



4th International Semantic Service Selection Contest

– Performance Evaluation of Semantic Service Matchmakers –

Matthias Klusch (DFKI, Germany)

Ulrich Küster, Birgitta König-Ries (University of Jena, Germany)

Alain Leger (France Telecom Research, France)

David Martin (SRI International, USA)

Massimo Paolucci (NTT DoCoMo Research Europe, Germany)

Abraham Bernstein (University of Zurich, Switzerland)



Track 1 OWL-S Service Matchmakers

1. iSeM 1.0 (DFKI, Germany)
2. OWLS-MX3 (DFKI, Germany)
3. SeMa² (TU Berlin, Germany)
4. OWLS-iMatcher (U Zurich, Switzerland)
5. SPARQLent (HP, Italy)
6. OWLS-SLR (lite) (Aristotle U of Thessaloniki, Greece)
7. XSSD (Beihang U, China)
8. EMMA (U Seville, Spain)



Track 2 SAWSDL Service Matchmakers

1. LOG4SWS.KOM (TU Darmstadt, Germany)
2. COV4SWS.KOM (TU Darmstadt, Germany)
3. iSeM 1.0 (DFKI, Germany)
4. SAWSDL-MX1 (DFKI, Germany)
5. URBE (Politecnico di Milano, Italy)
6. SAWSDL-iMatcher (U Zurich, Switzerland)



- **Service retrieval test collections**

- Track1: **OWLS-TC 4.0**

- 1083 services, 42 requests w/ binary & graded relevance sets, 38 ontologies
- Groundings in WSDL 1.1
- 160 services and 18 requests w/ preconditions + effects *each* in SWRL *and* PDDL 2
- @semwebcentral: 11026 downloads (Top 10, Nov 17 2010)

- Track2: **SAWSDL-TC 3.0**

- 1080 services, 42 requests w/ binary & graded relevance sets, 38 ontologies
- @semwebcentral: 387 downloads (Nov 17 2010)

- Development: DFKI (initial), U Jena, TU Darmstadt, Beihang U, U Thessaloniki, a.o.

- **Evaluation tool: SME² v2.1.1**

- Publicly available @semwebcentral.org since 2008, Developed @ DFKI
- Standard retrieval performance measures: **Macro-averaged recall/precision, Average precision, Q, nDCG (averaged cumulative gain); Average query response time** (Elapsed time per query execution)

Evaluation Tool SME² v2.1.1



Performance measures

- Macro-averaged recall/precision
- Average precision
- Q, nDCG [Graded relevance]
- Average query response time
- http-request analysis
- Precision@k, R-Precision (v2.2)

Easy-to-Use

- Load test collections +
Select matchmaker plugin(s) +
Configure evaluation
- Tailor your personal (printable)
evaluation result reports

Property	Value
service type	OWL-S 1.1
# of service off...	1083
# of requests	42
authors	Klusuch et al. (DFKI)
description	OWL-S test collection developed at DFKI, including graded rele...
htdocs root	testcollections\OWLS-TC4_PDDL\htdocs

Evaluation Tool SME² v2.1.1



Development

- Plug-in architecture
- Implemented in Java
- XML-based matchmaker plugin & TC configuration
- Embedded Jetty web server

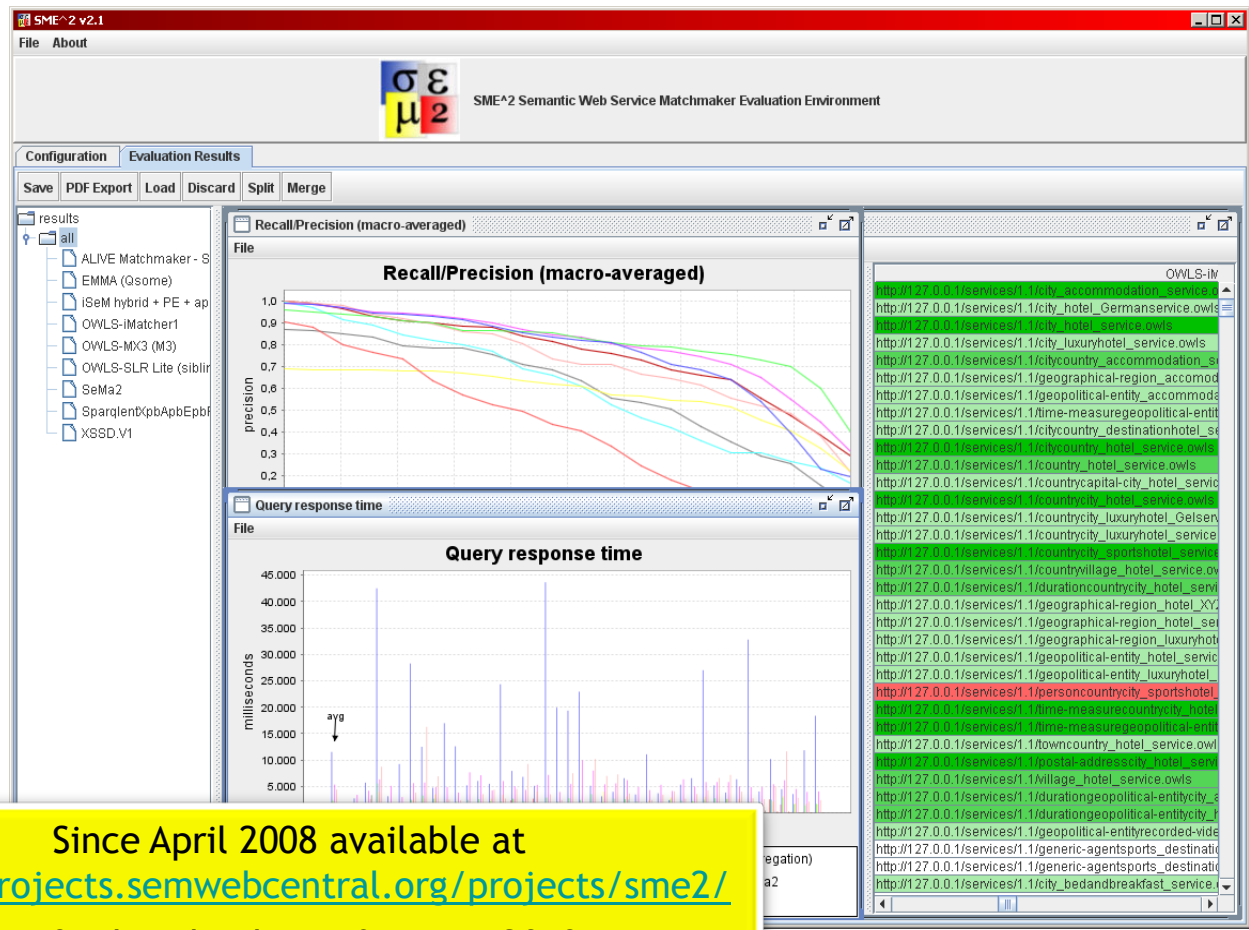
Developed @ DFKI:

Minko Dudev

Patrick Kapahnke

Josef Misutka

Martin Vasileski



Since April 2008 available at <http://projects.semwebcentral.org/projects/sme2/>
(1405 downloads as of Nov 17 2010)
Release of new version 2.2 in December 2010

Track 1: OWL-S Matchmakers in Brief



- **SeMa²**

- Selection: *Hybrid; Signature (I/O), Specification (P/E)*
 - **Logic-based match:** Logical I/O concept subsumption relation as numeric score
 - **Non-logic-based match:** String matching of I/O concept names (via `string.equal()`, `string.contains()`, AND-integrated with logic-based)
 - **Ranking:** Linear weighted aggregation of logical *and* string matching scores
- Dev: Nils Masuch (TU Berlin, Germany)

- **OWLS-SLR lite**

- Selection: *Hybrid; Signature (I/O), Non-functional parameters*
 - **Logic-based match:** Logical I/O concept subsumption relations as basis for ...
 - **Non-logic-based match:** ... Ontology-based structural match (edge distance, upward co-topic distance)
 - **Ranking:** Structural similarity
- Dev: Georgios Meditskos, Nick Bassiliades (U Thessaloniki, Greece)



- **OWLS-MX3**

- Selection: *Hybrid, adaptive; Signature (I/O)*
 - **Logic-based match** Logical I/O concept subsumption
 - **Non-logic-based match**: Text similarity of unfolded service signatures, Ontology-based structural match - Separated filters
 - **Adaptive (offline)**: SVM relevance classifier [TS = **10% OWLS-TC3**] for aggregation of (non-)logic-based matching degrees with subsequent **ranking**
- Dev: Matthias Klusch, Patrick Kapahnke (DFKI, Germany)

- **OWLS-iMatcher**

- Selection: *Syntactic; Signature (I/O)*
 - **Non-logic-based**: Vector-based text similarities of unfolded service signatures
 - **Ranking**: Text similarity
- Dev: Christoph Kiefer, Avi Bernstein (U Zurich, Switzerland)



- **SPARQLent**

- Selection: *Logic-Based*; *Signature (I/O)*, *Specification (P/E)*

- **Logic-based match**: P/E described in SPARQL, I/O concepts represented as additional constraints; I/O concept match via RDF entailment rules for RDF-encoded OWL

- **Ranking**: ?

- Dev: Marco Luca Sbodio (Hewlett-Packard EIC, Italy)



- **XSSD**

- Selection: *Hybrid; Signature (I/O), Service description tag*
 - **Logic-based match:** Logical I/O concept subsumption
 - **Non-logic-based match:** Text similarity match of service description tags
 - **Ranking:** Logic-based degree followed by text similarity-based ranking
- Dev: Jing Li, Dongjie Chu (U Beihang, China)

- **EMMA**

- Selection: *Logic-based semantic pre-filtering; Signature (I/O)*
 - **Logic-based pre-filtering:** SPARQL query in Jena RDF store using inference rules
 - **Hybrid match:** Based on pre-filtering using OWLS-MX3 (or other OWL-S MM plugins)
 - **Ranking:** Ranking procedure of internal OWLS-MX3 plugin
- Dev: José María García, David Ruiz, Antonio Ruiz-Cortés (U Seville, Spain)



- **iSeM 1.0**

- Selection: *Hybrid; Signature (I/O), Specification (P/E), Service description tag*
 - **Logic-based match** Logical I/O concept subsumption and information-theoretic valuation of approximated concept subsumption
 - **Non-logic-based match**: Text similarity of unfolded service signatures and service description tag, Ontology-based structural match - Separated filters
 - **Adaptive (offline)**: SVM relevance classifier with coherence-based weighting scheme [TS = **5% OWLS-TC4**] for aggregation of matching degrees with subsequent **ranking**
- Dev: Patrick Kapahnke, Matthias Klusch (DFKI, Germany)



Average Precision for Binary Relevance:

	1. iSeM 1.0	.922
	2. OWLS-MX3	.831
	3. EMMA	.803
	4. XSSD	.795
	5. SeMa ²	.741
	6. OWLS-iMatcher	.672
	7. SPARQLent	.612
	8. OWLS-SLR (lite)	.609

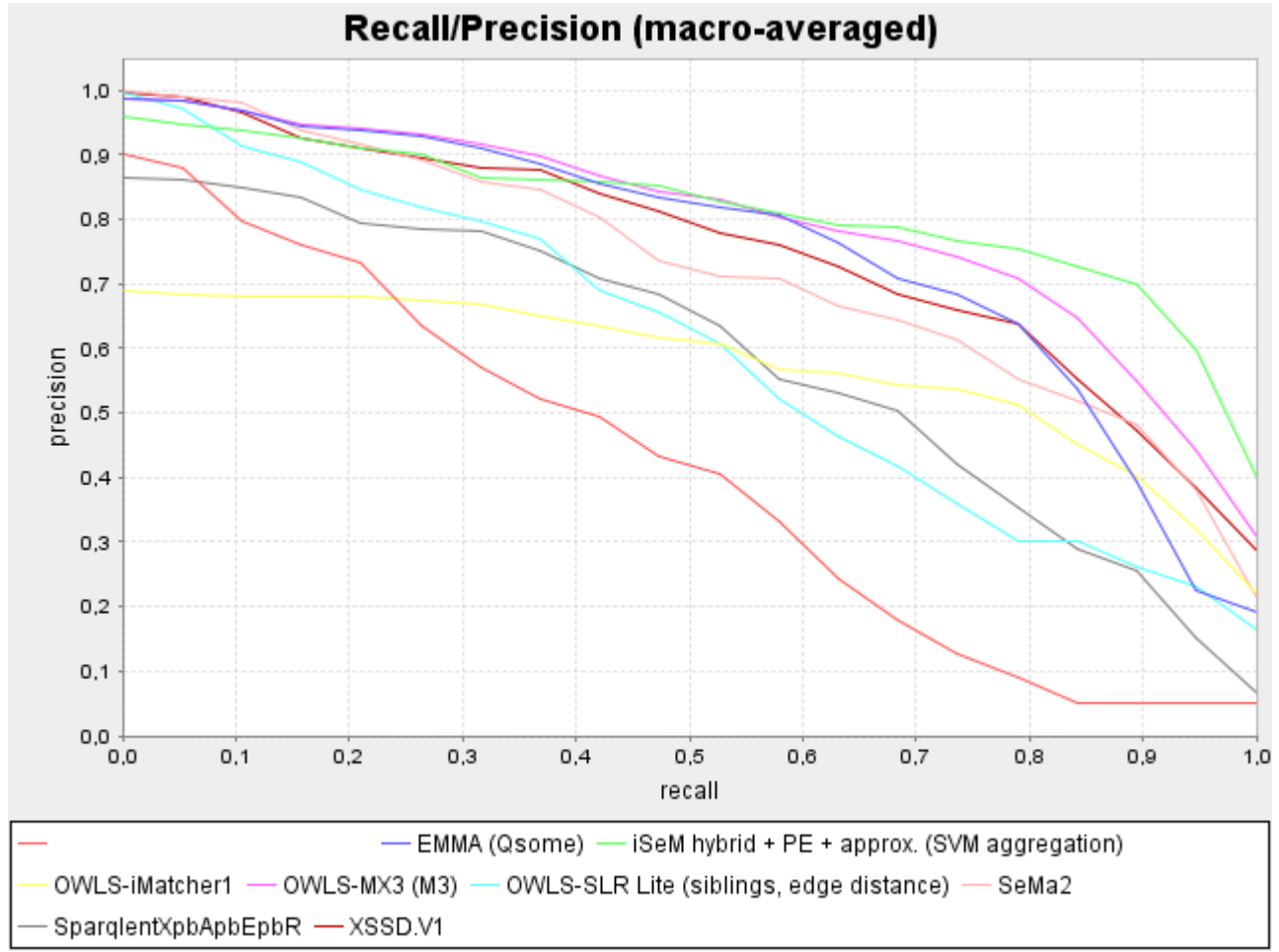
Please note: **For entries providing more than one variant, the variant with best AP was chosen.**

Detailed results *for all variants* can be found in the Appendix.

Performance Evaluation: MARP (Binary)



Macro-Averaged Recall/Precision for Binary Relevance:





SeMa²

- Structural comparison of SWRL rules + query containment (Abox)

SPARQLent

- SPARQL ASK [where] query containment (Abox)

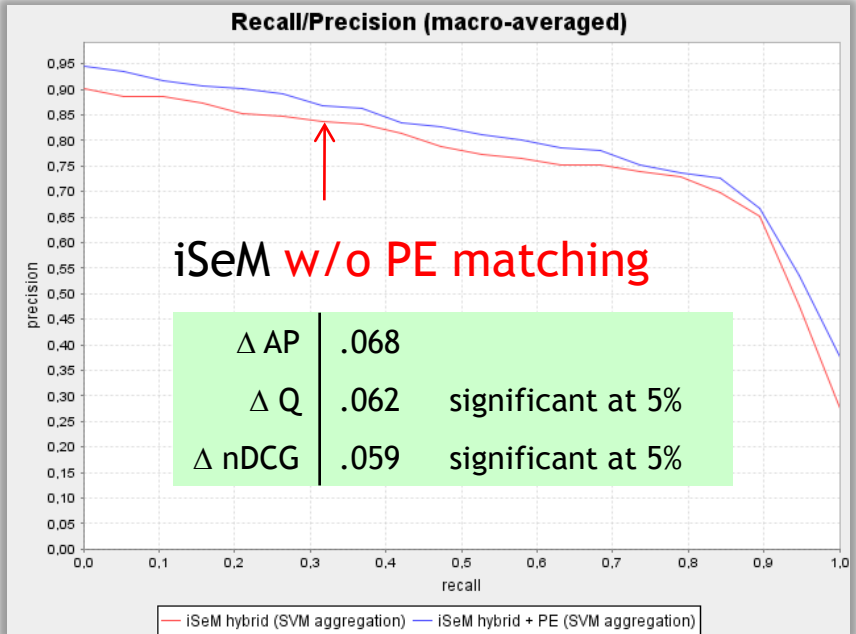
iSeM 1.0

- Approximated logical implication checking using \sqsubseteq -Subsumption between P/E in PDDL transformed to Prolog

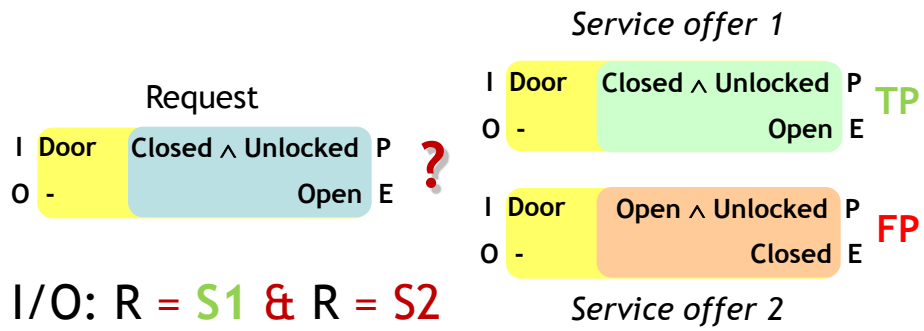
Problems

- No ABox in OWLS-TC4
 - Technical differences of SWRL syntax in OWL-S spec vs SWRL spec
 - No P/E described as SPARQL ASK [where] constraints
- = *iSeM only IOPE matchmaker used in S3 2010.*

Lesson Learned: Specification Matching



I/O matching pitfall example:



Rank:	1	2	3	4	
iSeM 1.0	TP	TP	FP	-	IOPE
OWLS-SLR	TP	FP	TP	FP	IO
SeMa2	TP	FP	FP	TP	IO(PE)
XSSD	TP	FP	TP	FP	IO


Problems

- Only 15% of OWLS-TC4 services have P/E: Low increase of precision w/ PE match
- „Solution“ of I/O pitfalls by „luck of random choice“ (S1 or S2) w/o PE matching
- Need: More services with complex P/E

Performance Evaluation: Response Time



AQRT - Average Query Response Time (in seconds):

	total	w/o HTTP response time	Vs. fastest variant [AQRT; diff AP]: diff rank AQRT
 1. XSSD	0.125	0.124	
2. OWLS-SLR lite	0.46	0.446	[0.169; - .023]: +1
3. SPARQLent	0.576	0.569	[0.201; - .423]: +2
4. OWLS-iMatcher	2.152	2.121	
5. iSeM 1.0	2.34	2.332	[1.828; - .097]: +1
6. SeMa ²	4.419	4.405	
7. OWLS-MX3	5.369	4.997	
8. EMMA	11.543*	11.089*	

* Caused by repeated plugin restart (see next slide)

Lesson Learned: Evaluation of Pre-Filtering



EMMA

- SPARQL-based pre-filtering of service offers for SME² plug-in of matchmakers

Idea

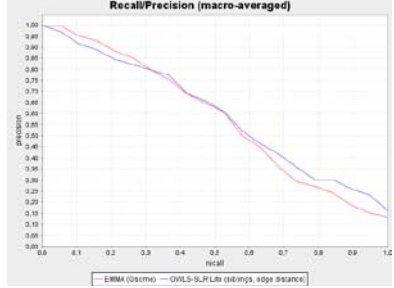
Fast pre-filtering techniques on top of heavy-weight matchmakers *to reduce AQRT while maintaining „good“ precision.*

Problem: „to reduce AQRT“

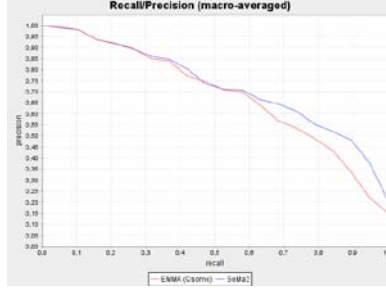
This kind of evaluation *not fully* possible with SME2 2.1.1 (but forthcoming SME2 2.2)

- Registration followed by querying phase: Requires registration of EMMA’s pre-filtering results (= service subsets) *for each query by means of plugin restart.*

EMMA + OWLS-SLR (lite)



EMMA + SeMa²



	AP	Q	nDCG
OWLS-SLR	.609	.57	.723
+ EMMA	.623	.588	.74
OWLS-MX3	.831	.834	.899
+ EMMA	.803	.815	.885
OWLS-iMatcher	.672	.671	.719
+ EMMA	.66	.666	.723
SeMa ²	.741	.73	.83
+ EMMA	.731	.728	.83

Significance at 5%: positive
negative

HTTP-Request Analysis



HTTP-Request Analysis of Query Phase

using SME2 with embedded Jetty Web server

	Average time (ms)	#http-requests min-max	average
1. XSSD	1	1-1	1
2. OWLS-SLR (lite)	14	1-13	1
3. SPARQLent	7	3-106	30
4. OWLS-iMatcher	31	1-1	1
5. iSeM 1.0	7*	2-8	3
6. SeMa ²	14	1-1	1
7. OWLS-MX3	372	1054 - 1085	1058
8. EMMA	454	390 - 2709	834

* HTTP-requests to *external* DIG API of approximative DL reasoner *not* included

Lesson Learned: Caching Strategies



Different caching strategies used by different matchmakers [exp. observation]

- Complete ontologies cached during service registration

Reduced #requests: Only queries but no ontologies

XSSD, OWLS-iMatcher, SeMa²

- Caching of self-contained (unfolded) concept definitions

Reduced #requests: *Some* queries require additional concept loading & classification

iSeM 1.0, OWLS-MX3

But: Services reloaded per query due to bug.

- Unknown strategy



OWLS-SLR lite, SPARQLent

- No caching

EMMA No caching for its SME² plug-ins possible



Precision for Graded Relevance:

	nDCG		Q
 1. OWLS-MX3	.899	 1. OWLS-MX3	.834
2. EMMA	.885	2. iSeM 1.0	.821
3. XSSD	.881	3. EMMA	.815
4. iSeM 1.0	.841	4. XSSD	.788
5. SeMa ²	.83	5. SeMa ²	.73
6. SPARQLent	.728	6. OWLS-iMatcher	.671
7. OWLS-SLR (lite)	.723	7. SPARQLent	.576
8. OWLS-iMatcher	.719	8. OWLS-SLR (lite)	.57



- **URBE**

- Selection: *Non-logic-based; Signature (I/O)*
 - **Non-logic-based match:** Bipartite graph-matching of service operations; Ontology-based structural I/O concept similarity (worst-case path length in given reference ontology); Text similarity (WordNet) for property-class and XSD data type matching
 - **Ranking:** Weighted aggregation of structural and text matching scores
- Dev: Pierluigi Plebani (Politecnico di Milano, Italy)

- **SAWSDL-MX1**

- Selection: *Hybrid; Signature (I/O)*
 - **Logic-based match:** Logical I/O concept subsumption
 - **Non-logic-based match:** Text similarity of unfolded concept definitions
 - **Ranking:** Logic-based sorted by text similarities
- Dev: Patrick Kapahnke, Matthias Klusch (DFKI, Germany)



• COV4SWS.KOM

- Selection: *Non-logic-based; Signature (I/O), Element names*
 - **Non-logic-based match:** Ontology-based semantic relatedness (Resnik, Lin); WordNet distance (fallback strategy for missing modelReference)
 - **Adaptive (offline):** Aggregated results using Ordinary Least Squares (OLS)
 - **Ranking:** Linear weighted average similarity of matched operations
- Dev: Stefan Schulte, Ulrich Lampe (TU Darmstadt, Germany)

• LOG4SWS.KOM

- Selection: *Hybrid; Signature (I/O), Element names*
 - **Logic-based match:** Logical I/O concept subsumption relation as numeric score
 - **Non-logic-based match:** Ontology-based structural I/O concept similarity (path length); WordNet distance (fallback strategy for missing modelReference)
 - **Adaptive (offline)/Ranking:** cf. COV4SWS.KOM
- Dev: Stefan Schulte, Ulrich Lampe (TU Darmstadt, Germany)



- **SAWSDL-iMatcher**

- Selection: *Non-logic-based; Signature (I/O)*
 - **Non-logic-based**: Vector-based text similarities of unfolded service signatures
 - **Ranking**: Text similarity
- Dev: Dengping Wei, Avi Bernstein (U Zurich, Switzerland)

- **iSeM 1.0 for SAWSDL**

- Selection: *Hybrid; Signature (I/O), Service name*
 - **Match**: cf. iSeM Track 1; but no P/E, service name instead of description tag
 - **Adaptive (offline)**: cf. iSeM Track 1; [TS = **5% SAWSDL-TC3**]
- Dev: Patrick Kapahnke, Matthias Klusch (DFKI, Germany)



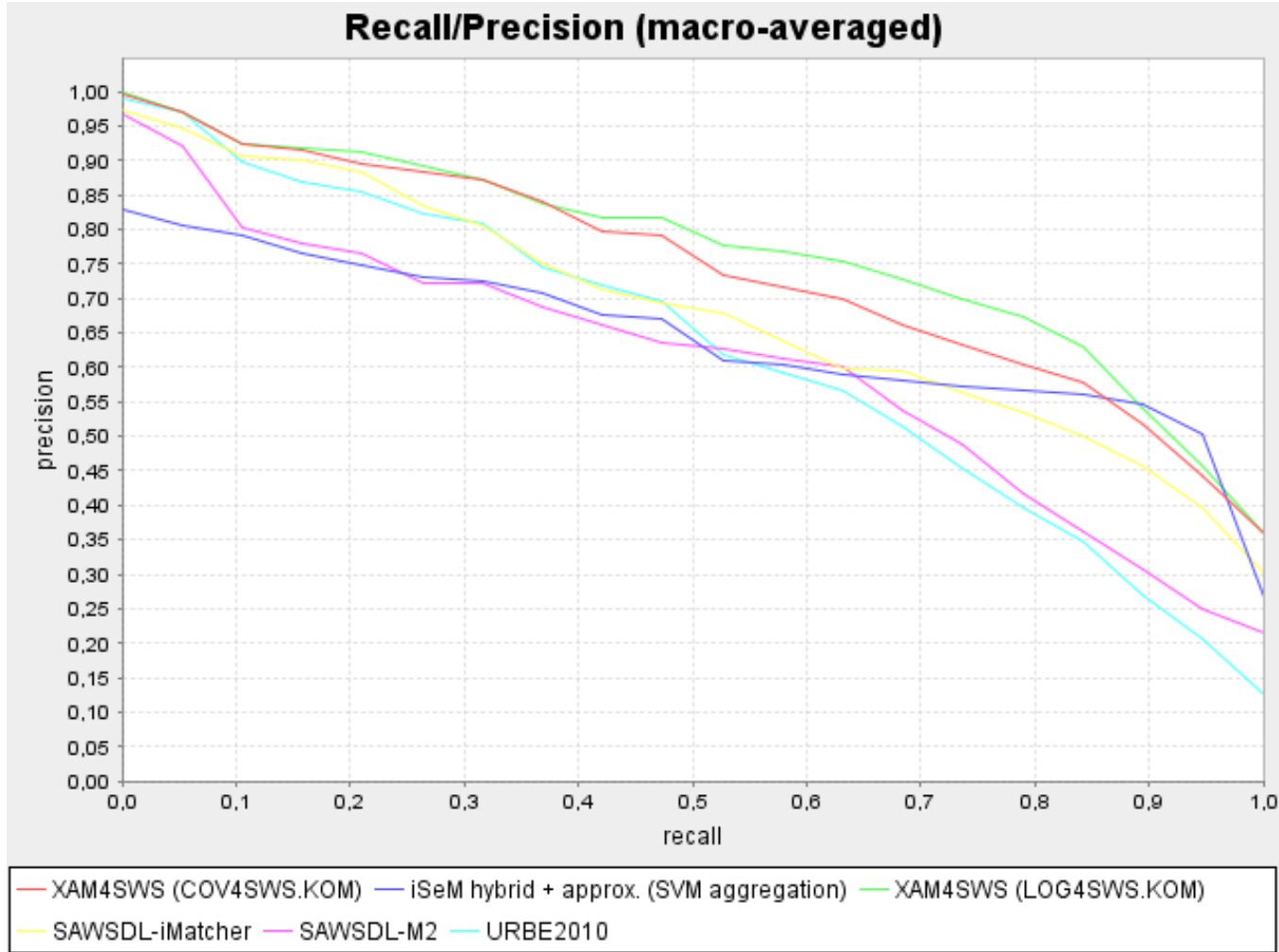
Average Precision for Binary Relevance:

	1. iSeM 1.0	.842
	2. LOG4SWS.KOM	.837
	3. COV4SWS.KOM	.823
	4. SAWSDL-iMatcher	.764
	5. URBE	.749
	6. SAWSDL-MX1	.747

Performance Evaluation: MARP (Binary)




Macro-Averaged Recall/Precision for Binary Relevance:



Performance Evaluation: Response Time



AQRT - Average Query Response Time (in seconds):

	total	w/o HTTP response time	Vs. fastest variant [AQRT; diff AP]: diff rank AQRT
 1. LOG4SWS.KOM	0.241	0.241	
2. COV4SWS.KOM	0.301	0.301	
3. SAWSDL-iMatcher	1.787	1.787	
4. SAWSDL-MX1	3.859	3.853	
5. iSeM 1.0	10.662	10.655	[1.584; - .018]: +2
6. URBE	40.01	39.941	

HTTP-Request Analysis



HTTP-Request Analysis of Query Phase

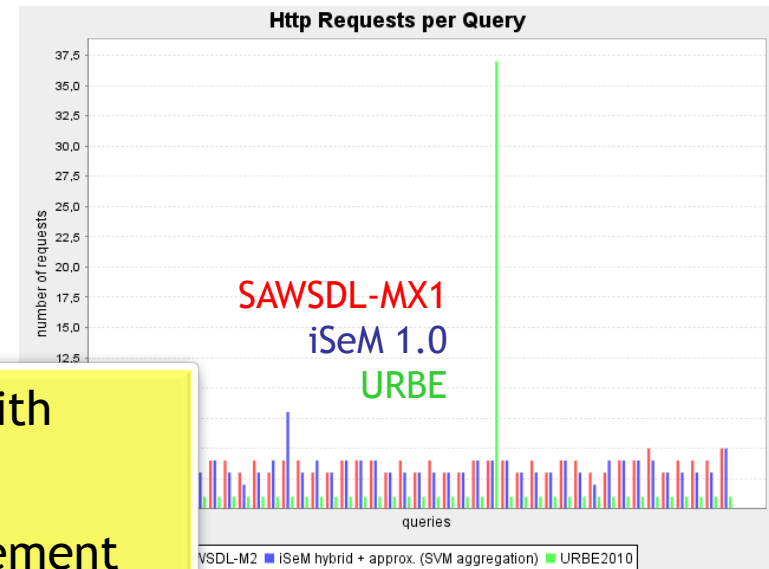
using SME2 with embedded Jetty Web server

	Average time (ms)	#http-requests min-max	average
1. LOG4SWS.KOM	0	0-2	0
2. COV4SWS.KOM	0	0-0	0
3. SAWSDL-iMatcher	0	0-0	0
4. SAWSDL-MX1	6	3-5	3
5. iSeM 1.0	7	2-8	3
6. URBE	69	1-37	1

Lesson Learned: Caching Strategies (2)





- Complete ontologies cached before service registration
 - **LOG4SWS.KOM, COV4SWS.KOM:**
Ontologies loaded and classified after plug-in initialization → global ontology assumed
 - **SAWSDL-iMatcher:**
Everything seems to be cached „out of the box“
- Caching of unfolded concepts
 - **SAWSDL-MX1, iSeM 1.0**
- Unknown strategy
 - **URBE**



SME² **2.2** (December 2010) with
Extended plug-in interface
for more detailed AQR measurement



Precision for Graded Relevance:

	nDCG		Q
 1. LOG4SWS.KOM	.896	 1. LOG4SWS.KOM	.851
2. COV4SWS.KOM	.884	2. COV4SWS.KOM	.825
3. SAWSDL-iMatcher	.855	3. SAWSDL-iMatcher	.784
4. URBE	.85	4. URBE	.777
5. SAWSDL-MX1	.839	5. SAWSDL-MX1	.767
6. iSeM 1.0	.803	6. iSeM 1.0	.762



- Principle

- **Specific Domain Test Collection:** Jena Geography Dataset JGD
 - 201 geoservices, 10 queries, graded relevance judgements
 - Semantic annotation of JGD services **in different formats provided by participants: Variants of the same JGD collection.**
 - Cross-evaluation: Comparative performance evaluation of entries over each JGD collection variant with graded relevance-based performance measures Q, nDCG and AQRT using the SME² tool.
 - **Initial cross-evaluation in 2009** organized by Ulrich Küster
- **No new entries in 2010 = No new results. See S3 in 2009.**
 - Submissions welcome at any time. Contact: Birgitta.Koenig-Ries@uni-jena.de



- **New** entries were indicated to be in preparation for 2011
- **Open call for location** of final presentation/discussion of S3 2011 results
- **New release of test collections:** OWLS-TC**5** and SAWSDL-TC**4**

But what about WSML-TC?

- **New release of evaluation tool SME² version 2.2** (December 2010)
 - ✓ Extended matchmaker plugin interface for even more detailed evaluation configuration
 - ✓ More performance measures: Precision@k, R-Precision
 - ✓ More matchmaker plugins included
 - ✓ Improved usability: Error handling, more configuration details

Contact: Patrick Kapahnke (DFKI) patrick.kapahnke@dfki.de



... Thanks for your attention !

Any **QUESTIONS?**



... Next year with ***your brand new ultra mega beat'em all matchmaker !?*** 😊

Appendix: Complete Results for Track 1



AP		Q			
1.	iSeM all	.922	1.	OWLS-MX3	.834
2.	iSeM no approx.	.893	2.	iSeM all	.821
3.	OWLS-MX3	.831	3.	EMMA Qsome	.815
4.	iSeM no approx., no PE	.825	4.	iSeM no approx.	.811
5.	EMMA Qsome	.803	5.	XSSD	.788
6.	XSSD	.795	6.	iSeM no approx., no PE	.749
7.	SeMa2	.741	7.	SeMa2	.73
8.	OWLS-iMatcher	.672	8.	OWLS-iMatcher	.671
9.	SPARQLent XpbApbEpbR	.612	9.	EMMA Qall	.67
10.	EMMA Qall	.61	10.	SPARQLent XpbApbEpbR	.576
11.	OWLS-slr (lite) siblings, edge distance	.609	11.	OWLS-slr (lite) siblings, edge distance	.57
12.	OWLS-slr (lite) siblings, cotopic distance	.586	12.	OWLS-slr (lite) siblings, cotopic distance	.553
13.	SPARQLent XpbApbEpbE	.495	13.	SPARQLent XpbApbEpbE	.462
	SPARQLent Xp_Ap_Ep_R	.495		SPARQLent Xp_Ap_Ep_R	.462
14.	OWLS-slr (lite) edge distance	.428	14.	OWLS-slr (lite) edge distance	.402
15.	SPARQLent Xp_Ap_Ep_E	.391	15.	ALIVE Composite 2	.399
16.	SPARQLent X__A__E__R	.262	16.	SPARQLent Xp_Ap_Ep_E	.363
17.	SPARQLent X__A__E__E	.189	17.	SPARQLent X__A__E__R	.24
			18.	SPARQLent X__A__E__E	.172

Appendix: Complete Results for Track 1



	nDC G		AQRT- http (ms)
1.	OWLS-MX3	.899	
2.	EMMA Qsome	.885	
3.	XSSD	.881	
4.	iSeM no approx.	.844	
5.	iSeM all	.841	
6.	SeMa2	.83	
7.	EMMA Qall	.802	
8.	iSeM no approx., no PE	.785	
9.	SPARQLent XpbApbEpbR	.728	
10.	OWLS-slr (lite) siblings, edge distance	.723	
11.	OWLS-iMatcher	.719	
12.	OWLS-slr (lite) siblings, cotopic distance	.712	
13.	SPARQLent Xp_Ap_Ep_R	.639	
14.	SPARQLent XpbApbEpbE	.625	
19.	OWLS-slr (lite) edge distance	.591	
20.	SPARQLent Xp_Ap_Ep_E	.556	
21.	SPARQLent X_A_E_R	.452	
22.	SPARQLent X_A_E_E	.366	
1.	XSSD		124
2.	OWLS-slr (lite) siblings, cotopic distance		169
3.	SPARQLent X_A_E_E		196
4.	SPARQLent X_A_E_R		277
8.	SPARQLent Xp_Ap_Ep_E		317
9.	ALIVE Composite 2		339
10.	OWLS-slr (lite) edge distance		379
11.	SPARQLent XpbApbEpbE		400
12.	OWLS-slr (lite) siblings, edge distance		446
13.	SPARQLent Xp_Ap_Ep_R		456
14.	SPARQLent XpbApbEpbR		569
15.	iSeM no approx., no PE		1828
16.	iSeM no approx.		1915
17.	OWLS-iMatcher		2121
18.	iSeM all		2332
19.	SeMa2		4405
20.	OWLS-MX3		4997
21.	EMMA Qsome		11089
22.	EMMA Qall		31814

Appendix: Complete Results for Track 2



		AP
1.	iSeM w. approx.	.842
2.	LOG4SWS.KOM	.837
3.	iSeM no approx.	.824
4.	COV4SWS.KOM	.823
5.	SAWSDL-iMatcher	.764
6.	URBE	.749
7.	SAWSDL-MX1	.747

		nDC G
1.	LOG4SWS.KOM	.896
2.	COV4SWS.KOM	.884
3.	SAWSDL-iMatcher	.855
4.	URBE	.85
5.	iSeM no approx.	.84
6.	SAWSDL-MX1	.839
7.	iSeM w. approx.	.803

		Q
1.	LOG4SWS.KOM	.851
2.	COV4SWS.KOM	.825
3.	iSeM no approx.	.799
4.	SAWSDL-iMatcher	.784
5.	URBE	.777
6.	SAWSDL-MX1	.767
7.	iSeM w. approx.	.762

		AQRT- http (ms)
1.	LOG4SWS.KOM	241
2.	COV4SWS.KOM	301
3.	iSeM no approx.	1584
4.	SAWSDL-iMatcher	1787
5.	SAWSDL-MX1	3853
6.	iSeM w. approx.	10655
7.	URBE	39941



XSSD: Jing Li li390603@gmail.com

SeMa²: Nils Masuch nils.masuch@dai-labor.de

OWLS-iMatcher2: Dengping Wei dengping@ifi.uzh.ch

SPARQLent: Marco Sbodio marco.sbodio@gmail.com

OWLS-SLR lite: Nick Bassiliadis nbassili@csd.auth.gr

iSeM 1.0, OWLS-MX3: Matthias Klusch klusch@dfki.de

EMMA: Jose Maria Garcia josemgarcia@us.es

OWL-S

SAWSDL

LOG4SWS, COM4SWS: Stefan Schulte schulte@kom.tu-darmstadt.de

iSeM 1.0, SAWSDL-MX1, SAWSDL-MX2: Matthias Klusch klusch@dfki.de

URBE: Pier Luigi Plebani plebani@elet.polimi.it

SAWSDL-iMatcher3: Dengping Wei dengping@ifi.uzh.ch