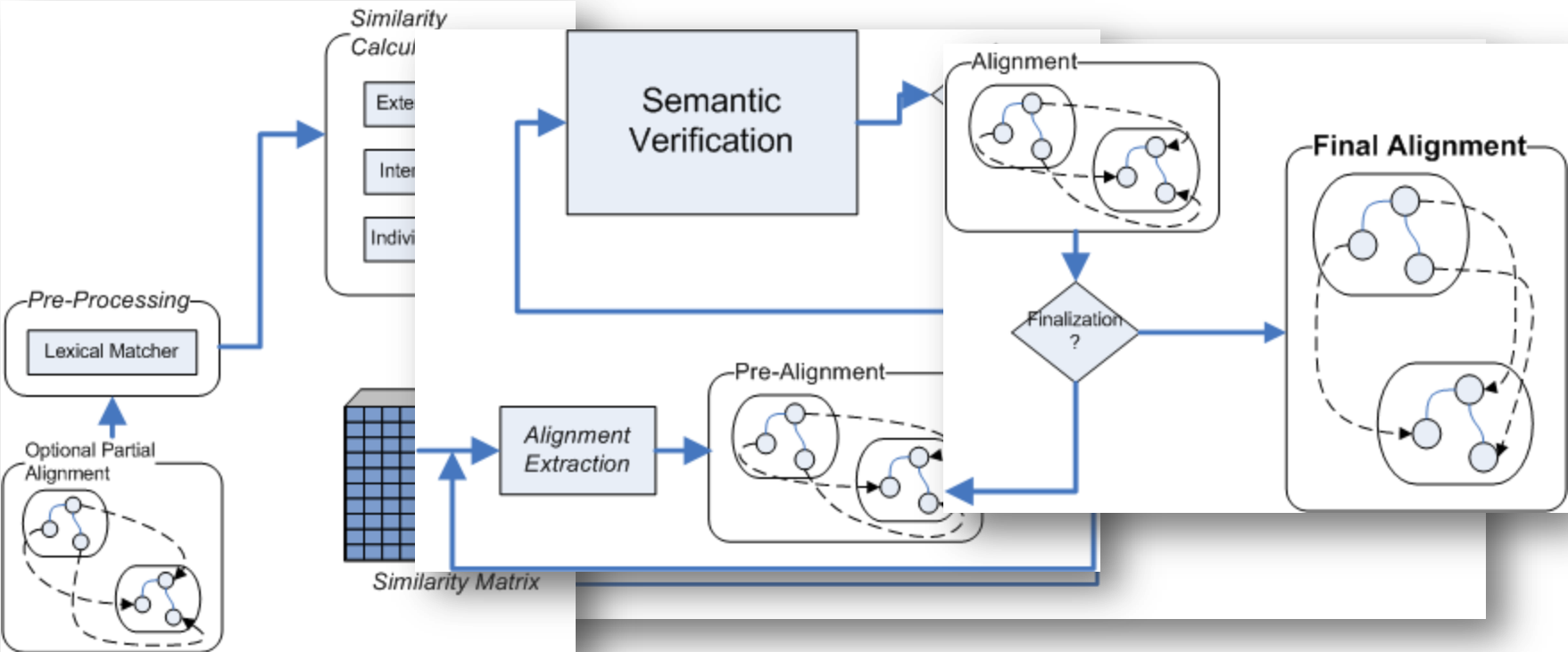 - Semantic cataloging (with Lockheed Martin)

Ontology Matching with Semantic Verification
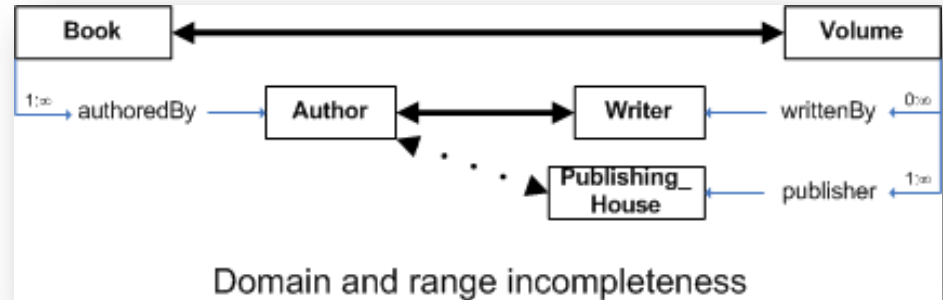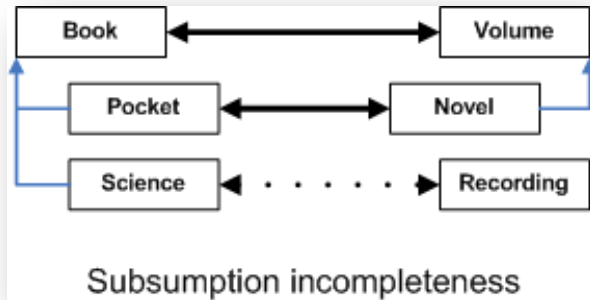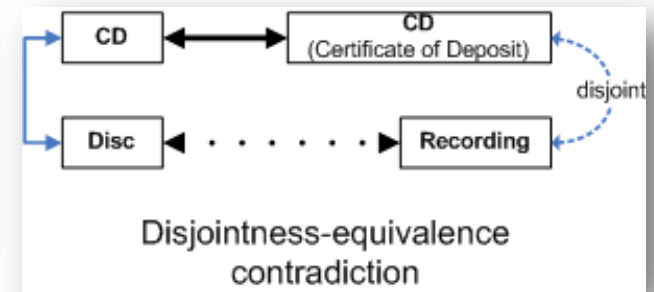Web Semantics: Science, Services and Agents on the World Wide Web
Yves R. Jean-Mary, Patrick Shironoshita, Mansur R. Kabuka
http://dx.doi.org/10.1016/j.websem.2009.04.001

INFOTECHSoFT
INNOVATIVE HEALTHCARE SOLUTIONS

ASMOV Algorithm

INFOTECHSOFT
INNOVATIVE HEALTHCARE SOLUTIONS

ASMOV Semantic Verification

Multiple-entity Correspondences

Crisscross correspondences

Disjointness-equivalence contradiction

Subsumption incompleteness

Domain and range incompleteness

INFOTECH SOFT
INNOVATIVE HEALTHCARE SOLUTIONS
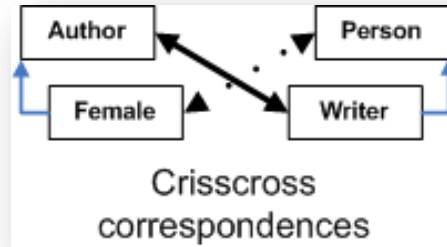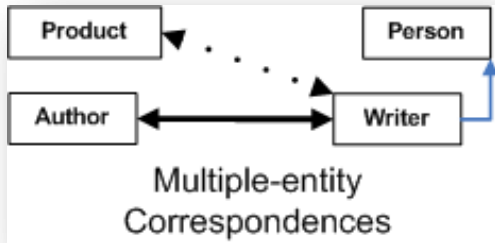
ASMOV: OAEI Participation

Tracks

2007 — 4
2008 — 5
2009 — 7

INFOTECHSOFT
INNOVATIVE HEALTHCARE SOLUTIONS
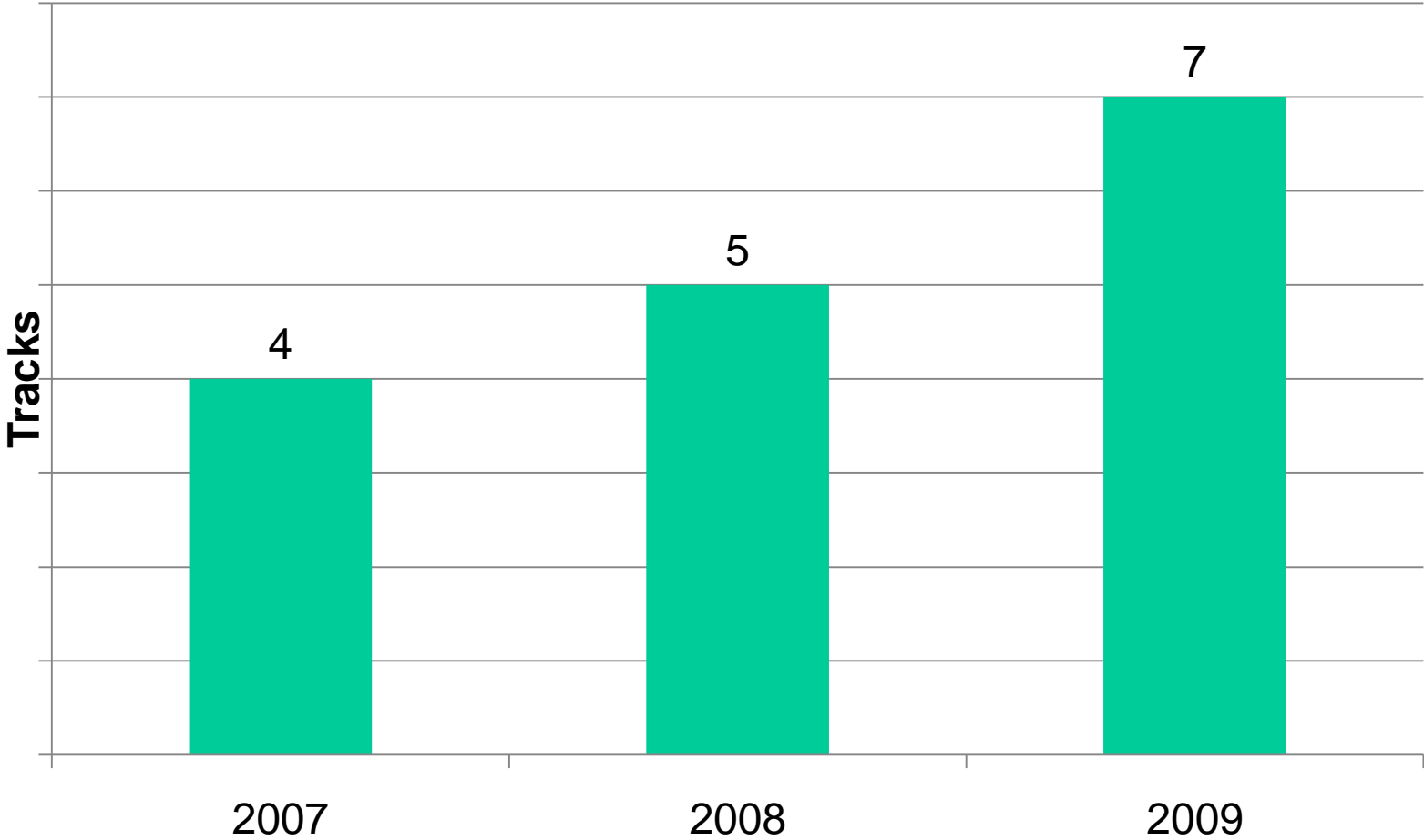
ASMOV OAEI 2009 – Overall Results

Results: http://support.infotechsoft.com/integration/ASMOV/OAEI-2009.
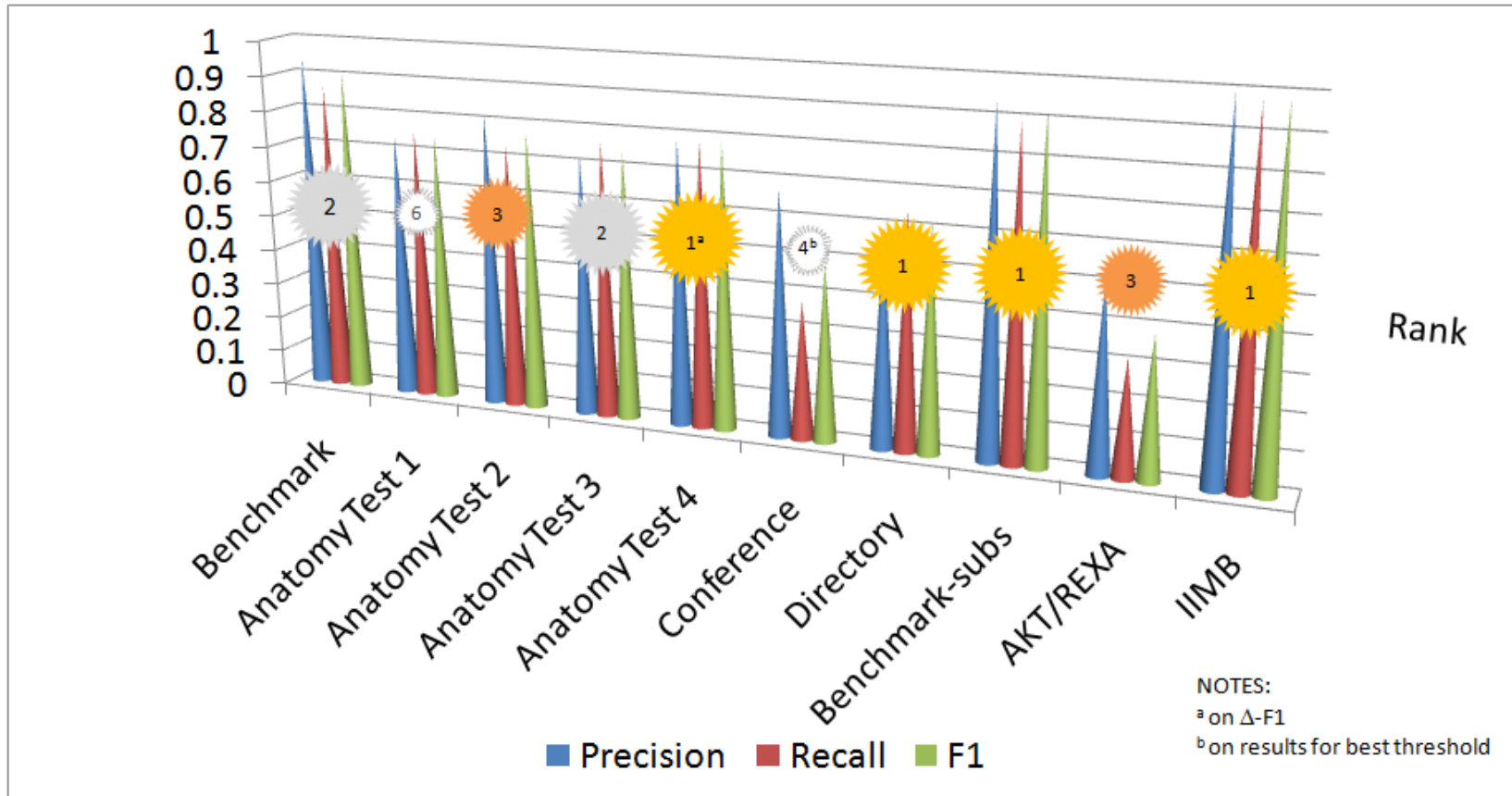
OAEI 2009 Benchmark Track

Level	ASMOV 2009			ASMOV 2008		
	Precision	**Recall**	**F-measure**	**Precision**	**Recall**	**F-measure**
0	1.00	1.00	1.00	1.00	1.00	1.00
1	1.00	1.00	1.00	1.00	1.00	1.00
2	1.00	0.99	0.99	1.00	0.99	0.99
3	0.99	0.98	0.98	0.98	0.97	0.97
4	0.99	0.98	0.98	0.99	0.98	0.98
5	0.97	0.93	0.95	0.96	0.93	0.94
6	0.95	0.88	0.91	0.94	0.88	0.91
7	0.94	0.83	0.88	0.93	0.83	0.88
8	0.91	0.71	0.80	0.90	0.71	0.79
9	*0.83*	*0.48*	*0.61*	*0.78*	*0.46*	*0.58*
10	0.4	0.04	0.07	0.40	0.04	0.07
3xx	0.81	0.82	0.81	0.81	0.77	0.79
All	**0.95**	**0.87**	**0.91**	**0.95**	**0.86**	**0.90**

INFOTECHSOFT
INNOVATIVE HEALTHCARE SOLUTIONS

OAEI 2009 Benchmark Track
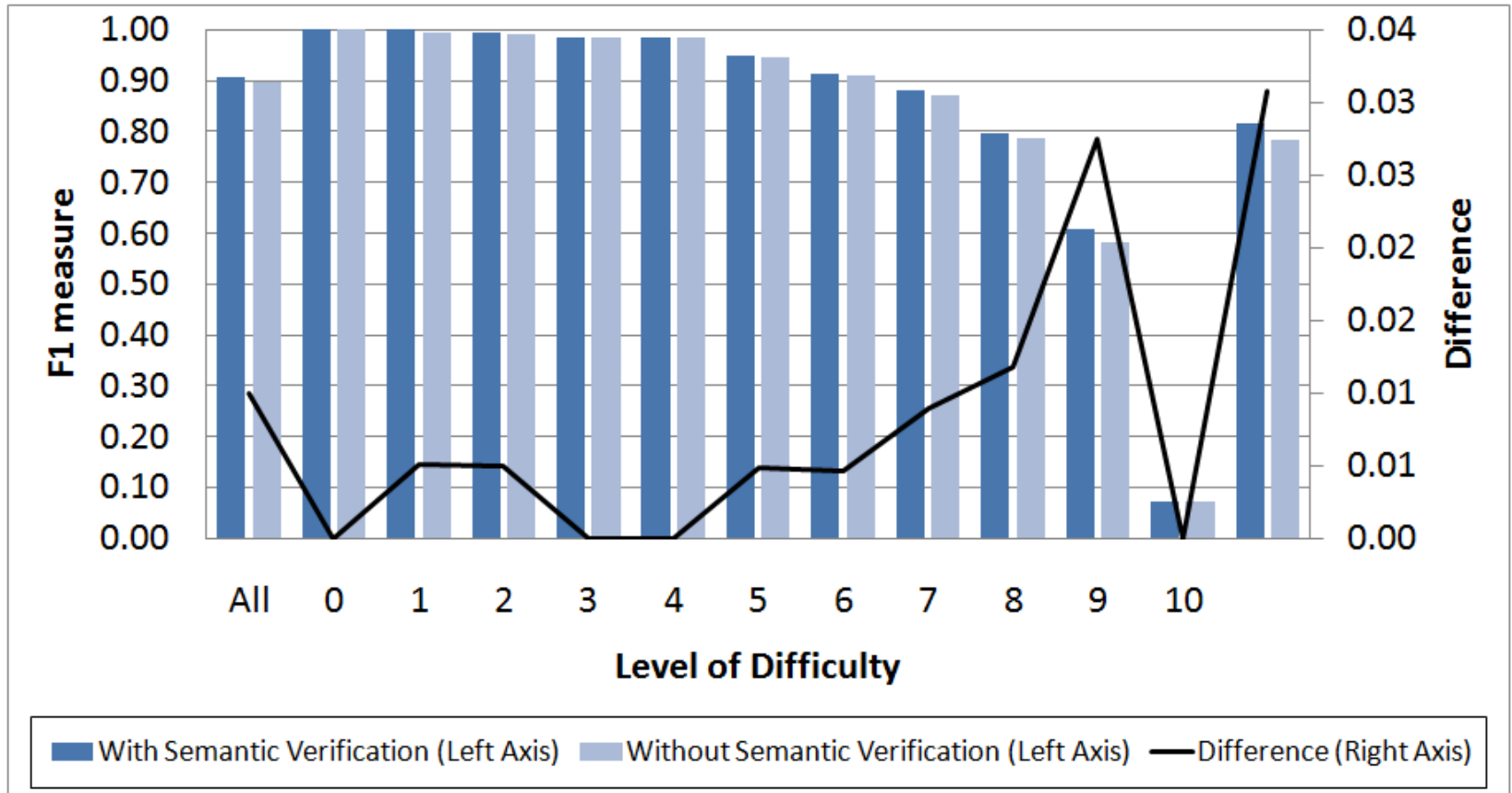
OAEI 2009 Anatomy Track

Configuration

- Lexical similarity calculations excludes the ids (names) of entities.
- UMLS is used as the thesaurus.

Improvements

- The time cost decreased from 4 hours to **5 minutes**.
- Significant increase in Recall+ leads to increase in overall F-measure.

Issues

- A problem with the UMLS Thesaurus Adapter implementation precluded the use of semantic distance.

Task	ASMOV 2009					ASMOV 2008				
	Time	P	R	R+	F	Time	P	R	R+	F
1	5 min	0.746	0.755	0.419	0.751	3h 50m	0.787	0.652	0.246	0.713
2		0.821	0.736		0.776		0.944	0.044		0.084
3		0.725	0.767	0.474	0.745		0.763	0.647	0.238	0.700
4		+0.034 0.759 → 0.792	-0.018 0.808 → 0.790		+0.009 0.782 → 0.791		+0.063 0.339 → 0.402	-0.004 0.258 → 0.254		+0.019 0.293 → 0.312

INFOTECHSOFT
INNOVATIVE HEALTHCARE SOLUTIONS

OAEI 2009 Results – Conference Track

- All 105 alignments evaluated
- Subsumption relations provided

Threshold	Precision	Recall	F-measure
0.2	0.58	0.40	0.47
0.5	0.22	0.03	0.04
0.7	0.05	0.01	0.01
Best: 0.23	0.68	0.38	0.47

INFOTECHSoFT
INNOVATIVE HEALTHCARE SOLUTIONS

OAEI 2009 Results – Directory Track

- Significant improvement in accuracy with respect to 2008
  - Due to improvement in codebase.

ASMOV 2009			ASMOV 2008		
Precision	Recall	F-Measure	Precision	Recall	F-Measure
0.60	0.65	0.63	0.64	0.12	0.20

OAEI 2009 Results –
Oriented matching track

- High accuracy in Benchmark track translates to subsumption mapping.

- *Equivalences heavily exploited*.

Test Set	Precision	Recall	F-measure
1xx	1.00	1.00	1.00
2xx	0.94	0.94	0.94
3xx	0.86	0.60	0.83
Average	0.94	0.90	0.93

INFOTECHSoft
INNOVATIVE HEALTHCARE SOLUTIONS

OAEI 2009 Results – Instance matching track

- High accuracy in instance matching
  - Slight decrease in 011-019 due to multiple possibilities.
- Availability of test enabled the enhancement of instance matching implementation.
- Need to improve scalability for large ontologies.

IIMB	Precision	Recall	F-measure
001-010	1.00	1.00	1.00
011-019	0.99	0.92	0.96
020-029	1.00	1.00	1.00
030-037	1.00	0.98	0.99
Overall	1.00	0.98	0.99
AKT / REXA	**0.52**	**0.32**	**0.39**

Observations & Future Work

- Both **accuracy** and **execution time** have been greatly improved in the last 2 years.

- Still need to improve scalability to very large ontologies and instance data:
  - Move from memory-based to datastore-based implementation.
  - Improve parallelization.

- Correct issues with UMLS thesaurus adapter.

- Complete implementation of user interface.

INFOTECHSoFT
INNOVATIVE HEALTHCARE SOLUTIONS

ASMOV: Results for OAEI 2009

Yves Reginald Jean-Mary
E. Patrick Shironoshita
Mansur R. Kabuka

INFOTECHSoft
INNOVATIVE HEALTHCARE SOLUTIONS

1201 Brickell Avenue, Suite 220
Miami, Florida, 33131 USA
Ph: +1 (305) 371 5111
Web: www.infotechsoft.com
Email: solutions@infotechsoft.com