

Concept Induction for Explainable AI



Pascal Hitzler

Data Semantics Laboratory (DaSe Lab)
Kansas State University

<http://www.daselab.org>

Neuro-symbolic AI



Publications on neuro-symbolic AI in major conferences (research papers only):

conference	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	total
ICML	0	0	0	0	0	1	3	2	5	6	17
NeurIPS	0	0	0	0	0	0	0	4	2	4	10
AAAI	0	0	0	0	0	1	0	1	1	1	4
IJCAI	1	0	0	0	0	0	2	2	0	2	7
ICLR	N/A	N/A	0	0	0	0	1	1	1	3	6
total	1	0	0	0	0	2	6	10	9	16	44

See

Md Kamruzzaman Sarker, Lu Zhou, Aaron Eberhart, Pascal Hitzler
Neuro-Symbolic Artificial Integration: Current Trends

AI Communications, to appear; <https://arxiv.org/abs/2105.05330>
for more analysis.



Computer Science perspective:

- **Let's try to get the best of both worlds:**
 - very powerful machine learning paradigm
 - robust to data noise
 - easy to understand and assess by humans
 - good at symbol manipulation
 - work seamlessly with background (domain) knowledge

- **How to do that?**
 - Endow connectionist systems with symbolic components?
 - Add connectionist learning to symbolic reasoners?

Some Background

**Workshop Series on Neural-Symbolic Learning and Reasoning, since 2005.
Joint with Artur d'Avila Garcez.**

<http://neural-symbolic.org/>

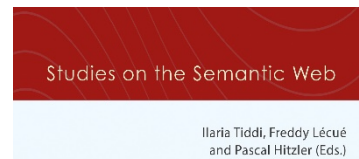
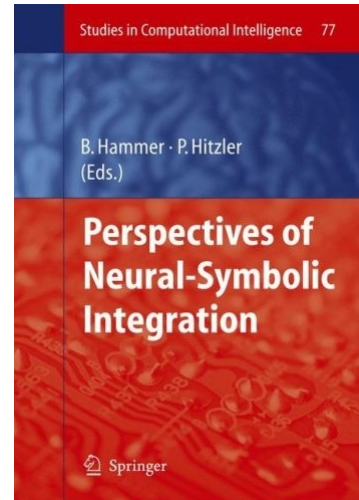
**Barbara Hammer and Pascal Hitzler (eds), Perspectives of
Neural-Symbolic Integration, Springer, 2007**

Neural-Symbolic Learning and Reasoning: A Survey and Interpretation

**Tarek R. Besold, Artur d'Avila Garcez, Sebastian Bader,
Howard Bowman, Pedro Domingos, Pascal Hitzler,
Kai-Uwe Kuehnberger, Luis C. Lamb, Daniel Lowd,
Priscila Machado Vieira Lima, Leo de Penning, Gadi Pinkas,
Hoifung Poon, Gerson Zaverucha**

<https://arxiv.org/abs/1711.03902> (2017)

**Ilaria Tiddi, Freddy Lecue, Pascal Hitzler (eds.), Knowledge Graphs
for eXplainable Artificial Intelligence: Foundations, Applications and
Challenges. Studies on the Semantic Web Vol. 47, IOS Press, 2020.**



Knowledge Graphs and Ontologies

Pascal Hitzler, Semantic Web: A Review of the Field.
Communications of the ACM 64 (2), 76-82, 2021.

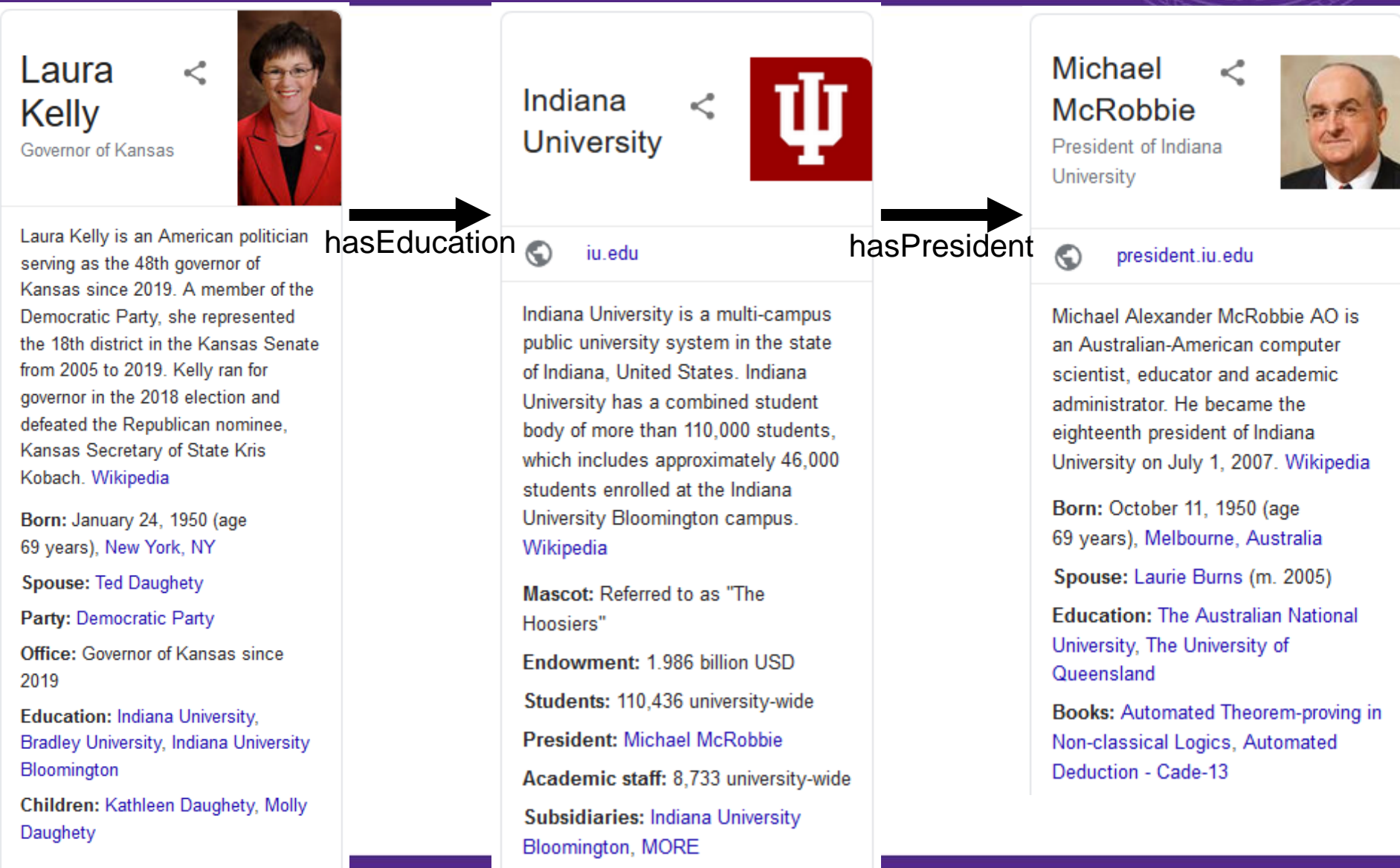
Knowledge Graphs and Ontologies (Schemas)



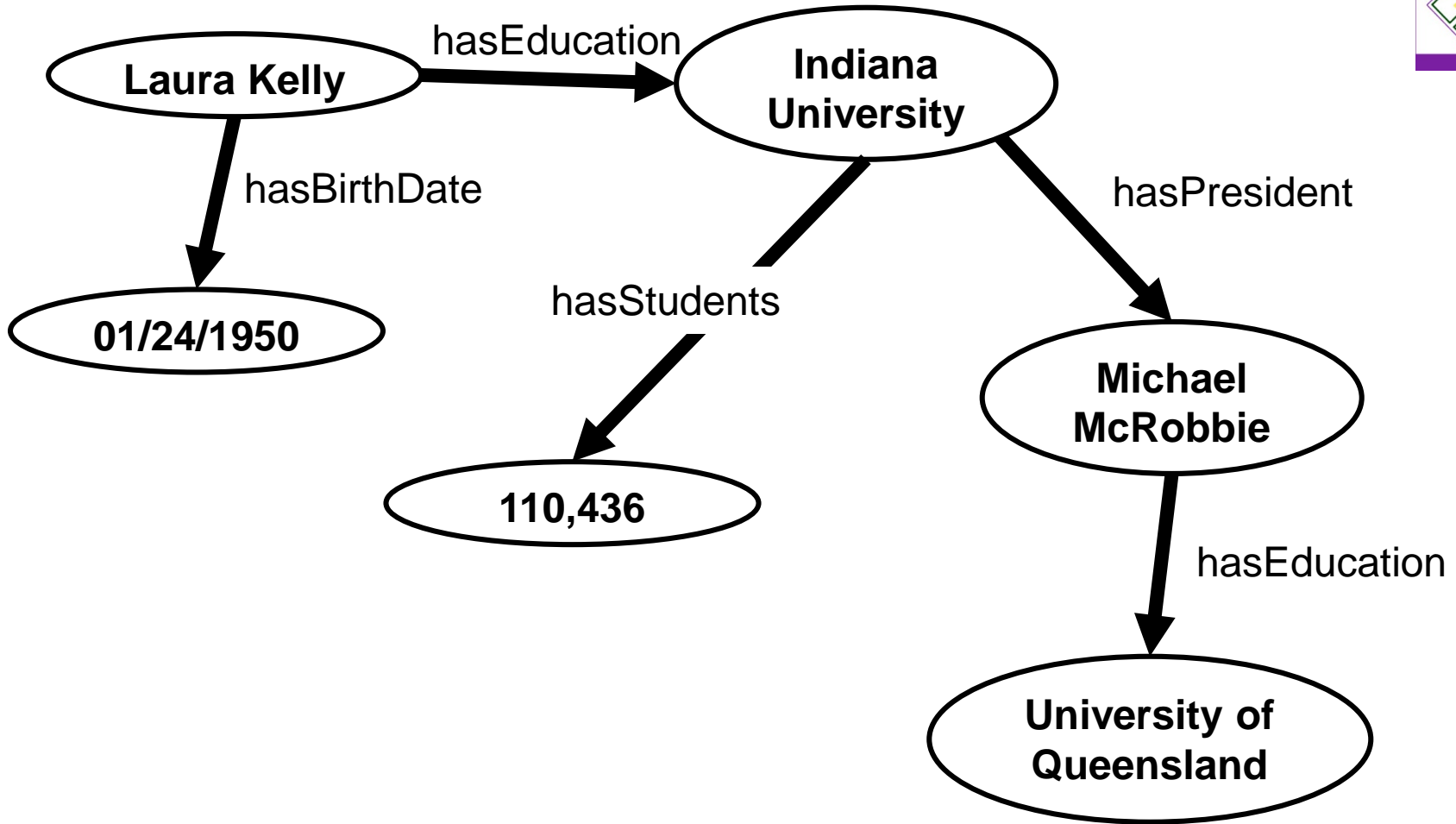
Knowledge Graphs (and their schemas) are made to enable easier

- **data sharing**
- **data discovery**
- **data integration**
- **data reuse**

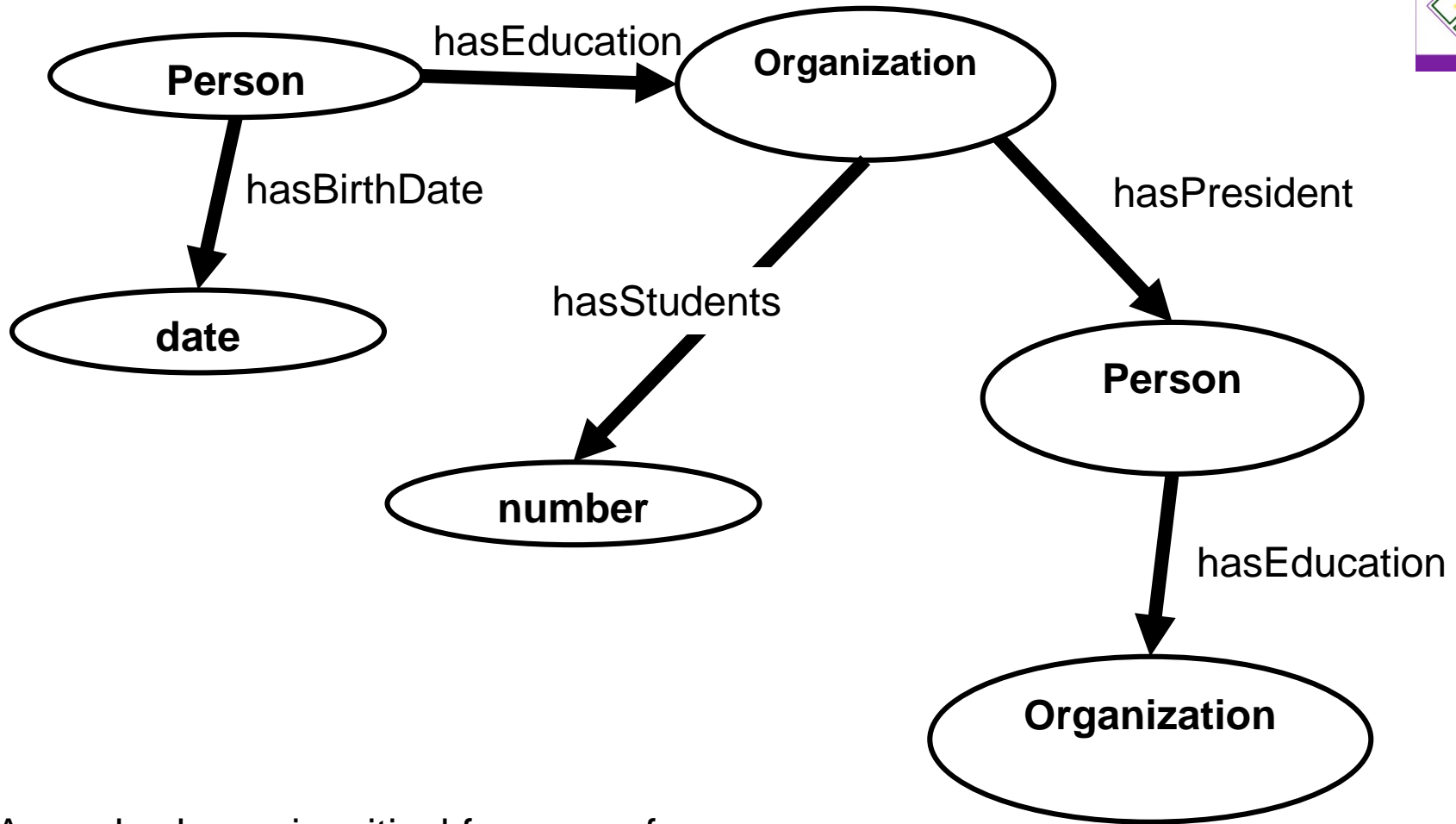
Google Knowledge Graph



Knowledge Graphs



Schema (as diagram)



A good schema is critical for ease of reuse

W3C Standards

RDF 1.1 Concepts and Abstract Syntax

W3C Recommendation 25 February 2014

This version:

<http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>

Latest published version:

<http://www.w3.org/TR/rdf11-concepts/>

Previous version:

<http://www.w3.org/TR/2014/PR-rdf11-concepts-20140109/>

Previous Recommendation:

<http://www.w3.org/TR/rdf-concepts>

Editors:

[Richard Cyganiak](#), [DERI](#), [NUI Galway](#)

[David Wood](#), [3 Round Stones](#)

[Markus Lanthaler](#), [Graz University of Technology](#)

Both established 2004
as versions 1.0.



OWL 2 Web Ontology Language Primer (Second Edition)

W3C Recommendation 11 December 2012

This version:

<http://www.w3.org/TR/2012/REC-owl2-primer-20121211/>

Latest version (series 2):

<http://www.w3.org/TR/owl2-primer/>

Latest Recommendation:

<http://www.w3.org/TR/owl-primer>

Previous version:

<http://www.w3.org/TR/2012/PER-owl2-primer-20121018/>

Editors:

[Pascal Hitzler](#), [Wright State University](#)

[Markus Krötzsch](#), [University of Oxford](#)

[Bijan Parsia](#), [University of Manchester](#)

[Peter F. Patel-Schneider](#), [Nuance Communications](#)

[Sebastian Rudolph](#), [FZI Research Center for Information](#)

PRACTICE

Industry-Scale Knowledge Graphs: Lessons and Challenges

By Natasha Noy, Yuqing Gao, Anshu Jain, Anant Narayanan, Alan Patterson, Jamie Taylor

Communications of the ACM, August 2019, Vol. 62 No. 8, Pages 36-43

10.1145/3331166

[Comments](#)

VIEW AS:



SHARE:



Credit: Adempercem / Shutterstock

in knowledge graphs by defining a *schema* or *ontology*. For example, a link from a movie to its director must connect an object of type *Movie* to an object of type *Person*. In some cases the links themselves might have their own properties: a link connecting an actor and a movie might have the name of the specific role the actor



Knowledge graphs are critical to many enterprises today: They provide the structured data and factual knowledge that drive many products and make them more intelligent and "magical."

In general, a knowledge graph describes objects of interest and connections between them. For example, a knowledge graph may have nodes for a movie, the actors in this movie, the director, and so on. Each node may have properties such as an actor's name and age. There may be nodes for multiple movies involving a particular actor. The user can then traverse the knowledge graph to collect information on all the movies in which the actor appeared or, if applicable, directed.

Many practical implementations impose constraints on the links

SIGN IN for Full Access

» [Forgot Password?](#)

» [Create an ACM Web Account](#)

SIGN IN

ARTICLE CONTENTS:

[Introduction](#)

[What's In a Graph? Design](#)

[Decisions](#)

[Challenges Ahead](#)

[Other Key Challenges](#)

[Conclusion](#)

[References](#)

[Authors](#)

MORE NEWS & OPINIONS

MIT Robot Could Help People



Explaining Deep Learning via Symbolic Background Knowledge

Md. Kamruzzaman Sarker, Ning Xie, Derek Doran, Michael Raymer, Pascal Hitzler, Explaining Trained Neural Networks with Semantic Web Technologies: First Steps. In: Tarek R. Besold, Artur S. d'Avila Garcez, Isaac Noble (eds.), Proceedings of the Twelfth International Workshop on Neural-Symbolic Learning and Reasoning, NeSy 2017, London, UK, July 17-18, 2017. CEUR Workshop Proceedings 2003, CEUR-WS.org 2017

Md Kamruzzaman Sarker, Pascal Hitzler, Efficient Concept Induction for Description Logics. In: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 – February 1, 2019. AAAI Press 2019 , pp. 3036-3043.

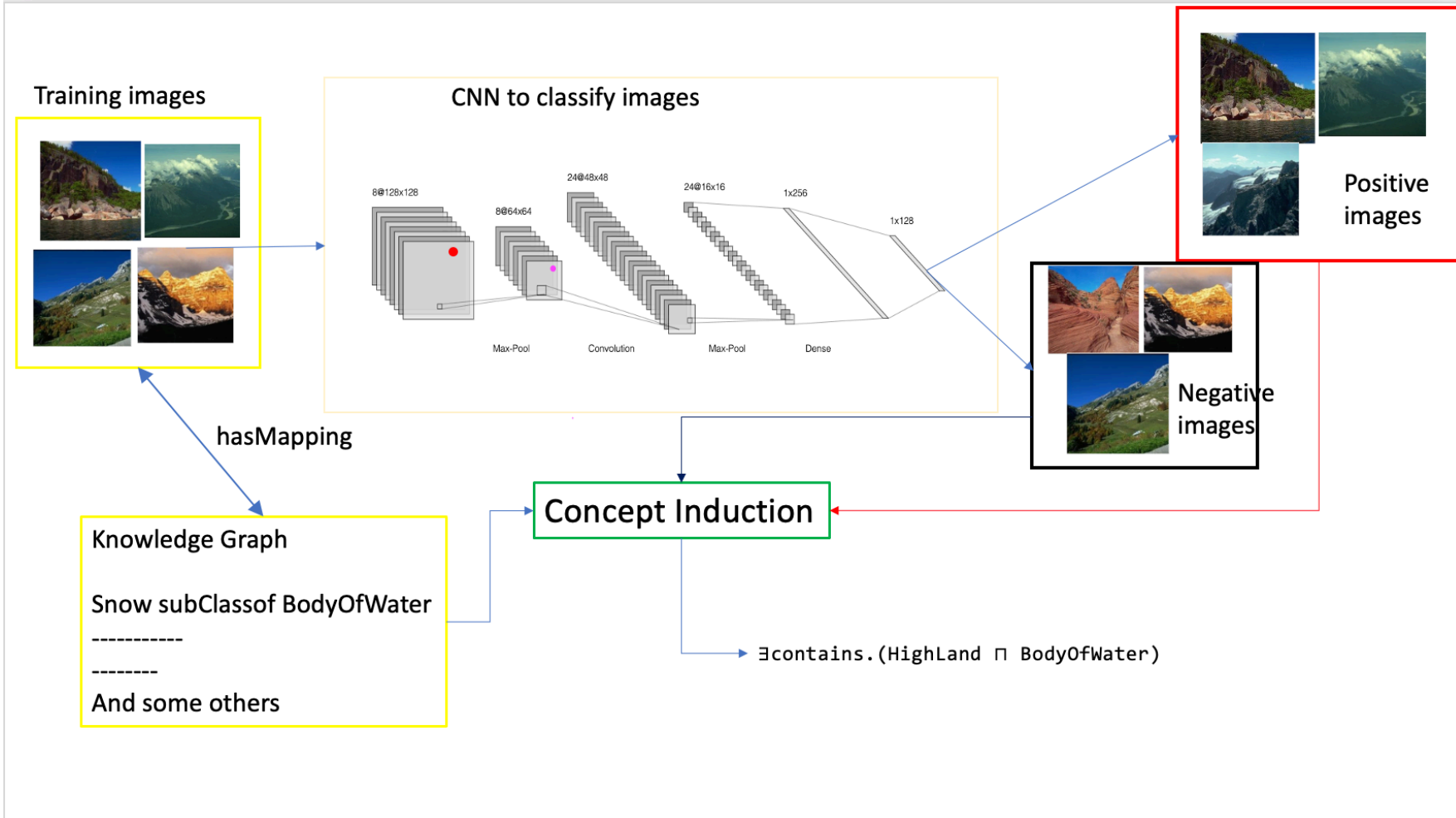
Md Kamruzzaman Sarker, Joshua Schwartz, Pascal Hitzler, Lu Zhou, Srikanth Nadella, Brandon Minnery, Ion Juvina, Michael L. Raymer, William R. Aue, Wikipedia Knowledge Graph for Explainable AI. In: Boris Villazón-Terrazas, Fernando Ortiz-Rodríguez, Sanju M. Tiwari, Shishir K. Shandilya (eds.), Knowledge Graphs and Semantic Web. Second Iberoamerican Conference and First Indo-American Conference, KGSWC 2020, Mérida, Mexico, November 26-27, 2020, Proceedings. Communications in Computer and Information Science, vol. 1232, Springer, Heidelberg, 2020, pp. 72-87.

Explainable AI



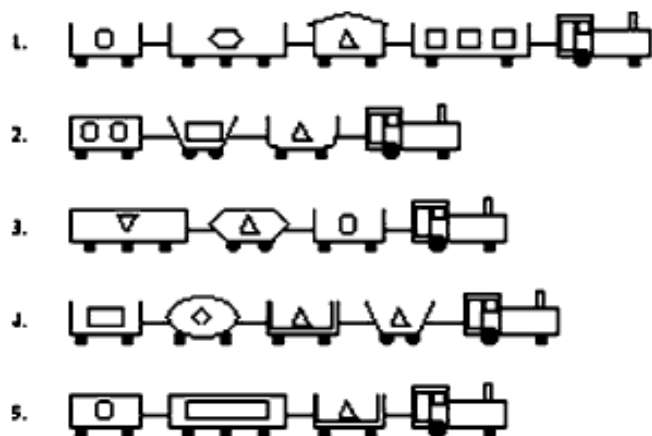
- **Explain behavior of trained (deep) NNs.**
- **Idea:**
 - **Use background knowledge in the form of linked data and ontologies to help explain.**
 - **Link inputs and outputs to background knowledge.**
 - **Use a symbolic learning system to generate an explanatory theory.**
- **We have key components for this now, but it's still early stages.**

Concept



Approach similar to inductive logic programming, but using Description Logics (the logic underlying OWL).

Positive examples:



negative examples:

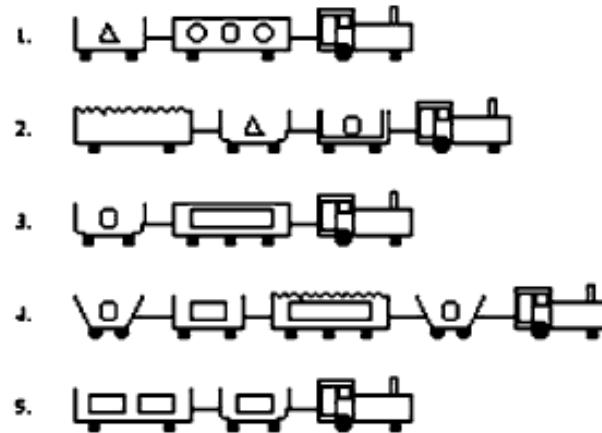


Task: find a class description (logical formula) which separates positive and negative examples.

Positive examples:



negative examples:



DL-Learner result:

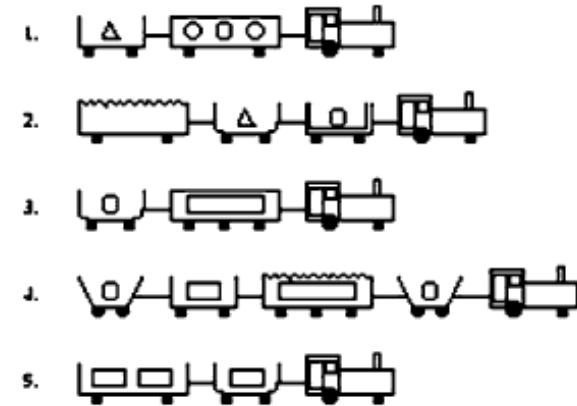
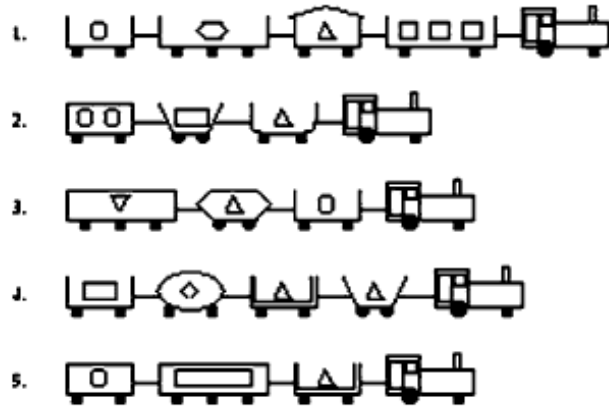
$\exists \text{hasCar} . (\text{Closed} \sqcap \text{Short})$

In FOL:

$$\{x \mid \exists y (\text{hasCar}(x, y) \wedge \text{Closed}(y) \wedge \text{Short}(y))\}$$



DL-Learner uses refinement operators to construct ever better approximations of a solution.



\top

Train – covers all examples.

$\exists \text{hasCar}.\top$

$\exists \text{hasCar}.\text{Closed}$ – covers all positives, two negatives

$\exists \text{hasCar}(\text{Closed} \sqcap \text{Short})$ – solution

Scalability Issues with DL-Learner



- For large-scale experiments, DL-Learner took 2 hours or more for one run.
- We knew we needed at least thousands of runs.
- So we needed a more scalable solution.
- The provably correct algorithms have very high complexity.
- Hence we had to develop a heuristic which trades (some) correctness for speed.
- It is also currently restricted to using a class hierarchy as underlying knowledge base.

ECII algorithm and system



- We thus implemented our own system, ECII (Efficient Concept Induction from Instances) which trades some correctness for speed. [Sarker, Hitzler, AAAI-19]

Experiment Name	Number of Logical Axioms	Runtime (sec)					Accuracy (α_3)		Accuracy α_2			
		DL ^a	DL FIC(1) ^b	DL FIC(2) ^c	ECII DF ^d	ECII KCT ^e	DL ^a	ECII DF ^d	DL FIC(1) ^b	DL FIC(2) ^c	ECII DF ^d	ECII KCT ^e
Yinyang_examples	157	0.065	0.0131	0.019	0.089	0.143	1.000	0.610	1.000	1.000	0.799	1.000
Trains	273	0.01	0.020	0.047	0.05	0.095	1.000	1.000	1.000	1.000	1.000	1.000
Forte	341	2.5	1.169	6.145	0.95	0.331	0.965	0.642	0.875	0.875	0.733	1.000
Poker	1,368	0.066	0.714	0.817	1	0.281	1.000	1.000	0.981	0.984	1.000	1.000
Moral Reasoner	4,666	0.1	3.106	4.154	5.47	6.873	1.000	0.785	1.000	1.000	1.000	1.000
ADE20k I	4,714	577.3 ^f	4.268	31.887	1.966	23.775	0.926	0.416	0.263	0.814	0.744	1.000
ADE20k II	7,300	983.4 ^f	16.187	307.65	20.8	293.44	1.000	0.673	0.413	0.413	0.846	0.900
ADE20k III	12,193	4,500 ^g	13.202	263.217	51	238.8	0.375	0.937	0.375	0.375	0.930	0.937
ADE20k IV	47,468	4,500 ^g	93.658	523.673	116	423.349	0.375	NA	0.608	0.608	0.660	0.608

^a DL : DL-Learner

^b DL FIC (1) : DL-Learner fast instance check with runtime capped at execution time of ECII DF

^c DL FIC (2) : DL-Learner fast instance check with runtime capped at execution time of ECII KCT

^d ECII DF : ECII default parameters

^e ECII KCT : ECII keep common types and other default parameters

^f Runtimes for DL-Learner were capped at 600 seconds.

^g Runtimes for DL-Learner were capped at 4,500 seconds.

ECII vs. DL-Learner

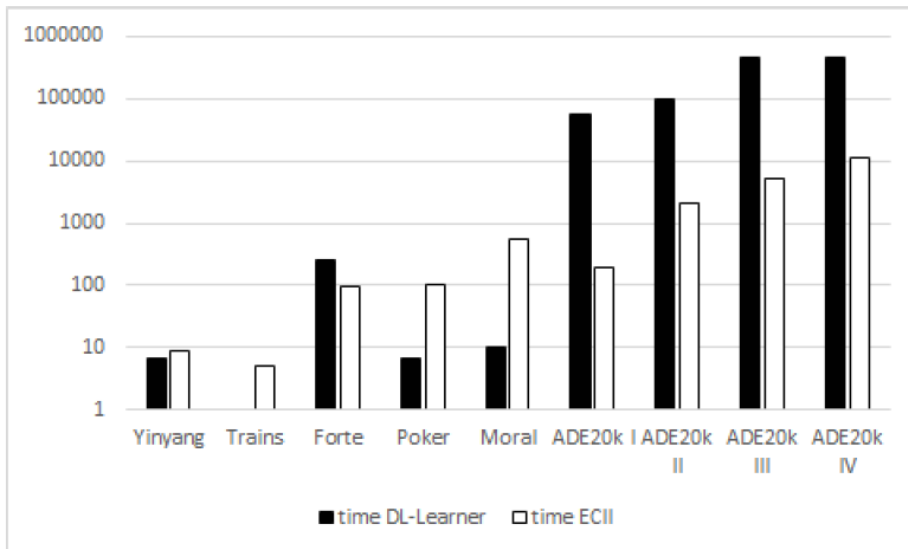


Figure 1: Runtime comparison between DL-Learner and ECII. The vertical scale is logarithmic in hundredths of seconds, and note that DL-Learner runtime has been capped at 4,500 seconds for ADE20k III and IV. For ADE20k I it was capped at each run at 600 seconds.

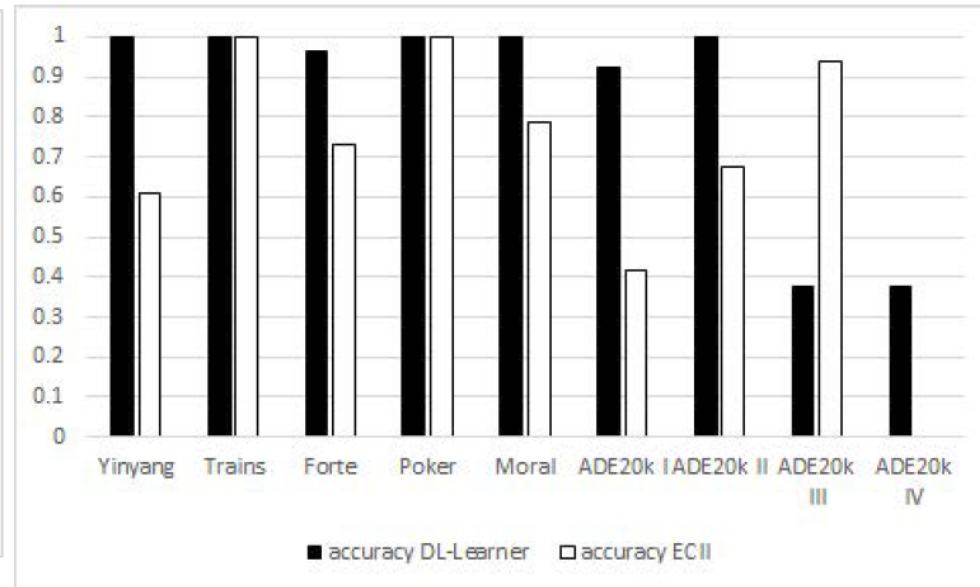


Figure 2: Accuracy (α_3) comparison between DL-Learner and ECII. For ADE20k IV it was not possible to compute an accuracy score within 3 hours for ECII as the input ontology was too large.

Reasons for Improvement



- **DL-Learner loops the following steps:**
 1. **Generate several (refined) candidate solutions.**
 2. **Test candidate solutions by calling a reasoner.**
 3. **Keep only the best solution(s).**
- **This results in many reasoner calls, which are expensive.**
- **ECII optimizes by introducing several (approximate) simplifications:**
 - **Partially materialize reasoning up-front: only one reasoner call required.**
 - **Allow only solutions of a restricted form/syntax.**
 - **Compose solution from pieces which are independently verified against the materialized data.**

Proof of Concept Experiment



Positive:



Negative:



Images



Come from the MIT ADE20k dataset

<http://groups.csail.mit.edu/vision/datasets/ADE20K/>

They come with annotations of objects in the picture:

```
001 # 0 # 0 # sky # sky # ""
002 # 0 # 0 # road, route # road # ""
005 # 0 # 0 # sidewalk, pavement # sidewalk # ""
006 # 0 # 0 # building, edifice # building # ""
007 # 0 # 0 # truck, motortruck # truck # ""
008 # 0 # 0 # hovel, hut, hutch, shack, shanty # hut # ""
009 # 0 # 0 # pallet # pallet # ""
011 # 0 # 0 # box # boxes # ""
001 # 1 # 0 # door # door # ""
002 # 1 # 0 # window # window # ""
009 # 1 # 0 # wheel # wheel # ""
```



Mapping to SUMO



Simple approach: for each known object in image, create an individual for the ontology which is in the appropriate SUMO class:

- contains road1**
- contains window1**
- contains door1**
- contains wheel1**
- contains sidewalk1**
- contains truck1**
- contains box1**
- contains building1**





- **Suggested Merged Upper Ontology**
<http://www.adampease.org/OP/>
- **Approx. 25,000 common terms covering a wide range of domains**
- **Centrally, a relatively naïve class hierarchy.**
- **Objects in image annotations became individuals (constants), which were then typed using SUMO classes.**

DL-Learner input



Positive:

img1: road, window, door, wheel, sidewalk, truck, box, building

img2: tree, road, window, timber, building, lumber

img3: hand, sidewalk, clock, steps, door, face, building, window, road

Negative:

img4: shelf, ceiling, floor

img5: box, floor, wall, ceiling, product

img6: ceiling, wall, shelf, floor, product

DL-Learner results include:

\exists contains.Transitway

\exists contains.LandArea

Proof of Concept Experiment



Positive:



Negative:



∃ contains. Transitway

∃ contains. Land Area

First 10 DL-Learner responses



- | | | | |
|--|-----|------------------------------------|------|
| \exists contains.Window | (1) | \exists contains.LandTransitway | (6) |
| \exists contains.Transitway | (2) | \exists contains.LandArea | (7) |
| \exists contains.SelfConnectedObject | (3) | \exists contains.Building | (8) |
| \exists contains.Roadway | (4) | \forall contains. \neg Floor | (9) |
| \exists contains.Road | (5) | \forall contains. \neg Ceiling | (10) |

Experiment 2



Positive (selection):



Negative (selection):



[-]contains.(DurableGood □ -ForestProduct)

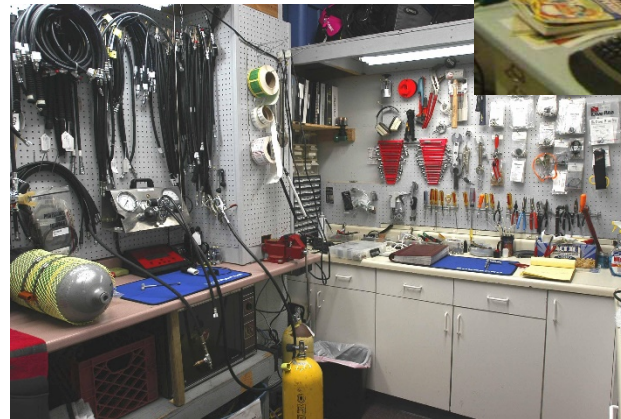
Experiment 3



Positive:



Negative:



$\forall \text{contains.} (\neg \text{Furniture} \sqcap \neg \text{IndustrialSupply})$

Experiment 4



Positive (selection):



Negative (selection):



Does not contain Sentient Agent

Experiment 5



Positive:



Negative (selection):

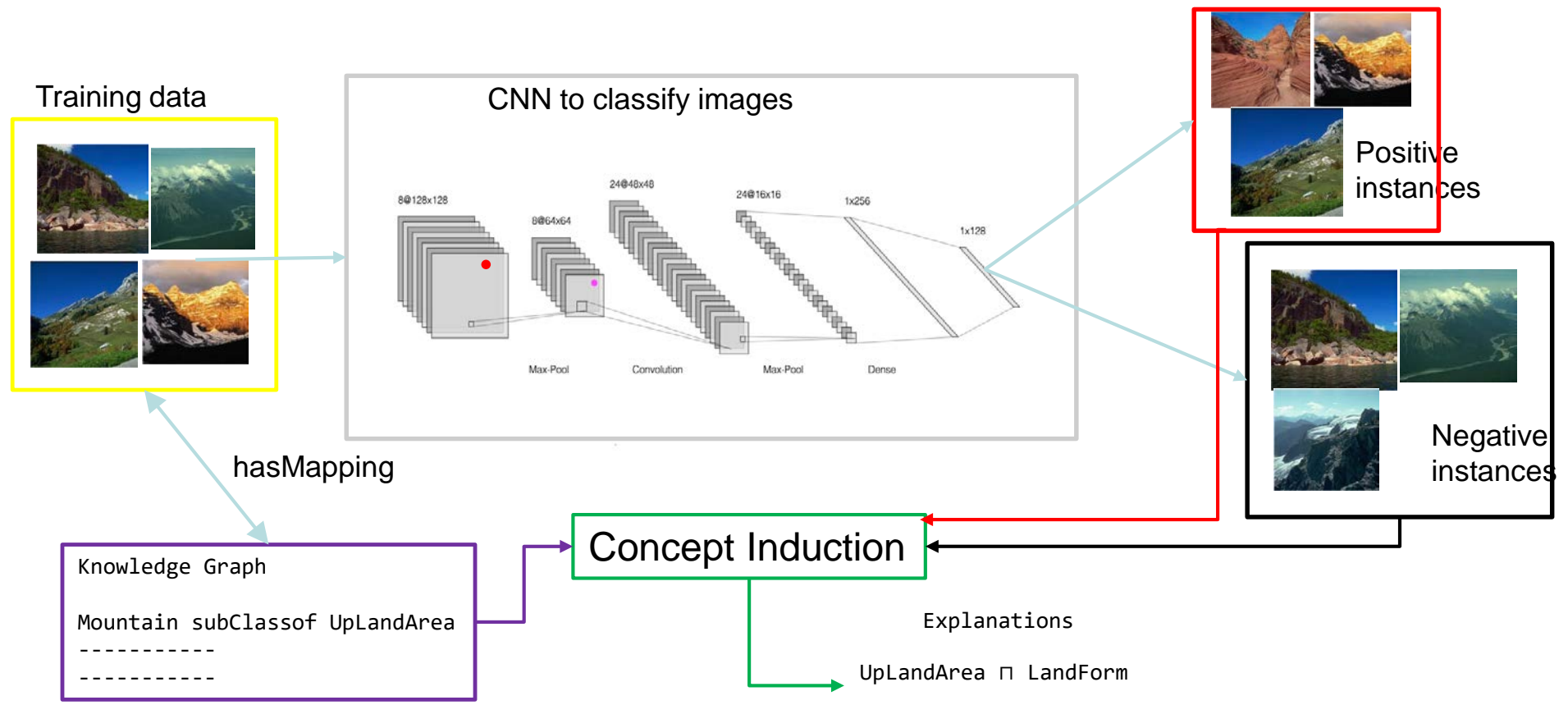


\exists contains.BodyOfWater

Idea Recap



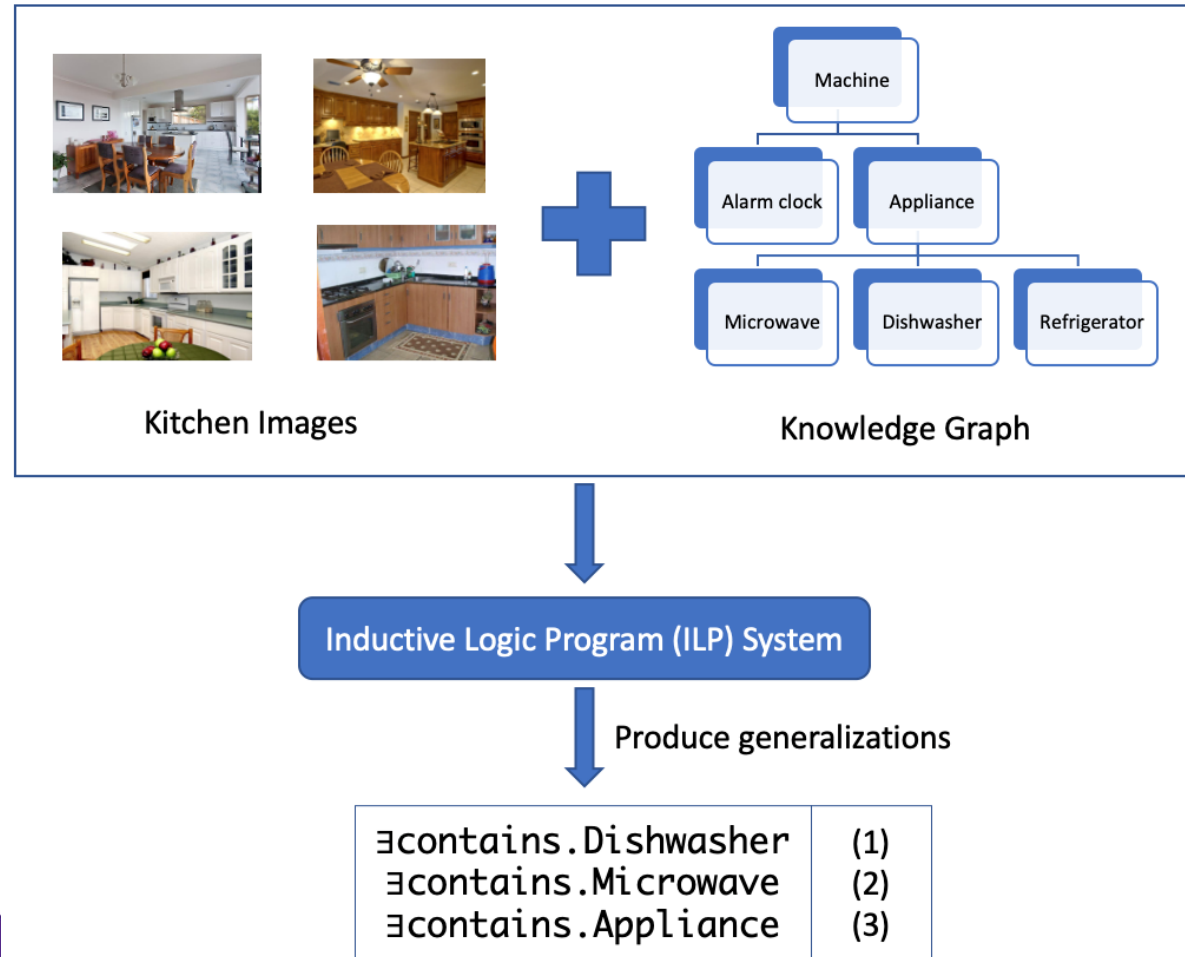
- Generate explanation of the whole model
- Global explanation



Motivation : Knowledge Graph in Explainable AI (XAI)

Generalization use the subclass relation

- Should not contain cyclic information

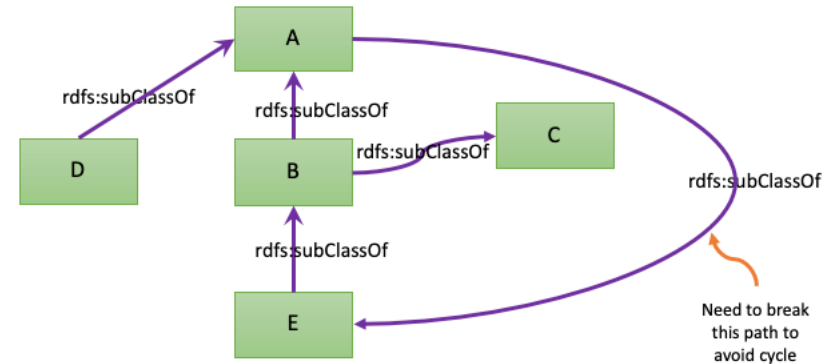


Wikipedia KG (WKG) : Breaking Cycle



Lost Significant Information

- 50% of the subclass relation
- 50% of the class assertion



Number of entities/facts	SUMO	DBpedia	Wikipedia cyclic	Wikipedia noncyclic
Concepts	4558	1183	1,901,708	1,860,342
Individuals	86,475	1	6,145,050	6,079,748
Object property	778	1144	2	2
Data property	0	1769	0	0
Axioms	175,208	7228	71,344,252	39,905,216
Class assertion axioms	167381	1	57,335,031	27,991,282
Subclass axioms	5330	769	5,962,463	3,973,845

Mapping with Knowledge Graph

Model:
Resnet-50

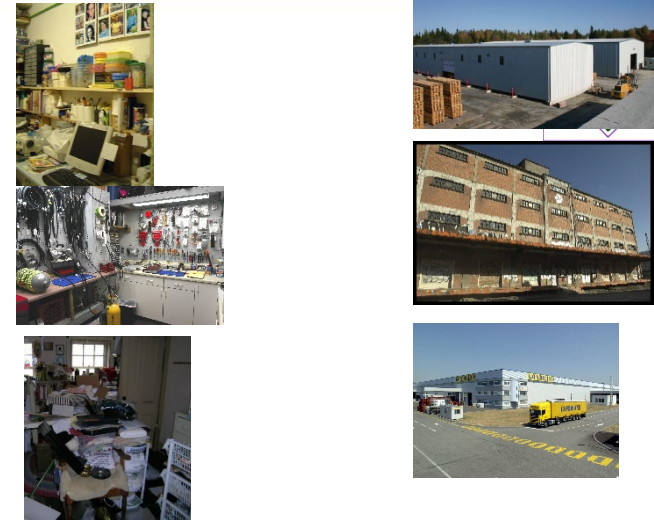
Data:
image p1: machinery, wall, desk, shelf,
pigeonhole, box, projector, computer, screen,
monitor, book

image p2:
image p3:

image n1: lumber, sky, road, sidewalk, building,
box, window, hutch

image n2:
image n3:

Mapping:
P1 imageContains machinery
machinery subclassof DurableGood
.....



Test images. **Workroom** as positive examples p_1, p_2, p_3 on the left, **Warehouse** as negative examples n_1, n_2, n_3 on the right (from top).

Evaluation : Knowledge Graph in XAI



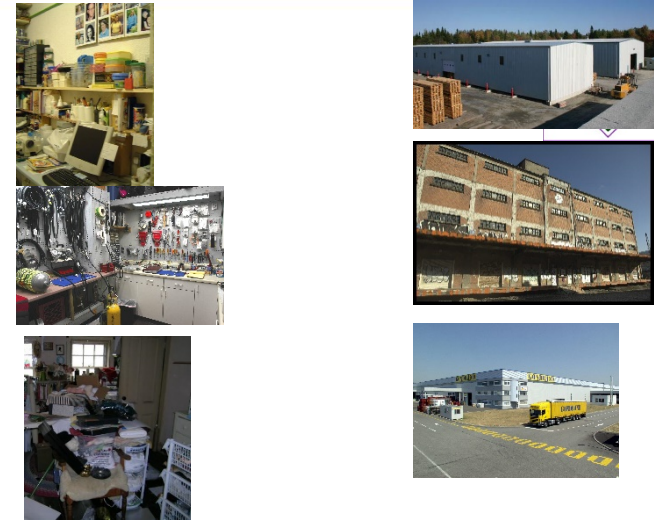
Workroom Explanations

SUMO

- $\exists \text{contains.}(\text{DurableGood} \sqcap \neg \text{ForestProduct})$
- $\exists \text{contains.}(\text{DurableGood} \sqcap \neg \text{Lumber})$
- $\exists \text{contains.} \text{Entity}$

Wikipedia

- $\exists \text{contains.}(\text{Wrenches} \sqcap \text{Tools} \sqcap \neg \text{Lumber})$
- $\exists \text{contains.}(\text{Mechanicaltools} \sqcap \neg \text{Lumber})$
- $\exists \text{contains.}(\text{Mechanicaltools} \sqcap \neg \text{Sky})$



e Lab

Test images. **Workroom** as positive examples p_1, p_2, p_3 on the left, **Warehouse** as negative examples n_1, n_2, n_3 on the right (from top).

Market Explanations

SUMO

- $\exists \text{contains.} \text{SentientAgent}$

Wikipedia

- $\exists \text{contains.}(\text{Structure} \sqcap \text{Life})$

Mountain Explanations

SUMO

- $\exists \text{contains.} \text{BodyOfWater}$

Wikipedia

- $\text{contains.}((\text{Life} \sqcap \text{Branches_of_botany}) \sqcap (\text{Nature}))$



Evaluation : Knowledge Graph in XAI



- **No numerical comparison method so far.**
- **Proposed to use the coverage score of ILP system to compare explanation.**

$$Coverage(S) = \frac{P_S + N_{NS}}{P_S + P_{NS} + N_S + N_{NS}} \tag{1}$$

where

- P_S = Number of positive individuals subsumed by the solution
- P_{NS} = Number of positive individuals not subsumed by the solution
- N_S = Number of negative individuals subsumed by the solution
- N_{NS} = Number of negative individuals not subsumed by the solution

$$Average\ coverage = \sum_{i=1}^n Coverage(S_i) \tag{2}$$



Evaluation : Knowledge Graph in XAI

- Wikipedia Knowledge graph producing better coverage score.
 - Reason behind this is the large number of concepts it has.

Experiment name	#Images	#Positive images	Wikipedia		SUMO	
			#Solution	Coverage	#Solution	Coverage
Market vs. WorkRoom and wareHouse	96	37	286	.72	240	.72
Mountain vs. Market and workRoom	181	85	195	.61	190	.53
OutdoorWarehouse vs. IndoorWarehouse	55	3	128	.94	102	.89
Warehouse vs. Workroom	59	55	268	.56	84	.24
Workroom vs. Warehouse	59	4	128	.93	93	.84



Contribution & Future Work

- **Contribution**
 - Indication that concept induction may be helpful for XAI
 - Scalable but approximate and limited concept induction system ECII
 - Wikipedia Category Hierarchy ready to use

- **Future Work**
 - Human evaluation (in progress)
 - Explore other background knowledge; non-image settings.
 - Use explanations to improve deep learning
 - Explain hidden layer activation patterns



Thanks!

References



Barbara Hammer and Pascal Hitzler (eds), Perspectives on Neural-Symbolic Integration. Springer, 2007

Tarek R. Besold, Artur d'Avila Garcez, Sebastian Bader, Howard Bowman, Pedro Domingos, Pascal Hitzler, Kai-Uwe Kuehnberger, Luis C. Lamb, Daniel Lowd, Priscila Machado Vieira Lima, Leo de Penning, Gadi Pinkas, Hoifung Poon, Gerson Zaverucha, Neural-Symbolic Learning and Reasoning: A Survey and Interpretation. <https://arxiv.org/abs/1711.03902> (2017)

Md Kamruzzaman Sarker, Lu Zhou, Aaron Eberhart, Pascal Hitzler, Neuro-Symbolic Artificial Integration: Current Trends. AI Communications, to appear.

References

Pascal Hitzler, Frank van Harmelen, A reasonable Semantic Web. Semantic Web 1 (1-2), 39-44, 2010.



Pascal Hitzler, Federico Bianchi, Monireh Ebrahimi, Md Kamruzzaman Sarker, Neural-Symbolic Integration and the Semantic Web. Semantic Web 11 (1), 2020, 3-11.

Sebastian Bader, Pascal Hitzler, Dimensions of neural-symbolic integration – a structured survey. In: S. Artemov, H. Barringer, A. S. d'Avila Garcez, L. C. Lamb and J. Woods (eds). We Will Show Them: Essays in Honour of Dov Gabbay, Volume 1. International Federation for Computational Logic, College Publications, 2005, pp. 167-194.

Pascal Hitzler, Semantic Web: A Review of the Field. Communications of the ACM 64 (2), 76-82, 2021.

References

Md. Kamruzzaman Sarker, Ning Xie, Derek Doran, Michael Raymer, Pascal Hitzler, Explaining Trained Neural Networks with Semantic Web Technologies: First Steps. In: Tarek R. Besold, Artur S. d'Avila Garcez, Isaac Noble (eds.), Proceedings of the Twelfth International Workshop on Neural-Symbolic Learning and Reasoning, NeSy 2017, London, UK, July 17-18, 2017. CEUR Workshop Proceedings 2003, CEUR-WS.org 2017

Md Kamruzzaman Sarker, Pascal Hitzler, Efficient Concept Induction for Description Logics. In: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 – February 1, 2019. AAAI Press 2019 , pp. 3036-3043.

Md Kamruzzaman Sarker, Joshua Schwartz, Pascal Hitzler, Lu Zhou, Srikanth Nadella, Brandon Minnery, Ion Juvina, Michael L. Raymer, William R. Aue, Wikipedia Knowledge Graph for Explainable AI. In: Boris Villazón-Terrazas, Fernando Ortiz-Rodríguez, Sanju M. Tiwari, Shishir K. Shandilya (eds.), Knowledge Graphs and Semantic Web. Second Iberoamerican Conference and First Indo-American Conference, KGSWC 2020, Mérida, Mexico, November 26-27, 2020, Proceedings. Communications in Computer and Information Science, vol. 1232, Springer, Heidelberg, 2020, pp. 72-87.





Thanks!