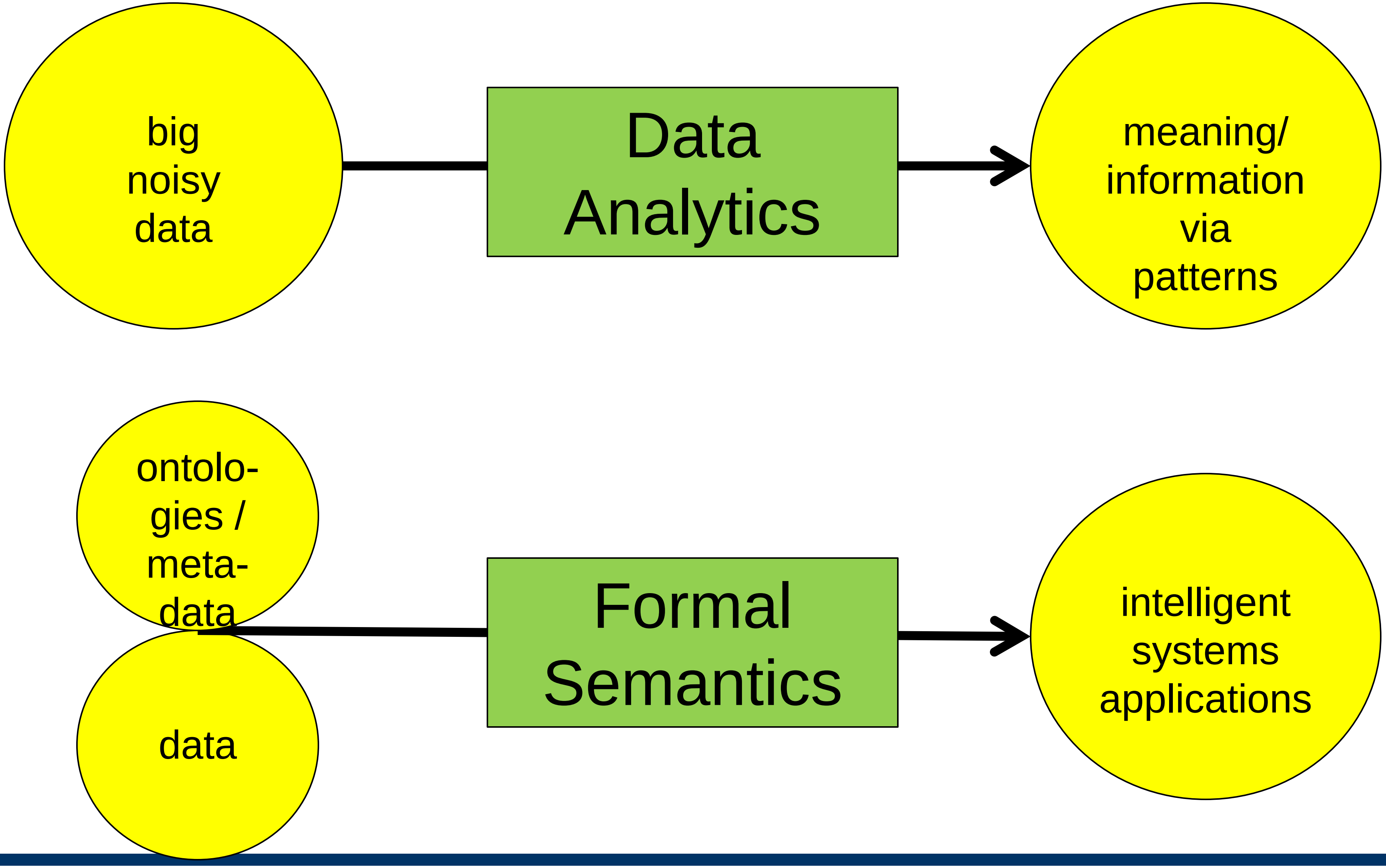
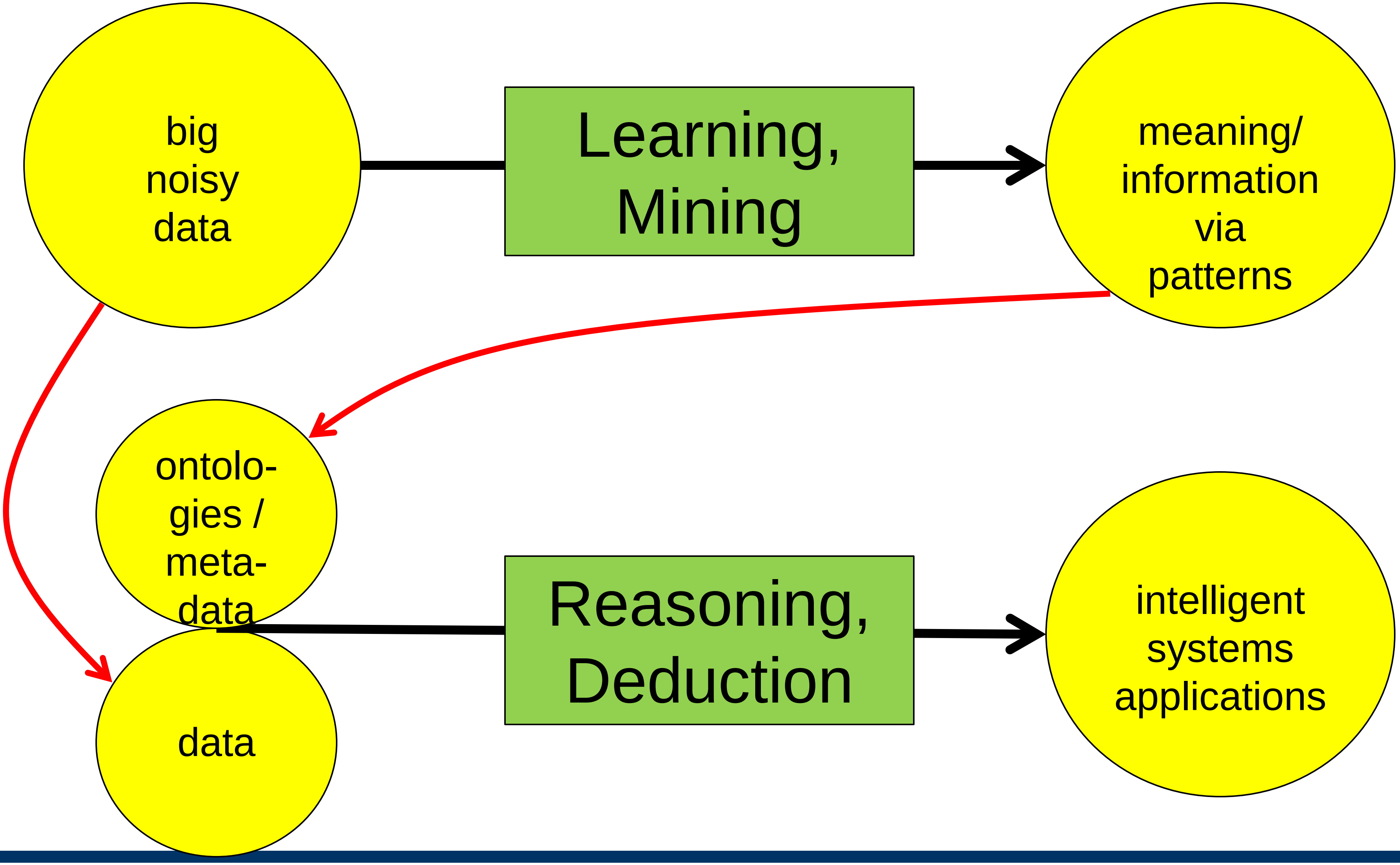
Semantics and Big Data – Opportunities and Challenges

Pascal Hitzler
Kno.e.sis Center
Wright State University, Dayton, OH
http://www.pascal-hitzler.de/

The Big Data Added Value Pipeline

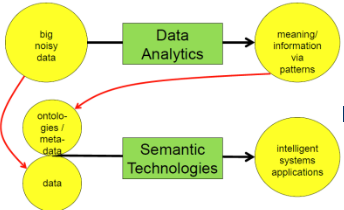


The Big Data Added Value Pipeline

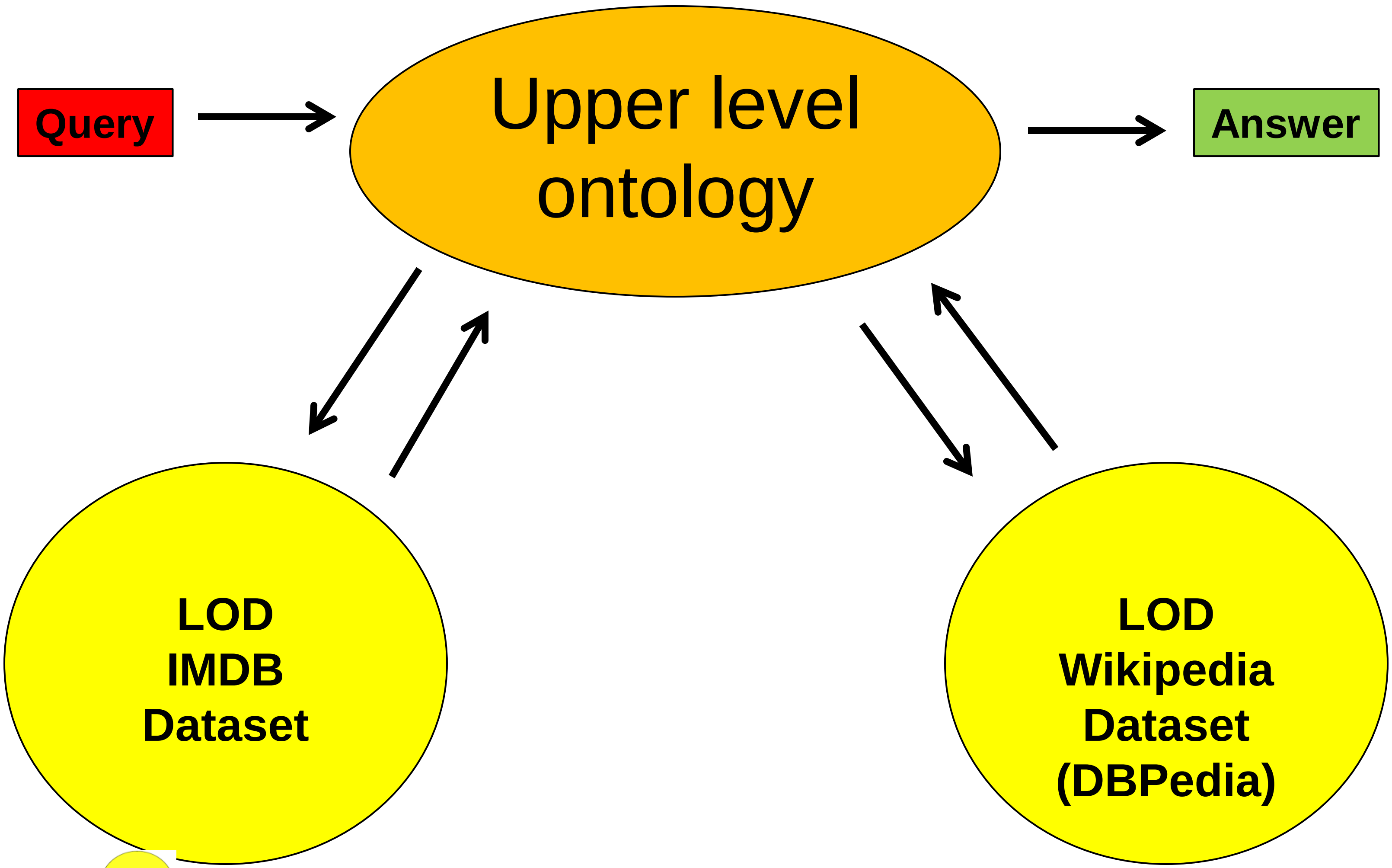


Closing the gap – key dimensions

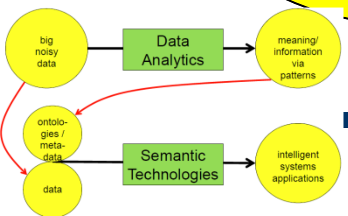
- **scaling tools to Big Data size**
- **learning / acquisition of metadata and ontologies**
- **tolerance for noise and heterogeneity in formal semantics**
- **learning and reasoning paradigms integration**



Linked Data federated querying



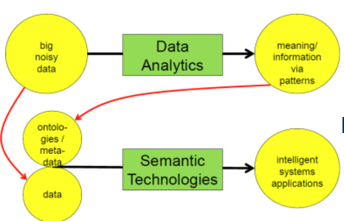
Joshi, Jain, Hitzler et al. ODBASE 2012



Bootstrapping-based alignment

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

Partonomy detection

Relation Type	Distinct Entity Pairs	Correctly Found	Precision
Stuff-Object-Part-Of	4178	3427	0.82
Component-Integral-Part-Of	3126	27931	0.89
Feature-Activity-Part-Of	1287	464	0.85
Member-Collection-Part-Of	1912	803	0.85
Portion-Mass-Part-Of	0	0	NA
Place Area-Part-Of	3350	1248	0.48
Total	13853	10557	0.76

Table 2: Precision of the six different relation types between DBpedia entities

Total # of Class Pairs	Correctly Identified	Precision
93	81	0.87

Table 4: Precision as measured on Schema Level Links Between DBpedia entities

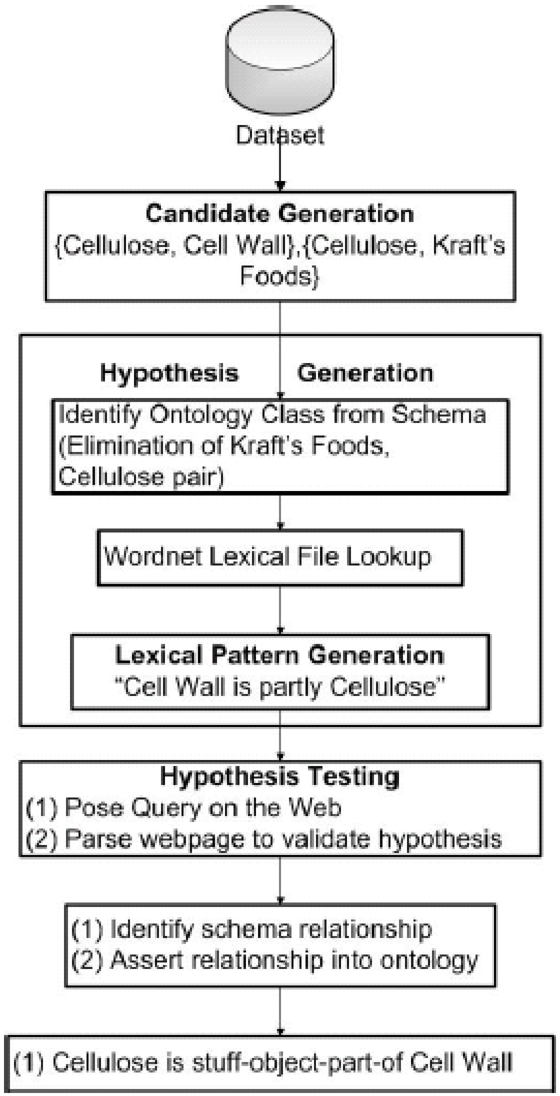
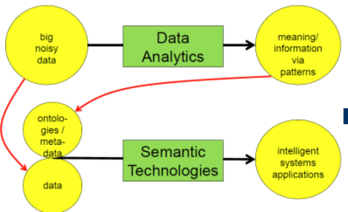


Figure 1: PLATO system flow chart

Jain, Hitzler et al, ACM Hypertext 2012



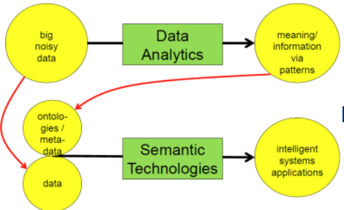
DL-Learner

Generation of schema knowledge from facts / raw data

Employs method from Inductive Logic Programming (ILP)
carried over to Description Logics / OWL

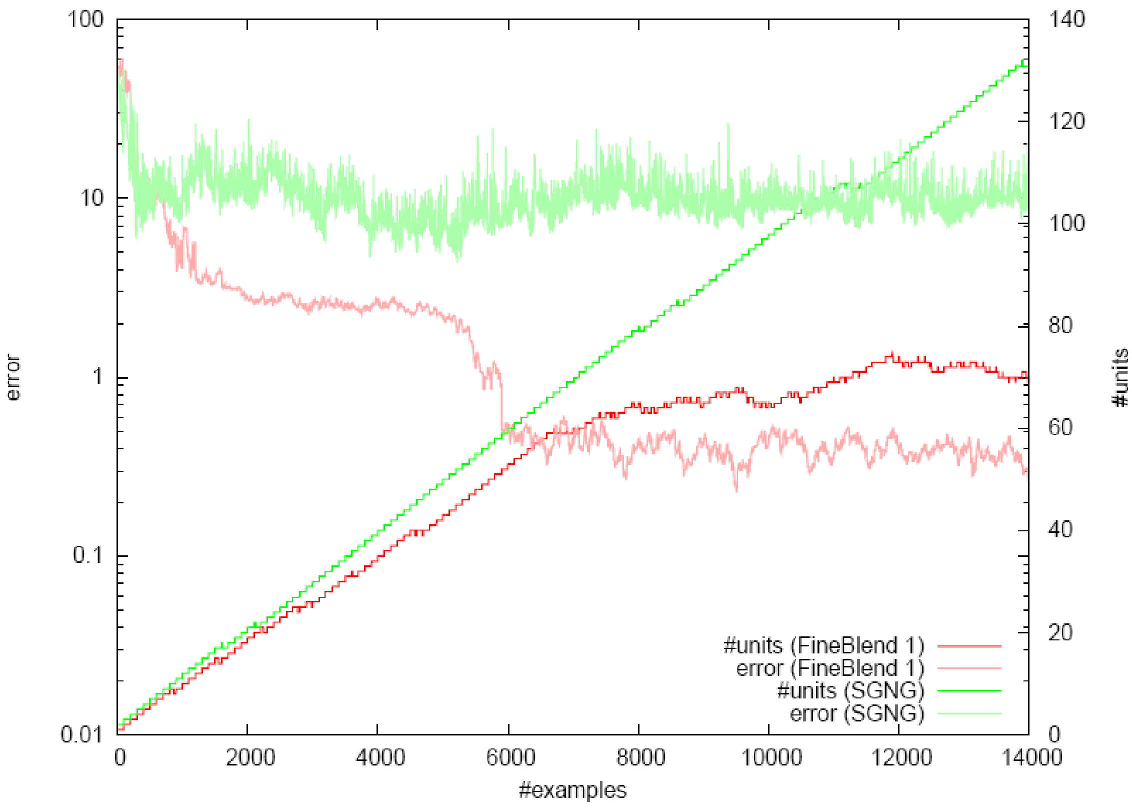
Resulting system DL-Learner is competetive on ILP benchmarks

- **Ontology Engineering Protégé Plugin**
- **DBPedia Navigator**
- **Traditional Machine Learning use cases**

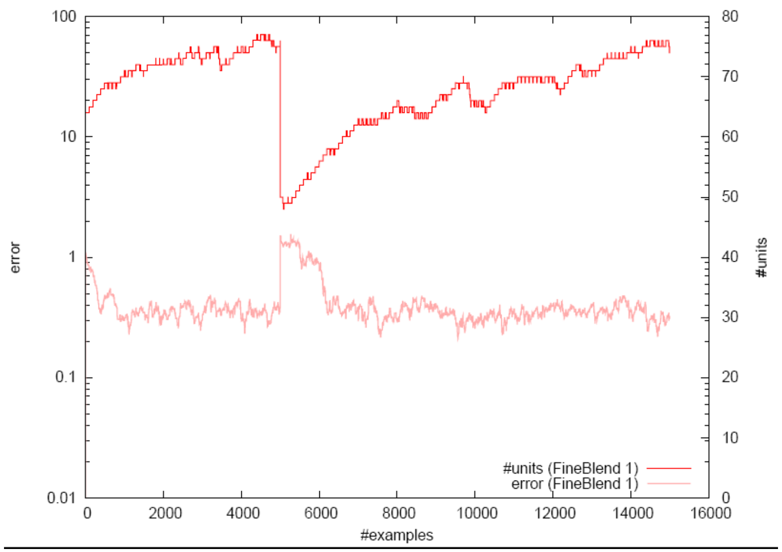


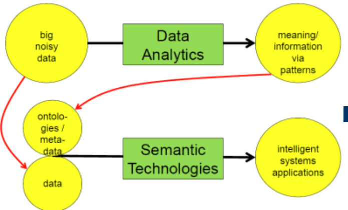
Lehmann, Hitzler, Machine Learning 78(1-2), 203-250, 2010

Neural-Symbolic Integration



unit failure

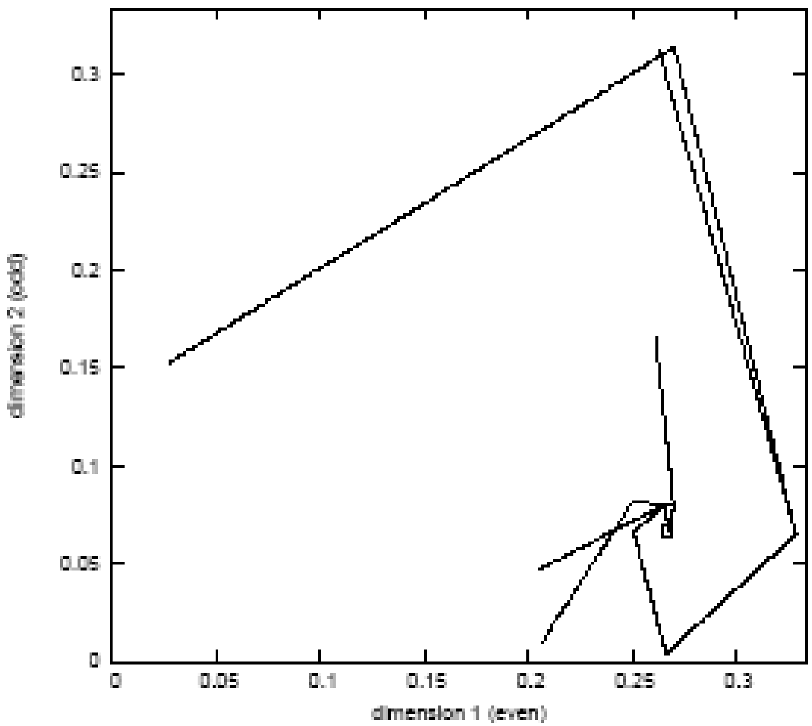


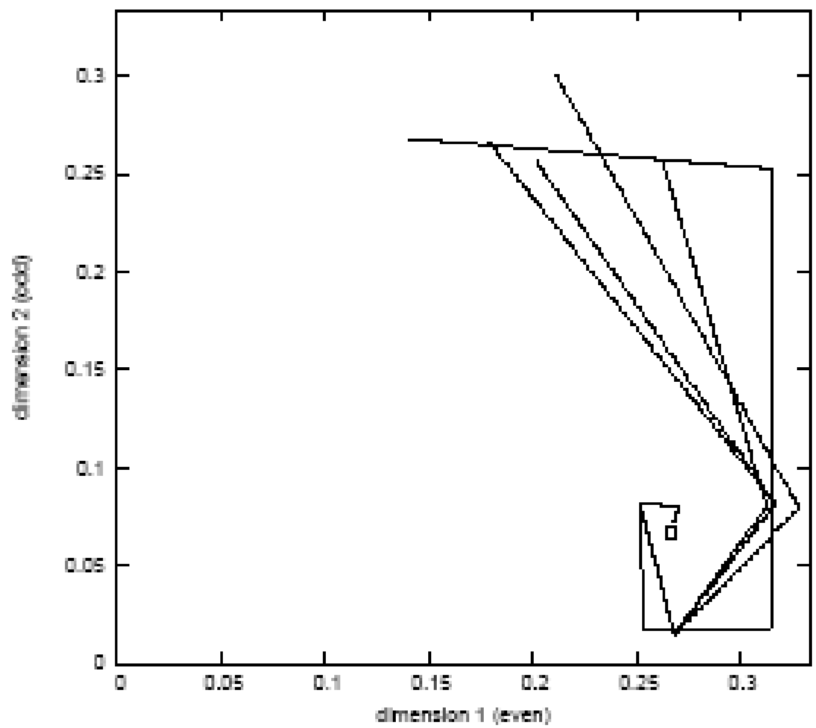


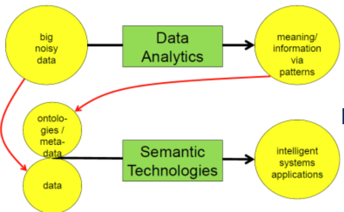
Bader, Hitzler, Hölldobler, Neurocomputing 71, 2420-2432, 2008.

Neural-Symbolic Integration

Iterating Random Inputs: We observe convergence to unique supported model of the program.







Bader, Hitzler, Hölldobler, Neurocomputing 71, 2420-2432, 2008.

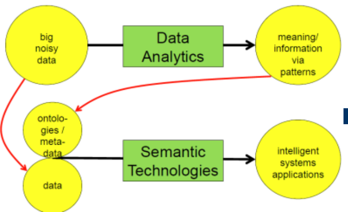
Linked Data Structure Mining

Currently used as a compression method for Linked Data.

Frequent Itemset Mining approach.

Data Set	#triples	#predicates	#transactions	compression ratio	
				intra-property	inter-property
Dog Food	130,178	132	12,695	0.99	0.93
CN 2012	137,484	26	14,553	0.82	0.51
ArchiveHub	431,088	141	51,411	0.92	0.77
Jamendo	1,047,950	25	335,925	0.99	0.83
LinkedMdb	6,147,996	222	694,400	0.97	0.77
DBpedia rdftypes	9,237,320	1	9,237,320	0.49	0.49
RDF About	17,188,323	108	3,132,667	0.97	0.86
DBLP	46,597,620	27	2,840,639	0.96	0.88
Geonames	119,416,854	26	7,711,126	0.97	0.52

Table 1. Compression ratio (based on triple counts) for various linked open datasets



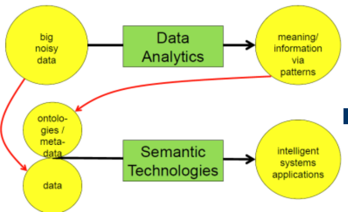
Joshi, Hitzler, Dong, SemQuant 2012

MapReduce for ontology reasoning

Normal Form	Completion Rule	Key
$A_1 \sqcap A_2 \sqsubseteq B$	**R1-1** If $A_1 \in S(X)$ and $A_1 \sqcap A_2 \sqsubseteq B \in \mathcal{O}$ then $P(X) := P(X) \cup \{(A_2, B)\}$	A_1
$(A, B) \in P(X)$	**R1-2** If $A \in S(X)$ and $((A, B) \in P(X)$ or $A \sqsubseteq B \in \mathcal{O})$ then $S(X) := S(X) \cup \{B\}$	A
$A \sqsubseteq \exists r.B$	**R2** If $A \in S(X)$ and $A \sqsubseteq \exists r.B \in \mathcal{O}$ then $R(r) := R(r) \cup \{(X, B)\}$	A
$\exists r.A \sqsubseteq B$ for A	**R3-1** If $A \in S(X)$ and $\exists r.A \sqsubseteq B \in \mathcal{O}$ then $\mathcal{O} := \mathcal{O} \cup \{\exists r.X \sqsubseteq B\}$	A
$\exists r.A \sqsubseteq B$ for r	**R3-2** If $(X, Y) \in R(r)$ and $\exists r.Y \sqsubseteq B \in \mathcal{O}$ then $S(X) := S(X) \cup \{B\}$	r (or Y)
$r \sqsubseteq s$	**R4** If $(X, Y) \in R(r)$ and $r \sqsubseteq s \in \mathcal{O}$ then $R(s) := R(s) \cup \{(X, Y)\}$	r
$r \circ s \sqsubseteq t$	**R5-1** If $(X, Z) \in R(r)$ and $(Z, Y) \in R(s)$ then $R(r \circ s) := R(r \circ s) \cup \{(X, Y)\}$	Z

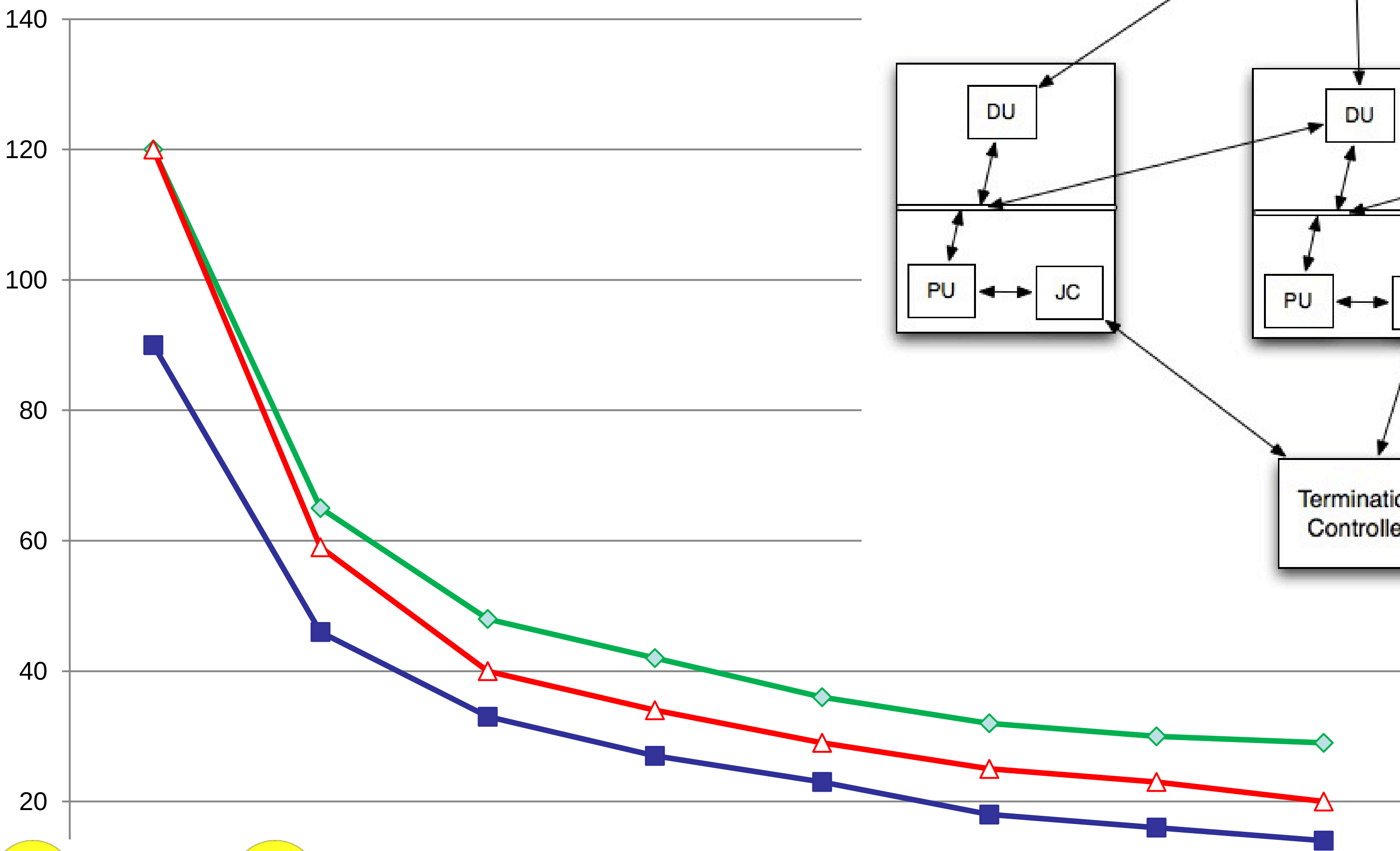
Fig. 2. Revised CEL algorithm for $\mathcal{EL}^+$. The keys are used in the MapReduce algorithm. Note that in R4, r is allowed to be compound, i.e., of the form $s \circ t$.

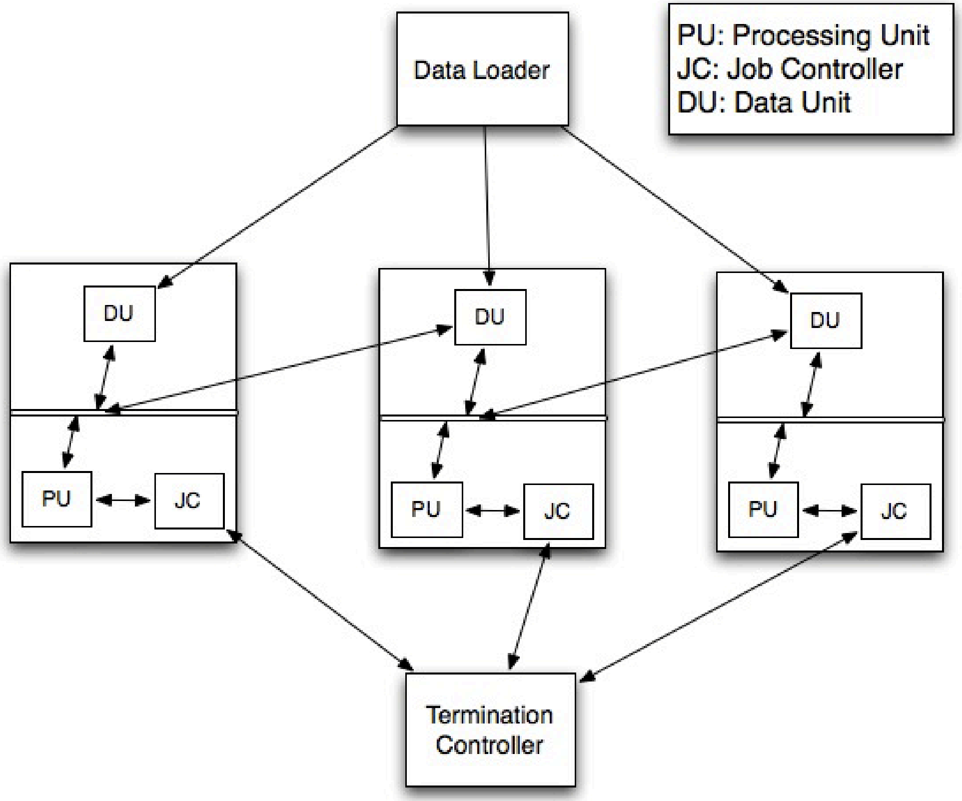
Mutharaju, Maier, Hitzler, DL2010
Zhou, Qi, Liu, Hitzler, Mutharaju, ECAI 2012

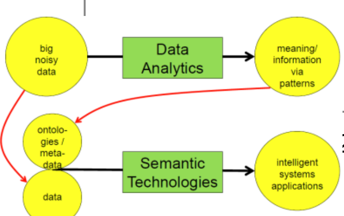


Parallelized ontology reasoning

work in progress





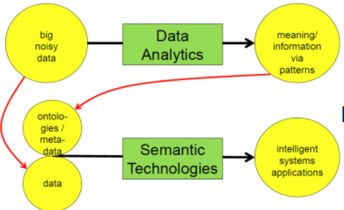


Inconsistency-tolerant reasoning

- **Inconsistency-tolerant semantics for the Web Ontology Language.**
- **Compatible with the standard semantics.**

- **Implementation by linear pre-processing, using off-the-shelf OWL reasoner.**

Maier, Ma, Hitzler, Semantic Web journal, to appear
Huang, Li, Hitlzer, Logic Journal of the IGPL, to appear



Multi-perspective alignment

$$a{:}hasWife \sqsubseteq a{:}hasSpouse$$
$$\mathrm{symmetric}(a{:}hasSpouse)$$
$$\exists a{:}hasSpouse.a{:}Female \sqsubseteq a{:}Male$$
$$\exists a{:}hasSpouse.a{:}Male \sqsubseteq a{:}Female$$
$$a{:}hasWife(a{:}john, a{:}mary)$$
$$b{:}Male(a{:}john)$$
$$b{:}Female(a{:}mary)$$
$$a{:}Male \sqcap a{:}Female \sqsubseteq \bot$$

$$\mathrm{symmetric}(b{:}hasSpouse)$$
$$b{:}hasSpouse(b{:}mike, b{:}david)$$
$$b{:}Male(b{:}david)$$
$$b{:}Male(b{:}mike)$$
$$b{:}Female(b{:}anna)$$

- **Bridging semantic heterogeneity by employing defeasible alignment rules:** $a{:}hasSpouse(x, y) \overset{\text{default}}{\longleftrightarrow} b{:}hasSpouse(x, y)$

Hitzler, Janowicz et al. work in progress
Knorr, Alferes, Hitzler, Artificial Intelligence 175(9-10), 1528-1554, 2011
Knorr, Hitzler, Maier, ECAI 2012
Sengupta, Krisnadhi, Hitzler, ISWC 2011

Reasoning as IR

- **Deductive reasoning can be viewed as a classification task:**

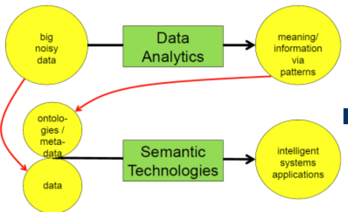
  Input: Knowledge base (Ontology + Data)
  plus a (ground) query

  Output: yes or no

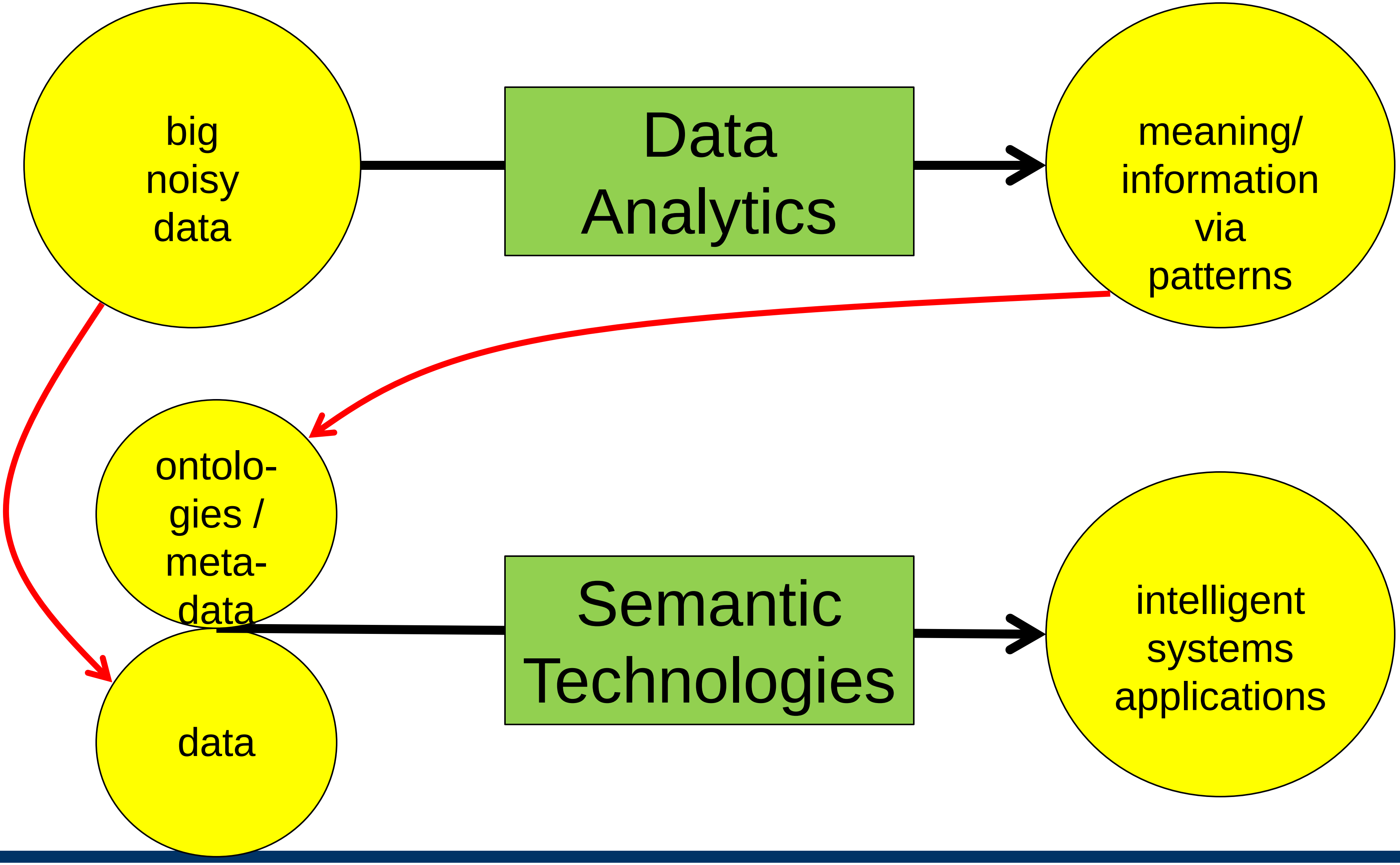
- **To what extent can non-deductive IR methods be used to achieve this?**
- **Deductive reasoning systems available as benchmarks.**

- **IR methods could potentially lift “deductive” reasoning to noisy data?**

Hitzler, van Harmelen, Semantic Web 1(1-2), 39-44, 2010 (vision paper)



Thanks!


big
noisy
data
Data
Analytics
meaning/
information
via
patterns
ontolo-
gies /
meta-
data
data
Semantic
Technologies
intelligent
systems
applications

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.

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.