5th International Semantic Service Selection Contest

– Performance Evaluation of Semantic Service Matchmakers –

Matthias Klusch (DFKI, Germany)
Birgitta König-Ries (University of Jena, Germany)
David Martin (Apple Inc., USA)
Massimo Paolucci (NTT DoCoMo Research Europe, Germany)
Abraham Bernstein (University of Zurich, Switzerland)
Ulrich Lampe (TU Darmstad, Germany)
Stefan Schulte (TU Vienna, Austria)
Terry Payne (U Liverpool, UK)

Summary Report for 2012 Edition. April 12, 2013.

Outline

- Semantic Selection
- Evaluation Framework
- Evaluation Results & Lessons Learned

- Klusch, M. (2012): Overview of the S3 Contest: Performance Evaluation of Semantic Service Matchmakers. In: Blake, M.B.; Cabral, L.; König-Ries, B.; Küster, U.; Martin, D. (Eds.): Semantic Web Services – Advancement through Evaluation; Chapter 2; Springer.

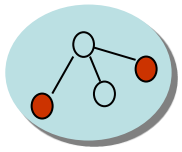
Semantic Service Selection

- **Service discovery**
  - Centralized in Web service registries (W3C SOA) or with search engines
  - Decentralized in P2P service networks
- **Semantic selection** (aka semantic matchmaking)

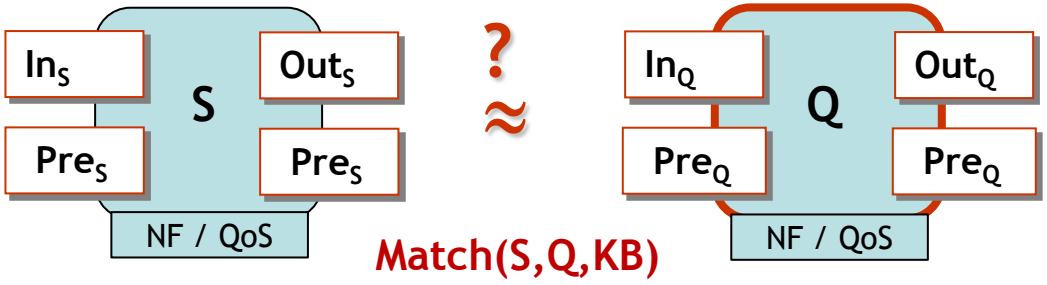
  (1) Semantic **matching** of registered services S with desired service description Q

  (2) Relevance **ranking** of S (answer set) for final selection of services by user

  NO brokerage (composition, publish/subscribe negotiation, execution handling)



Shared ontology KB
for semantic annotation:
In/Out, Non-func params



Evaluation of Semantic Selection

(1) Support of service description languages

- OWL-S, WSML, SAWSDL, SA-REST, USDL, hRESTS
- Agnostic: Semantic-preserving transformations, metamodels

(2) Support of composition

- Pruning of composition search space by selection
- Iterative selection for forward/backward chaining

(3) Security (data privacy)

(4) Usability and configuration efforts

(5) Performance of selection

Correctness: *Precision, Recall, MAP, F1, etc.*

Speed: *Average query response time*

Other Evaluation Initiatives

- Comparison with other service evaluation initatives:

	SWS Challenge	WS Challenge	S3 Contest
Scope	composition (given scenarios)		**discovery**
Performance	-	runtime	IR measures, runtime
Usability/effort	adaptation effort	-	description effort (cross-eval track in 2009)
Correctness	Alg. correctness		Retrieval correctness

- Blake, M.B.; Cabral, L.; Koenig-Ries, B.; Küster, U.; Martin, D. (2012): Introduction. In: (dito), Semantic Web Services – Advancement through Evaluation; Chapter 1. Springer.
- Kuester, U.; Koenig-Ries, B.; Klusch, M. (2009): Evaluating Semantic Web Service Technologies: Criteria, Approaches and Challenges. In: Progressive Concepts for Semantic Web Evolution: Applications and Developments; Advances in Semantic Web Information Systems series. IGI Global.

Participants of S3 Contest 2012

Track 1 OWL-S Service Matchmakers

1. iSeM 1.1 (DFKI, Germany)
2. OWLS-MX3 (DFKI, Germany)
3. SeMa^2 v2 (TU Berlin, Germany)
4. Nuwa-OWLS (URJC Madrid, Spain)
5. OWLS-iMatcher (U Zurich, Switzerland)
6. SPARQLent (HP, Italy)
7. OWLS-SLR (Aristotle U of Thessaloniki, Greece)
8. XSSD (Beihang U, China)
9. EMMA (U Seville, Spain)
10. iSeM-TSM1 (Payame Noor U, Iran / DFKI)

Service Selection By Participants in Brief

- **iSeM 1.1** [analog iSeM 1.1 for SAWSDL]
  - Selection: *Hybrid; Signature (I/O), Specification (P/E), Service description tag*
    - Logic-based matching
      - Logical I/O concept subsumption + information-theoretic valuation of approximated logical I/O concept subsumption
      - Logical P/E plugin checking (theta-subsumption)
    - Non-logic-based matching
      - Text similarity of unfolded service signatures (I/O) and service description tags,
      - Ontology-based structural I/O 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), 2010

Service Selection By Participants in Brief

- **iSeM-TSM1**

  Selection: *Non-logic-based; Signature (I/O), Service description tag*

  - Non-logic-based matching
    - Text similarity of unfolded service signatures (I/O) and service description tags,
    - Ontology-based structural I/O match – Separated filters

  -Dev: Elyad Alaei, Ahmad Faraahi (Payame Noor U, Iran),
  Mohammad-Reza Feizi-Derakhshi (U Tabriz, Iran)

Service Selection By Participants in Brief

Nuwa-OWLS

[analog Nuwa-SAWSDL]

Selection: *Hybrid; Signature (I/O), Service description text*

- Logic-based matching: Logical concept subsumption
- Non-logic-based matching:
  - Ontology-based (WordNet) structural I/O concept label match
    (I/O concept label e.g. http://foo/bar.owl#door --> label: “door“)
  - Text similarity (Cosine TF-IDF) of keywords extracted from:
    Semantic I/O concept URI fragments, labels
    Service textual description
    Service name and service URI fragment
- Ranking: Weighted sum of results of both matching types

Dev: Zije Cong, Alberto Fernandez (URJC Madrid, Spain)

Service Selection By Participants in Brief

- SeMa^2 v2
  - Selection: *Hybrid; Signature (I/O), Specification (P/E)*
    - Logic-based matching:
      - Logical I/O concept subsumption relation as numeric score
      - Logical P/E (SWRL rule) plugin matching with theta-subsumption (no ABox) + separated precondition checking over given ABox
    - Non-logic-based matching:
      - String matching of I/O concept names (string.equal() / .contains())
      - Structural and taxonomic matching of variable types in SWRL (P/E) rules
    - Ranking: Linear weighted aggregation of all matching scores
  - Dev: Nils Masuch (TU Berlin, Germany)

Service Selection By Participants in Brief

- **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-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)

Service Selection By Participants in Brief

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

Service Selection By Participants in Brief

- **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)

Participants of S3 Contest 2012

Track 2 SAWSDL Service Matchmakers

1. LOG4SWS.KOM (TU Darmstadt, Germany)
2. COV4SWS.KOM (TU Darmstadt, Germany)
3. iSeM 1.1 (DFKI, Germany)
4. SAWSDL-MX1 (DFKI, Germany)
5. URBE (Politecnico di Milano, Italy)
6. SAWSDL-iMatcher (U Zurich, Switzerland)
7. Nuwa-SAWSDL (URJC Madrid, Spain)

Service Selection By Participants

- **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): Aggregated results using Ordinary Least Squares (OLS)
    - Ranking: Linear weighted average similarity of matched operations
- **COV4SWS.KOM**
  - Selection: *Non-logic-based (see LOG4SWS.KOM); Signature (I/O), Element names*

Dev: Stefan Schulte, Ulrich Lampe (TU Darmstadt, Germany)

Service Selection By Participants

- **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)

Track 2: SAWSDL Matchmakers in Brief

- **SAWSDL-iMatcher**
  - Selection: *Non-logic-based; Signature (I/O)*
    - Non-logic-based: Vector-based text similarities of unfolded service signatures
    - Ranking: Text similarity

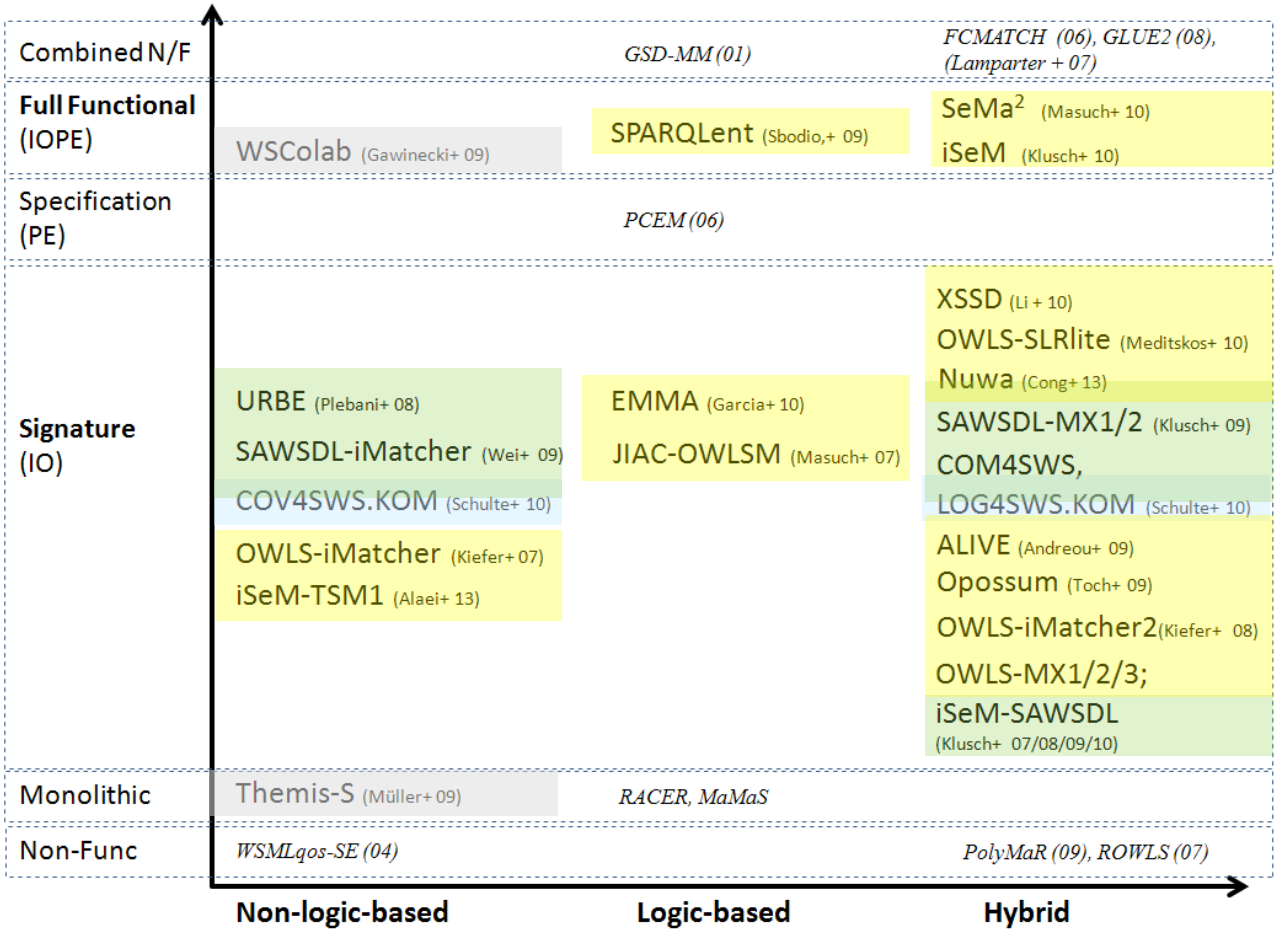
Dev: Dengping Wei, Avi Bernstein (U Zurich, Switzerland)

- **iSeM 1.1 for SAWSDL**
  - Selection: *Hybrid; Signature (I/O), Service name*
    - Match: [cf. iSeM 1.1 for OWL-S, slide 7]

      but no P/E match; uses service name instead of description tag

- **Nuwa-OWLS**
  - Selection: *Hybrid; Signature (I/O), Service description text*
    - Match: [cf. Nuwa-OWLS, slide 9]

Classification



- **Tracks** [#participants]
  - OWL-S [11]
  - SAWSDL [7]
  - hREST/WSML-lite [2]
  - Others [3]

Outline

- Semantic Selection
- **Evaluation Framework**
- Evaluation Results & Lessons Learned

Framework Components in Brief

- **Service retrieval test collections**
  - Track1: **OWLS-TC 4.0**
    - 1.083 services, 42 requests w/ binary & graded relevance sets, 38 ontologies
    - Groundings in WSDL 1.1, 7 domains (Communication, Economy, Education, Food, Medical Care, Travel, Military)
    - 160 services and 18 requests w/ preconditions + effects *each* in SWRL *and* PDDL2
    - @semwebcentral: 14.339 downloads (in Top-10 as of March 7, 2012)
  - Track2: **SAWSDL-TC 3.0**
    - 1.080 services, 42 requests w/ binary & graded relevance sets, 38 ontologies
    - @semwebcentral: 760 downloads (March 7 2012)
  - Track3: **hRESTS 1.0**
  - Development: DFKI, U Jena, TU Darmstadt, U Beihang, U Thessaloniki, a.o.
- **Evaluation tool: SME^2 v2.2**
  - Open source publicly available @semwebcentral.org since 2008: 2.816 downloads (March 7 2012)
  - Plugin interface for contested matchmakers; standard retrieval performance measures

Service Relevance

- **Relevance assessment of services**
  - **Binary relevance** value: Relevant (**1**), or Irrelevant (**0**)
  - Standard NTCIR **4-graded relevance** scale used @TREC:

Relevance Grade	Gain value	Intuitive Meaning of Relevance Grade
Highly relevant	**3**	„Perfectly satisfies request **(S ≡ R)**“
Relevant	**2**	„Relevant to request with some conditions of its conditions not satisfied **(S ⊂ R)**“
Partially relevant	**1**	„Helpful to satisfy request by providing related information **(S ∩ R ≠ ∅, S ⊄ R)**“
Not Relevant	**0**	„Not relevant at all **(S ∩ R = ∅)**“

  - Relevance sets defined by union average pooling of assessments:
    - >> Service relevant if judged relevant by *at least one* user (TREC).
    - >> Services not yet rated, or not in relevance set are irrelevant.

Evaluation Tool SME² v2.2

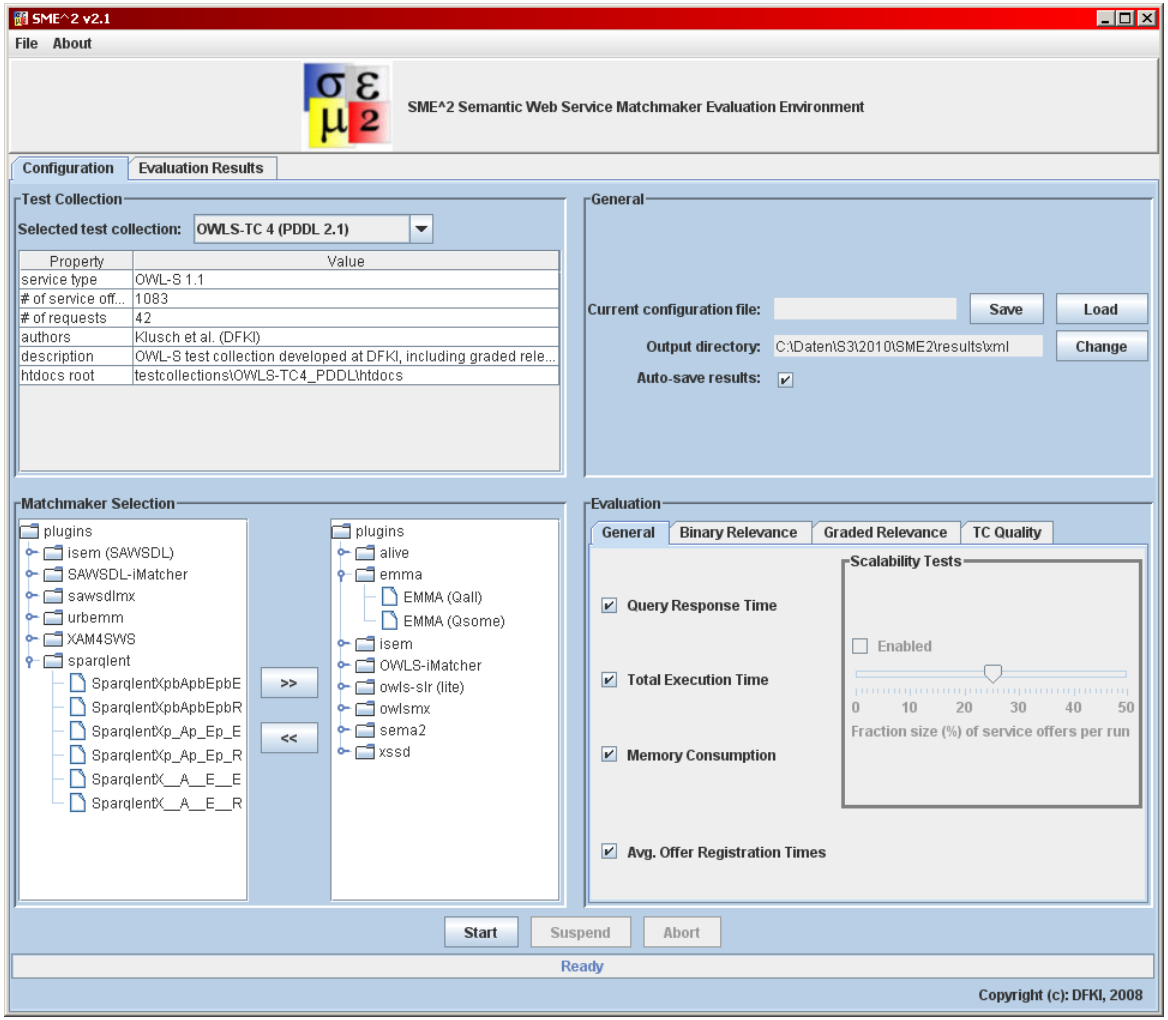
Performance measures

- Macro-averaged precision@recall MAP
- Average precision AP
- Q, nDCG [Graded relevance]
- Average query response time AQRT (elapsed time per query execution)
- http-request analysis
- Precision@k, R-Precision

Easy handling

→ Load test collections +
Select matchmaker plugin(s) +
Configure evaluation run

→ Tailor your (printable) report of evaluation results



Evaluation Tool SME² v2.2

Implementation

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

Developed @ DFKI:

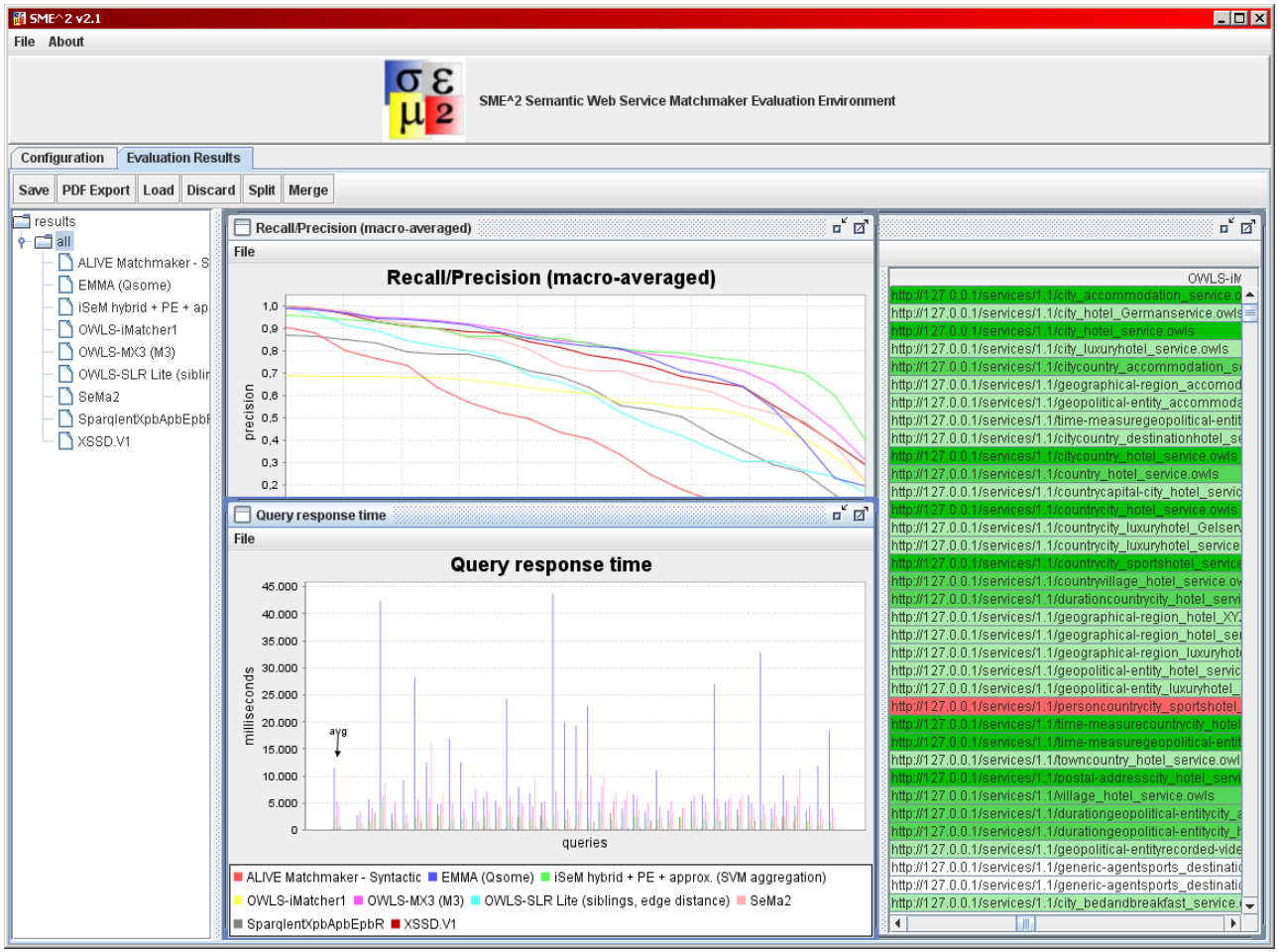
Minko Dudev

Patrick Kapahnke

Josef Misutka

Martin Vasileski

Matthias Klusch



Outline

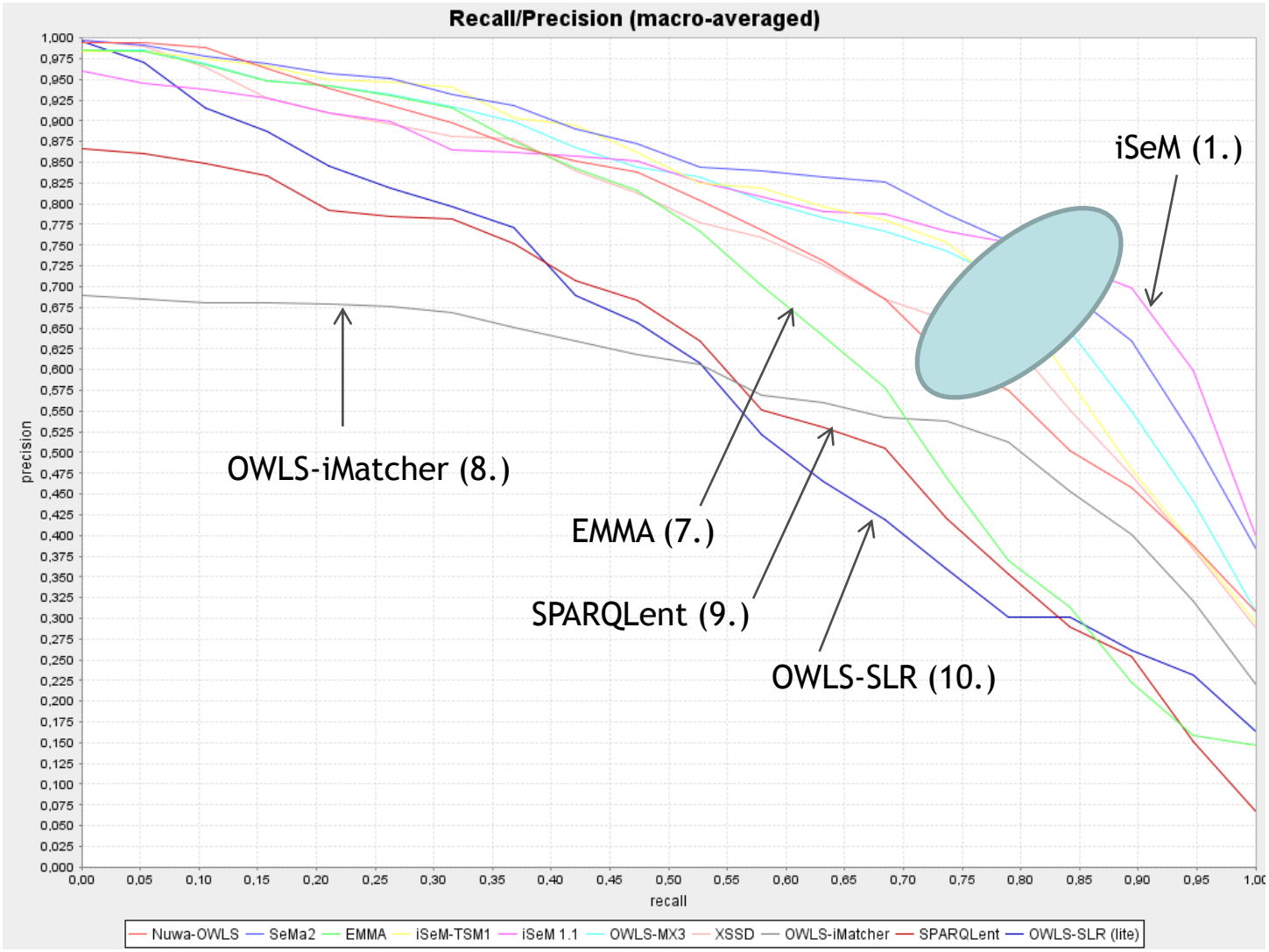
- Semantic Selection
- Evaluation Framework
- **Evaluation Results & Lessons Learned**

OWL-S Selection: Average Precision (Bin)

	Matchmaker	AP		Dev.
1.	iSeM 1.1	.922		DFKI, Germany
2.	SeMa² v2	.877	v1: .741	TU Berlin, Germany
3.	iSeM-TSM1	.861		Payame Noor U, Iran / DFKI
4.	Nuwa-OWLS	.853	-	URJC Madrid, Spain
5.	OWLS-MX3	.831		DFKI, Germany
6.	XSSD	.795		U Beijing, PR China
7.	EMMA	.762		U Seville, Spain
8.	OWLS-iMatcher	.672		U Zurich, Switzerland
9.	SPARQLent	.612		HP, Italy
10.	OWLS-SLR (lite)	.609		Aristotle U, Greece

Please note: **For matchmakers with more than one variant, the one with best AP is shown.**

OWL-S Selection: Macro-Averaged Precision for Binary Relevance



OWL-S Selection: Average Precision (Grad)

Matchmaker	AP:nDCG
1. SeMa² v2	.927
2. iSeM-TSM1	.916
3. Nuwa-OWLS	.911
4. OWLS-MX3	.899
5. XSSD	.881
6. EMMA	.87
7. iSeM 1.1	.841
8. SPARQLent	.728
9. OWLS-SLR (lite)	.723
10. OWLS-iMatcher	.719

Matchmaker	AP:Q
1. SeMa² v2	.883
2. iSeM-TSM1	.855
3. Nuwa-OWLS	.846
4. OWLS-MX3	.834
5. iSeM 1.1	.821
6. EMMA	.7884
7. XSSD	.7881
8. OWLS-iMatcher	.671
9. SPARQLent	.576
10. OWLS-SLR (lite)	.57

Lesson Learned: Specification Matching

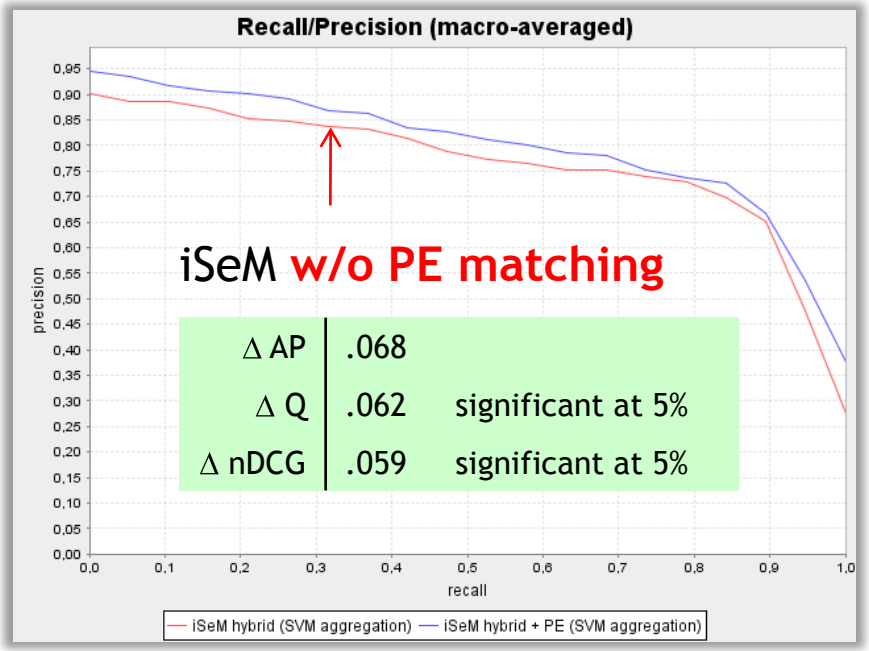
Only very few matchmakers perform specification (P/E) matching

- SeMa^2 v2 (TU Berlin)
  - ➢ Structural + logical plugin (no Abox) + precondition satisfaction (ABox)
- SPARQLent (HP Italy)
  - ➢ SPARQL ASK [where] query containment (ABox)
- iSeM 1.1 (DFKI)
  - ➢ Logical plugin (no Abox)

Current problems:

- Test collection OWLS-TC has no ABoxes
- P/E in PDDL and SWRL: SWRL syntax in OWL-S spec and SWRL spec differ

Lesson Learned: Specification Matching



Example:

Service request R

I	Door	Closed ∧ Unlocked	P
O	-	Open	E

?

Service offer S1

I	Door	Closed ∧ Unlocked	P	TP
O	-	Open	E	

I	Door	Open ∧ Unlocked	P	FP
O	-	Closed	E	

Service offer S2

I/O match: R = S1 = S2

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/Es. Low increase of precision with P/E match.
- „Solution“ of I/O pitfalls by „luck of random choice“ (S1 or S2) w/o PE matching
- Collections require more services with (complex) P/E descriptions

OWL-S Selection: Average Response Time

Matchmaker	AQRT (s)	w/o http	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 1.1	2.34	2.332	[1.828; - .097]: +1
6. iSeM-TSM1	4.447	4.437	
7. OWLS-MX3	5.369	4.997	
8. SeMa² v2	5.084 (v1: 4.4)	5.063	
9. EMMA	9.644	9.335	
10. Nuwa-OWLS	18.356	18.334	

more http traffic during query phase for OWLS-MX3: not everything's cached → cf. next slide

Repeated restart of plugin!

Lessons Learned: Caching Strategies

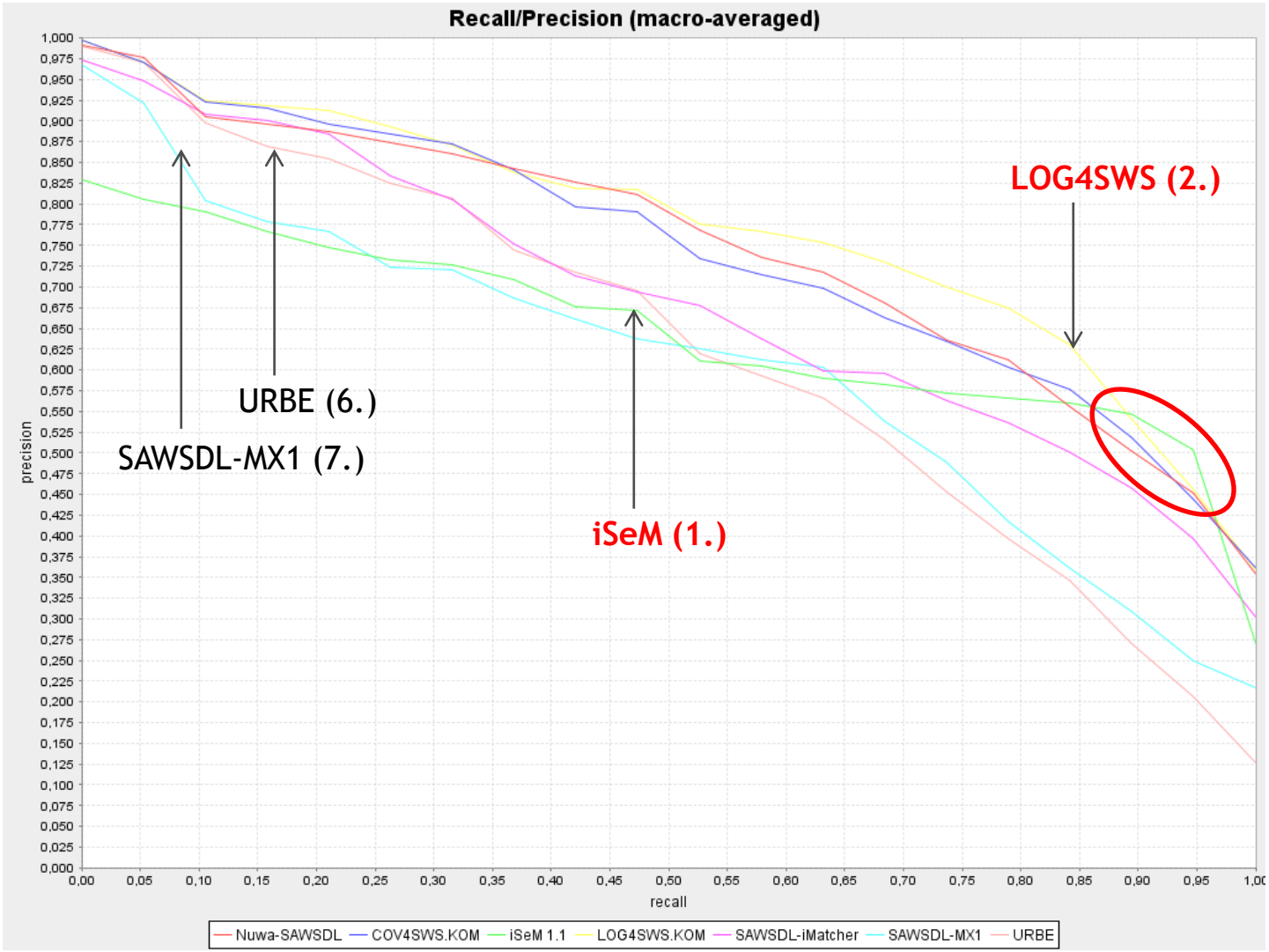
OWL-S matchmakers deal with required service ontologies quite differently

- Caching of *complete* ontologies during service registration
  - Reduces #http-requests: Only queries but no ontology d/l required for Q/A
  - Used by **XSSD, OWLS-iMatcher, SeMa²**
- Caching of self-contained (unfolded) concept definitions
  - Reduces #http-requests: No additional classification of concepts required for Q/A
  - Used by **iSeM 1.1** (and **iSeM-TSM1**), **OWLS-MX3**
- No caching at all
  - **EMMA** – restarts internally used pugin for every query

SAWSDL Selection: Average Precision

Matchmaker	AP (B)	AP (G): nDCG,	Q	Dev.
1. iSeM 1.1	.842	.803	.762	DFKI, Germany
2. LOG4SWS.KOM	.837	.896	.851	TU Darmstadt, Germany
3. COV4SWS.KOM	.823	.884	.825	TU Darmstadt, Germany
4. Nuwa-SAWSDL	.819	.884	.817	URJC Madrid, Spain
5. SAWSDL-iMatcher	.764	.855	.784	U Zurich, Switzerland
6. URBE	.749	.85	.777	Politecnico Milano, Italy
7. SAWSDL-MX1	.747	.839	.767	DFKI, Germany

SAWSDL Selection: Macro-Averaged Precision for Binary Relevance



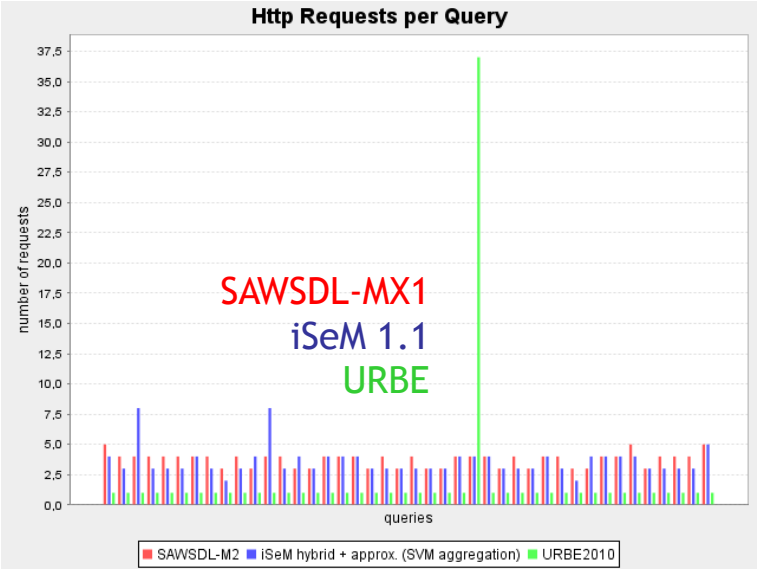
SAWSDL Selection: Average Response Time

Matchmaker	AQRT (s)	w/o http	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. Nuwa-SAWSDL	9.009	8.986	
6. iSeM 1.1	10.662	10.655	[1.584s; - .018]: +3
7. URBE	40.01	39.941	

Lesson Learned: Caching Strategies

SAWSDL matchmakers deal with required service ontologies quite differently

- Caching of complete ontologies *before* service registration
  - Ontologies are loaded and classified right after matchmaker plug-in initialization
  - Used by **LOG4SWS.KOM, COV4SWS.KOM, SAWSDL-iMatcher**
- Caching of self-contained (unfolded) concept definitions
  - Used by **SAWSDL-MX1, iSeM 1.1**
- Unknown strategy: URBE



Lesson Learned: Performance

- **Highest precision (AP):**

	Binary	Graded (nDCG)	
Hybrid + Adaptive			
OWL-S	**0.92**	0.84	iSeM 1.1
SAWSDL	**0.84**	0.8	iSeM 1.1
Hybrid			
OWL-S	0.88	**0.93**	SeMa² v2
SAWSDL	**0.84**	**0.90**	LOG4SWS
Logic-based			
OWL-S	0.76	0.87	EMMA
SAWSDL	-	-	
Non-logic-based			
OWL-S	0.87	0.92	iSeM-TSM1
SAWSDL	0.82	0.88	COV4SWS, NUWA-SASWDL

- **Fastest response (AQRT): 0.12s** XSSD (OWL-S), **0.24s** LOG4SWS (SAWSDL)
- **Best trade-off (AP_B/AQRT; SAW,$w_{1,2}$=.5): iSeM** (.939, OWL-S), **LOG4SWS** (.973, SAWSDL)