COIN: COmmonsense INference in Natural Language Processing
Workshop to be held in conjunction with EMNLP-IJCNLP in Hong Kong
November 3, 2019

Learning to Reason: from Question Answering to Problem Solving

Michael Witbrock

Broad AI Lab, University of Auckland School of Computer Science

m.witbrock@auckland.ac.nz

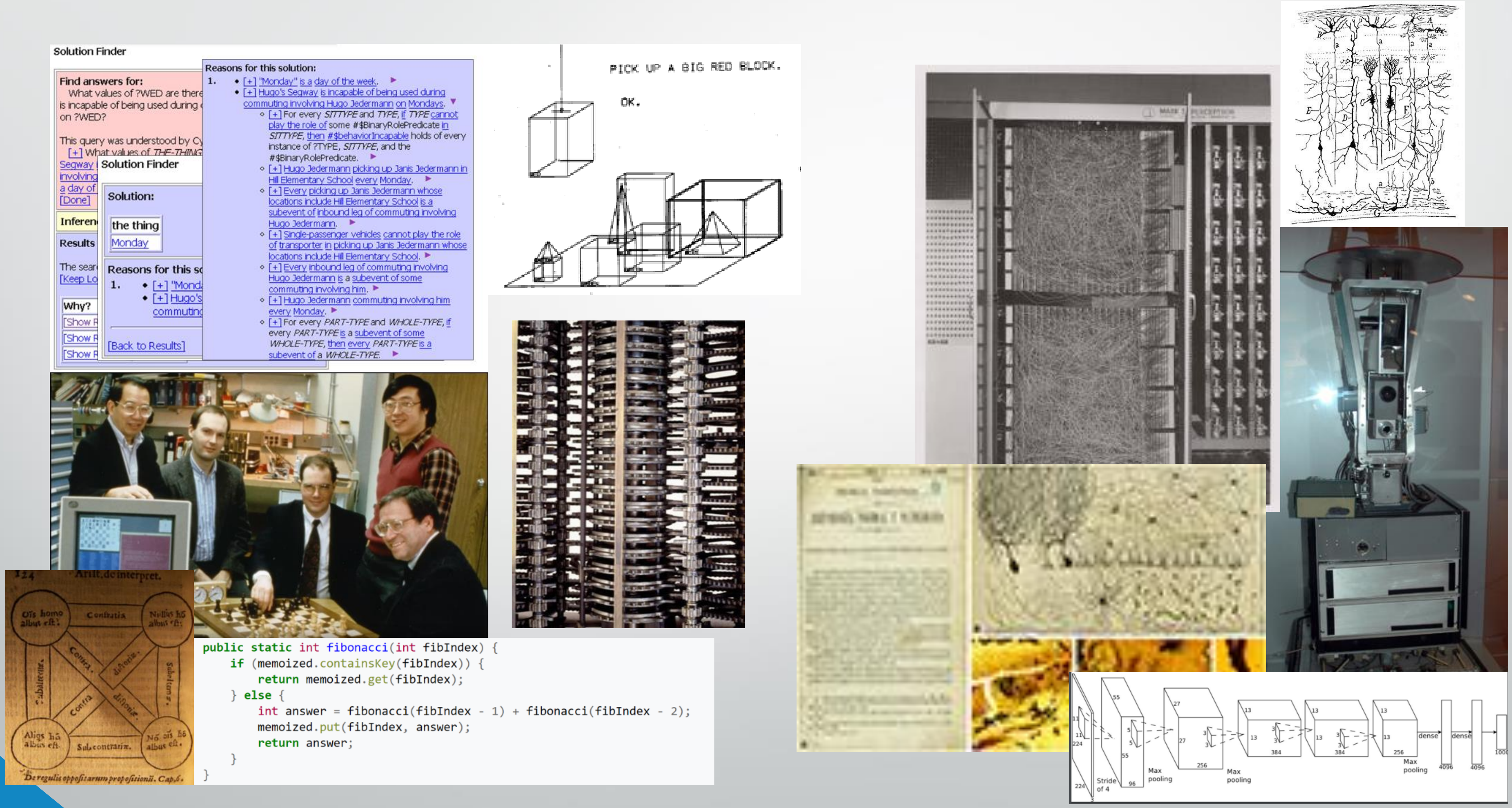
Presented Remotely

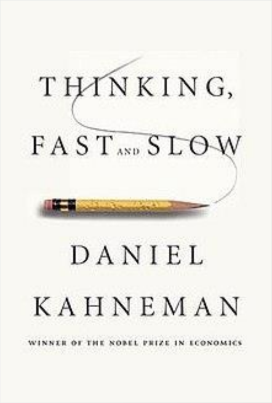
THE UNIVERSITY OF AUCKLAND NEW ZEALAND

SCIENCE SCHOOL OF COMPUTER SCIENCE

V6

History of AI





Two Kinds of Processing

System 1: skill.

- Fast,
- Frequent,
- Automatic,
- Unconscious,
- Stereotyped
- Invariant,
- Emotional

- Mammals, Birds, Fish, Insects, Reptiles

System 2: reason.

- Slow
- Infrequent,
- Deliberate,
- Conscious,
- Compositional,
- Flexible
- Rational

- Mostly Humans, apes, cetaceans, octopuses, parrots

Two Kinds of Learning

Skills.

- Training
- Require many examples
- Transfer is limited
- Performed

Knowledge

- Instruction
- Can be acquired by reading, listening, watching, looking
- Can be reused widely
- Reasoned with

SKILLS

Classification

Generative

Reinforcement Learning

Deep and Convolutional

But what happened to knowledge and reasoning?
&
How can we combine them with skills and learning?

"This falls significantly short of human-problem solving, including question-answering: it does not recursively decompose problems for solution, it does not follow that decomposition to assemble answers, and it does not store and apply salient background knowledge for decomposition, partial solution, or answer composition."

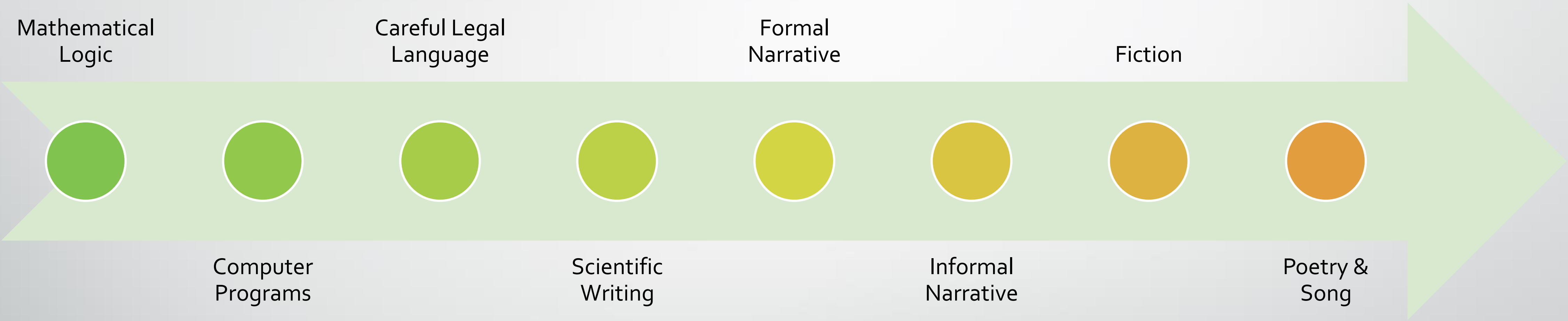
What does that mean for progress towards broader and general AI?

Symbolic AI

Applying knowledge that is stored in a use-agnostic form to solve new problems

One way to think about knowledge, from a machine-learning point-of-view, is stored inductive bias: it reduces the amount of new data required to learn an applicable solution method

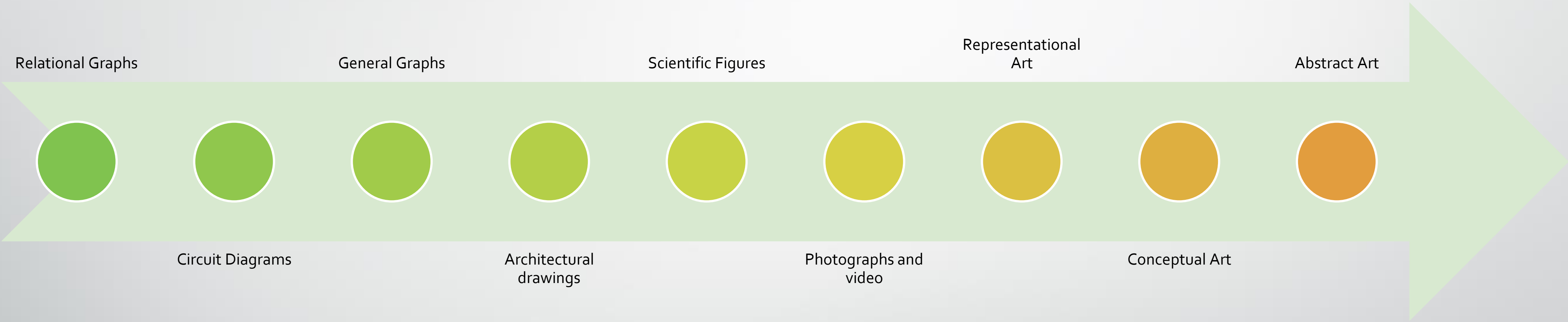
Types of Symbolic Knowledge (Text Like)



High precision and reliable reuse,
Low expressiveness

Low precision and usability
High expressiveness

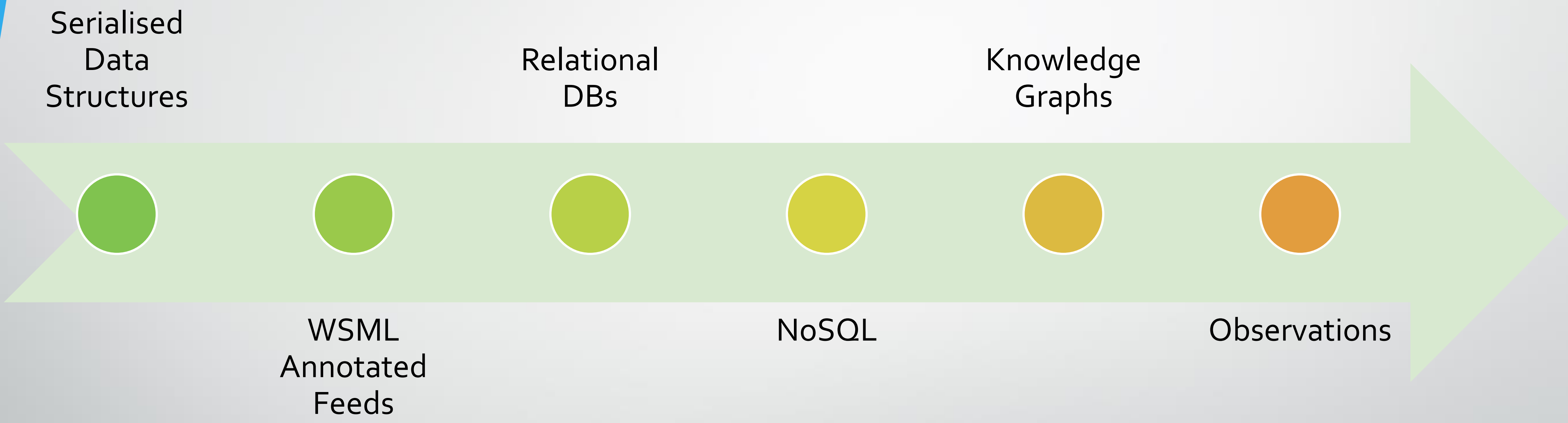
Types of Symbolic Knowledge (picture-like)



High precision and reliable reuse,
Low expressiveness

Low precision and usability
High expressiveness

Types of Symbolic Knowledge (data-like)



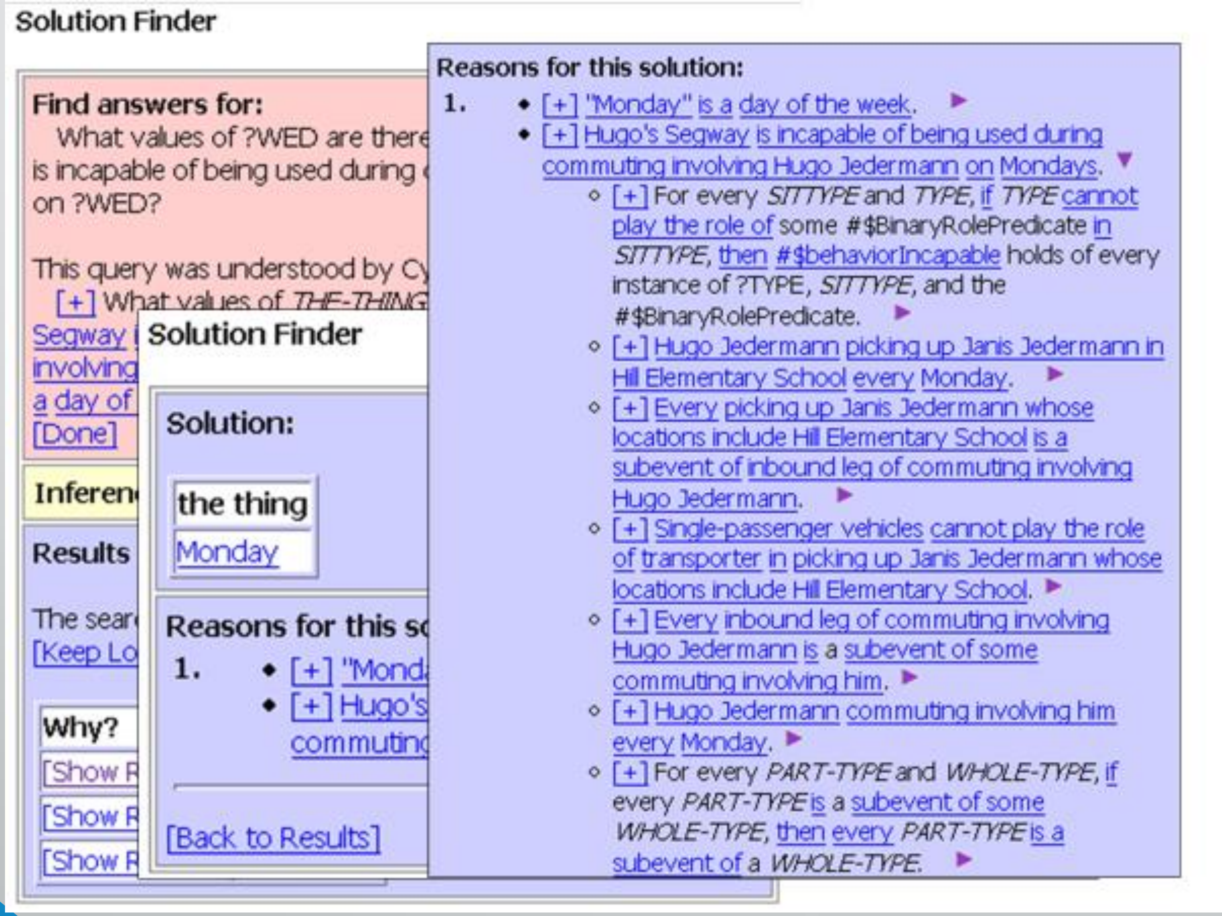
High precision and reliable reuse,
Low expressiveness

Low precision and usability
High expressiveness

Fundamental Operation of Reasoning (recursive and exploratory)

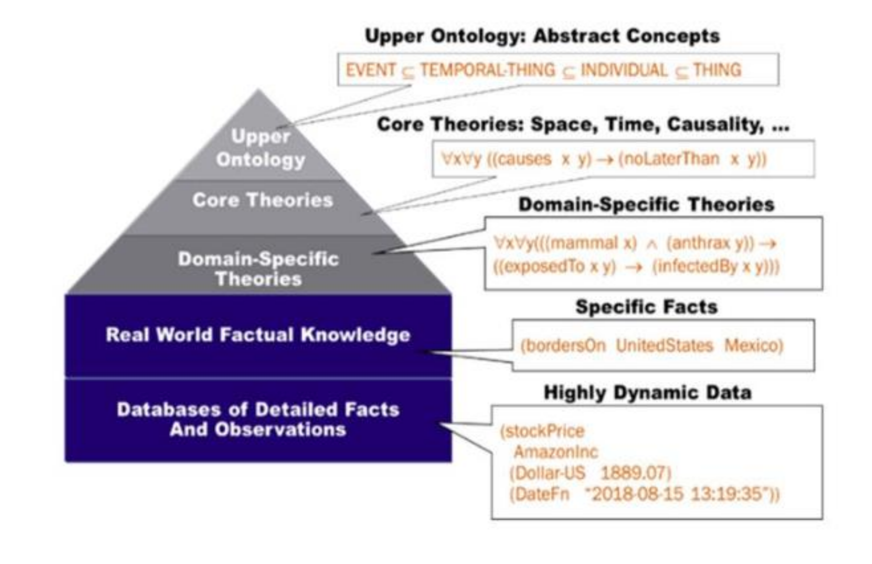
- Interpret problem to be solved
- Transform it into a set of simpler problems, using knowledge
- Solve the simpler problems (by knowing the answer, applying a skill, or breaking them down)
- Use the solutions of the simple problems to assemble the solution to the larger problem

What does reasoning look like to a computer



What days of the week
can't Hugo Jederman use *his Segway*
to commute on.

(read Lime Scooter)



https://www.cyc.com/cyc-technology-overview/

Is Honolulu a City?

Yes

or

It's a city if it's the capital of a country

It's a city if it has more than 70000 inhabitants

It's a city because >70000

and

Is Honolulu the capital of a country?

X is number of inhabitants of Honolulu

X > 70000

Nope

351800 is the number of inhabitants of Honolulu (2016)

351800 > 70000

Between 1845 and 1893 yes.

Kind of trivial in pseudo logic, BUT

- Hundreds of thousands of millions of facts or problem transformation methods might apply at each step [intractable]
- Much as computers might like it, humans have not converted most facts and problem solving methods into mathematical logic (or computer programs) [impractical]

What might we do?

- Pretend reasoning is a game, and apply reinforcement learning to it
- Pretend logic is a natural language, and translate into it

- Learn to reason directly with text, as we did (kind of) in the example
- Learn to modify text so it's more like logic

- Work on problems that *are* already in logic (Mizar & HOLStep mathematical formalisation, WikiData, parts of systems biology, e.g. OBO)

Question Answering as a Proxy for NLU and Reasoning

Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task

Tao Yu **Rui Zhang** **Kai Yang** **Michihiro Yasunaga**
Dongxu Wang **Zifan Li** **James Ma** **Irene Li**
Qingning Yao **Shanelle Roman** **Zilin Zhang** **Dragomir R. Radev**
Department of Computer Science, Yale University
{tao.yu, r.zhang, k.yang, michihiro.yasunaga, dragomir.radev}@yale.edu

Abstract

We present *Spider*, a large-scale, complex and cross-domain semantic parsing and text-to-SQL dataset annotated by 11 college students. It consists of 10,181 questions and 5,693 unique complex SQL queries on 200 databases with multiple tables, covering 138 different domains. We define a new complex

Easy

What is the number of cars with more than 4 cylinders?

```
SELECT COUNT(*)
FROM cars_data
WHERE cylinders > 4
```

Meidum

For each stadium, how many concerts are there?

```
SELECT T2.name, COUNT(*)
FROM concert AS T1 JOIN stadium AS T2
ON T1.stadium_id = T2.stadium_id
GROUP BY T1.stadium_id
```

Hard

Which countries in Europe have at least 3 car manufacturers?

```
SELECT T1.country_name
FROM countries AS T1 JOIN continents
AS T2 ON T1.continent = T2.cont_id
JOIN car_makers AS T3 ON
T1.country_id = T3.country
WHERE T2.continent = 'Europe'
GROUP BY T1.country_name
HAVING COUNT(*) >= 3
```

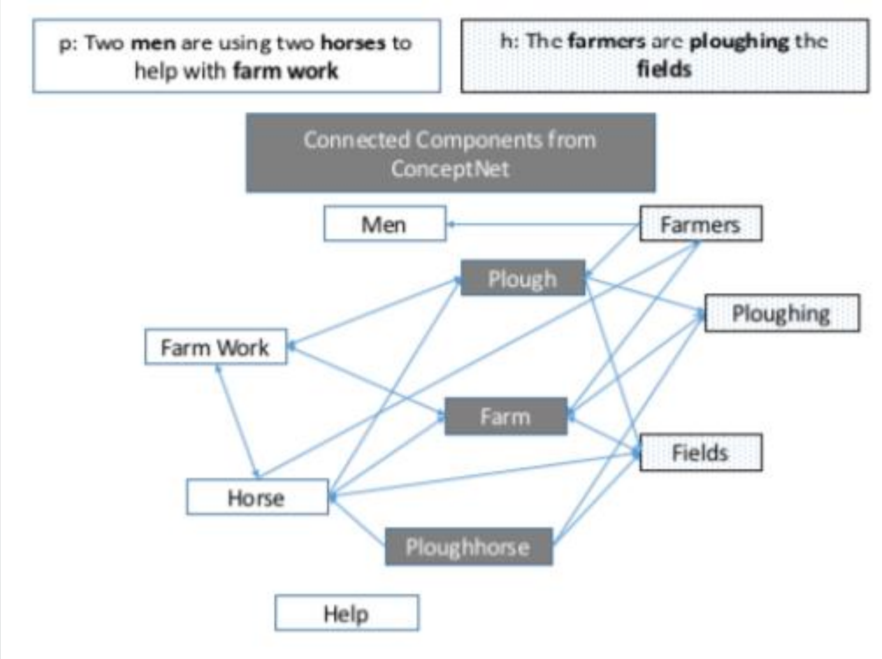
	Test					Dev
	Easy	Medium	Hard	Extra Hard	All	All
Example Split						
Seq2Seq	22.0	7.8	5.5	1.3	9.4	10.3
Seq2Seq+Attention (Dong and Lapata, 2016)	32.3	15.6	10.3	2.3	15.9	16.0
Seq2Seq+Copying	29.3	13.1	8.8	3.0	14.1	15.3
SQLNet (Xu et al., 2017)	34.1	19.6	11.7	3.3	18.3	18.4
TypeSQL (Yu et al., 2018)	47.5	38.4	24.1	14.4	33.0	34.4
Database Split						
Seq2Seq	11.9	1.9	1.3	0.5	3.7	1.9
Seq2Seq+Attention (Dong and Lapata, 2016)	14.9	2.5	2.0	1.1	4.8	1.8
Seq2Seq+Copying	15.4	3.4	2.0	1.1	5.3	4.1
SQLNet (Xu et al., 2017)	26.2	12.6	6.6	1.3	12.4	10.9
TypeSQL (Yu et al., 2018)	19.6	7.6	3.8	0.8	8.2	8.0

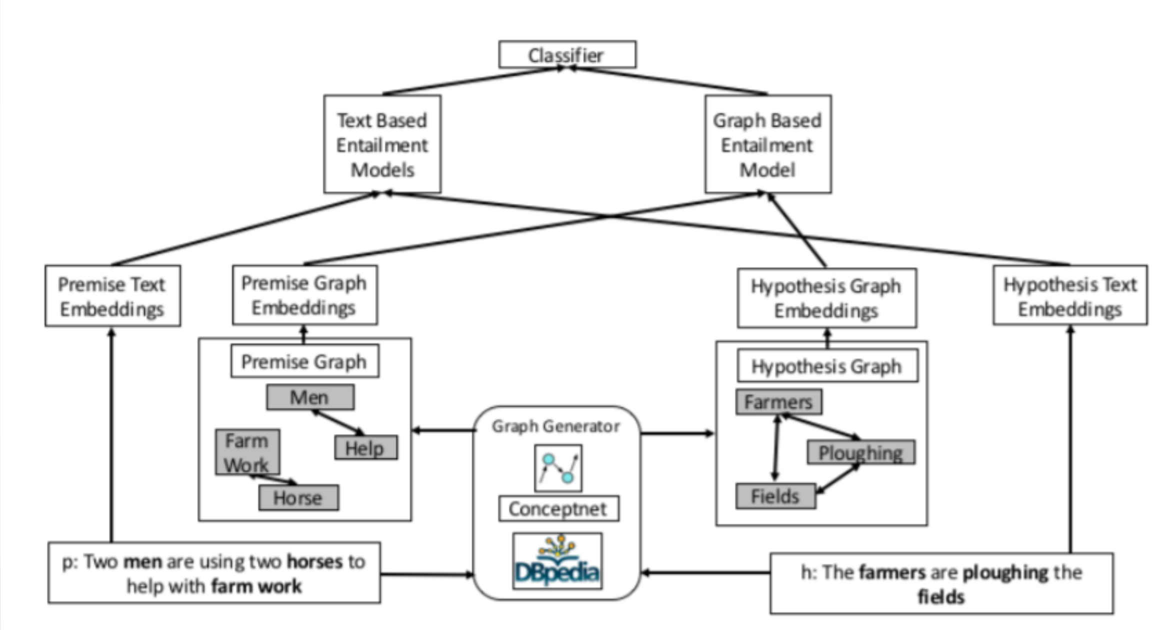
Table 2: Accuracy of Exact Matching on SQL queries with different hardness levels.

Tao Yu, Rui Zhang, Kai Yang, Michihiro Yasunaga,Dongxu Wang, Zifan Li, James Ma, Irene Li, Qingn-ing Yao, Shanelle Roman, et al. 2018c. Spider: A large-scale human-labeled dataset for complex andcross-domain semantic parsing and text-to-sql task. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 3911–3921

Improving Natural Language Inference Using External Knowledge in the Science Questions Domain

Xiaoyan Wang[§], Pavan Kapanipathi[†], Ryan Musa[†], Mo Yu[†],
Kartik Talamadupula[†], Ibrahim Abdelaziz[†], Maria Chang[†],
Achille Fokoue[†], Bassem Makni[†], Nicholas Mattei[†], Michael Witbrock[†]





Model — SciTail dataset	Dev	Test
Decomp-Attn (Parikh et al. 2016)	75.4	72.3
DGEM* (Khot, Sabharwal, and Clark 2018)	79.6	77.3
DeIsTe (Yin, Roth, and Schütze 2018)	82.4	82.1
BiLSTM-Maxout (Mihaylov et al. 2018)	-	84.0
match-LSTM (Wang and Jiang 2015)	88.2	84.1
Our implementation		
match-LSTM (GRU)	88.5	84.2
match-LSTM+WordNet* (Chen et al. 2018)	88.8	84.3
match-LSTM+Gmatch-LSTM* (ConSeqNet)	**89.6**	**85.2**

Answering Science Exam Questions Using Query Reformulation with Background Knowledge

Automated Knowledge Base Construction (2019)

Ryan Musa[†], Xiaoyan Wang[§], Achille Fokoue[†], Nicholas Mattei[*], Maria Chang[†], Pavan Kapanipathi[†], Bassem Makni[†], Kartik Talamadupula[†], Michael Witbrock[†]

AI2 Reasoning Challenge (ARC) multi-choice selected to be unanswerable by IR. Knowledge added to query representation, inputs rewritten to improve performance, entailment of answer-augmented query by retrieved evidence.

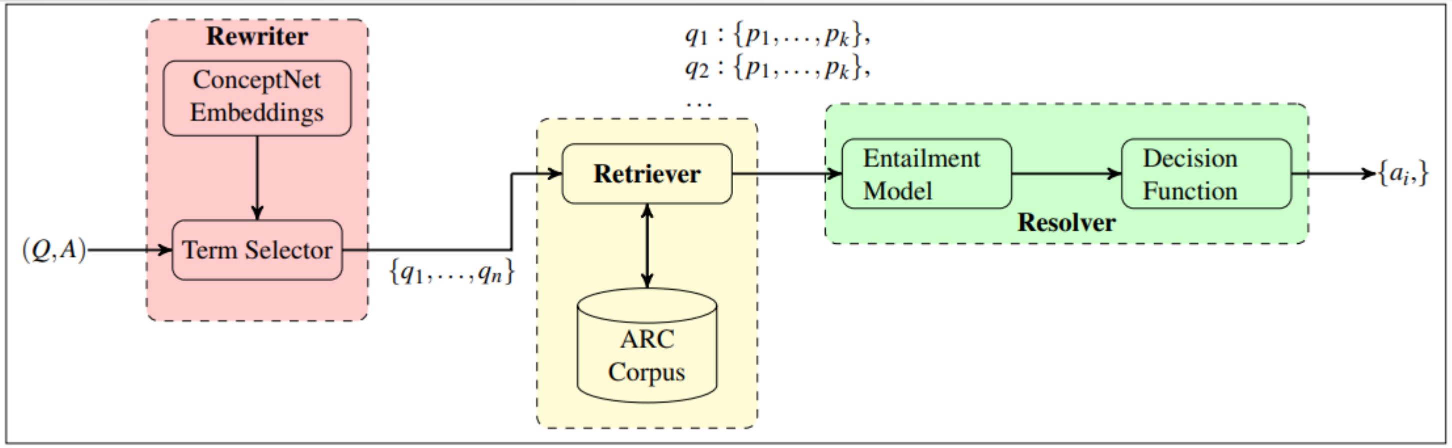


Figure 1: Our overall system architecture. The Rewriter module reformulates a natural-language question into queries by selecting salient terms. The Retriever module executes these queries to obtain a set of relevant passages. Using the passages as evidence, the Resolver module computes entailment probabilities for each answer and applies a decision function to determine the final answer set.

Multi-hop Reading Comprehension through Question Decomposition and Rescoring

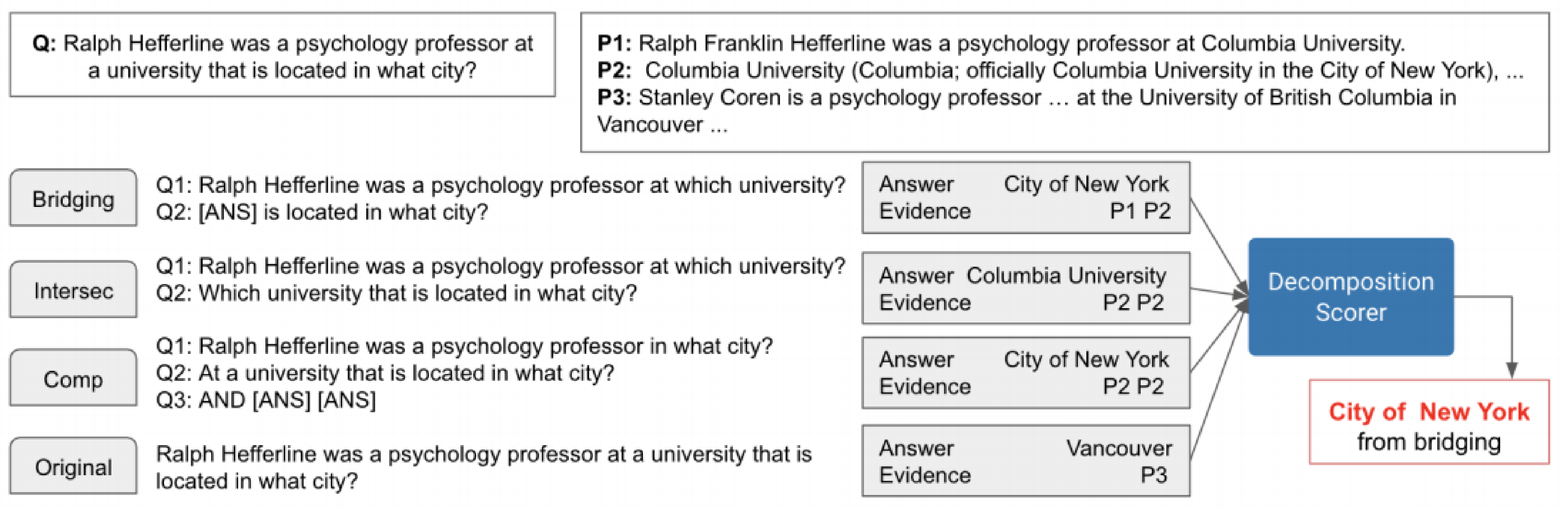
Sewon Min[1], Victor Zhong[1], Luke Zettlemoyer[1], Hannaneh Hajishirzi[1,2]
[1]University of Washington
[2]Allen Institute for Artificial Intelligence
{sewon,vzhong,lsz,hannaneh}@cs.washington.edu

Q Which team does the player named 2015 Diamond Head Classic's MVP play for?
P1 The 2015 Diamond Head Classic was ... Buddy Hield was named the tournament's MVP.
P2 Chavano Rainier Buddy Hield is a Bahamian professional basketball player for the Sacramento Kings ...

Q1 Which player named 2015 Diamond Head Classic's MVP?
Q2 Which team does ANS play for?

Table 1: An example of multi-hop question from HOTPOTQA. The first cell shows given question and two of given paragraphs (other eight paragraphs are not shown), where the red text is the groundtruth answer. Our system selects a span over the question and writes two sub-questions shown in the second cell.

BERT encoding; pointers into sentences to decompose;
fixed decomposition patterns; exactly two sub-questions;
base QA system is BERT reading comprehension model (Devlin et al., 2019)
algorithmic answer composition; choses best one to answer



The reasoning these systems do is very limited

NeurIPS 2017

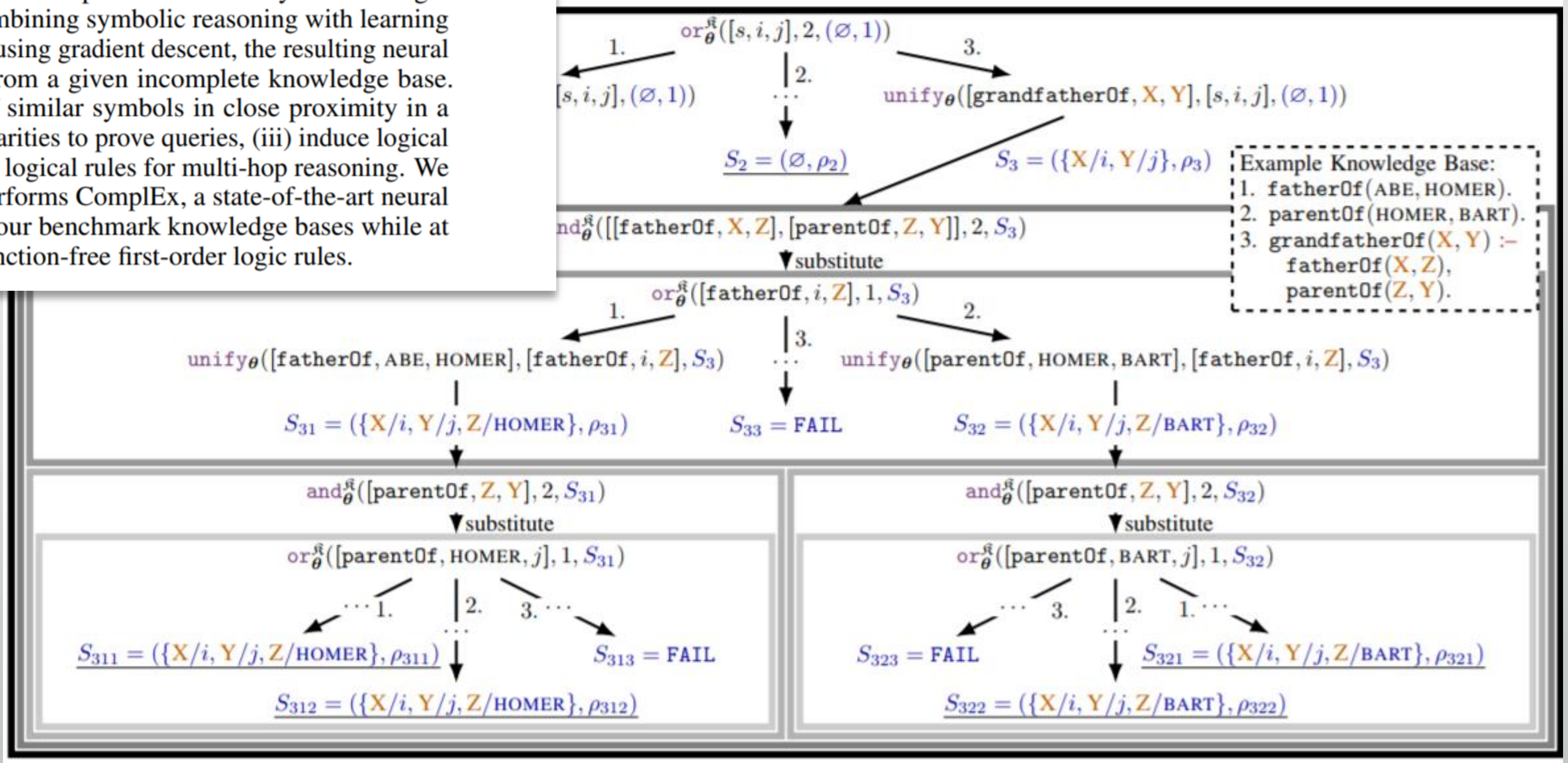
End-to-End Differentiable Proving

Tim Rocktäschel
University of Oxford
tim.rocktaschel@cs.ox.ac.uk

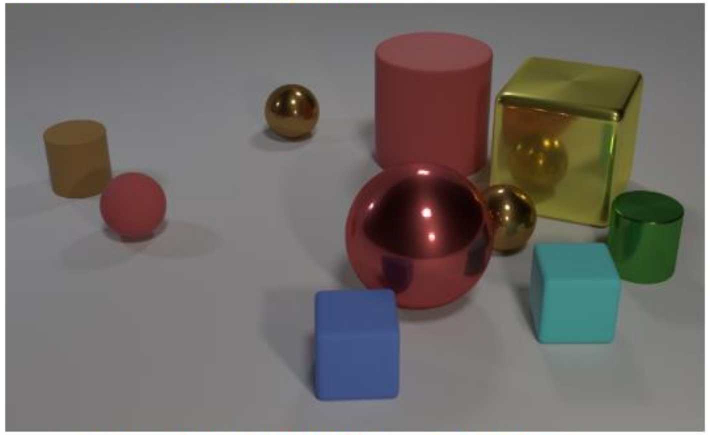
Sebastian Riedel
University College London & Bloomsbury AI
s.riedel@cs.ucl.ac.uk

Abstract

We introduce neural networks for end-to-end differentiable proving of queries to knowledge bases by operating on dense vector representations of symbols. These neural networks are constructed recursively by taking inspiration from the backward chaining algorithm as used in Prolog. Specifically, we replace symbolic unification with a differentiable computation on vector representations of symbols using a radial basis function kernel, thereby combining symbolic reasoning with learning subsymbolic vector representations. By using gradient descent, the resulting neural network can be trained to infer facts from a given incomplete knowledge base. It learns to (i) place representations of similar symbols in close proximity in a vector space, (ii) make use of such similarities to prove queries, (iii) induce logical rules, and (iv) use provided and induced logical rules for multi-hop reasoning. We demonstrate that this architecture outperforms ComplEx, a state-of-the-art neural link prediction model, on three out of four benchmark knowledge bases while at the same time inducing interpretable function-free first-order logic rules.

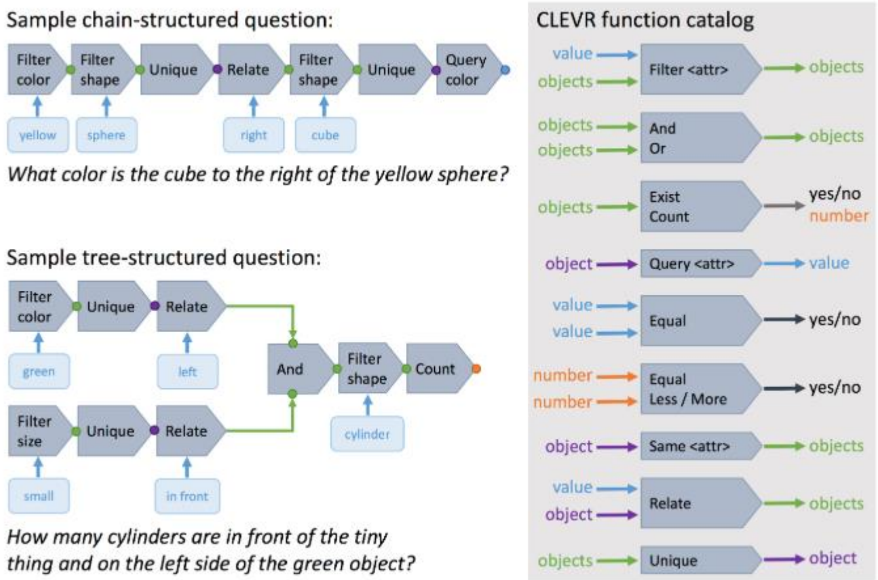


Questions in CLEVR test various aspects of visual reasoning including attribute identification, counting, comparison, spatial relationships, and logical operations.



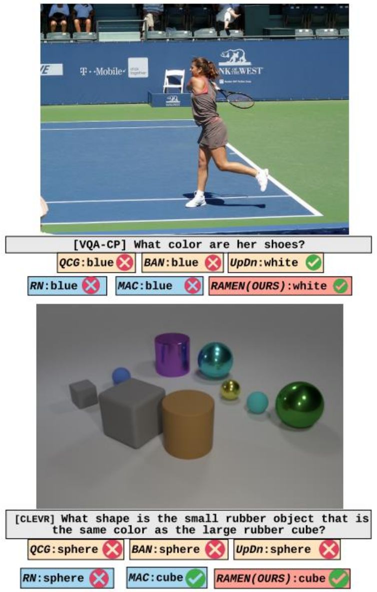
Q: Are there an equal number of large things and metal spheres?
Q: What size is the cylinder that is left of the brown metal thing that is left of the big sphere?
Q: There is a sphere with the same size as the metal cube; is it made of the same material as the small red sphere?
Q: How many objects are either small cylinders or red things?

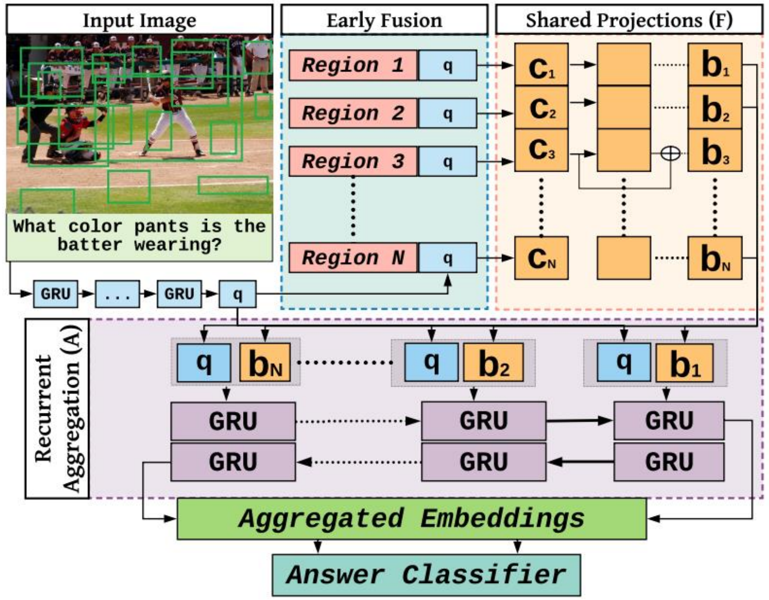
Each question in CLEVR is represented both in **natural language** and as a **functional program**. The functional program representation allows for precise determination of the reasoning skills required to answer each question.



https://cs.stanford.edu/people/jcjohns/clevr/

"Answer Them All! Toward Universal Visual Question Answering Models", Robik Shrestha, Kushal Kafle, Christopher Kanan, CVPR 2019 – RAMEN model





Deep Reasoning for QA

- Recursively transform queries into subproblems to be solved, over as many levels as necessary
- Learn the transformation operations
- Learn the answer assembly operations

Enabling Inference at Scale

- Treat deduction, abduction and inductive steps, and other reliable skills, as steps in a game
- Learn to be very good at that game

Freeing inference from logic

- Recalling the slide earlier about the precision / expressiveness trade-off
- Learn to produce inferences that are as reliable as possible using language-like representations
- Where possible, bound the unreliability and report it with explanations

Interoperable Symbolic Representations

- Using modern Deep Learning techniques for language (sequence) modelling (e.g. GPT2), form representations for logic, natural language, and programming and query languages that aid downstream tasks
  - E.G. Optimise BERT or GPT2 for entailment tasks
- Apply to tasks in understanding legal documents, question answering, and understanding biomedical texts and biological systems

Automatic Production of Precise Representations

- Exact, logical or programmatic knowledge representations have strong efficiency and accuracy benefits when available.
- How can computers build these representations themselves, with little or no human intervention, from existing knowledge resources, text, and other training data

Knowledge Transformation and Capture for Problem Solving

- Learn to transform/rewrite text sources so that they can be more reliably used for problem solving steps
- Learn to elicit material in reasoning-ready form from human beings in task contexts

What’s it for?

- Broad AI: solving these challenges should move us significantly closer to AI systems that can be applied to a broad and heterogeneous set of problems. Move beyond the narrow (but still often superhuman AI skills of today).

Michael Witbrock

m.witbrock@auckland.ac.nz

THE UNIVERSITY OF AUCKLAND NEW ZEALAND

SCIENCE SCHOOL OF COMPUTER SCIENCE

Graph2seq: Graph to sequence learning with attention-based neural networks K Xu, L Wu, Z Wang, Y Feng, M Witbrock, V Sheinin
arXiv preprint arXiv:1804.00823

Broad AI Lab

- Learning-based general artificial intelligence
- Complex problem solving with Natural Language (NL)
- Combine Deep Learning revolution in AI with symbolic AI to give computers the powers of understanding and integration
- Challenging area of AI research, with high potential commercial impact.
  - Near-term advances in **understanding text**, diagrams and tables so they can be **automatically repurposed and combined** to **answer questions**
- Secondary Focus: AI for the benefit of human civilisation
- Michael Witbrock
  - Ph.D, Computer Science, Carnegie Mellon University
  - Previous:
    - Distinguished Research Staff Member & Manager of Reasoning Group at IBM Research, Yorktown New York
    - Vice President for Research, Cycorp Inc
    - Principal Scientist, Lycos Inc.
    - Research Scientist, JustSystem Pittsburgh Research Center
  - Research: Deep Learning, Automated Reasoning, Natural Language Understanding
  - Current affiliations: University of Auckland, Robust.AI, Epistemic.AI, AI for Good Foundation AI4Good.org (founder)
  - Current collaborations: IBM Research AI (USA), Northwestern University (USA)