



Intermezzo: Reflection on semantics and “real” understanding of natural language, common sense, and scientific knowledge.

Global Reasoning over Semantic Abstractions and Semi-Formal Knowledge Bases

Ashish Sabharwal

Allen Institute for Artificial Intelligence (AI2)

AI Seminar, Cornell

November 3, 2017

INTRO ONLY

Bart Selman



Founded in 2014 by Paul Allen (Microsoft co-founder)

Non-profit research organization

Mission: contribute to humanity through high-impact AI research
and engineering

Hiring full-time researchers, postdocs, engineers and interns!

<http://allenai.org/jobs.html>



Oren Etzioni
CEO



AI for the Common Good.

Our mission is to contribute to humanity through high-impact
AI research and engineering.





ALLEN INSTITUTE
for ARTIFICIAL INTELLIGENCE



ARISTO

An intelligent system that reads,
learns, and reasons about science.



Euclid

Solving math and geometry problems.



Plato

Project Plato is focused on extracting visual knowledge from images, diagrams, and videos to enrich knowledge bases that are conventionally derived from textual resources.



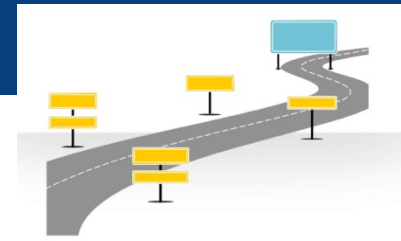
Semantic Scholar

Cut through the clutter.

Home in on key papers, citations, and results.

AllenNLP

An open-source NLP research library,
built on PyTorch



- Reasoning over natural language
 - Question answering (QA) as an AI challenge
- Reasoning with semi-formal knowledge bases
- Reasoning with semantic abstractions of language
- The role of deep learning
- Efficiently assessing the quality of massive datasets
- Summary



IJCAI-2016

ACL-2017

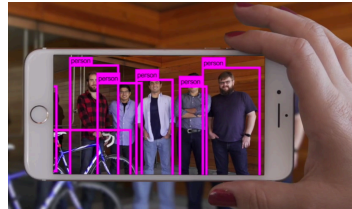
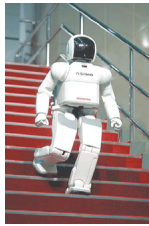
xxxx-2018

xxxx-2018

UAI-2017

Reasoning over Natural Language

- Remarkable feats of AI and ML, particularly deep learning

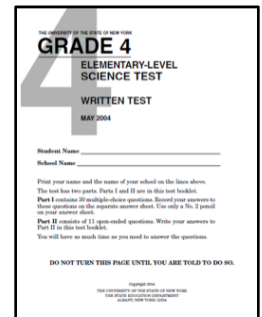


- Huge surge in conversational “AI”
- **Interacting intelligently with humans remains challenging!**
 - Requires rich linguistic understanding, common sense, domain knowledge, situational awareness, conversational memory, ...

Question Answering as an AI Challenge

- Working hypothesis: **Real-world language and reasoning capabilities can be assessed via “well-designed” QA tasks**
- Surge of large scale QA datasets
 - bAbI, SQuAD, Story Cloze, MC Test, TriviaQA, TQA, ...
- Corresponding surge of Deep Learning architectures
 - MemNet, BiDAF, r-net, attention-over-attention, ...

Despite remarkable progress on these datasets,
best systems fail or barely pass (“D”) 4th grade science



Real questions by which humans are assessed remain challenging!

Question Answering as an AI Challenge

Science QA: Elementary and Middle-School Level

Which physical structure would best help a bear to **survive a winter** in New York State?

(A) big ears (B) black nose (C) **thick fur** (D) brown eyes



Much more heterogeneous than popular large-scale QA datasets!

- Appeal to a **wide variety of topics**
- Require various types of knowledge
 - Taxonomic, cause-effect, common sense / **world knowledge**, ...
- Involve **diverse linguistic constructs**
 - Even the best systems are easily fooled by simple textual variations

Afford **little training data** --- generating good quality questions is hard!

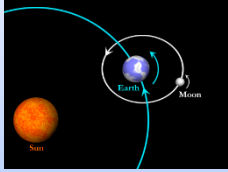
- Limits the viability of popular end-to-end learning approaches

- **Would an IBM Watson (Jeopardy!) style system work?**
- **Jeopardy:** Questions ask for a factual piece of information (a factoid) or combination of factoids.
Example: What is the capital of country X? Or combination of factual pieces of information: What US airport is named after a US general?
Jeopardy! Is hard for humans because they need to know an enormous amount of factual knowledge. *Looking up facts is easy for a machine.* The hard part for the Watson team was understanding the questions. Luckily, a limited set of question types.
- **Science test: requires deeper understanding of information.** Generally, not factoids you can look up. You need to build “reasoning chains” from more basic understanding.

Broad Domain Knowledge

Celestial Phenomena

sun
moon
stars
day/night,
rotation
revolution



The Earth

air
water
land
weather
precipitation
erosion



Matter

solid/liquid/gas
properties
conductivity
texture
temperature
measuring tools



Energy

forms
energy transfer
heat
electricity
chemical energy
energy conversion



Forces

gravity
magnetism
force
friction
pull/pushing
attraction



Living things

living
nonliving
characteristics
animals
plants
fish



Inheritance

inherited traits
resemblance
acquired traits
learned traits
body features
skills



The Environment and Adaptation

senses
habitats
behavior
camouflage
survival



Continuity of Life

life cycle
life span
offspring
reproduction
coloration
mating



Life Functions

breathing
growing
eating
food
air
water

input

output

O_2 →

H_2O →

food →

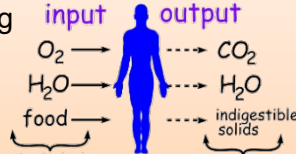
→ CO_2

→ H_2O

→ indigestible solids

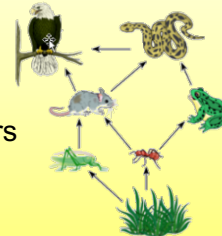
included in launch payload

(discarded)



Interdependence

food web
producers
consumers
decomposers
predators
prey



Human Impact

human activities
environment
ecosystem
pollution
conservation
deforestation

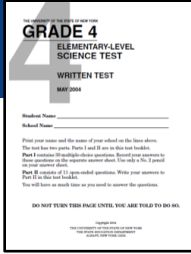


Two Motivating QA Examples

- A. Memorization / retrieval vs. **Reasoning via general principles**
 - Setup: Multiple-choice, requiring external knowledge

- B. Brittleness of the best systems to **linguistic variability**
 - Setup: Paragraph comprehension

Example A: Science Question Answering



New Zealand

shortest

night

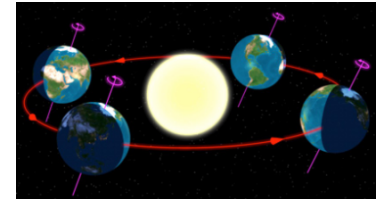
In ~~New York State~~, the ~~longest~~ period of ~~daylight~~ occurs during which month?

- (A) June
- (B) March
- (C) December
- (D) September

Premise: a system that “understands” this phenomenon can correctly answer many variations!

- **Sophisticated physics model** of planetary movement
 - very powerful for complex reasoning
 - but difficult to implement, scale up, or learn automatically

- **Information retrieval / statistical association**
 - generalize well, often effective
 - but limited reasoning, expect answers explicitly written



Example A: Semi-Structured Inference

New Zealand

shortest

night

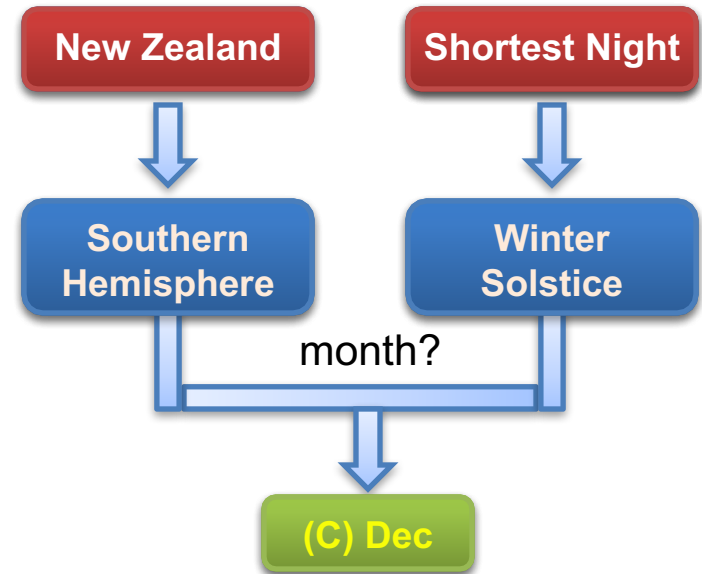
In ~~New York State~~, the ~~longest~~ period of ~~daylight~~ occurs during which month?

- (A) June
- (B) March
- (C) December
- (D) September



- Structured, Multi-Step Reasoning
 - Use science **knowledge in small, reusable, swappable pieces**:
regions, hemispheres, solstice
 - ✓ **principled** approach
 - ✓ **explainable** answers
 - ✓ **robust** to variations

How can we achieve this?



Example B: Brittleness to Linguistic Variation



P: Teams are under pressure after PSG purchased Neymar this season. Chelsea purchased Morata. The Spaniard looked like he was set for a move to Old Trafford for the majority of the summer only for Manchester United to sign Romelu Lukaku instead, paving the way for Morata to finally move to Chelsea for an initial £56m.

Q: *Who did Chelsea purchase this season?*

A: *{Alvaro Morata, Neymar, Romelu Lukaku}*

Example B: Brittleness to Linguistic Variation



P: Teams are under pressure after PSG purchased Neymar this season. **Chelsea purchased Morata**. The Spaniard looked like he was set for a move to Old Trafford for the majority of the summer only for Manchester United to sign Romelu Lukaku instead, paving the way for Morata to finally move to Chelsea for an initial £56m.

Q: *Who did Chelsea purchase this season?*

A: *{Alvaro Morata, Neymar, Romelu Lukaku}*

Simple “lookup” based on proximity to question words, answer type

- Basic word overlap suffices
- Neural methods (e.g., BiDAF) excel at

Example B: Brittleness to Linguistic Variation

Rephrasing #1

P: Teams are under pressure after PSG purchased Neymar this season. **Morata is the recent acquisition by Chelsea.** The Spaniard looked like he was set for a move to Old Trafford for the majority of the summer only for Manchester United to sign Romelu Lukaku instead, paving the way for Morata to finally move to Chelsea for an initial £56m.

Q: *Who did Chelsea purchase this season?*

A: {*Alvaro Morata, Neymar, Romelu Lukaku*}

Simple rewording can confuse solvers

- E.g., BiDAF outputs *“Neymar this season. Morata”*

Example B: Brittleness to Linguistic Variation

Rephrasing #1

nominal

P: Teams are under pressure after PSG purchased Neymar this season. **Morata is the recent acquisition by Chelsea.** The Spaniard looked like he was set for a move to Old Trafford for the majority of the summer only for Manchester United to sign Romelu Lukaku instead, paving the way for Morata to finally move to Chelsea for an initial £56m.

Q: *Who did Chelsea purchase this season?*

A: {*Alvaro Morata, Neymar, Romelu Lukaku*}

verb

Linguistic understanding can help!

- Verbs and their nominalization
- Domain agnostic => can use pre-trained NLP modules!

Example B: Brittleness to Linguistic Variation

Rephrasing #2

P: Teams are under pressure after PSG purchased Neymar this season. **Morata, the recent acquisition by Chelsea, will start for the team tomorrow.** The Spaniard looked like he was set for a move to Old Trafford for the majority of the summer only for Manchester United to sign Romelu Lukaku instead, paving the way for Morata to finally move to Chelsea for an initial £56m.

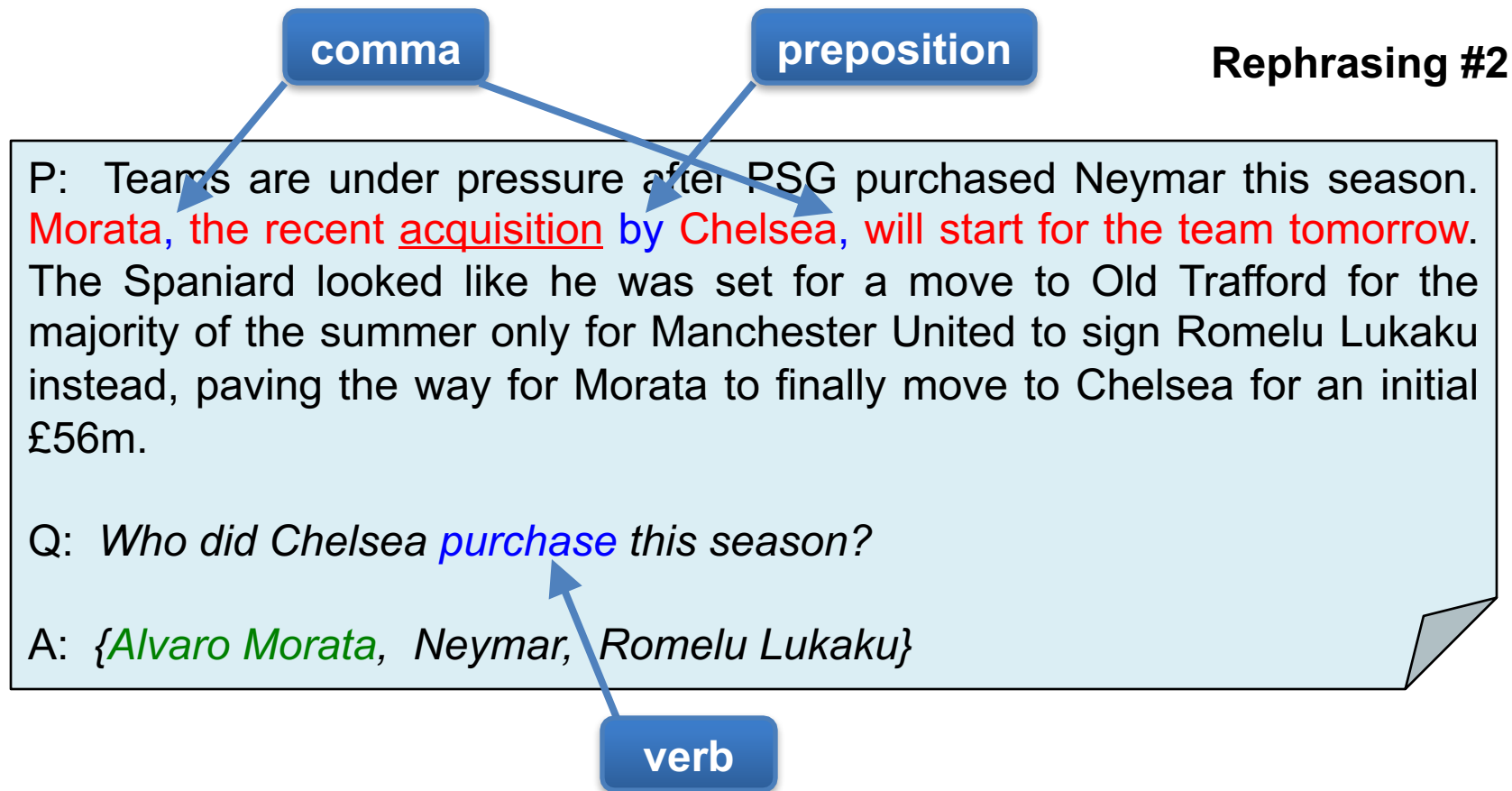
Q: *Who did Chelsea purchase this season?*

A: {*Alvaro Morata, Neymar, Romelu Lukaku*}

Simple rewording can confuse solvers

- E.g., BiDAF outputs “Neymar”

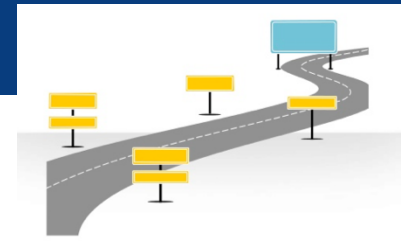
Example B: Brittleness to Linguistic Variation



Linguistic understanding can help!

- Verbs, preposition, punctuation
- Domain agnostic => can use pre-trained NLP modules!

Roadmap



- Reasoning over natural language
 - Question answering (QA) as an AI challenge
- Reasoning with semi-formal knowledge bases
[not here]
- Reasoning with semantic abstractions of language
[not here]
- The role of deep learning
- Efficiently assessing the quality of massive datasets
[not here]
- Summary



The Role of Deep Learning

“Blind” end-to-end learning may not be the best option

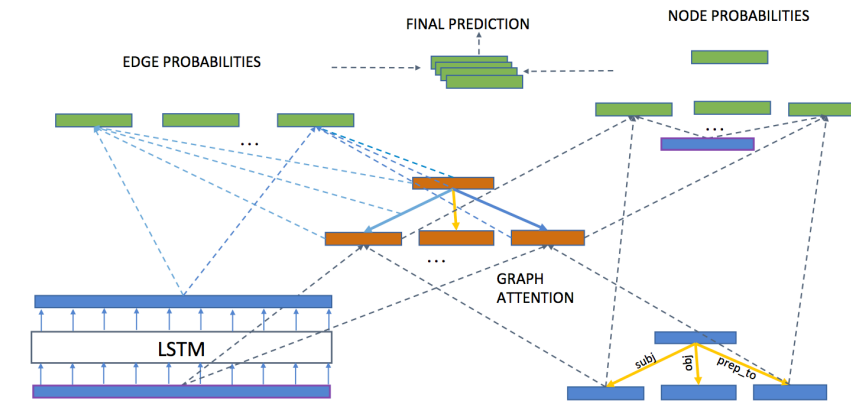
- Limited training data in “interesting” QA domains
- Unclear how to incorporate **general principles**
 - Very prominent in science: $F = m \cdot a$, materials, camouflage, etc.
 - Succinctly summarize a wealth of human knowledge
 - Learning general principles from 100K examples appears:
 - (a) challenging
 - (b) not very human-like

The Role of Deep Learning

Still highly valuable for many pieces of the puzzle!

- Individual components (e.g., NLP modules) of an orchestrated system
- Outer layer on top of the system: Training/tuning the ILP
- **SciTail Dataset (26K): Sentence-to-sentence entailment**

xxxx-2018



DGEM: Decomposable Graph Entailment Model

Summary

- It's an exciting time to be doing research in AI, in particular in **reasoning with natural language**
- Many **fascinating challenges**
 - In language understanding, knowledge representation, reasoning
 - Not quite captured by popular large-scale QA datasets
- **Deep learning** has an important role to play, but **is not everything** (yet)



EXTRA SLIDES

Aristo Demo

- <http://aristo-demo.allenai.org>
- <http://aristo-demo-internal.allenai.org>
- <http://tableviz.dev.ai2:8090>
- <http://aristo-controller.dev.ai2:8080>