

Question Answering as Global Reasoning over Semantic Abstractions

Daniel Khashabi, Dan Roth (*University of Pennsylvania*)

Tushar Khot, Ashish Sabharawal (*Allen Institute for Artificial Intelligence*)

Feb 6, 2018





Standardized science exams (Clark et al, 2015):

- Broad knowledge (general and scientific)
- Connecting question to

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

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

P: ... Polar bears, saved from the bitter cold by their thick fur coats, are among the animals in danger ...



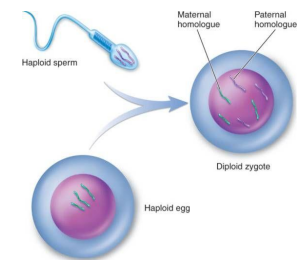
Biology exams (Berant et al, 2014):

- Technical terms and answer not easy to find.
- Requires understanding complex relations.

Q: What does meiosis directly produce?

(A) Gametes (B) **Haploid cells**

P: ... Meiosis produces not gametes but haploid cells that then divide by mitosis and give rise to either unicellular descendants or a haploid multicellular adult organism. Subsequently, the haploid organism carries out further **mitoses, producing the cells** that develop into gametes.



Linguistic variability



Which physical structure would best **help a bear to survive a winter?**
(A) big ears (B) black nose (C) **thick fur** (D) brown eyes

Thick fur helps a bear survive a winter.

A thick coat of white fur helps bears survive in these cold latitudes.

Polar bears, saved from the bitter cold by their thick fur coats, are among the animals in danger of extinction because of the global warming and human activities.

A given “meaning” can be phrased many surface forms!

QA is a language understanding problem!



verb

Which physical structure would best help a bear to survive a winter?
(A) big ears (B) black nose (C) **thick fur** (D) brown eyes

comma

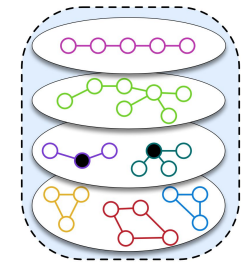
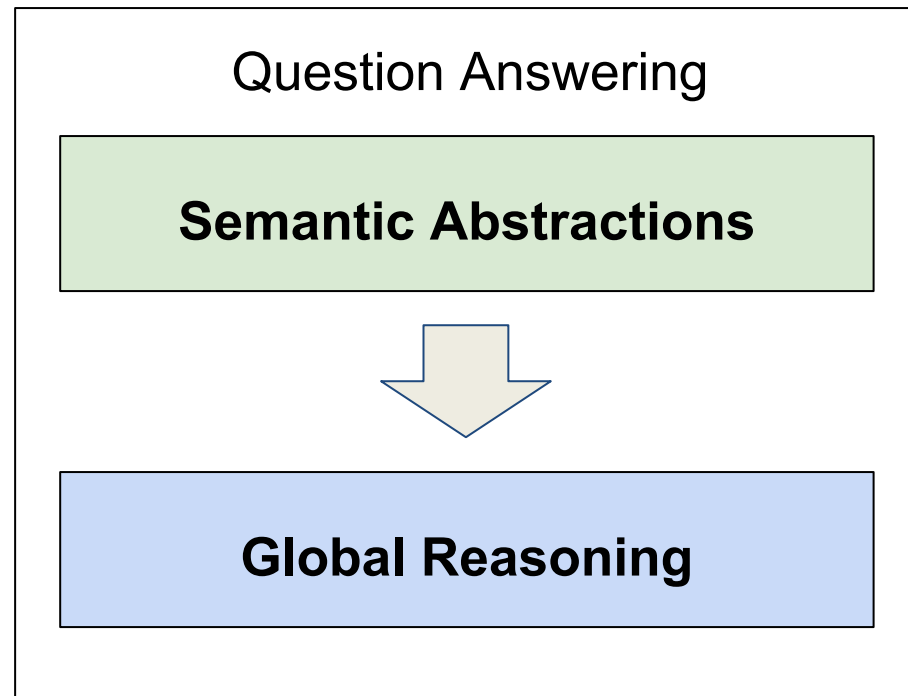
preposition

Polar bears, saved from the bitter cold **by** their thick fur coats, are among the animals in danger of extinction because of the global warming and human activities.

QA is fundamentally a NLU problem

A single abstraction is not enough

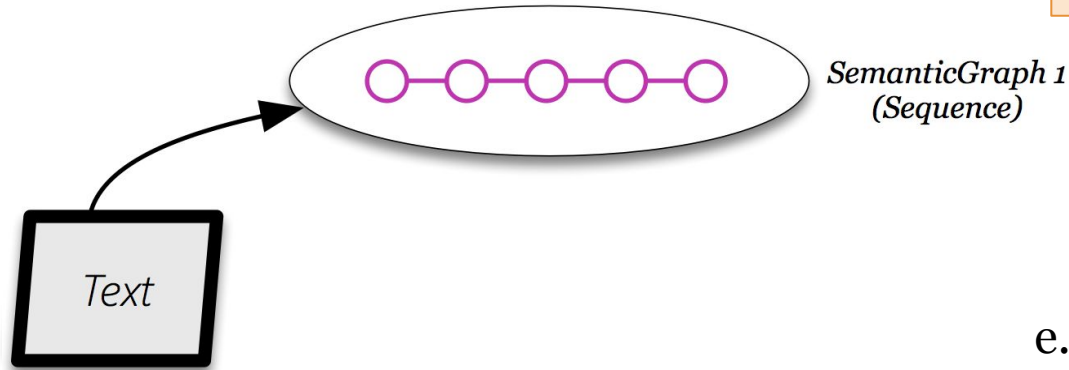
Question Answering as **Global Reasoning** over **Semantic Abstractions**



Collections of semantic graphs

Create a **unified representation** as a **family of graphs**

- predicate-argument, trees, clusters, sequences



A single representation is not enough to capture complexity of language

e.g named-entities

e.g dependency parse

e.g semantic role labeling
(verb, preposition, comma)

e.g co-reference

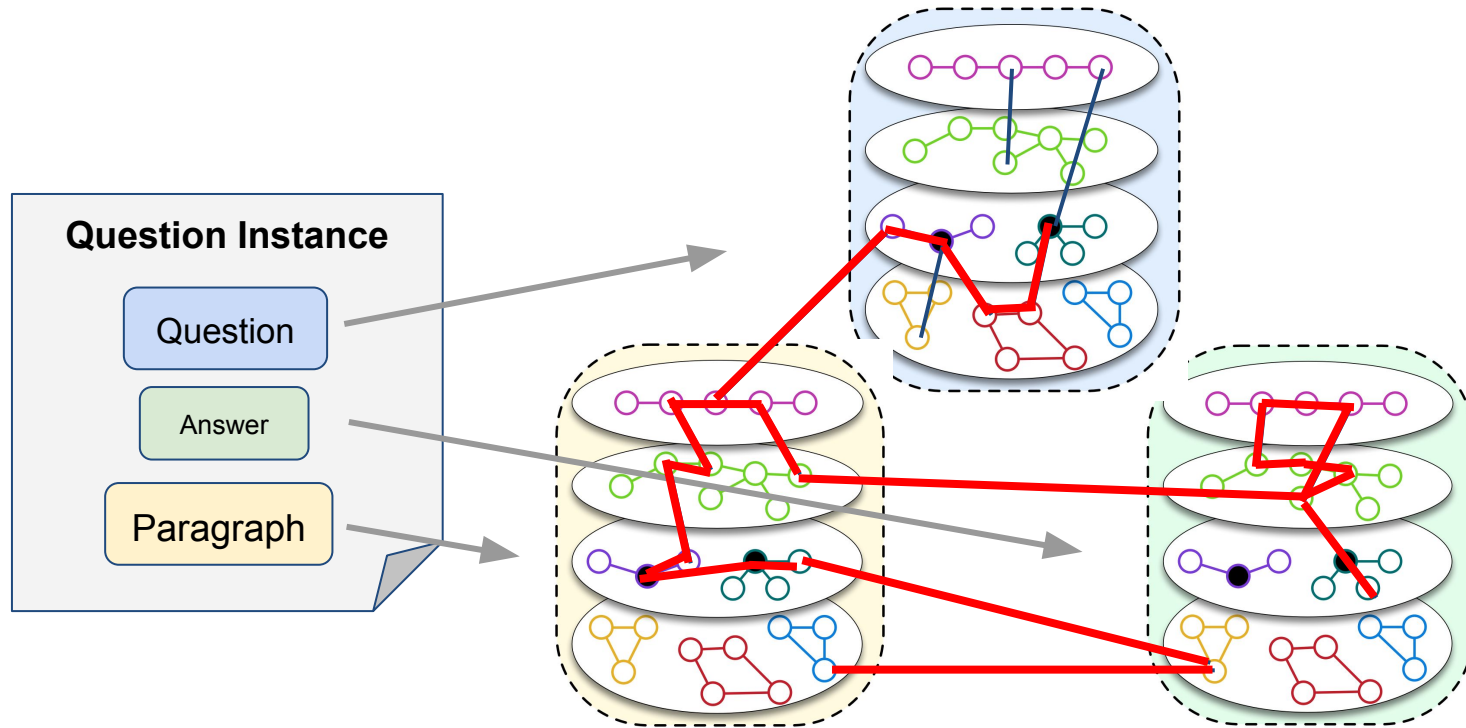
e.g tables

Our representation has nothing to do with the QA task. It reflects our understanding of the language

Reasoning With a Meaning Representation

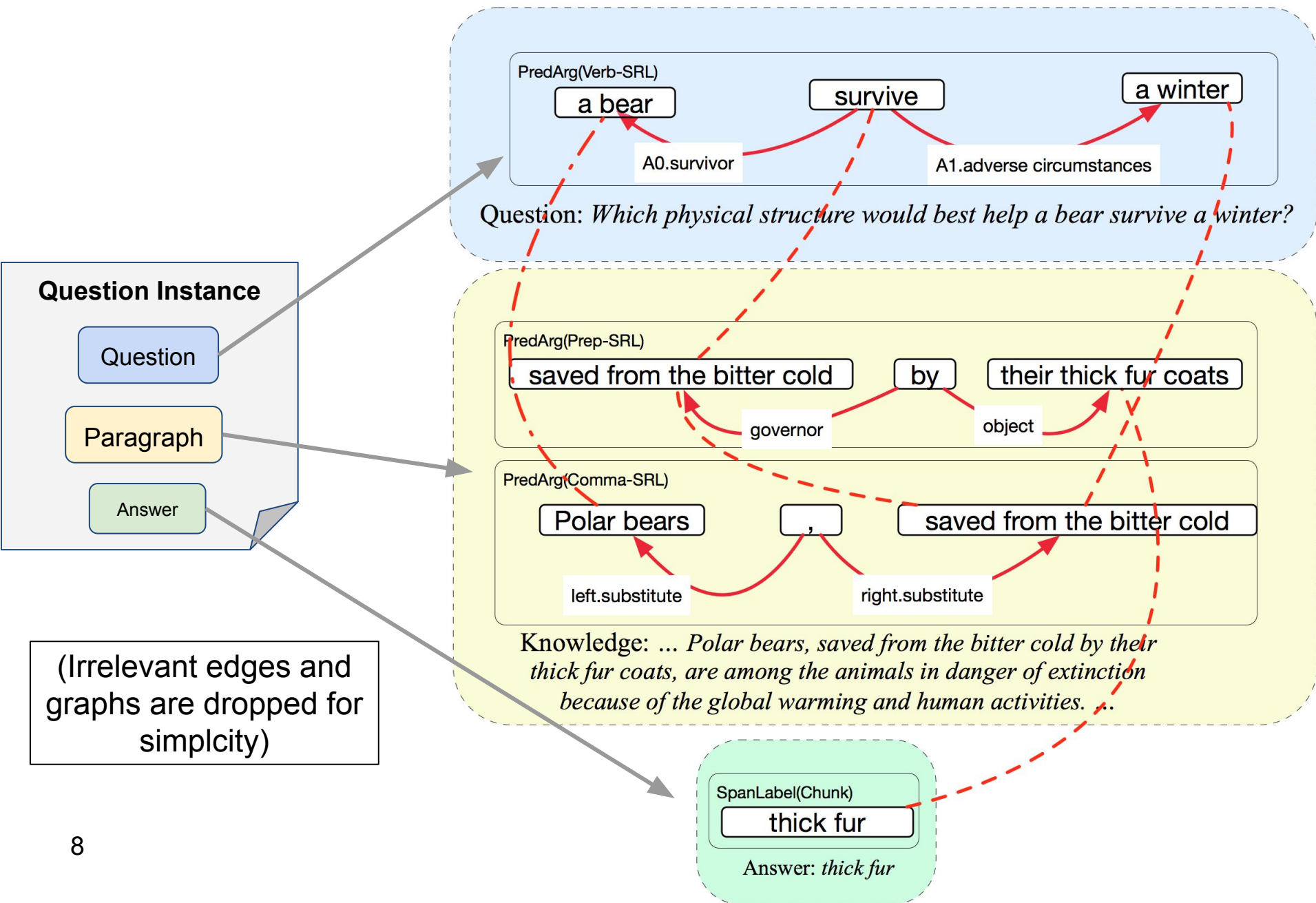
- **Augmented Graph** is the graph which contains potential alignments between elements of any two graphs

Connections via similarity / entailment



Reasoning formulated as best subgraph reasoning

Example subgraph

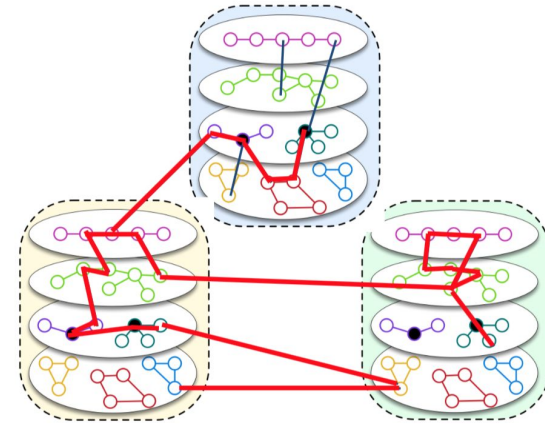


SemanticLP, some details.

Translate QA into a **search for an optimal** subgraph

Constraint: Incorporate **global** and **local** constraints

- **Global** e.g.
 - Have ends in question and paragraph
 - Connected graph
- **Local** e.g.
 - If using a pred-arg graphs,
 - use at least predicate and argument, or
 - use at least two arguments



Objective: Capture what's a valid reasoning, what's preferred

- **Preferences** e.g.
 - Use sentences nearby
 - If using a pred-arg graph, give priority to the subject

Formulate as Integer Linear Program (**ILP**) **optimization**

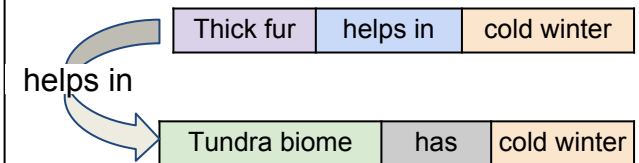
- Solution points to the best supported answer

Evaluation: notable baselines

- IR (Clark et al, AAI'15)
 - Information retrieval baseline (Lucene)
 - Using 280 GB of plain text
- TupleINF (Khot et al, ACL'17)
 - Inference over **independent rows**
 - Auto-generated short triples**
 - And type-constrained rules

Thick white fur is an animal adaptation **most needed** for **the climate** in which biome?
(A) deserts (B) taiga (C) deciduous forest (D) **tundra**

Type constrained rules:
(X, **helps in**, Y), (Z, **has**, Y) => (X, **helps in**, Z)

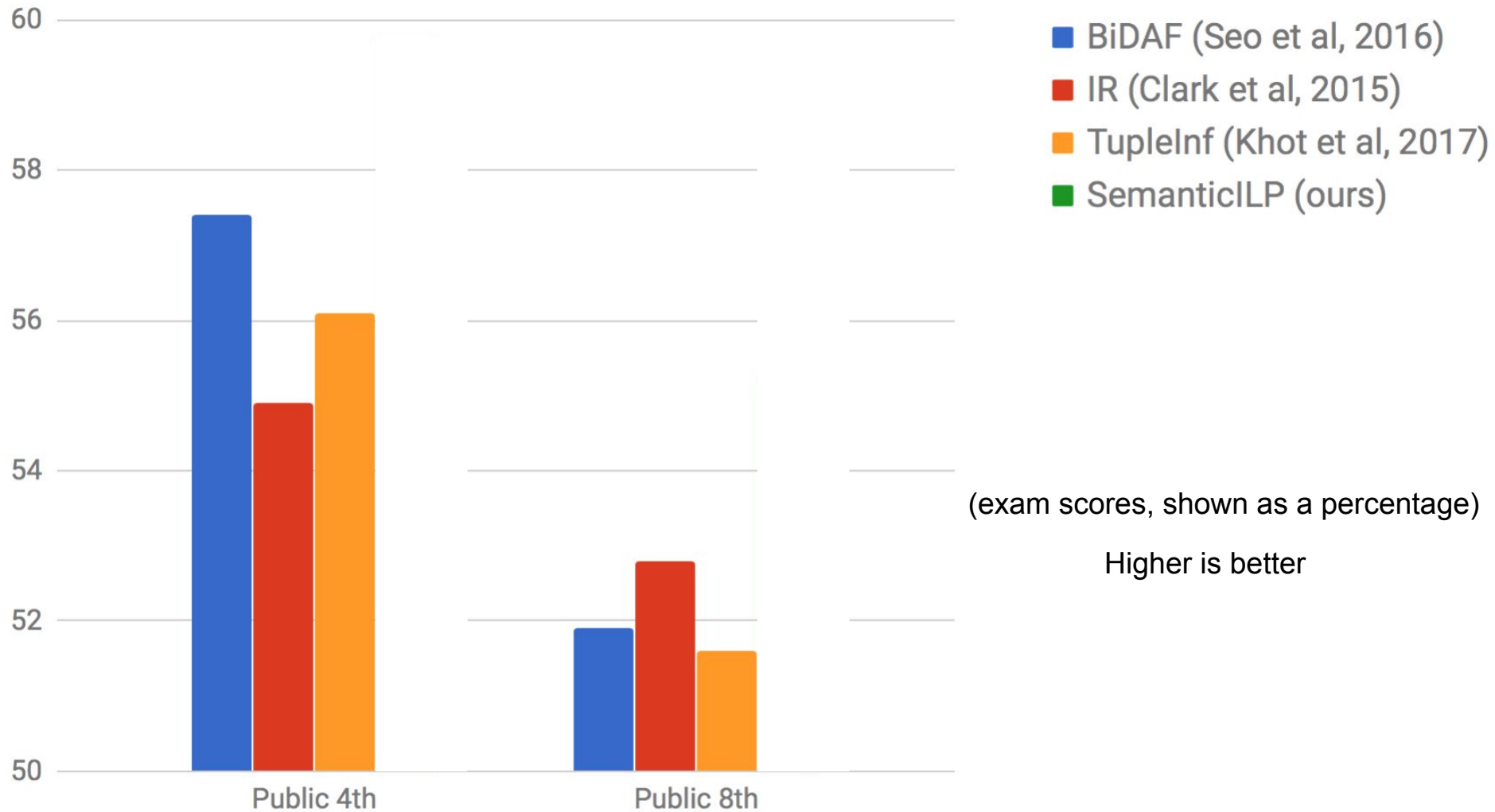


- BiD
 - We compare with the best baseline on each domain.
 - However we use one version of our systems across all the datasets.

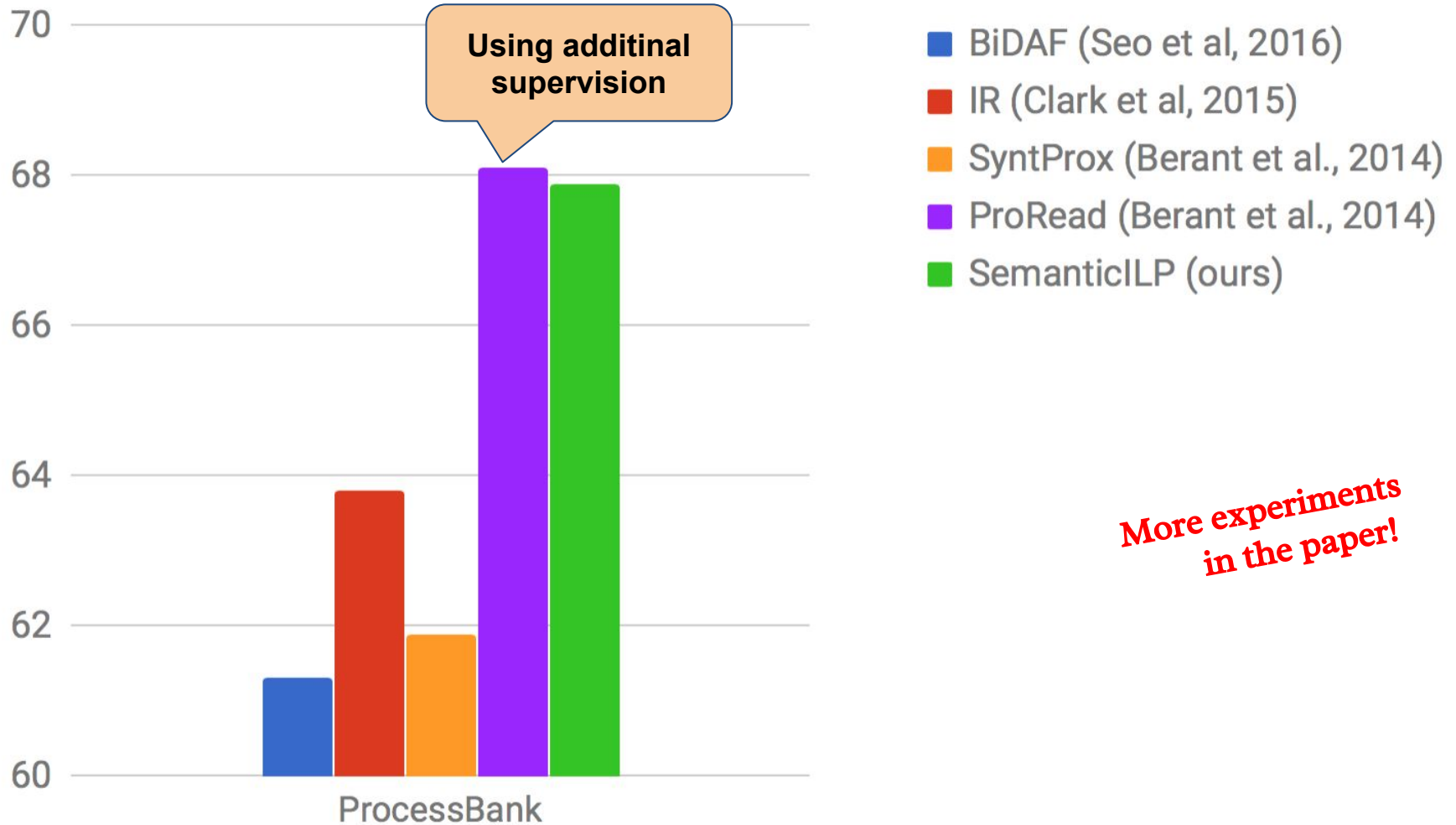
Barak Obama is the president of the U.S.

Who leads the United States?

Results #1: Science Questions



Results #2: Biology Questions



One single system tested on different datasets.

Concluding remarks

- Reasoning over language requires dealing with diverse set of linguistic phenomena.
- Linguistic variability \Rightarrow collection of semantic abstraction that are linguistically informed
- We decoupled “reasoning for QA” from “abstraction”
- Strong performance on two domains simultaneously

Give it a try:

<https://github.com/allenai/semantici1p>

CogComp-NLP:

<https://github.com/CogComp/cogcomp-nlp>

Extra slides

- Reasoning over language requires dealing with linguistic phenomena.
- Reasons over a wide range of semantic abstractions of the text
- Strong performance on two domains simultaneously

Give it a try:

<https://github.com/allenai/semanticilp>

CogComp-NLP:

<https://github.com/CogComp/cogcomp-nlp>

Concluding

Linguistic variability



Which physical structure would best **help a bear to survive a winter?**

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

Thick fur helps a bear survive a winter.

A thick coat of white fur helps bears survive in these cold latitudes.

Polar bears, saved from the bitter cold by their thick fur coats, are among the animals in danger of extinction because of the ~~global warming~~ and human activities.

brown noses

A given “meaning” can be phrased many surface forms!

Semantic understanding can help!

verb

Which physical structure would best help a bear to **survive** a winter?

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

■ Verb predicate: survive

- Survivor: “a bear”
- Adverse circumstances: “a winter”

comma

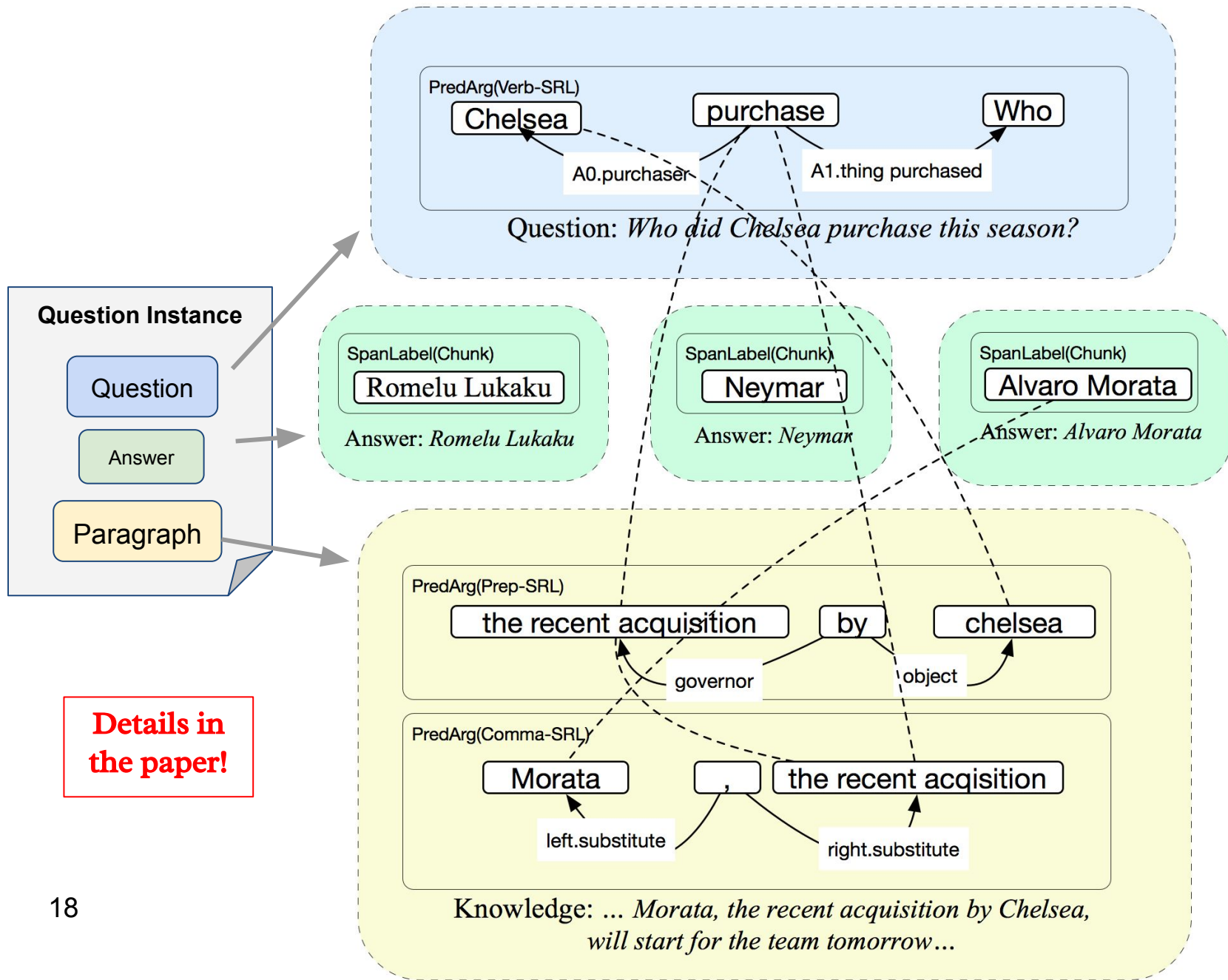
preposition

- A single abstraction is not enough:
 - We use off-the-shelf linguistically-informed annotators
 - Coref, Shallow semantic representations, etc.

■ Comma predicate: , (sense: substitute)

- indicates an apposition structure

SemanticLP: Example



Linguistic Complexity

how many soldiers died in valley forge by typhus and typhoid



All

News

Images

Shopping

Videos

More

Settings

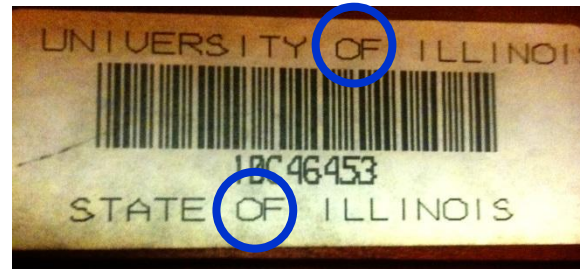
Tools

About 7,220 results (0.55 seconds)

Yet cold and starvation were not the most dangerous threats to soldiers at Valley Forge: Diseases like influenza, dysentery, typhoid and typhus killed two-thirds of the nearly **2,000 soldiers** who died during the encampment. Dec 19, 2012

[235 Years Ago, Washington's Troops Made Camp at Valley Forge ...](http://www.history.com/news/235-years-ago-washingtons-troops-made-camp-at-valley-forge)
www.history.com/news/235-years-ago-washingtons-troops-made-camp-at-valley-forge

Typhus	
killed	
nearly	
two-thirds	Governor
of	PartWhole (of)
the	
2000	
soldiers	Object

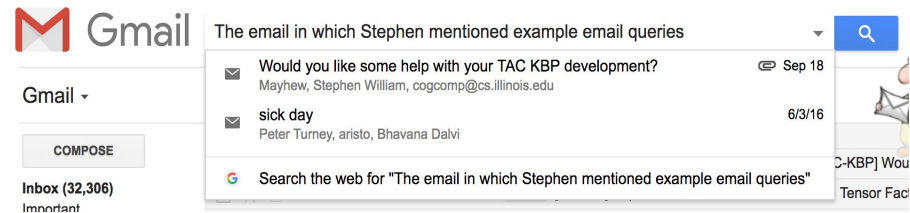
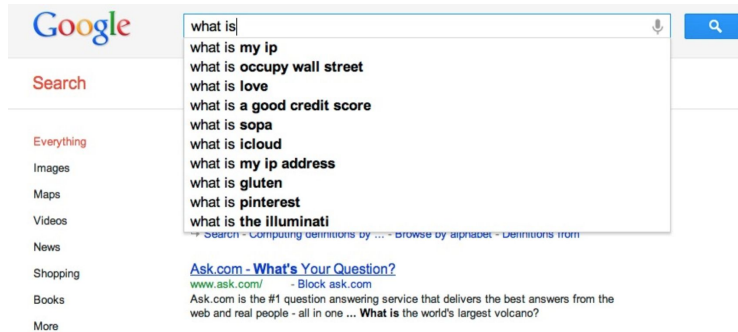


Not all "of"s are the same

- One argument is a part of another.
 - The governor is a number
 - The object is a group modified by the governor.

QA is everywhere

- One of the oldest problems in AI
- Remarkable features of QA



QA systems are still far from exhibiting human-like intelligence, even in relatively simple ways (vs. human-level)

Results #2: Biology Questions

Dataset	BiDAF	BiDAF tuned	IR	SyntProx Baseline*	ProRead* (structural supervision)	SemanticLP
Process Bank**	58.7	61.3	63.8	61.9	68.1	67.9

SemanticLP does not rely on domain-specific process structure annotation

- Close to the specialized, state-of-the-art ProRead system
- Substantially better than syntax-based and neural baselines

One single system tested on different datasets.

**More experiments
in the paper!**

* Berant et al. (EMNLP, 2014)

** ~70% of the original dataset; true/false and temporal questions currently out of scope

QA solvers; getting lucky

Widely accepted that QA systems get lucky when answering questions, because they fail with small *variations*.

A screenshot of a Google search result for the query "what's the biggest airport in moscow". The search bar contains the query, and the word "smallest" is circled in red. Below the search bar, there are tabs for "All", "Maps", "Images", "News", "Shopping", and "More". The search results show "About 4,310,000 results (0.88 seconds)". The main result is "Domodedovo International Airport" with the subtitle "Russia's busiest airports by passenger traffic in 2016". Below this is a table with columns "Rank" and "Airport".

Rank	Airport
1	Sheremetyevo International Airport
2	Domodedovo International Airport
3	Vnukovo International Airport
4	Pulkovo International Airport

74 more rows

A screenshot of a Google search result for the query "what's the smallest airport in moscow". The search bar contains the query, and the word "smallest" is circled in red. Below the search bar, there are tabs for "All", "Maps", "Images", "Videos", "News", and "More". The search results show "About 375,000 results (0.63 seconds)". The main result is "List of airports in Russia - Wikipedia" with the URL "https://en.wikipedia.org/wiki/List_of_airports_in_Russia". Below this is a snippet of text: "List of airports in Russia (Russian Federation), sorted by location. There are 270 airports ... Moscow, UDDD, DME, Domodedovo International Airport · Moscow · Khodinka Airport · Moscow, UUMO, OSF, Ostafyevo International Airport · Moscow, UUEE ...". Below this is another result: "List of the busiest airports in Russia - Wikipedia" with the URL "https://en.wikipedia.org/wiki/List_of_the_busiest_airports_in_Russia". Below this is a snippet of text: "This is a list of the busiest airports in Russia, using data from the Federal Air Transport Agency. 7.4%, Steady, 2, Sheremetyevo International Airport · Moscow". Below this is a third result: "Sheremetyevo International Airport - Wikipedia" with the URL "https://en.wikipedia.org/wiki/Sheremetyevo_International_Airport". Below this is a snippet of text: "Sheremetyevo International Airport (IATA: SVO, ICAO: UUEE) is an international airport located The Moscow Oblast government has reserved adjacent land for a future third runway. Tools. What links here · Related changes · Upload file · Special pages · Permanent link · Page information · Wikidata item · Cite this page ...".

Results #1: Science Questions

- Paragraphs obtained by concatenating top k Lucene-retrieved sentences for various answer options

best overall

best baseline

Dataset
Regents 4th
Public 4th
Regents 8th
Public 8th

(exam scores, shown as a percentage)



How many ways can we encode the knowledge required for the following question?

Which physical structure would best **help** a bear to **survive** a winter?
(A) big ears (B) black nose (C) **thick fur** (D) brown eyes

verb

comma

preposition

Polar bears, saved from the bitter cold **by** their thick fur coats, are among the animals in danger of extinction because of the global warming and human activities.

Linguistic understanding can help!

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