Semantic Web and Machine Learning

Time to re-sync

Filip Železný, Czech Technical University in Prague

Bottom Line of the Talk

“Ask not what machine learning can do for you,
ask what you can do for machine learning.”

I will show how people try to make the semantic web understandable to machine learning algorithms.

Machine Learning

Data: ground observations, facts

Knowledge: patterns allowing *inference*

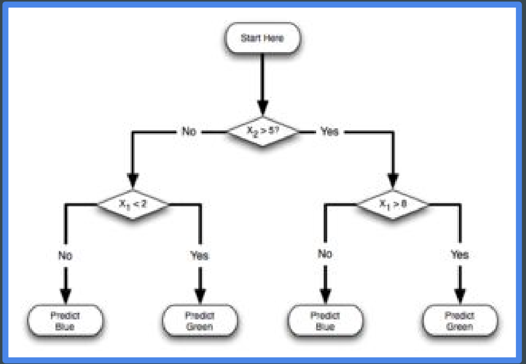
How data is (usually) encoded in Machine Learning

Name	Thread pitch (mm)	Minor diameter tolerance	Nominal diameter (mm)	Head shape	Price for 50 screws	Available at factory outlet?	Number in stock	Flat or Phillips head?
M4	0.7	4g	4	Pan	$10.08	Yes	276	Flat
M5	0.8	4g	5	Round	$13.89	Yes	183	Both
M6	1	5g	6	Button	$10.42	Yes	1043	Flat
M8	1.25	5g	8	Pan	$11.98	No	298	Phillips
M10	1.5	6g	10	Round	$16.74	Yes	488	Phillips
M12	1.75	7g	12	Pan	$18.26	No	998	Flat
M14	2	7g	14	Round	$21.19	No	235	Phillips
M16	2	8g	16	Button	$23.57	Yes	292	Both
M18	2.1	8g	18	Button	$25.87	No	664	Both
M20	2.4	8g	20	Pan	$29.09	Yes	486	Both
M24	2.55	9g	24	Round	$33.01	Yes	982	Phillips
M28	2.7	10g	28	Button	$35.66	No	1067	Phillips
M36	3.2	12g	36	Pan	$41.32	No	434	Both
M50	4.5	15g	50	Pan	$44.72	No	740	Flat

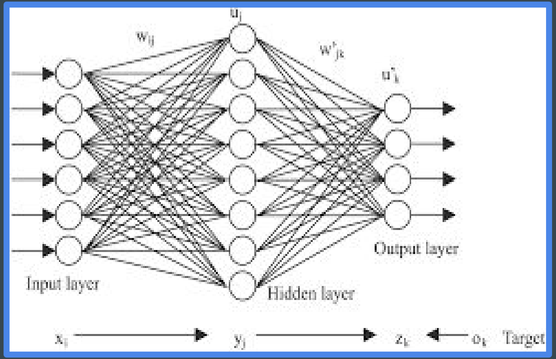
Attributes (features)

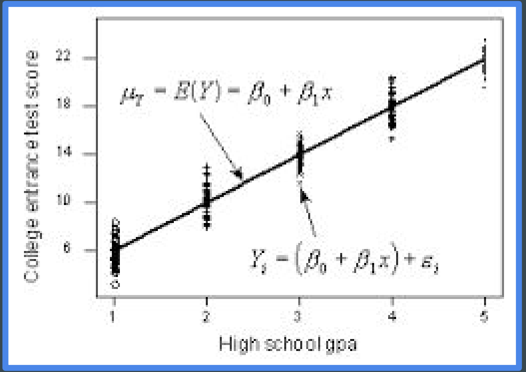
Independent samples

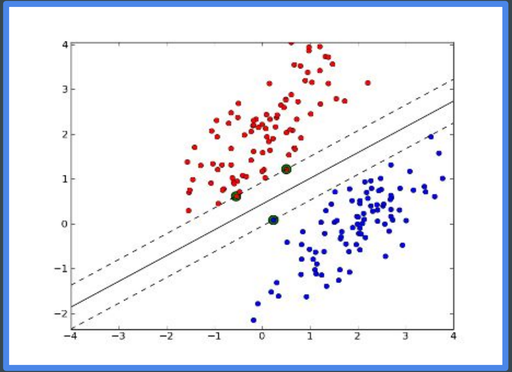
How knowledge is encoded in Machine Learning

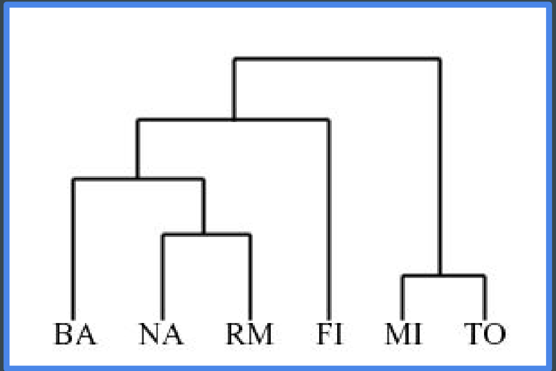


- IF (humidity = high) and (outlook = sunny) THEN play=no (3.0/0.0)
- IF (outlook = rainy) and (windy = TRUE) THEN play=no (2.0/0.0)
- OTHERWISE play=yes (9.0/0.0)

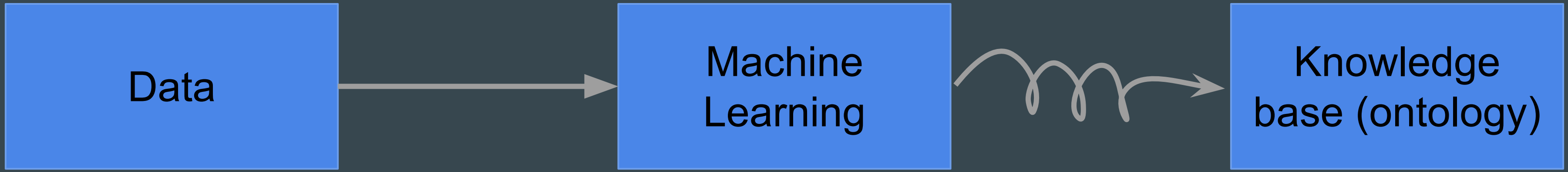






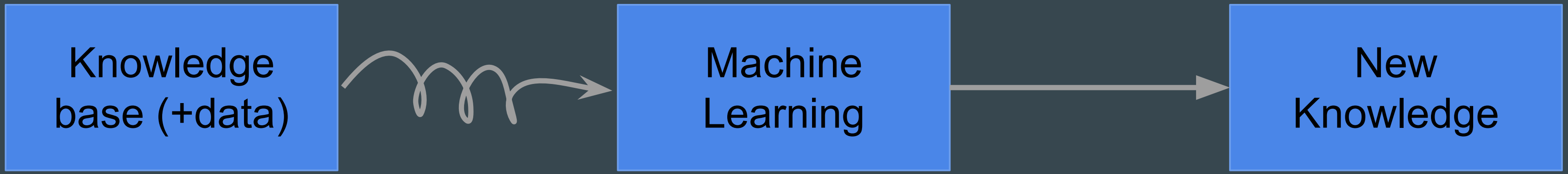


Semantic Web and Machine Learning (1)



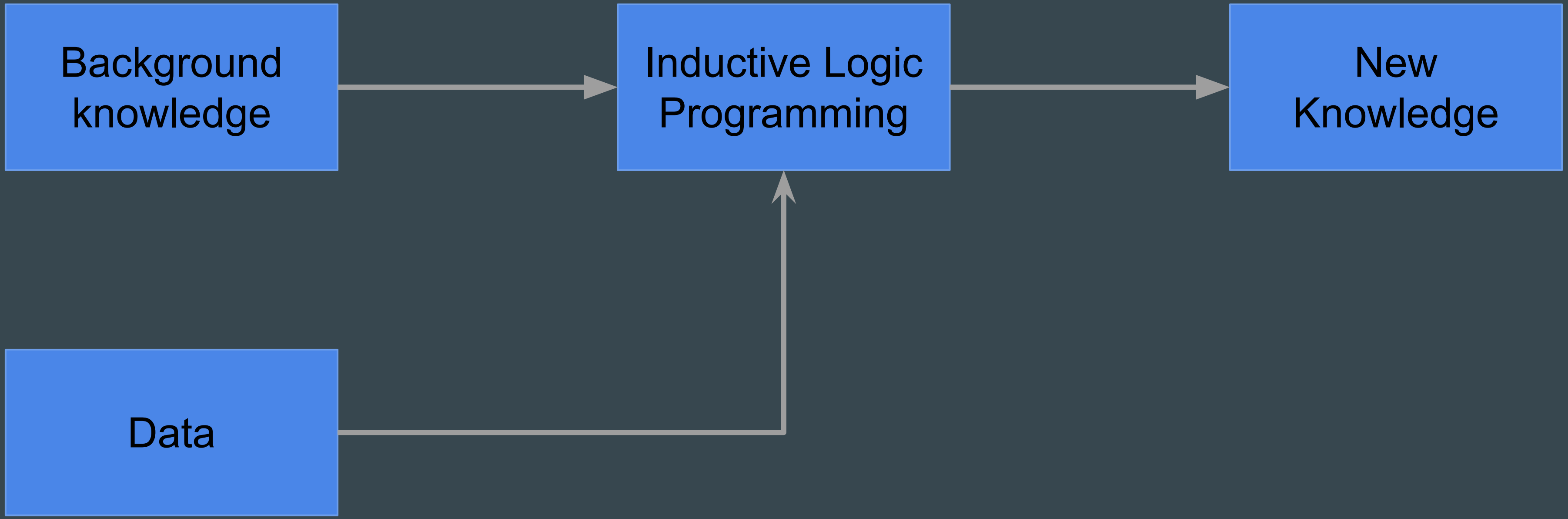
- Learning **concept taxonomies** through **hierarchical clustering**
- Learning **deep annotation** rules, concept/relation population
- Ontology **alignment**
- **Duplicate** detection
- Etc.
- Not the scope of this talk

Semantic Web and Machine Learning (2)

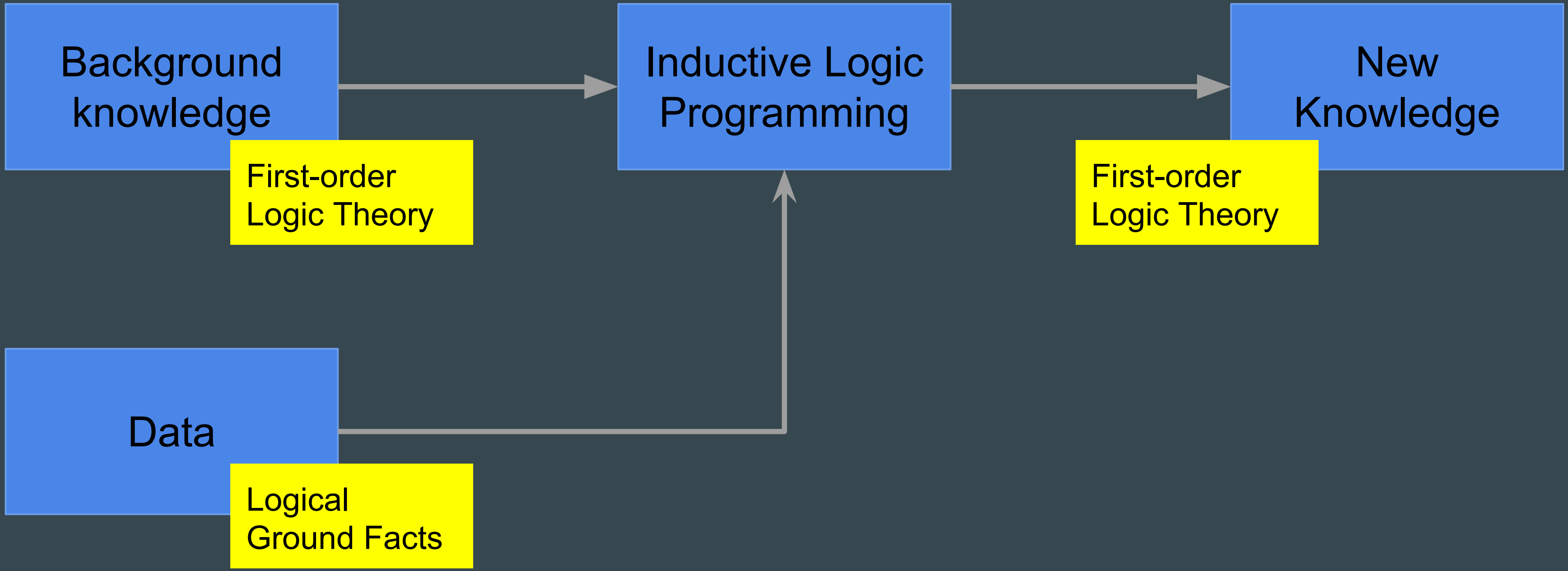


- Lots of motivation: FreeBase, DBpedia, Google Knowledge Graph, ...
- Calls for learning algorithms which, apart from data, understand formal knowledge
- Taxonomies, rules, etc.

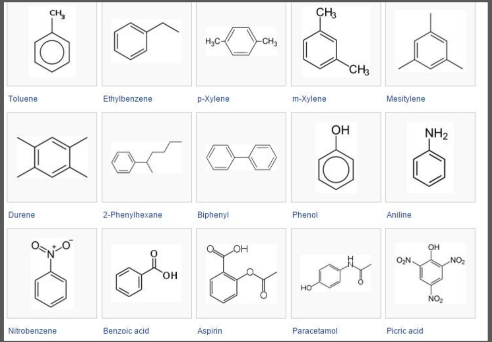
Inductive Logic Programming



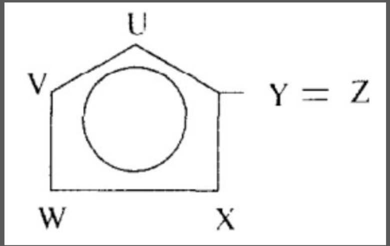
Inductive Logic Programming

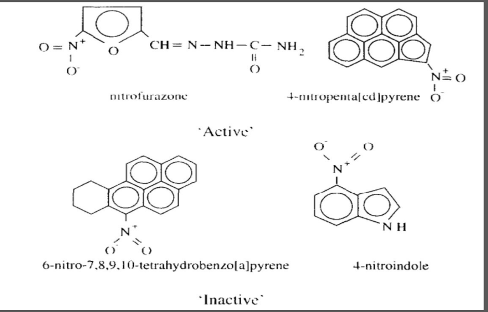


ILP: example

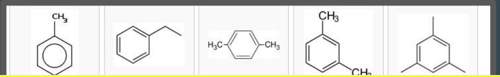


Inductive Logic Programming



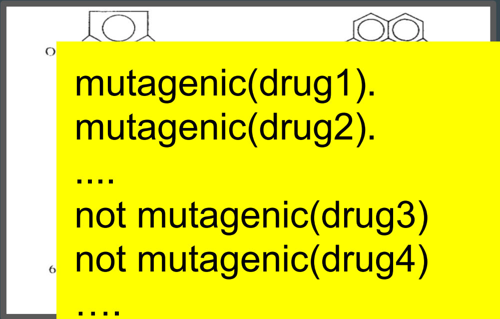


ILP: example



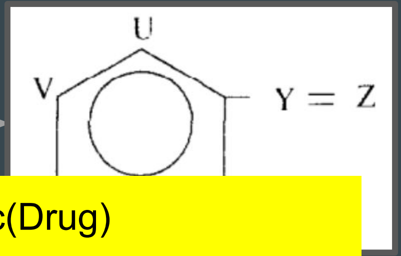
atom(drug1, c1, carbon).
dbond(drug1, c1, c2).

anthracene(Drug, [Ring1,Ring2])
IF
benzene(Drug,Ring1),
benzene(Drug,Ring2).



mutagenic(drug1).
mutagenic(drug2).
....
not mutagenic(drug3)
not mutagenic(drug4)
….

Inductive Logic Programming

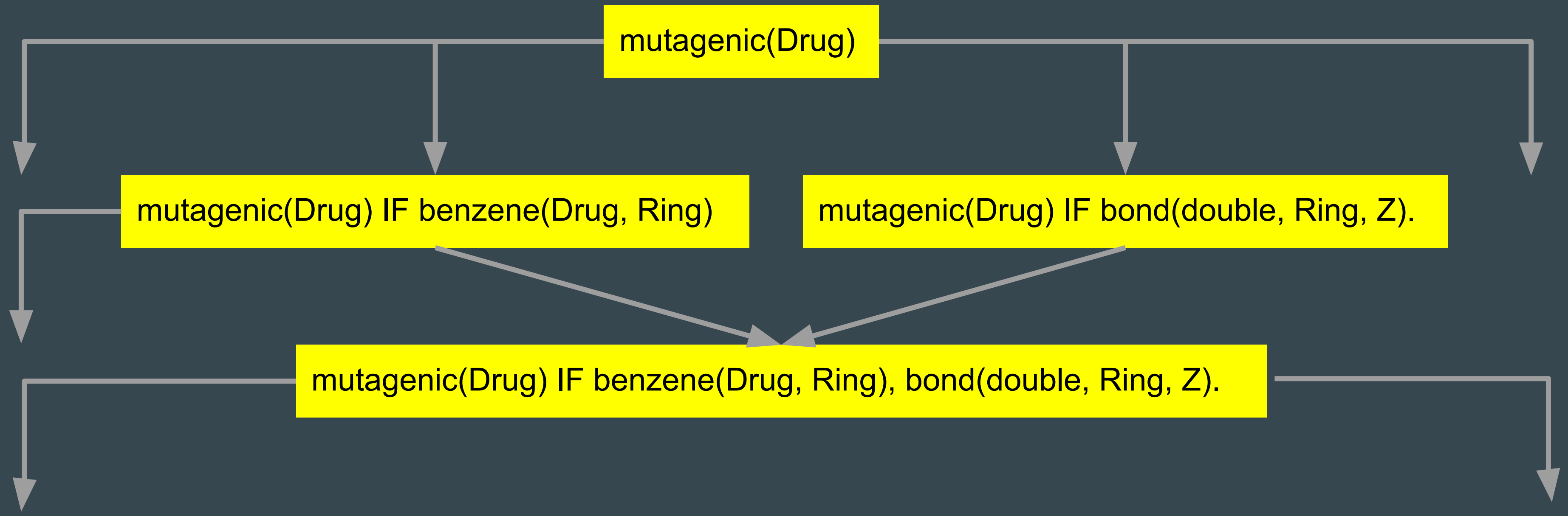


mutagenic(Drug)
IF
benzene(Drug, Ring),
bond(double, Ring, Z).

Srinivasan et al., Artif. Intell. 1996

ILP: basic principle

Search through a subsumption lattice



ILP: discussion

- Energetic research since the 90’s
- Mainly in Europe
- Annual ILP conferences (cca 50 people)
    - ILP’08: Semantic Web Keynote by Frank Harmelen
- Current focus: combining with probabilistic inference
    - Probabilistic ILP, Statistical Relational Learning

- Main issues of ILP:
    - **Scalability**, handling **uncertainty**, **numerical** reasoning
    - **First-order** logic (mainly Datalog) not a semantic web standard

ILP with Description Logics

DL-Learner: learns concept descriptions from examples (concept instances) and counterexamples, and an OWL-DL ontology

Principle: search through a subsumption lattice as in ILP

$$\mathrm{T} \rightarrow \mathrm{Person} \rightarrow \mathrm{Person} \sqcap \exists\ \mathrm{ParticipatesIn.Event} \rightarrow \mathrm{Person} \sqcap \exists\ \mathrm{ParticipatesIn.Conference}$$

Hellman et al. ISWC 2008
Lehman et al, Jr. Mach. Learn. Res., 2009

Taxonomies as Guidance in ILP

- Rule search co-guided by the Gene Ontology

expressed(Gene) IF **ion_binding**(Gene)

expressed(Gene) IF
ion_binding(Gene), interacts(Gene, cd4)

zinc_ion_binding ⊆ ion_binding

expressed(Gene) IF **zinc_ion_binding**(Gene)

M. Zakova et al, ECML 2007

- General mechanism implemented in ILP system Aleph

A. Vavpetic, PhD thesis, IJS Ljubljana 2016

A Full Hybrid Approach

Learning views (rules) from a Datalog base AND an ontology

```
famous(Mary)
famous(Paul)
famous(Joe)
scientist(Joe)
```

$\mathtt{RICH(X)} \leftarrow \mathtt{famous(X), not\ scientist(X)}$

$\mathtt{RICH} \sqcap \mathtt{UNMARRIED} \sqsubseteq \exists\, \mathtt{WANTS\text{-}TO\text{-}MARRY}^{-}.\top$

$\mathtt{WANTS\text{-}TO\text{-}MARRY} \sqsubseteq \mathtt{LOVES}$

$\mathtt{UNMARRIED(Mary)}$

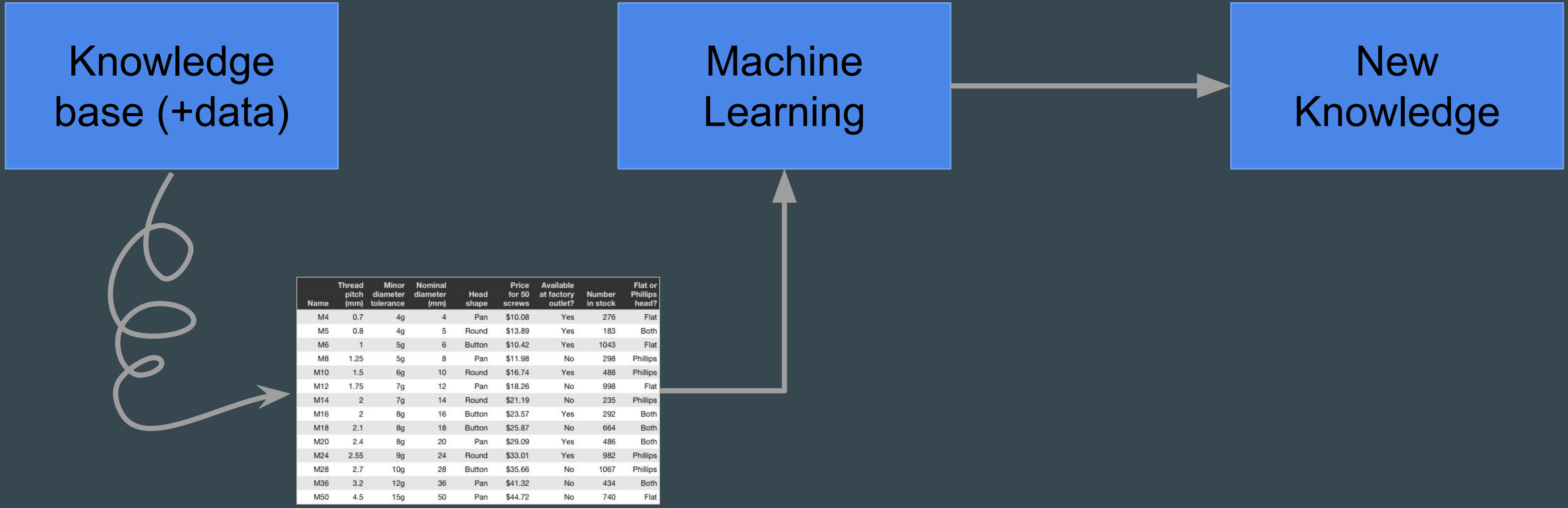
$\mathtt{UNMARRIED(Joe)}$

Example e.g. happy(Mary), counter-example happy(Paul)

$\longrightarrow$ $\mathtt{happy(X)} \leftarrow \mathtt{famous(X), WANTS\text{-}TO\text{-}MARRY(Y,X)}$

F. Lisi: TPLP, 2010

Semantic Web and Machine Learning: Workaround



Name	Thread pitch (mm)	Minor diameter tolerance	Nominal diameter (mm)	Head shape	Price for 50 screws	Available at factory outlet?	Number in stock	Flat or Phillips head?
M4	0.7	4g	4	Pan	$10.08	Yes	276	Flat
M5	0.8	4g	5	Round	$13.89	Yes	183	Both
M6	1	5g	6	Button	$10.42	Yes	1043	Flat
M8	1.25	5g	8	Pan	$11.98	No	298	Phillips
M10	1.5	6g	10	Round	$16.74	Yes	488	Phillips
M12	1.75	7g	12	Pan	$18.26	No	998	Flat
M14	2	7g	14	Round	$21.19	No	235	Phillips
M16	2	8g	16	Button	$23.57	Yes	292	Both
M18	2.1	8g	18	Button	$25.87	No	664	Both
M20	2.4	8g	20	Pan	$29.09	Yes	486	Both
M24	2.55	9g	24	Round	$33.01	Yes	982	Phillips
M28	2.7	10g	28	Button	$35.66	No	1067	Phillips
M36	3.2	12g	36	Pan	$41.32	No	434	Both
M50	4.5	15g	50	Pan	$44.72	No	740	Flat

Produce a data table which also *accounts for* background knowledge.
Then use your favorite, fantastic machine learning algorithm.

Propositionalization

- An ILP-inspired technique to automate the workaround
- First-order features constructed via combinatorial search
- Each is true or false for a particular sample

	Feature 1	Feature 2	Feature 3	class
Sample 1	+	-	-	+
Sample 2	-	+	-	-

Propositionalization

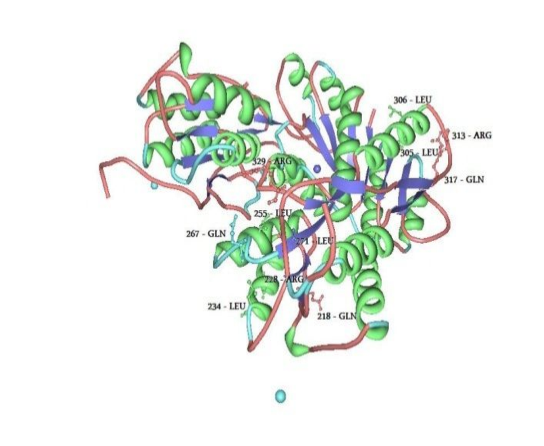
- A technique to automate the workaround
- First-order features constructed via *combinatorial search*
- Each is true or false for a particular sample

benzene(Drug, Ring1), benzene(Drug, Ring2), not Ring1=Ring2

	Feature 1	Feature 2	Feature 3	mutagenic?
drug 1	+	-	-	+
drug 2	-	+	-	-

Propositionalization: state of the art

- **Treeliker** : a fast algorithm for propositionalization
- Open SW, comes with documentation and example data, google it
- Produces (up to) very complex features efficiently, from very complex data/knowledge
- Prolog/Datalog representation, no support for SW standards (yet)



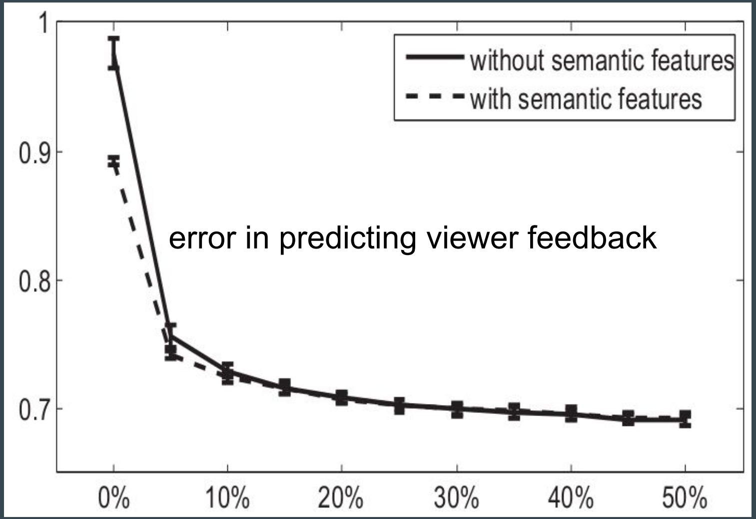
res(res_seq1_1_A, lys), res(res_seq1_2_A , trp), dist(res_seq1_1_A, res_seq1_2_A, 4.0), res(res_seq1_3_A , lys), dist(res_seq1_1_A, res_seq1_3_A, 6.0), res(res_seq1_4_A , leu), dist(res_seq1_1_A, res_seq1_4_A, 6.0), res(res_seq1_5_A, trp), …

Kuzelka et al., Machine Learning 2011; ECML 2011
Szaboova et al., Proteome Sci. 2012

Ontopropositionalization (?)

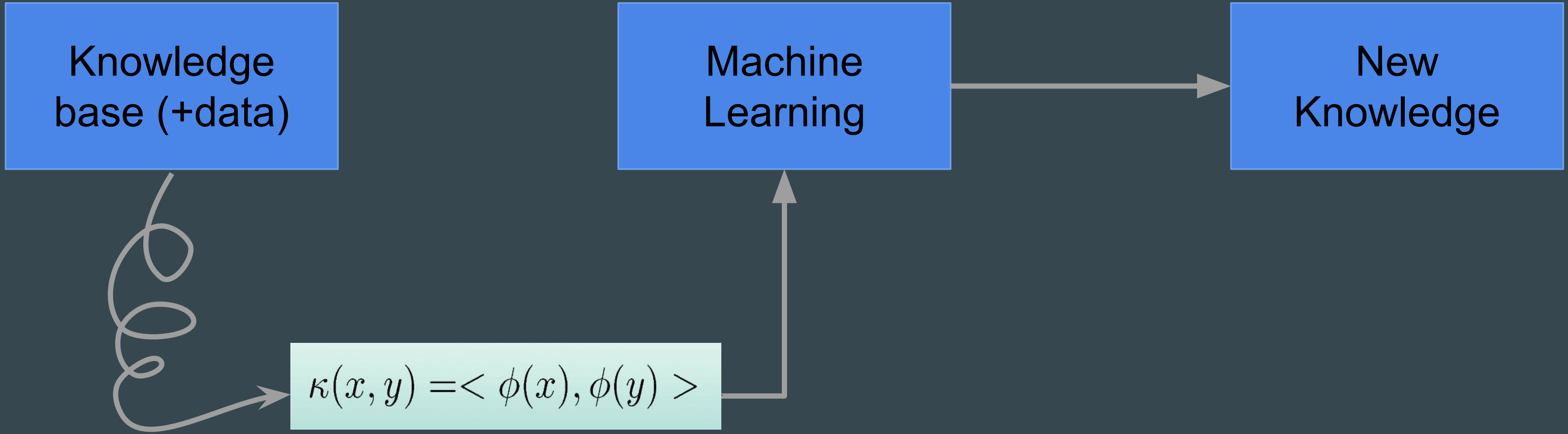
```
SELECT ?e ?r
WHERE { (m  ?r  ?e) UNION (?e  ?r  m) }
```

	Ø(Nikita)	Ø(Léon)	Ø(Black Swan)
European Movie, subClass	1	1	0
American Movie, subClass	0	0	1
Action Movie, hasGenre	1	1	0
Thriller Movie, hasGenre	0	0	1
French Movie, type	1	1	0
Hollywood Movie, type	0	0	1
Jean Reno, actedIn	1	1	0
Anne Parillaud, actedIn	1	0	0
Patrice Ledoux, produced	1	1	0
Luc Besson, directed	1	1	0
Scott Franklin, produced	0	0	1
Mark Heyman, directed	0	0	1
Natalie Portman, actedIn	1	0	1



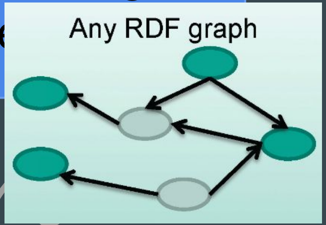
W. Cheng, CKMI 2011

Kernels for RDF Data: another Workaround

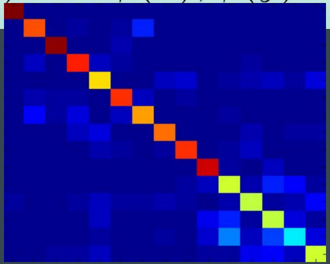


Kernels for RDF Data

Knowledge base

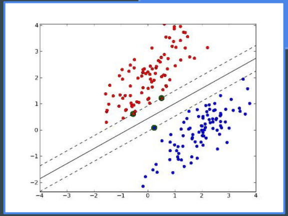


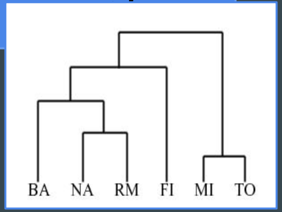
$$\kappa(x, y) = < \phi(x), \phi(y) >$$



Machine Learning

New Knowledge



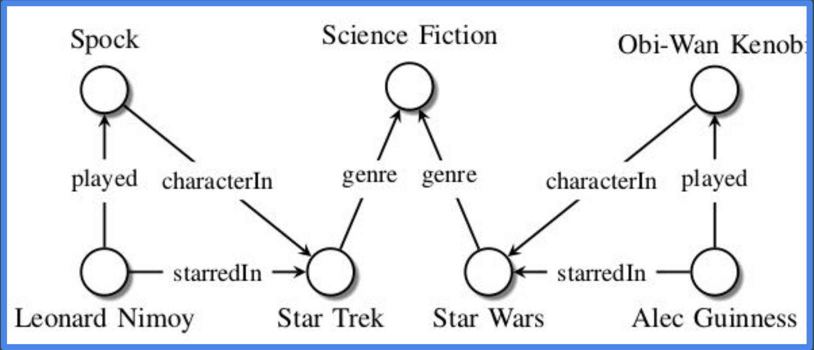


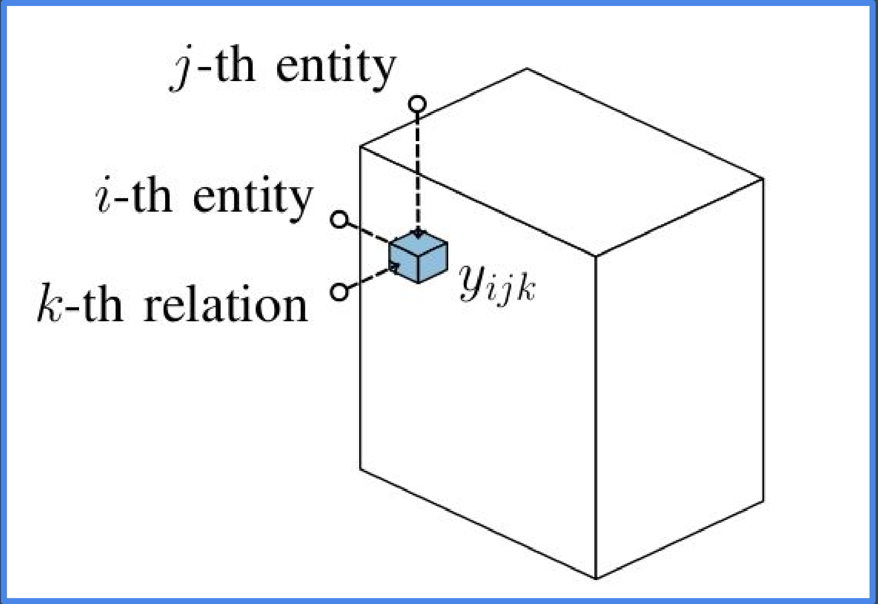
Used successfully for property-value and link prediction.

U. Loesch: ESWC 2012

Knowledge Graph as a Tensor

subject	*predicate*	*object*
(LeonardNimoy,	*profession,*	*Actor)*
(LeonardNimoy,	*starredIn,*	*StarTrek)*
(LeonardNimoy,	*played,*	*Spock)*
(Spock,	*characterIn,*	*StarTrek)*
(StarTrek,	*genre,*	*ScienceFiction)*

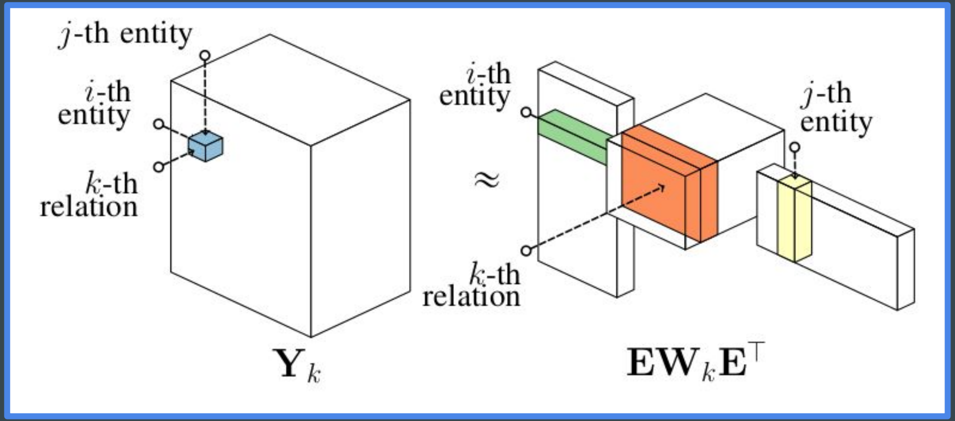




Pictures from: Nickel et al.: A Review of Relational ML for Knowledge Graphs, ArXiv 2015

Probabilistic Model

- The RESCAL bilinear model



$$f_{ijk}^{\text{RESCAL}} := \mathbf{e}_i^\top \mathbf{W}_k \mathbf{e}_j = \sum_{a=1}^{H_e} \sum_{b=1}^{H_e} w_{abk} e_{ia} e_{jb}$$

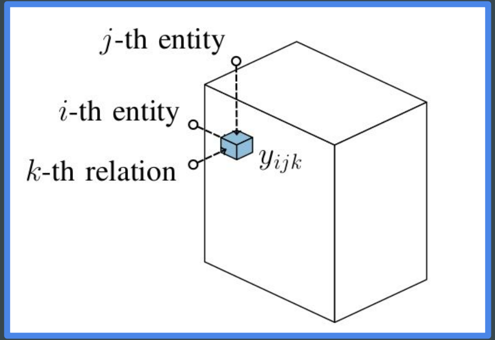
Learned vector embeddings (latent variables)

Result: Similar vectors - similar semantics

Nickel et al., ICML 2011
Pictures from: Nickel et al.: A Review of Relational ML for Knowledge Graphs, ArXiv 2015

Probabilistic Model

- The ER-MLP model (Google's "Knowledge Vault")



$$
\begin{aligned}
f_{ijk}^{\text{ER-MLP}} &:= \mathbf{w}^\top \mathbf{g}(\mathbf{h}_{ijk}^{c}) \\
\mathbf{h}_{ijk}^{c} &:= \mathbf{C}^\top \phi_{ijk}^{\text{ER-MLP}} \\
\phi_{ijk}^{\text{ER-MLP}} &:= [\mathbf{e}_i; \mathbf{e}_j; \mathbf{r}_k].
\end{aligned}
$$

Prediction

Interactions

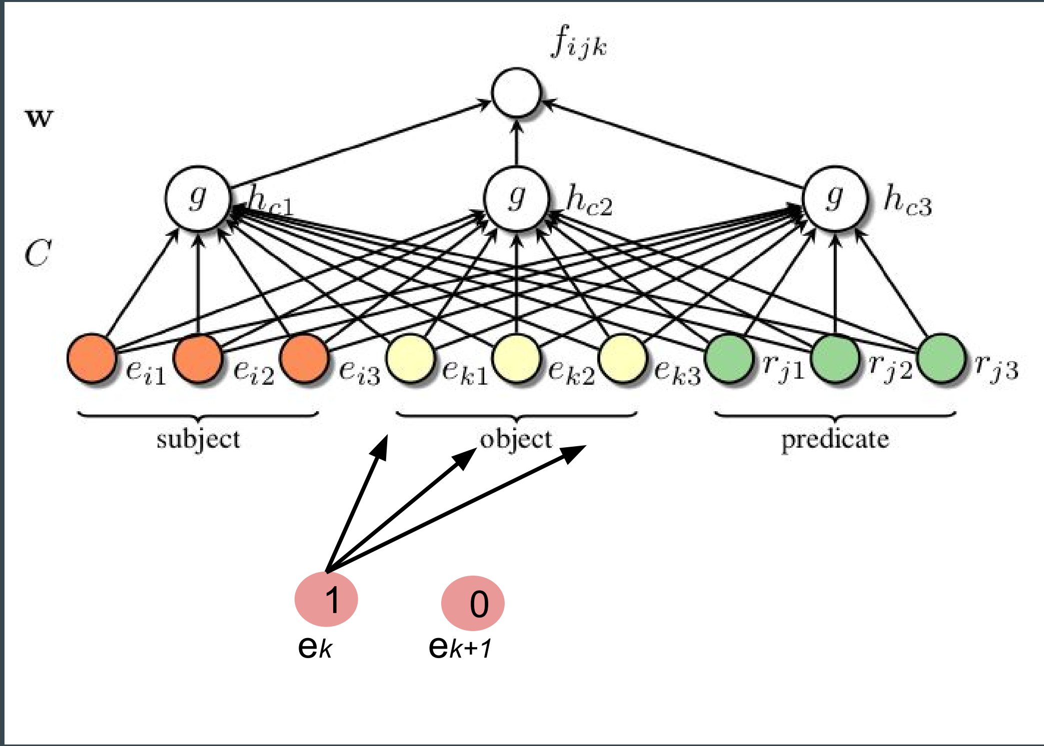
Embeddings

Learned vectors
(latent variables)

Dong et al., KDD 2014
Pictures from: Nickel et al.: A Review of Relational ML for Knowledge Graphs, ArXiv 2015

Probabilistic Model

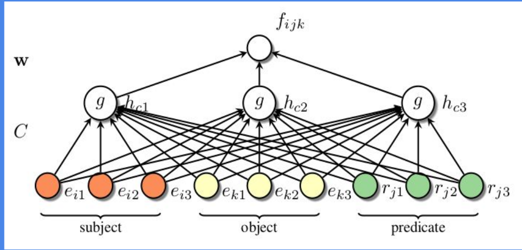
$$\begin{aligned} f_{ijk}^{\text{ER-MLP}} &:= \mathbf{w}^\top \mathbf{g}(\mathbf{h}_{ijk}^c) \\ \mathbf{h}_{ijk}^c &:= \mathbf{C}^\top \phi_{ijk}^{\text{ER-MLP}} \\ \phi_{ijk}^{\text{ER-MLP}} &:= [\mathbf{e}_i; \mathbf{e}_j; \mathbf{r}_k]. \end{aligned}$$



Dong et al., KDD 2014
Pictures from: Nickel et al.: A Review of Relational ML for Knowledge Graphs, ArXiv 2015

Embeddings Capture Semantics

$$
\begin{aligned}
f_{ijk}^{\text{ER-MLP}} &:= \mathbf{w}^\top \mathbf{g}(\mathbf{h}_{ijk}^c) \\
\mathbf{h}_{ijk}^c &:= \mathbf{C}^\top \phi_{ijk}^{\text{ER-MLP}} \\
\phi_{ijk}^{\text{ER-MLP}} &:= [\mathbf{e}_i; \mathbf{e}_j; \mathbf{r}_k].
\end{aligned}
$$



Relation	Nearest Neighbors					
children	parents	(0.4)	spouse	(0.5)	birth-place	(0.8)
birth-date	children	(1.24)	gender	(1.25)	parents	(1.29)
edu-end[10]	job-start	(1.41)	edu-start	(1.61)	job-end	(1.74)

Dong et al., KDD 2014

Pictures from: Nickel et al.: A Review of Relational ML for Knowledge Graphs, ArXiv 2015

Other Recent Interesting Approaches

- Gaussian embeddings
  - Dos Santos et al, ECML 2016
  - Embed entities as normal random vars
  - Conflicting “gradient forces” result in greater variance
- “Injecting” logical knowledge into conventional learning paradigms
  - Diligenti et al, MLJ 2012: Horn rules to regularize SVM’s
  - Rocktaeschel et al. NAACL 2015: Tensor factorization guided by logical rules
- General, encompassing SRL frameworks
  - Such as Markov Logic Networks [Richardson, Domingos - MLJ 2006]
  - Scalability issues

Other Recent Interesting Approaches (2)

- Lifted Relational Neural Networks
    - Hybrid FOL-Neural paradigm
    - A weighted FOL theory is a template to derive ground neural networks
    - Sourek et al., NIPS Neural Symb Integr. Workshop 2015, ILP 2016
- Differentiable databases
    - Compiles a deductive database s.t. inference is a differentiable function
    - Cohen, NIPS 2016
- Machine Learning ontologies
    - Used to plan learning processes
    - Zakova et al., IEEE T-ASE, 2007; Vavpetic, PhD thesis 2016

Conclusions

- Exciting recent research in ML triggered by the growing SW
- Some great novel concepts include
  - Vector and other embeddings of knowledge graphs
  - Tensor and neural learning approaches
  - Hybrid symbolic-neural (or, vector space) approaches
- Some standing challenges
  - E.g. embedding non-ground knowledge