Semantic Web – State of the Art

Pascal Hitzler

Kno.e.sis Center

Wright State University, Dayton, OH

http://www.knoesis.org/pascal/

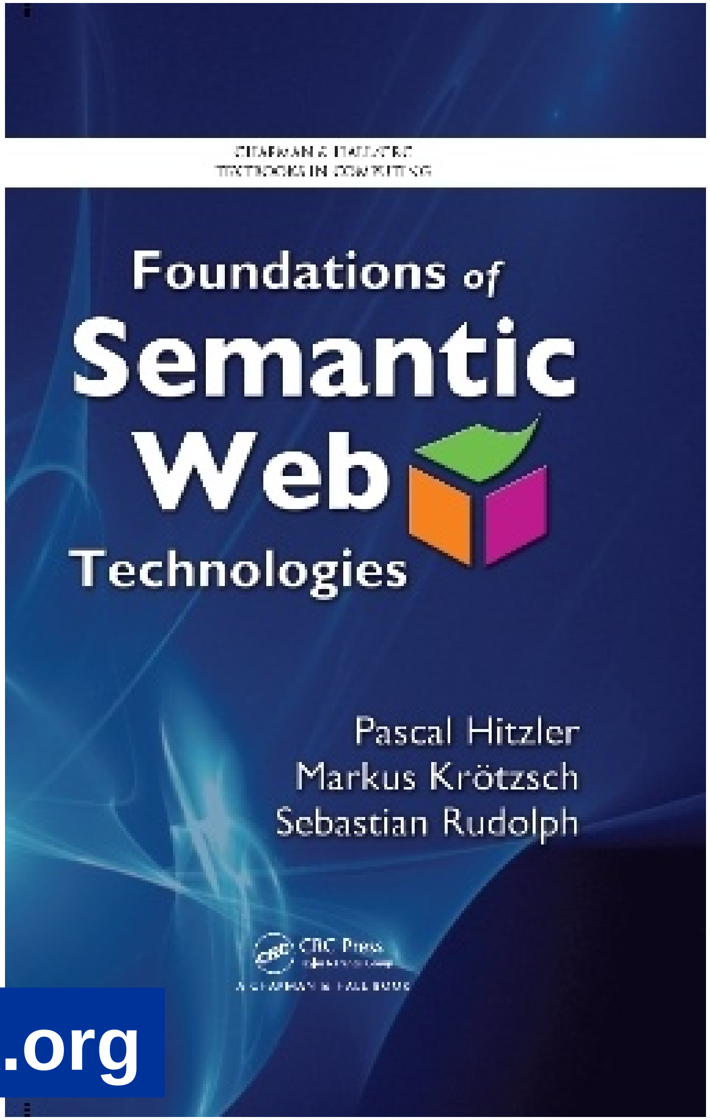
Introduction to key foundations

Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph

Foundations of Semantic Web Technologies

Chapman & Hall/CRC, 2010

Choice Magazine Outstanding Academic Title 2010 (one out of seven in Information & Computer Science)



http://www.semantic-web-book.org

Textbook (Chinese translation)

Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph

语义Web技术基础

Tsinghua University Press (清华大学出版社），2011, to appear

Translators:

Yong Yu, Haofeng Wang, Guilin Qi (俞勇，王昊奋，漆桂林)

http://www.semantic-web-book.org

Contents

- What is Semantic Web?
  - Limitations of the current World Wide Web
  - The basic Semantic Web idea
  - Semantic Web Semantics
- Semantic Data Web (state of the art)
  - its limitations
  - and how to overcome them
- Some current work

The current (World Wide) Web

- Immensely successful.
- Huge amounts of data.
- Syntax standards for transfer of structured data.
- Machine-processable, human-readable documents.

BUT:

- Content/knowledge cannot be accessed by machines.
  Meaning (semantics) of transferred data is not accessible.

Examples

- **Find that landmark article on data integration written by an Indian researcher in the 1990s.**
  [If you manage this without knowing the answer, let me know how you did it.]

- **Which car is called a “duck” in German?**
  [This needs some intelligent integration of content from different websites plus background knowledge.]

Another example

"Identify congress members, who have voted "No" on pro environmental legislation in the past four years, with high-pollution industry in their congressional districts."

In principle, all the required knowledge is on the Web – most of it even in machine-readable form.

However, without automated processing and reasoning we cannot obtain a useful answer.

Very brief history of the Semantic Web

- invented ca. 1989.
- 1990s: W3C metadata activity (lead to RDF(S))
- W3C semantic web activity: chartered 2001.

- USA: DAML-Programme 2000-2005
  approx. $90M.
- Many large scale EU projects since 2002 and ongoing.
  ! FP6/FP7
- Major IT companies and
  venture capital now investing.

Semantic Technologies in the US

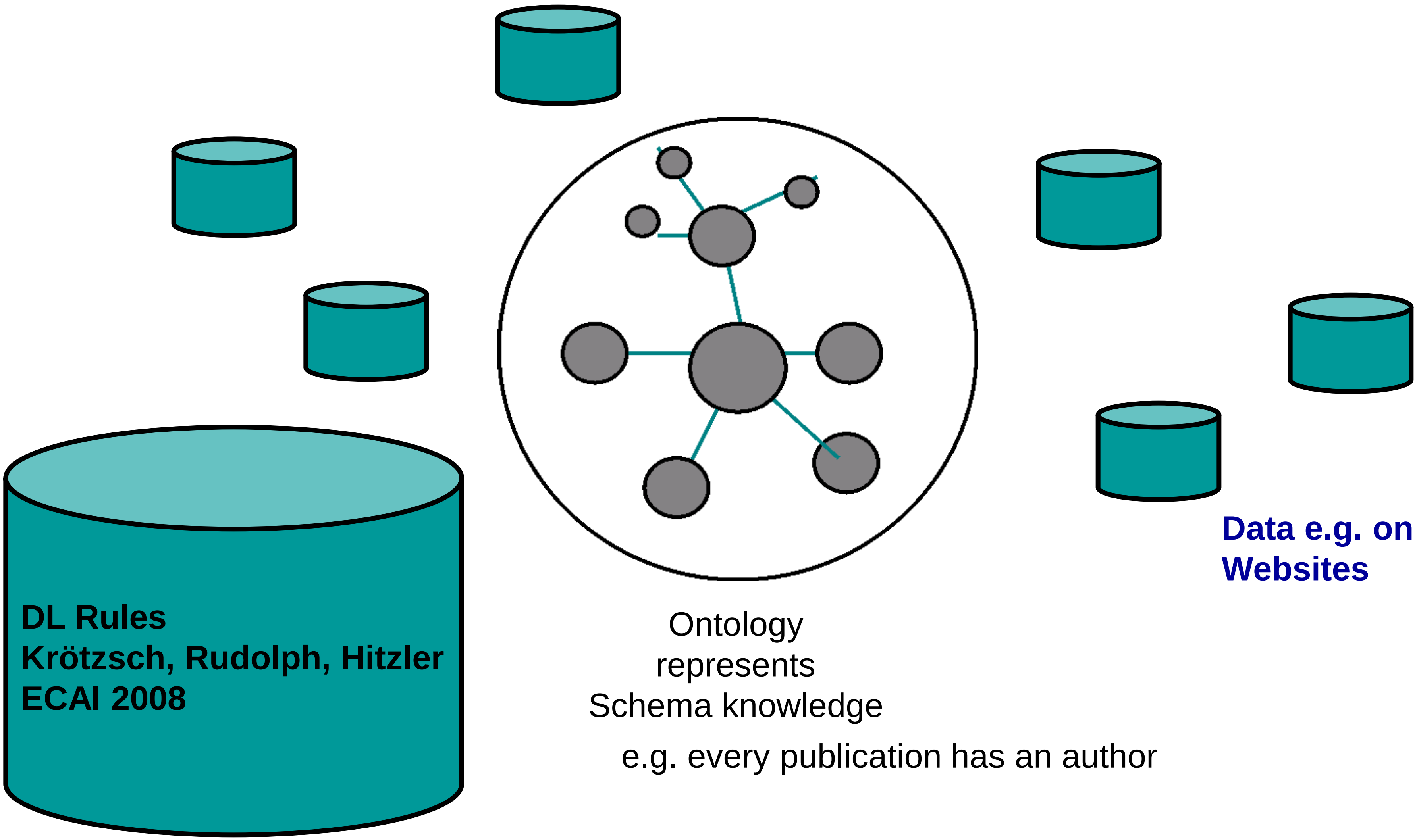
- Funding available e.g. via
  - NIH
  - NSF
  - DoD, DoE, AFRL
  - IARPA, DARPA
  - ...

- Considerable industrial take-up
  - Annual Semantic Technology Conference in CA
    Taylored towards industry
  - Major IT players (Oracle, IBM, HP, ...) invest
  - Major government contractors (BBN, Lockheed, ...)
  - Venture capital (e.g. Vulcan, Inc.).

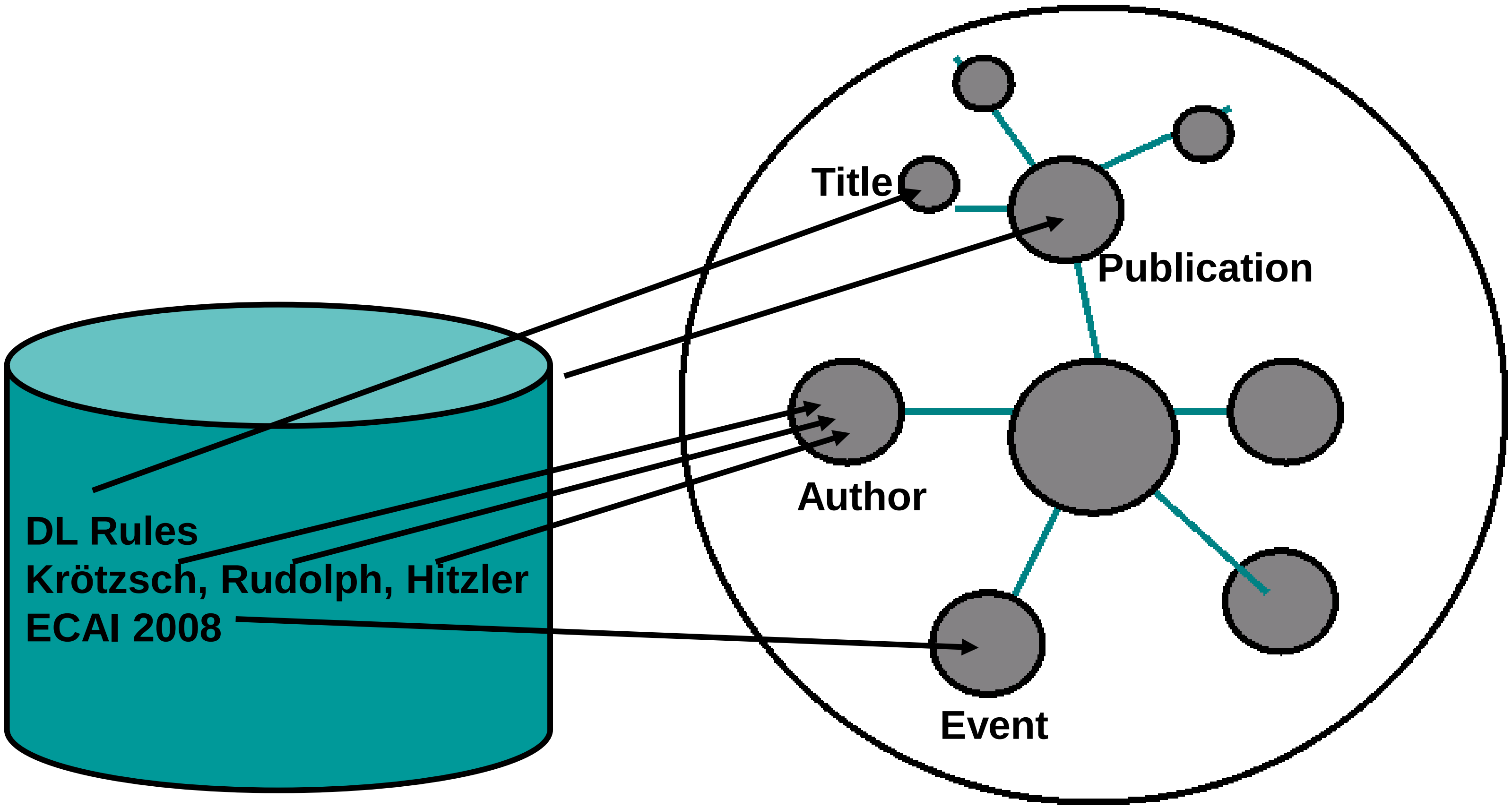
Contents

- What is Semantic Web?
  - Limitations of the current World Wide Web
  - The basic Semantic Web idea
  - Semantic Web Semantics
- Semantic Data Web (state of the art)
  - its limitations
  - and how to overcome them
- Some current work

Basic Idea of the Semantic Web

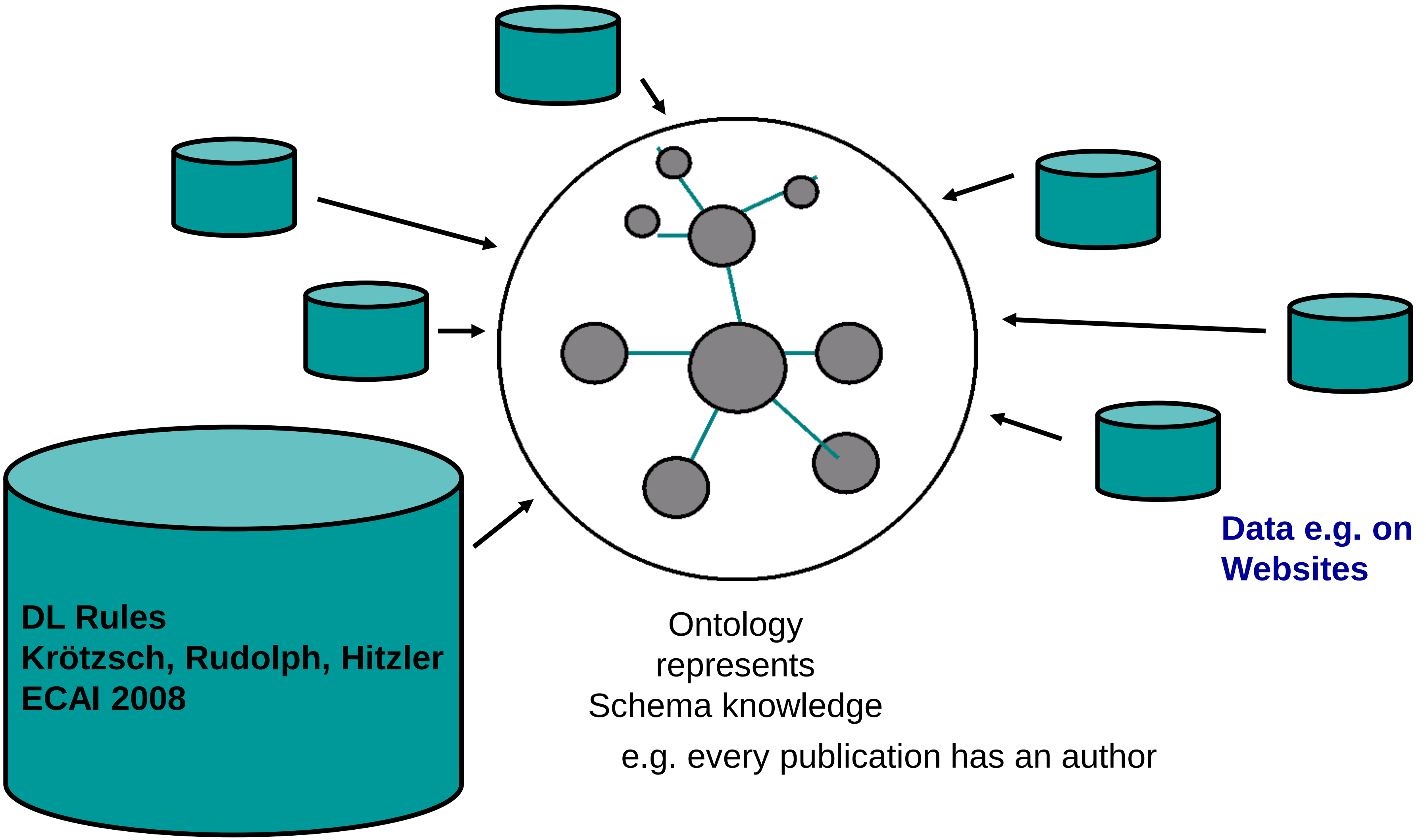


Basic Idea of the Semantic Web



e.g. every publication has an author

Basic Idea of the Semantic Web



Contents

- What is Semantic Web?
  - Limitations of the current World Wide Web
  - The basic Semantic Web idea
  - Semantic Web Semantics
- Semantic Data Web (state of the art)
  - its limitations
  - and how to overcome them
- Some current work

What Is Semantic Web Semantics?

- **Opinions Differ. Here's my take.**

- **Semantic Web requires a shareable, declarative and *computable* semantics.**
- **I.e., the semantics must be a formal entity which is clearly defined and automatically computable.**

- **Ontology languages provide this by means of their formal semantics.**
- **Semantic Web Semantics is given by a relation – the *logical consequence* relation.**

- **Note: This is considerably more than saying that the semantics of an ontology is the set of its logical consequences!**

In other words

We capture the meaning of information

not by specifying its meaning (which is impossible)
but by specifying

how information interacts with other information.

We describe the meaning indirectly through its effects.

Simple Logical Reasoning

If I ask for soccer team members, I also want to get the goalkeepers listed ...

If I ask for cities, I also want all capitals listed ...

inheritance reasoning

Less Simple Reasoning

answering requires merging of knowledge from many websites and using background knowledge.

What was again the name of that russian researcher who worked on resolution-based calculi for EL?

Which car is called „duck“ in German?

What is "Käuzchen" in english?

SNOMED CT

- **SNOMED CT: commercial ontology, medical domain ca. 300,000 axioms**

- **InjuryOfFinger ´ Injury u 9site.Finger$_S$**
  InjuryOfHand ´ Injury u 9site.Hand$_S$
  Finger$_S$ v Hand$_P$
  Hand$_P$ v Hand$_S$ u 9part.Hand$_E$

- **Reasoning has been used e.g. for**
  - **classification (computing the hidden taxonomy) e.g., InjuryOfFinger v InjuryOfHand**
  - **bug finding**

So what happened?

- In 2004, two W3C Recommendations were completed:
  - RDF + RDF Schema with formal model-theoretic semantics
  - OWL with formal model-theoretic semantics

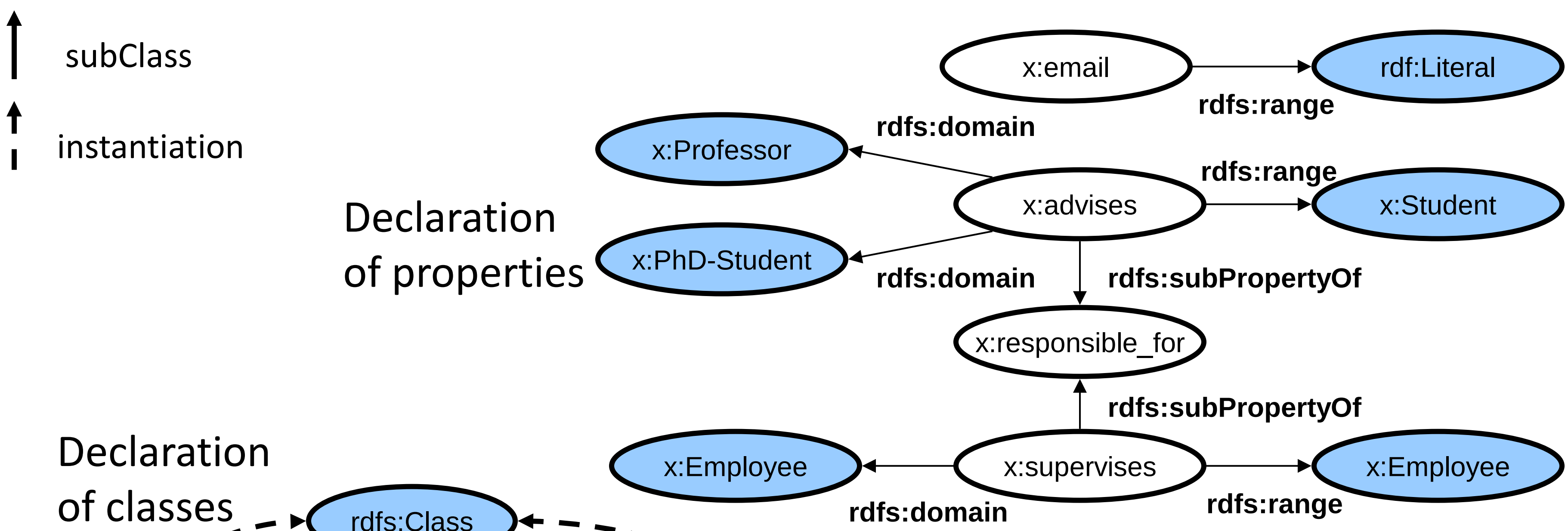
- OWL 2 update emerged 2009.
- RDF update is being discussed right now.

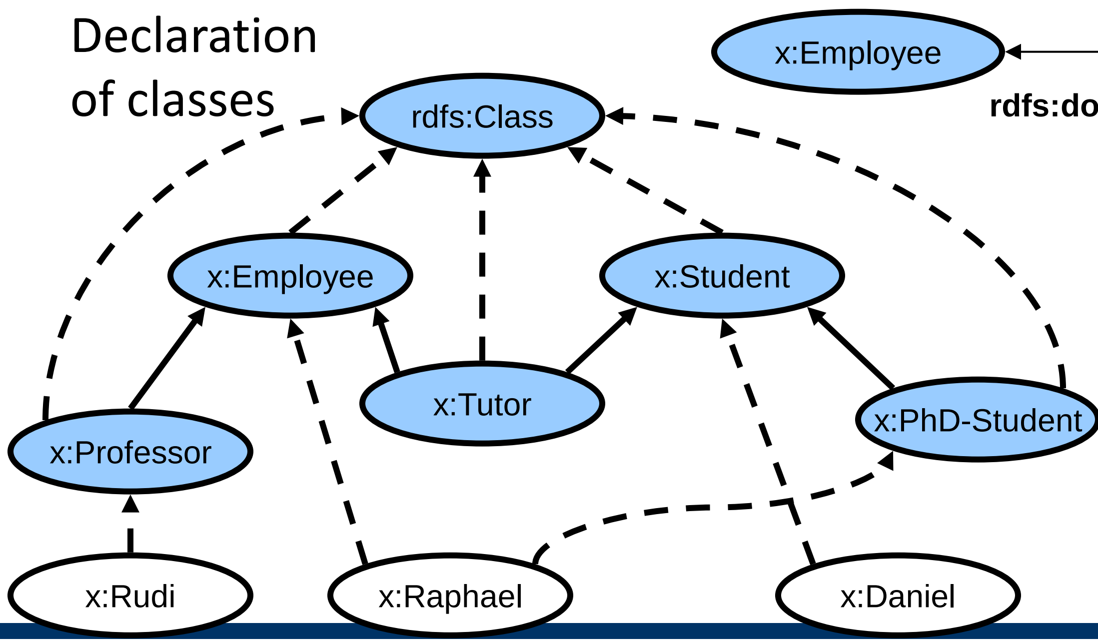
Ontology languages

- Of central importance for the realisation of Semantic Technologies are suitable representation languages.
- Meaning (semantics) provided via logic and deduction algorithms.
- Scalability is a challenge.

Language standards recommended by W3C

Ontology Example





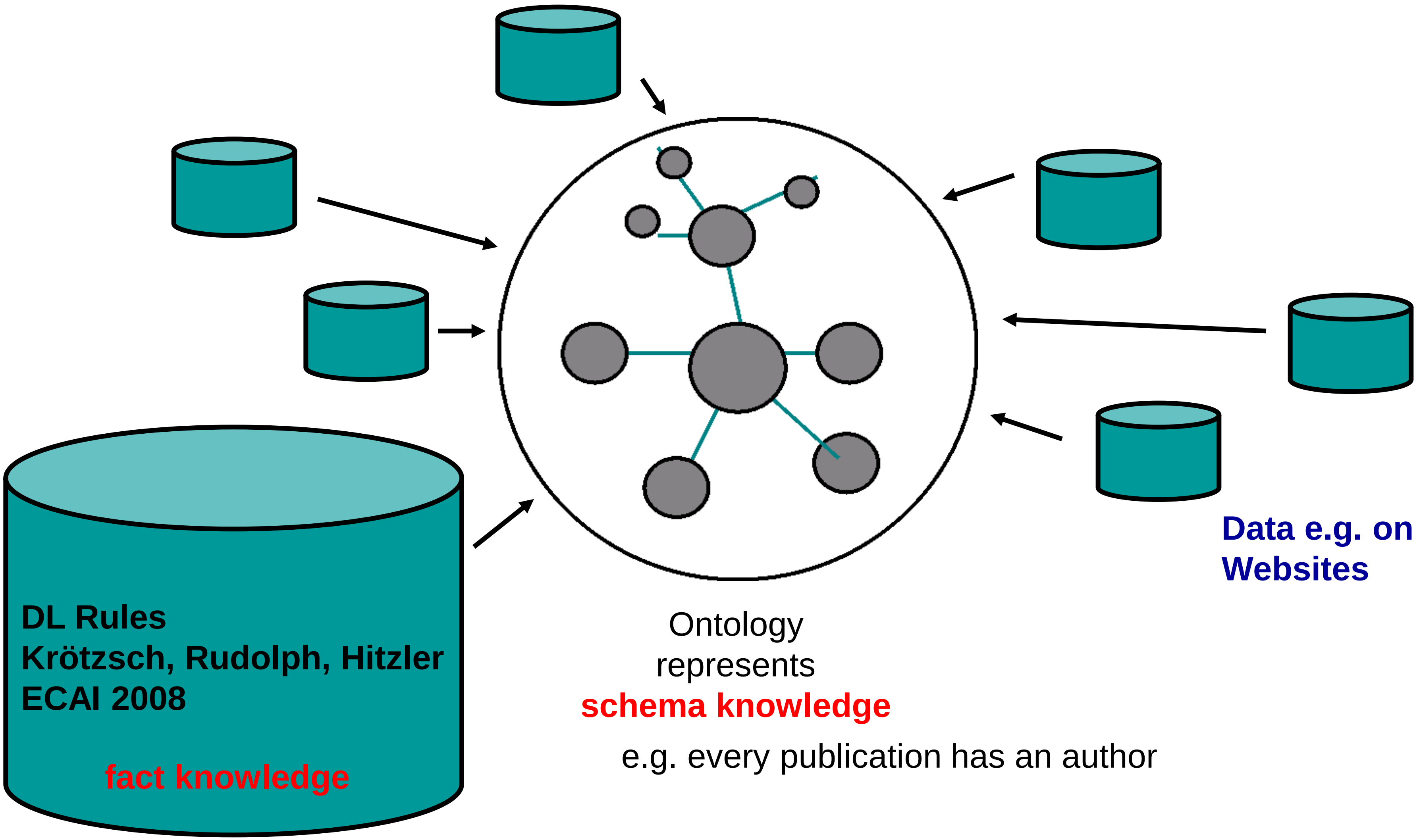
schema knowledge
PhDStudent v 9advisedBy.Professor

rules
responsible_for(y,x) ÆProfessor(y)
! Employee(x)

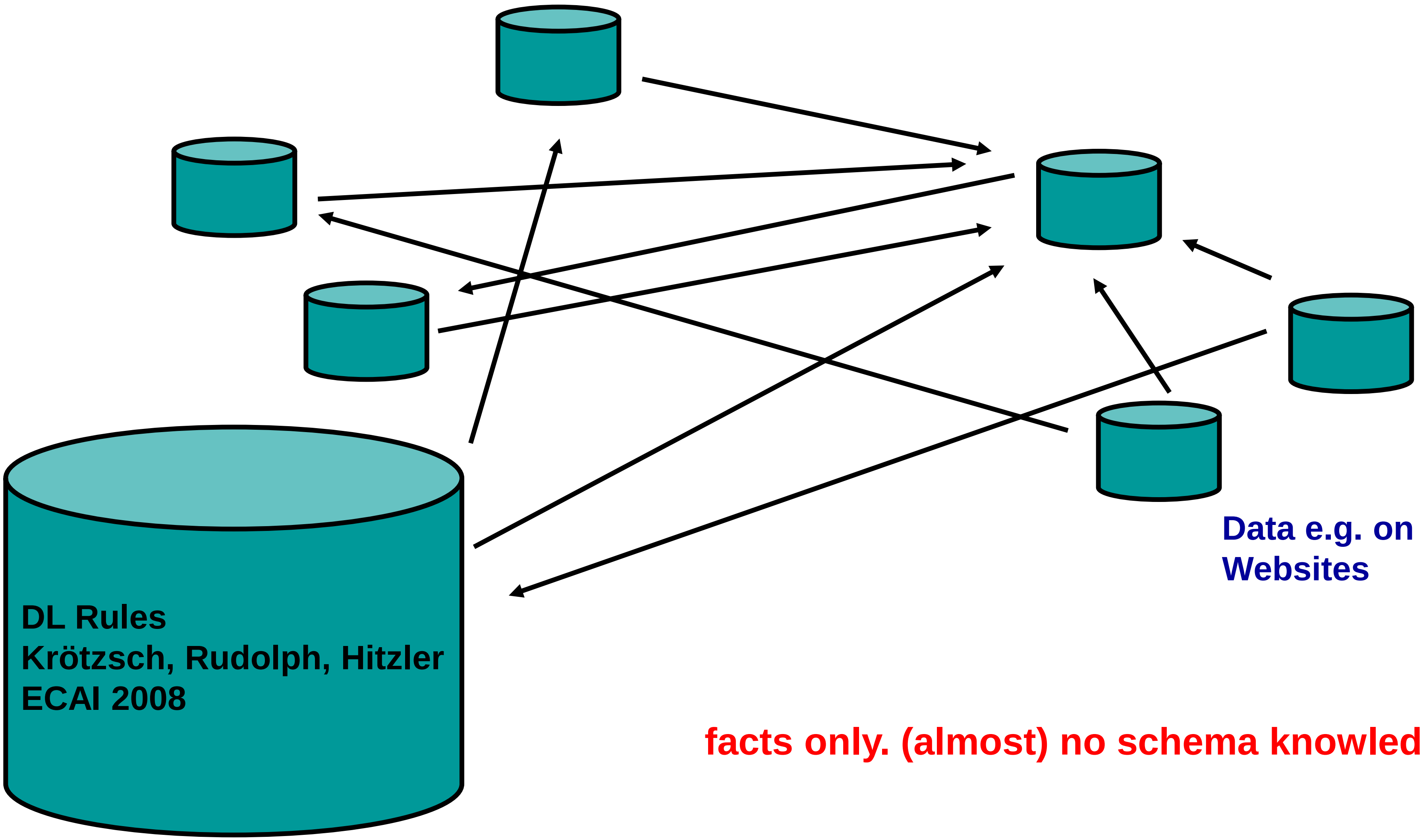
Contents

- What is Semantic Web?
  - Limitations of the current World Wide Web
  - The basic Semantic Web idea
  - Semantic Web Semantics
- Semantic Data Web (state of the art)
  - its limitations
  - and how to overcome them
- Some current work

Basic Idea of the Semantic Web

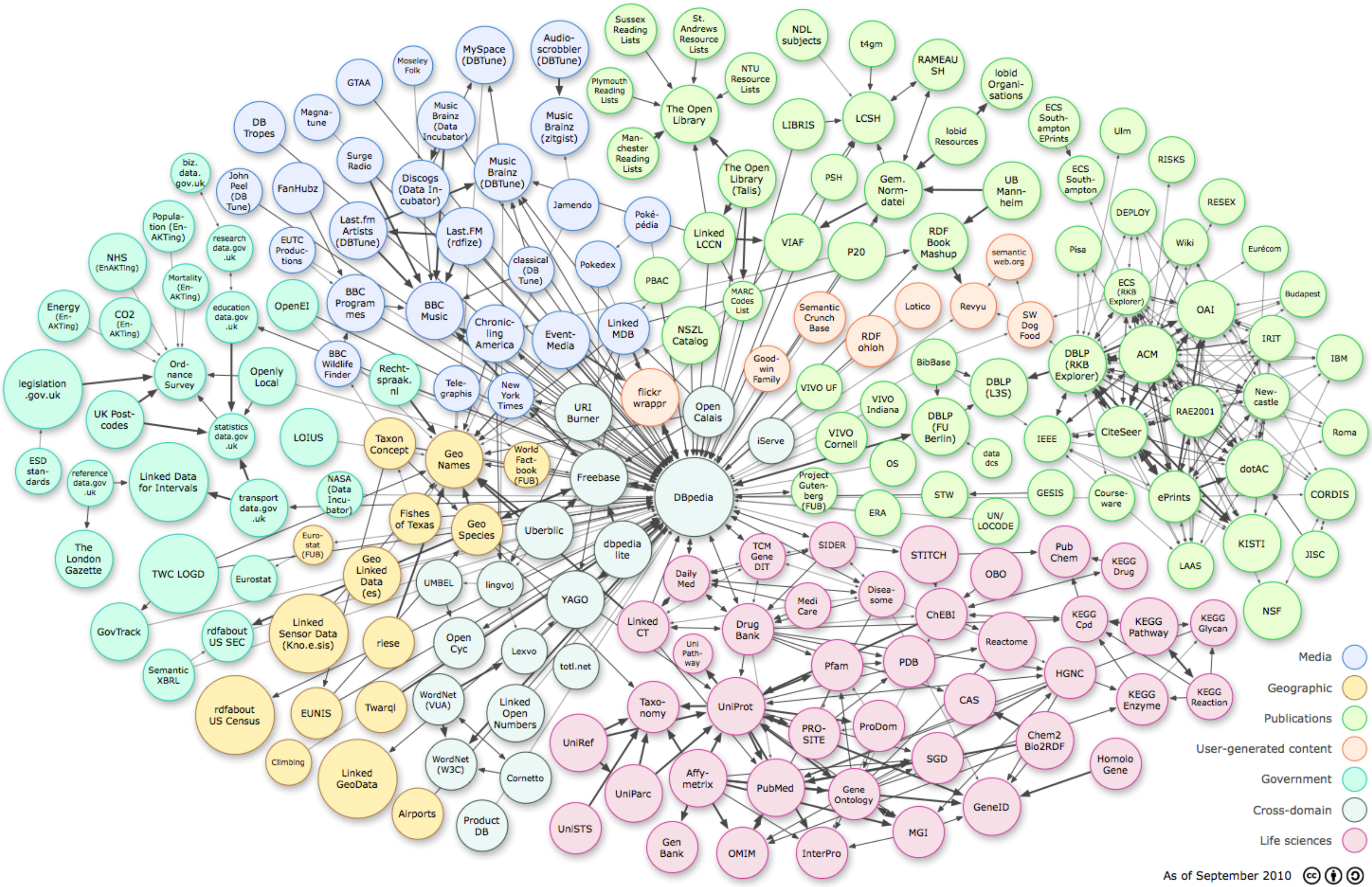


Currently it's looking like this

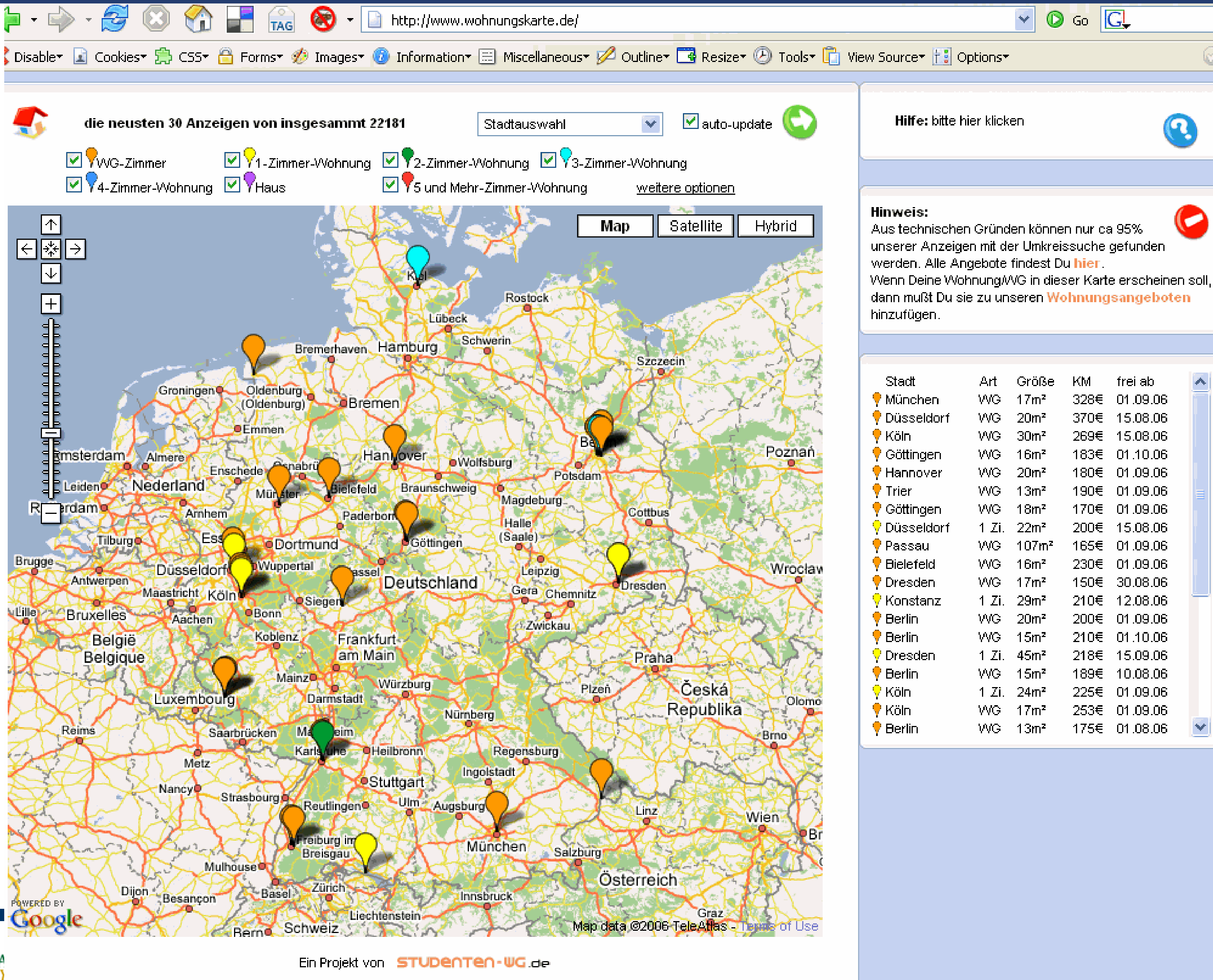


facts only. (almost) no schema knowledge

Linked Open Data



Mashups


http://www.wohnungskarte.de/
Disable Cookies CSS Forms Images Information Miscellaneous Outline Resize Tools View Source Options
die neusten 30 Anzeigen von insgesammt 22181
Stadtauswahl
auto-update
WG-Zimmer
1-Zimmer-Wohnung
2-Zimmer-Wohnung
3-Zimmer-Wohnung
4-Zimmer-Wohnung
Haus
5 und Mehr-Zimmer-Wohnung
weitere optionen
Map Satellite Hybrid
Hilfe: bitte hier klicken
Hinweis:
Aus technischen Gründen können nur ca 95% unserer Anzeigen mit der Umkreissuche gefunden werden. Alle Angebote findest Du hier.
Wenn Deine Wohnung/WG in dieser Karte erscheinen soll, dann mußt Du sie zu unseren Wohnungsangeboten hinzufügen.
Stadt Art Größe KM frei ab
München WG 17m² 328€ 01.09.06
Düsseldorf WG 20m² 370€ 15.08.06
Köln WG 30m² 269€ 15.08.06
Göttingen WG 16m² 183€ 01.10.06
Hannover WG 20m² 180€ 01.09.06
Trier WG 13m² 190€ 01.09.06
Göttingen WG 18m² 170€ 01.10.06
Düsseldorf 1 Zi. 22m² 200€ 15.08.06
Passau WG 107m² 165€ 01.09.06
Bielefeld WG 16m² 230€ 01.09.06
Dresden WG 17m² 150€ 30.08.06
Konstanz 1 Zi. 29m² 210€ 12.08.06
Berlin WG 20m² 200€ 01.09.06
Berlin WG 15m² 210€ 01.10.06
Dresden 1 Zi. 45m² 218€ 15.09.06
Berlin WG 15m² 189€ 10.08.06
Köln 1 Zi. 24m² 225€ 01.09.06
Köln WG 17m² 253€ 01.09.06
Berlin WG 13m² 175€ 01.08.06
Map data ©2006 TeleAtlas - Terms of Use
Ein Projekt von STUDENTEN-WG.de

Contents

- What is Semantic Web?
  - Limitations of the current World Wide Web
  - The basic Semantic Web idea
  - Semantic Web Semantics
- Semantic Data Web (state of the art)
  - its limitations
  - and how to overcome them
- Some current work

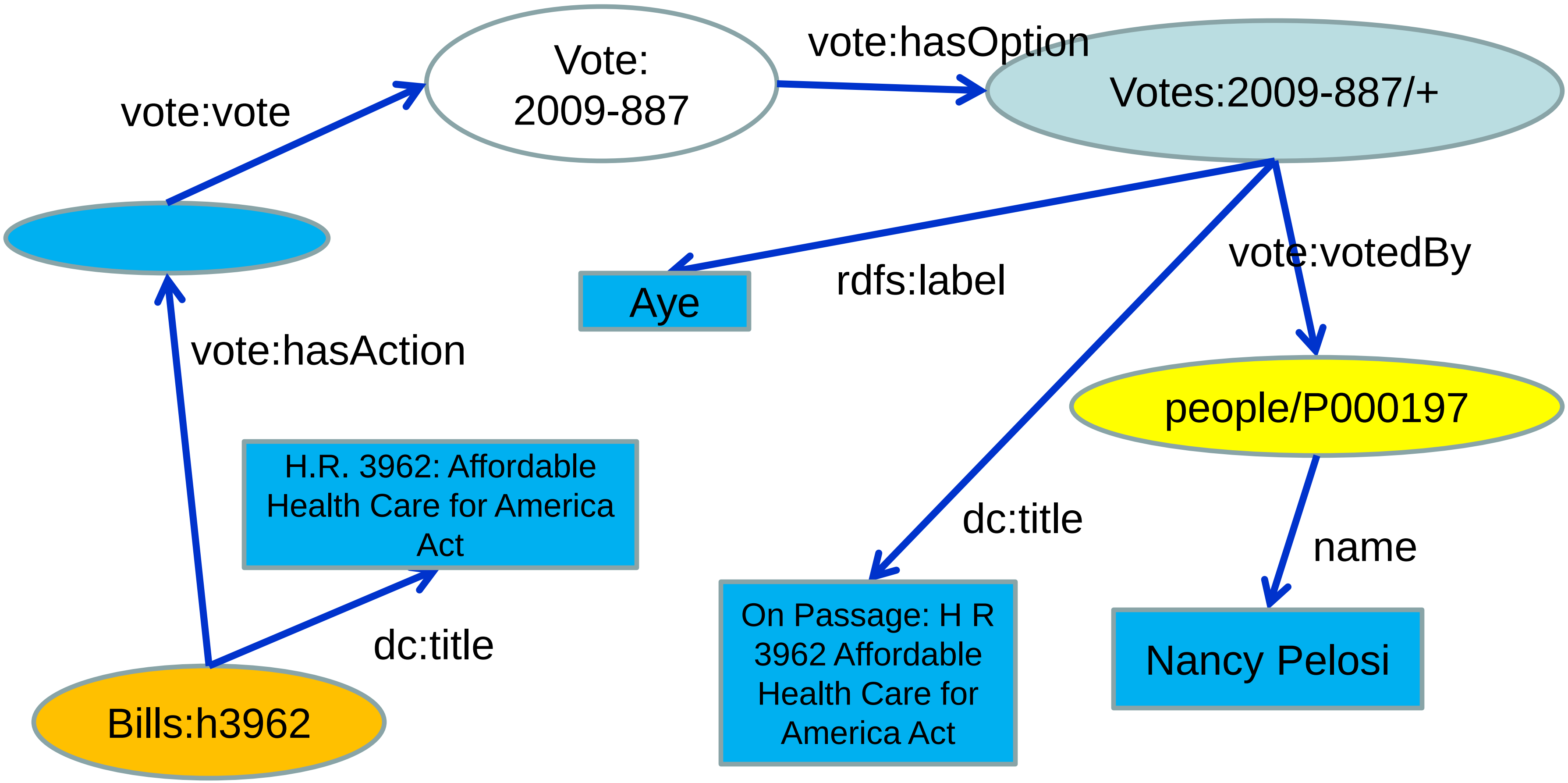
Example: GeoNames

Populated Place Features (city, village,...)			
2,518,403	P.PPL	populated place	a city, town, village, or other agglomeration of buildings where people live and work
48,483	P.PPLX	section of populated place	
39,336	P.PPLL	populated locality	an area similar to a locality but with a small group of dwellings or other buildings
13,306	P.PPLQ	abandoned populated place	
2,684	P.PPLA4	seat of a fourth-order administrative division	
2,028	P.PPLA	seat of a first-order administrative division	seat of a first-order administrative division (PPLC takes precedence over PPLA)
1,847	P.PPLW	destroyed populated place	a village, town or city destroyed by a natural disaster, or by war
1,006	P.PPLF	farm village	a populated place where the population is largely engaged in agricultural activities
930	P.PPLA3	seat of a third-o administrative d	
695	P.PPLA2	seat of a second administrative division	
253	P.PPLS	populated places	cities, towns, villages, or other agglomerations of buildings where people live and work
249	P.STLMT	israeli settlement	
235	P.PPLC	capital of a political entity	
57	P.		
29	P.PPLR	religious populated place	a populated place whose population is largely engaged in religious occupations
6	P.PPLG	seat of government of a political entity	
2,629,547	Total for P		

rdfs:subClassOf?

Example: GovTrack

"Nancy Pelosi voted in favor of the Health Care Bill."



Example querying LoD

"Identify congress members, who have voted "No" on pro environmental legislation in the past four years, with high-pollution industry in their congressional districts."

In principle, all the knowledge is there:

- GovTrack
- GeoNames
- DBPedia
- US Census

But even with LoD we cannot answer this query.

Example querying LoD

"Identify congress members, who have voted "No" on pro environmental legislation in the past four years, with high-pollution industry in their congressional districts."

Some missing puzzle pieces:

- **Where is the data?**
  - **GovTrack**
    GeoNames
    US Census
    requires intimate knowledge of the LoD data sets

Example querying LoD

"Identify congress members, who have voted "No" on pro environmental legislation in the past four years, with high-pollution industry in their congressional districts."

Some missing puzzle pieces:

- Where is the data?
  (smart federation needed)
- Missing background (schema) knowledge.
  (enhancements of the LoD cloud)
- Crucial info still hidden in texts.
  (ontology learning from texts)
- Added reasoning capabilities (e.g., spatial).
  (new ontology language features)

Don’t get me wrong

Linked Open Data is great, useful, cool, and a very important step.

But we need to make use of the added value of formal semantics in order to advance towards the Semantic Web vision!

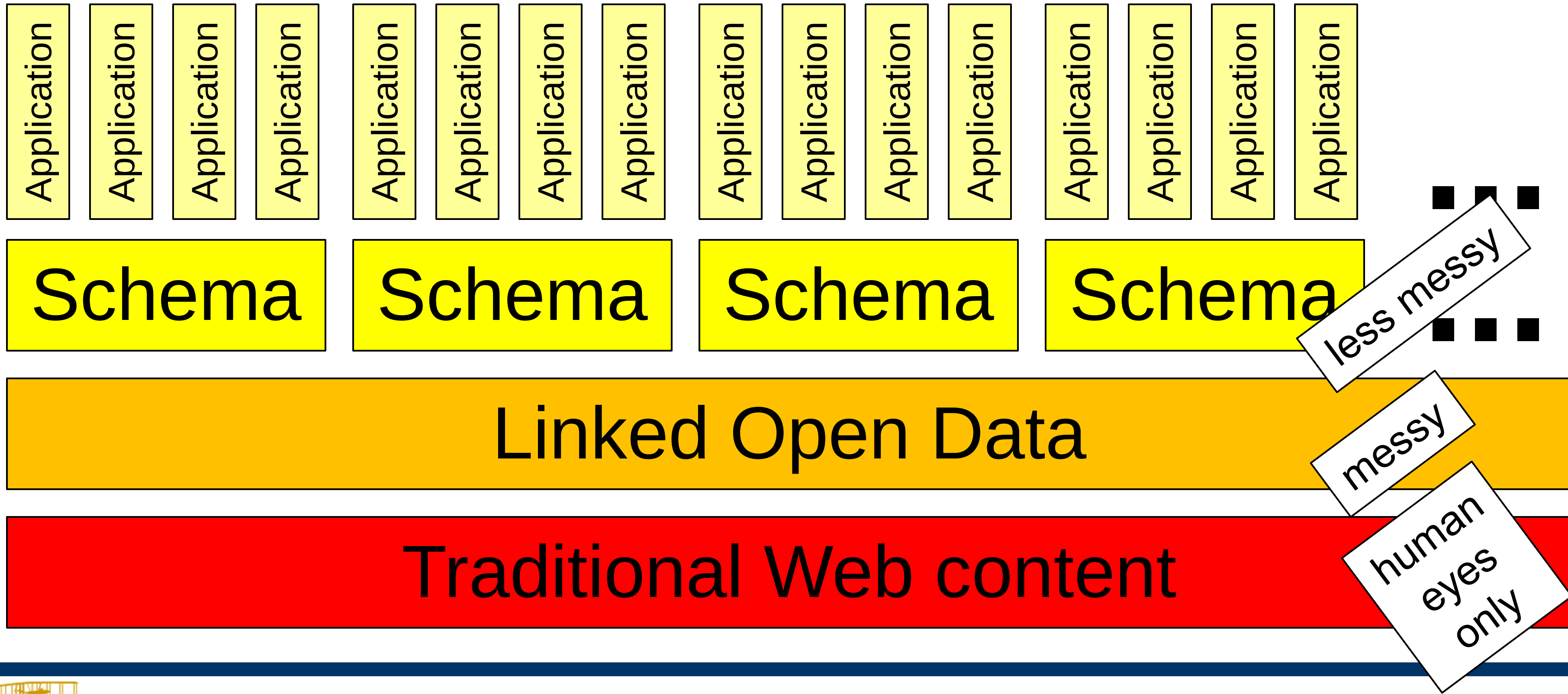
Contents

- What is Semantic Web?
  - Limitations of the current World Wide Web
  - The basic Semantic Web idea
  - Semantic Web Semantics
- Semantic Data Web (state of the art)
  - its limitations
  - and how to overcome them
- Some current work

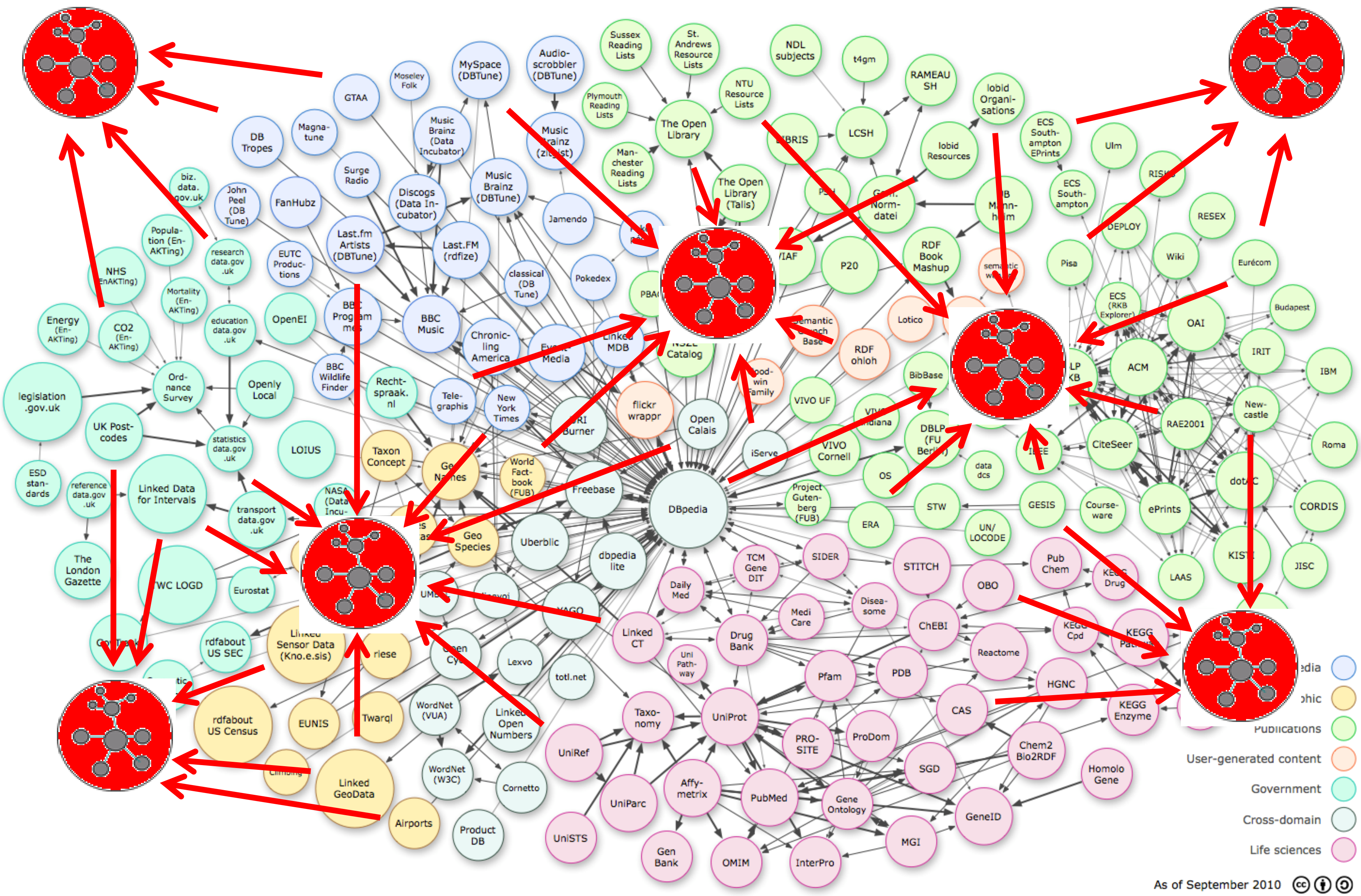
The Semantic Data Web Layer Cake

To leverage LoD, we require schema knowledge

- **application-type driven (reusable for same kind of application)**
- **less messy than LoD (as required by application)**
- **overarching several LoD datasets (as required by application)**



Schema on top of the LoD cloud



Contents

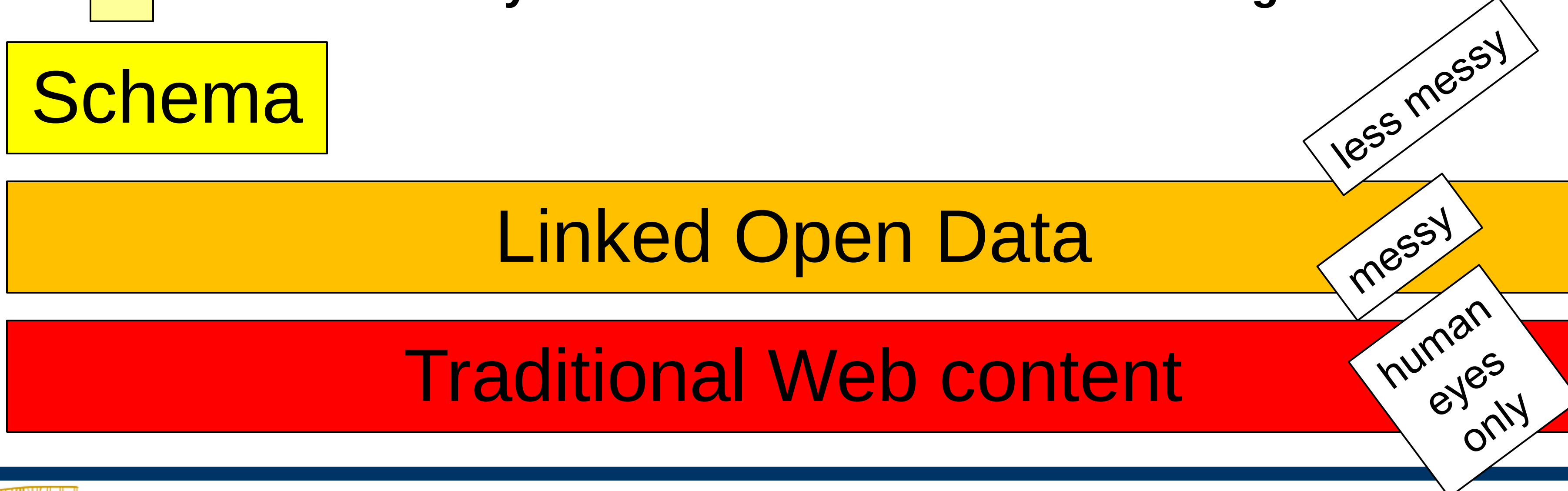
- What is Semantic Web?
  - Limitations of the current World Wide Web
  - The basic Semantic Web idea
  - Semantic Web Semantics
- Semantic Data Web (state of the art)
  - its limitations
  - and how to overcome them
- Some current work

LOQuS – Querying Linked Open Data

Work in progress.

- **Schema creation for**
  - **query federation**
  - **utilizing background knowledge**
  - **compilation of LOD knowledge into reason-able form**
- **Reasoning algorithm (on suitable language) for very efficient data-intensive reasoning**

LOD querying



BLOOMS

Table 4. Results of various systems for LOD Schema Alignment. Legends: Prec=Precision, Rec=Recall, M=Music Ontology, B=BBC Program Ontology, F=FOAF Ontology, D=DBpedia Ontology, G=Geonames Ontology, S=SIOC Ontology, W=Semantic Web Conference Ontology, A=AKT Portal Ontology, err=System Error, NA=Not Available

Linked Open Data Schema Ontology Alignment												
	Alignment API		OMViaUO		RiMoM		S-Match		AROMA		BLOOMS	
Test	Prec	Rec	Prec	Rec	Prec	Rec	Prec	Rec	Prec	Rec	Prec	Rec
M,B	0.4	0	1	0	err	err	0.04	0.28	0	0	0.63	0.78
M,D	0	0	0	0	err	err	0.08	0.30	0.45	0.01	0.39	0.62
F,D	0	0	0	0	err	err	0.11	0.40	0.33	0.04	0.67	0.73
G,D	0	0	0	0	err	err	0.23	1	0	0	0	0
S,F	0	0	0	0	0.3	0.2	0.52	0.11	0.30	0.20	0.55	0.64
W,A	0.12	0.05	0.16	0.03	err	err	0.06	0.4	0.38	0.03	0.42	0.59
W,D	0	0	0	0	err	err	0.15	0.50	0.27	0.01	0.70	0.40
Avg.	0.07	0.01	0.17	0	NA	NA	0.17	0.43	0.25	0.04	0.48	0.54

Jain, Hitzler et al, ISWC2010

BLOOMS

Table 1. Results on the oriented matching track. Results for RiMOM and AROMA have been taken from the OAEI 2009 website. Legends: Prec=Precision, A-API=Alignment API, OMV=OMViaUO, NaN=division by zero, likely due to empty alignment.

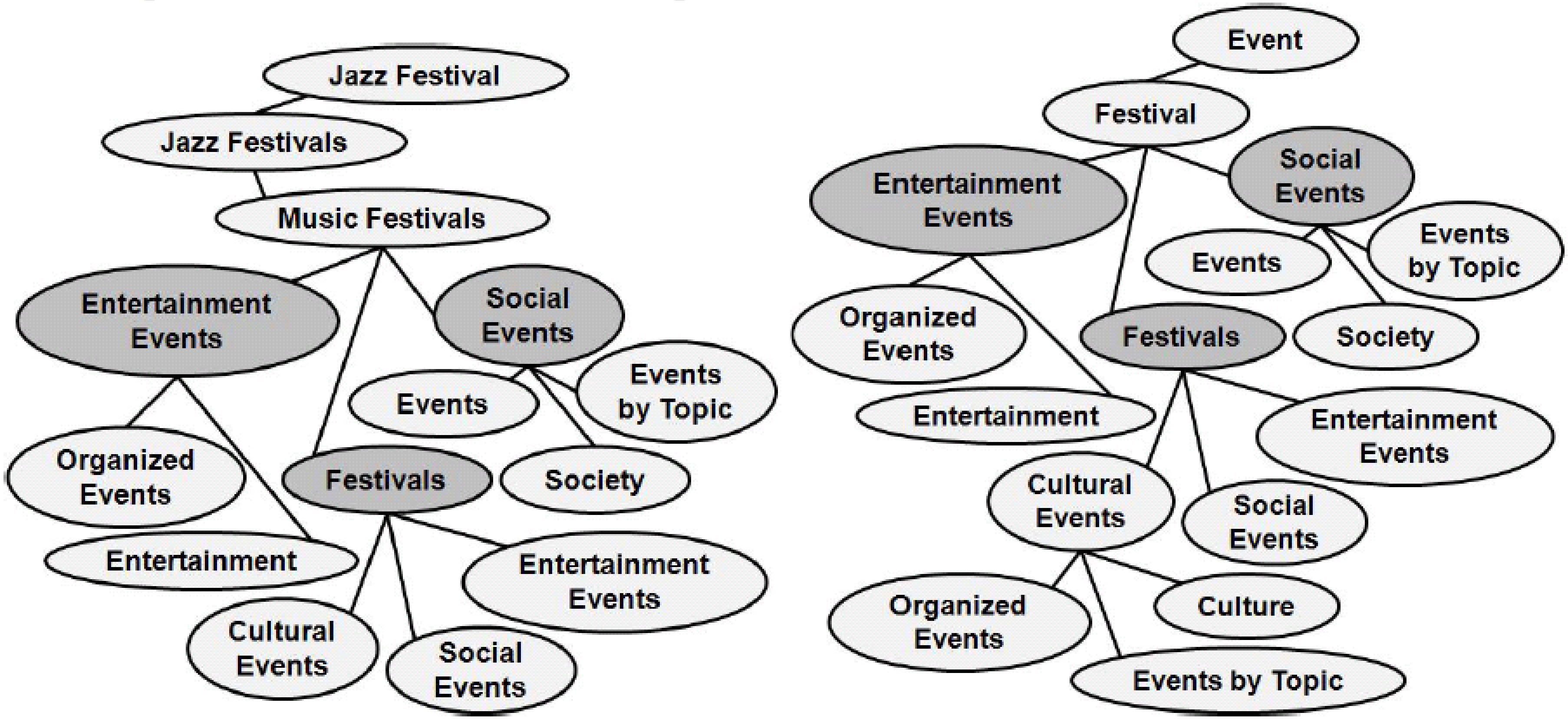
Ontology Alignment Initiative—Oriented Matching Track												
	A-API		OMV		S-Match		AROMA		RiMoM		BLOOMS	
Test	Prec	Rec	Prec	Rec	Prec	Rec	Prec	Rec	Prec	Rec	Prec	Rec
1XX	0	0	0.02	0.06	0.01	0.71	NaN	0	1	1	1	1
2XX	0	0	0.01	0.03	0.05	0.30	0.84	0.08	0.67	0.85	0.52	0.51
3XX	0.01	0.03	0.02	0.047	0.01	0.14	0.72	0.11	0.59	0.81	1	0.84
Avg.	0.00	0.01	0.02	0.04	0.03	0.38	0.63	0.07	0.75	0.88	0.84	0.78

BLOOMS

1. **Pre-processing of the input ontologies** in order to (i) remove property restrictions, individuals, and properties, and to (ii) tokenize composite class names to obtain a list of all simple words contained within them, with stop words removed.
2. **Construction of the BLOOMS forest T_C** for each class name C, using information from Wikipedia.
3. **Comparison of constructed BLOOMS forests**, which yields decisions which class names are to be aligned.
4. **Post-processing** of the results with the help of the Alignment API and a reasoner.

BLOOMS trees

Fig. 1. BLOOMS trees for Jazz Festival with sense Jazz Festival and for Event with sense Event. To save space, some categories are not expanded to level 4.



BLOOMS

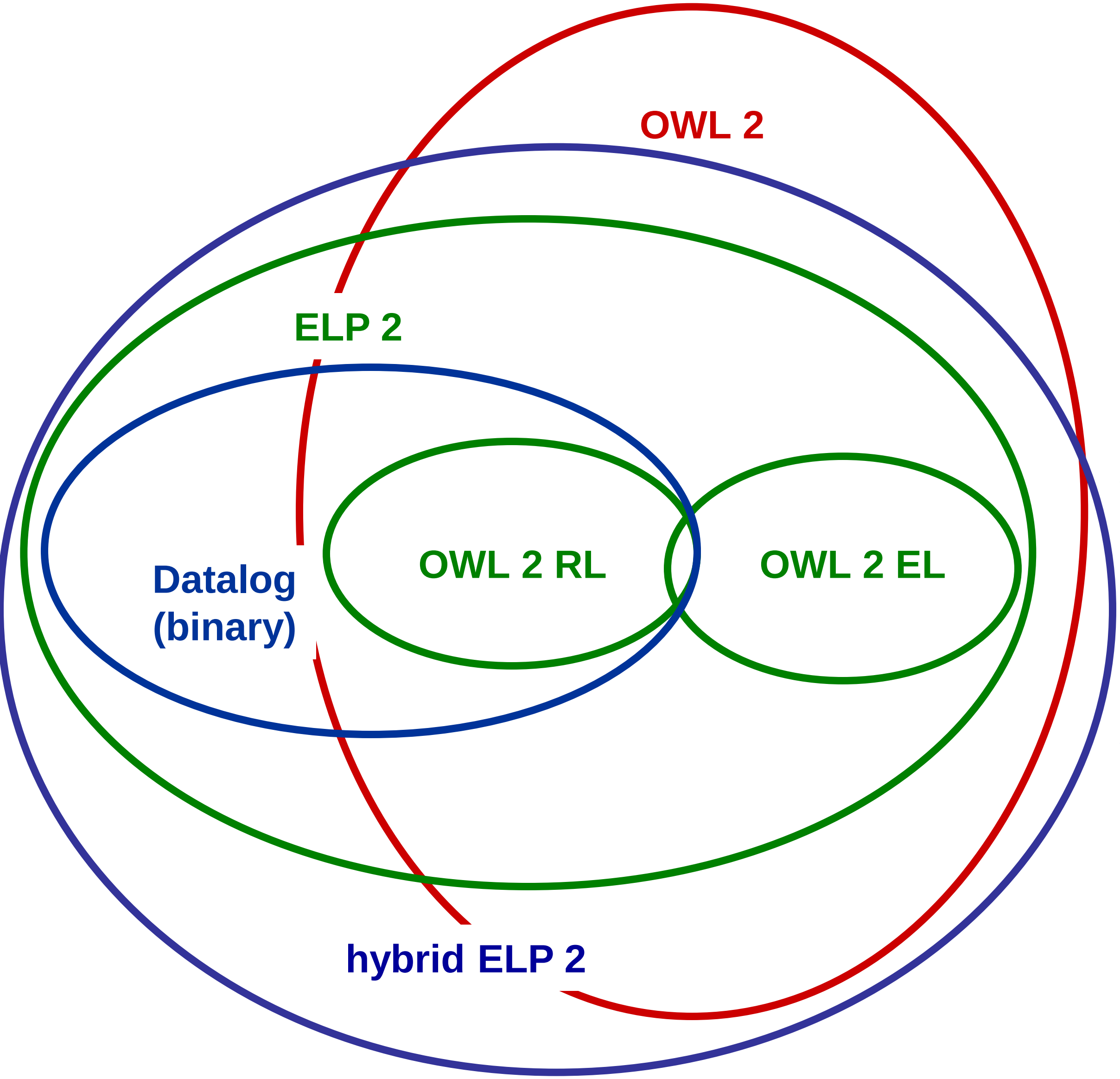
1. **Pre-processing of the input ontologies** in order to (i) remove property restrictions, individuals, and properties, and to (ii) tokenize composite class names to obtain a list of all simple words contained within them, with stop words removed.
2. **Construction of the BLOOMS forest** T_C for each class name C, using information from Wikipedia.
3. **Comparison of constructed BLOOMS forests**, which yields decisions which class names are to be aligned.
4. **Post-processing** of the results with the help of the Alignment API and a reasoner.

BLOOMS and LOQuS

1. **Pre-processing of the input ontologies** in order to (i) remove property restrictions, individuals, and properties, and to (ii) tokenize composite class names to obtain a list of all simple words contained within them, with stop words removed.
2. **Construction of the BLOOMS forest** T_C for each class name C, using information from Wikipedia.
3. **Comparison of constructed BLOOMS forests**, which yields decisions which class names are to be aligned.
4. **Post-processing** of the results with the help of the Alignment API and a reasoner.

) **We're currently evaluating the LOQuS querying approach while utilizing BLOOMS.**

Reasoning: useful scalable languages

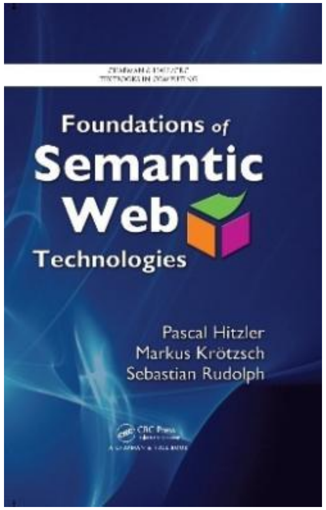


- **OWL 2: complexity > exponential**
- **ELP 2: complexity = polynomial [WWW2011]**
- **OWL 2 EL and RL: complexity = polynomial**
- **hybrid ELP 2 + Datalog: data complexity = polynomial [follows from ECAI2008]**

Thanks!

Collaborators on the covered topics:

Kno.e.sis:	**Prateek Jain, Adila Alfa Krisnadhi, Frederick Maier, Raghava Mutharaju, Amit Sheth**
Accenture:	**Kunal Verma, Peter Z. Yeh**
Karlsruhe:	**Sebastian Rudolph**
Oxford:	**Markus Krötzsch**
Lisboa:	**Matthias Knorr, Jose J. Alferes**



http://www.semantic-web-book.org
http://www.semantic-web-journal.net

References

- Krzysztof Janowicz, Pascal Hitzler, *The Digital Earth as Knowledge Engine*. Semantic Web 3 (3), 213-221, 2012.
- Prateek Jain, Pascal Hitzler, Peter Z. Yeh, Kunal Verma, Amit P. Sheth, *Linked Data is Merely More Data*. In: Dan Brickley, Vinay K. Chaudhri, Harry Halpin, Deborah McGuinness: Linked Data Meets Artificial Intelligence. Technical Report SS-10-07, AAAI Press, Menlo Park, California, 2010, pp. 82-86. ISBN 978-1-57735-461-1. Proceedings of LinkedAI at the AAAI Spring Symposium, March 2010.
- Pascal Hitzler, Frank van Harmelen, *A reasonable Semantic Web*. Semantic Web 1(1-2), 39-44, 2010.
- Pascal Hitzler, Krzysztof Janowicz, *What's Wrong with Linked Data?* http://blog.semantic-web.at/2012/08/09/whats-wrong-with-linked-data/ , August 2012.
- Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph, *Foundations of Semantic Web Technologies*. Chapman and Hall/CRC Press, 2009.

References

- Pascal Hitzler, Markus Krötzsch, Bijan Parsia, Peter F. Patel-Schneider, Sebastian Rudolph, OWL 2 Web Ontology Language: Primer. W3C Recommendation, 27 October 2009.
- Prateek Jain, Pascal Hitzler, Amit P. Sheth, Kunal Verma, Peter Z. Yeh, Ontology Alignment for Linked Open Data. In P. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, J. Pan, I. Horrocks, B. Glimm (eds.), The Semantic Web - ISWC 2010. 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010, Revised Selected Papers, Part I. Lecture Notes in Computer Science Vol. 6496. Springer, Berlin, 2010, pp. 402-417.
- Prateek Jain, Pascal Hitzler, Kunal Verma, Peter Yeh, Amit Sheth, Moving beyond sameAs with PLATO: Partonomy detection for Linked Data. In: Ethan V. Munson, Markus Strohmaier (Eds.): 23rd ACM Conference on Hypertext and Social Media, HT '12, Milwaukee, WI, USA, June 25-28, 2012. ACM, 2012, pp. 33-42.

References

- Amit Krishna Joshi, Prateek Jain, Pascal Hitzler, Peter Z. Yeh, Kunal Verma, Amit P. Sheth, Mariana Damova, Alignment-based Querying of Linked Open Data. In: Meersman, R.; Panetto, H.; Dillon, T.; Rinderle-Ma, S.; Dadam, P.; Zhou, X.; Pearson, S.; Ferscha, A.; Bergamaschi, S.; Cruz, I.F. (eds.), On the Move to Meaningful Internet Systems: OTM 2012, Confederated International Conferences: CoopIS, DOA-SVI, and ODBASE 2012, Rome, Italy, September 10-14, 2012, Proceedings, Part II. Lecture Notes in Computer Science Vol. 7566, Springer, Heidelberg, 2012, pp. 807-824.
- Shasha Huang, Qingguo Li, Pascal Hitzler, Reasoning with Inconsistencies in Hybrid MKNF Knowledge Bases. Logic Journal of the IGPL. To appear.
- Frederick Maier, Yue Ma, Pascal Hitzler, Paraconsistent OWL and Related Logics. Semantic Web journal. To appear.

References

- Barbara Hammer, Pascal Hitzler (eds.), Perspectives of Neural-Symbolic Integration. Studies in Computational Intelligence, Vol. 77. Springer, 2007, ISBN 978-3-540-73952-1.
- Matthias Knorr, Jose Julio Alferes, Pascal Hitzler, Local Closed-World Reasoning with Description Logics under the Well-founded Semantics. Artificial Intelligence 175(9-10), 2011, 1528-1554.
- Jens Lehmann, Pascal Hitzler, Concept Learning in Description Logics Using Refinement Operators. Machine Learning 78(1-2), 203-250, 2010.
- Sebastian Bader, Pascal Hitzler, Steffen Hölldobler, Connectionist Model Generation: A First-Order Approach. Neurocomputing 71, 2008, 2420-2432.

References

- Matthias Knorr, David Carral Martinez, Pascal Hitzler, Adila A. Krisnadhi, Frederick Maier, Cong Wang, Recent Advances in Integrating OWL and Rules (Technical Communication). In: Markus Krötzsch, Umberto Straccia (eds.), Web Reasoning and Rule Systems, 6th International Conference, RR2012, Vienna, Austria, September 10-12, 2012, Proceedings. Lecture Notes in Computer Science Vol. 7497, Springer, Heidelberg, 2012, pp. 225-228.
- Matthias Knorr, Pascal Hitzler, Frederick Maier, Reconciling OWL and Non-monotonic Rules for the Semantic Web. In: De Raedt, L., Bessiere, C., Dubois, D., Doherty, P., Frasconi, P., Heintz, F., Lucas, P. (eds.), ECAI 2012, 20th European Conference on Artificial Intelligence, 27-31 August 2012, Montpellier, France. Frontiers in Artificial Intelligence and Applications, Vol. 242, IOS Press, Amsterdam, 2012, pp. 474-479.
- Markus Krötzsch, Frederick Maier, Adila Alfa Krisnadhi, Pascal Hitzler, A Better Uncle For OWL – Nominal Schemas for Integrating Rules and Ontologies. In: S. Sadagopan, Krithi Ramamritham, Arun Kumar, M.P. Ravindra, Elisa Bertino, Ravi Kumar (eds.), WWW '11 20th International World Wide Web Conference, Hyderabad, India, March/April 2011. ACM, New York, 2011, pp. 645-654.

References

- **Zhangquan Zhou, Guilin Qi, Chang Liu, Pascal Hitzler, Raghava Mutharaju, Reasoning with Fuzzy-EL+ Ontologies Using MapReduce. In: De Raedt, L., Bessiere, C., Dubois, D., Doherty, P., Frasconi, P., Heintz, F., Lucas, P. (eds.), ECAI 2012, 20th European Conference on Artificial Intelligence, 27-31 August 2012, Montpellier, France. Frontiers in Artificial Intelligence and Applications, Vol. 242, IOS Press, Amsterdam, 2012, pp. 933-934.**
- **Raghava Mutharaju, Frederick Maier, Pascal Hitzler, A MapReduce Algorithm for EL+. In: Volker Haarslev, Davind Toman, Grant Weddell (eds.), Proceedings of the 23rd International Workshop on Description Logics (DL2010), Waterloo, Canada, 2010. CEUR Workshop Proceedings Vol. 573, pp. 464-474.**
- **Prateek Jain, Pascal Hitzler, Kunal Verma, Peter Yeh, Amit Sheth, Moving beyond sameAs with PLATO: Partonomy detection for Linked Data. In: Ethan V. Munson, Markus Strohmaier (Eds.): 23rd ACM Conference on Hypertext and Social Media, HT '12, Milwaukee, WI, USA, June 25-28, 2012. ACM, 2012, pp. 33-42.**

References

- Kunal Sengupta, Adila Krisnadhi, Pascal Hitzler, Local Closed World Reasoning: Grounded Circumscription for OWL. In: L. Aroyo, C. Welty, H. Alani, J. Taylor, A. Bernstein, L. Kagal, N. F. Noy, E. Blomqvist (Eds.): The Semantic Web - ISWC 2011 - 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011, Proceedings, Part I. Lecture Notes in Computer Science Vol. 7031, Springer, Heidelberg, 2011, pp. 617-632.
- Prateek Jain, Peter Z. Yeh, Kunal Verma, Reymonrod G. Vasquez, Mariana Damova, Pascal Hitzler, Amit P. Sheth, Contextual Ontology Alignment of LOD with an Upper Ontology: A Case Study with Proton. In: Grigoris Antoniou, Marko Grobelnik, Elena Paslaru Bontas Simperl, Bijan Parsia, Dimitris Plexousakis, Pieter De Leenheer, Jeff Pan (Eds.): The Semantic Web: Research and Applications - 8th Extended Semantic Web Conference, ESWC 2011, Heraklion, Crete, Greece, May 29-June 2, 2011, Proceedings, Part I. Lecture Notes in Computer Science 6643, Springer, 2011, pp. 80-92.