Compositional Semantic Parsing on Semi-Structured Tables

Ice Pasupat and Percy Liang
Stanford University

ACL 2015
Tuesday, July 28, 2015
Task

Question answering given a knowledge source

In which city was Ada Lovelace born?
Semantic Parsing

Parse questions into executable **logical forms**

In which city was Ada Lovelace born?

$$\text{Type.City} \cap \text{PeopleBornHere.AdaLovelace}$$

(Lambda DCS)
Semantic Parsing

Logical forms can be executed on the knowledge source to get denotations

Type.City ⊓ PeopleBornHere.AdaLovelace
Semantic Parsing

Logical forms can be executed on the knowledge source to get denotations

Type.City \sqcap \text{PeopleBornHere.AdaLovelace}

Shanghai
  Type → City

London
  Type → England
  Capital

Database
Semantic Parsing

Logical forms can be executed on the knowledge source to get denotations.

Type.City ⊓ PeopleBornHere.AdaLovelace
Semantic Parsing

Logical forms can be executed on the knowledge source to get denotations

Type.City ⊓ PeopleBornHere.AdaLovelace

Database

Shanghai
London ...

...
Compositionality

We can compose logical forms into bigger ones with logical operations

Type.City \sqcap PeopleBornHere.AdaLovelace

Shanghai
London ...

England
London ...

Compositionality

We can **compose** logical forms into bigger ones with logical operations

**Intersection**

Type.City $\cap$ PeopleBornHere.AdaLovelace
Compositionality

We can compose logical forms into bigger ones with logical operations

Intersection

Type.City \cap PeopleBornHere.AdaLovelace

London
Compositionality

We can compose logical forms into bigger ones with logical operations

- $\text{Type.City} \sqcup \text{Type.State} \rightarrow \text{cities and / or states}$
- $\text{count}(\text{Type.City}) \rightarrow \text{how many cities}$
- $\text{argmax}(\text{Type.City}, \text{Area}) \rightarrow \text{largest city}$
- $\text{sum} (\text{AreaOf.Type.City}) \rightarrow \text{total area of all cities}$
- $\text{AreaOf.London} – \text{AreaOf.Paris} \rightarrow \text{how much bigger is London than Paris?}$
Related Work

Early systems: Parse very compositional questions into database queries

How many rivers are in the state with the largest population?

\[
\text{answer}(A, \\
\quad \text{count}(B, \\
\quad (\text{river}(B), \text{loc}(B, C), \\
\quad \text{largest}(D, (\text{state}(C), \text{population}(C, D)))), \\
\quad A))
\]

Compositionality: High

**Related Work**

**Early systems:** Parse very compositional questions into database queries

How many rivers are in the state with the largest population?

answer(A, count(B, (river(B), loc(B, C), largest(D, (state(C), population(C, D)))), A))

**Compositionality:** High

**Knowledge source:** Database
- few entities / relations
- fixed schema

Related Work

Depth
(compositionality)

Early Systems

Breadth
(domain size)
Related Work

Scaling to large knowledge bases (KBs): Answer open-domain questions using curated KBs

In which comic book issue did Kitty Pryde first appear?

Knowledge source: Large KBs
- lots of entities / relations
- fixed schema

[Cai + Yates, 2013 / Berant et al., 2013 + 2014 / Fader et al., 2014 / Reddy et al., 2014 / ...]
Related Work

Scaling to large knowledge bases (KBs): Answer open-domain questions using curated KBs

In which comic book issue did Kitty Pryde first appear?

\[ R[\text{FirstAppearance}.\text{KittyPryde}] \]

Compositionality: Lower

Knowledge source: Large KBs
- lots of entities / relations
- fixed schema

[Cai + Yates, 2013 / Berant et al., 2013 + 2014 / Fader et al., 2014 / Reddy et al., 2014 / ...]
Related Work

Scaling to large knowledge bases (KBs): Answer open-domain questions using curated KBs

In which comic book issue did Kitty Pryde first appear?

R[FirstAppearance].KittyPryde

Still, only < 10% of general questions can be answered by Freebase [Berant et al., 2013]

Compositionality: Lower

Knowledge source: Large KBs
  ▶ lots of entities / relations
  ▶ fixed schema

[Cai + Yates, 2013 / Berant et al., 2013 + 2014 / Fader et al., 2014 / Reddy et al., 2014 / ...]
Related Work

Breadth
(domain size)

Depth
(compositionality)

Early Systems

Scale to KBs
Related Work

**Web search:** Keyword search over the whole Web (information retrieval / not semantic parsing)

- stanford cs professors

**Compositionality:** None

**Knowledge source:** Internet
- open-domain
- unstructured (no schema)
Related Work

Depth
(compositionality)

Breadth
(domain size)

Early Systems

Scale to KBs

Web Search
Motivation

Web text in general is too unstructured

However, the Web also contains semi-structured data (tables, lists, repeated headings, ...)
stanford cs professors
http://cs.stanford.edu/faculty

<table>
<thead>
<tr>
<th>Name</th>
<th>Phone</th>
<th>Office</th>
<th>email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maneesh Agrawala</td>
<td>5-3359</td>
<td>GATES 411</td>
<td>aiken</td>
</tr>
<tr>
<td>Alex Aiken</td>
<td>3-3334</td>
<td>Clark S266</td>
<td>serafim</td>
</tr>
<tr>
<td>Peter Bailis</td>
<td>650 723-7666</td>
<td>Beckman B321</td>
<td>pballis</td>
</tr>
<tr>
<td>Serafim Batzoglou</td>
<td>4-1248</td>
<td>Gates 384</td>
<td>msb</td>
</tr>
<tr>
<td>Gill Bejerano</td>
<td>5-3897</td>
<td>GATES 475</td>
<td>dabo</td>
</tr>
<tr>
<td>Michael Bernstein</td>
<td>5-9145</td>
<td>GATES 301</td>
<td></td>
</tr>
<tr>
<td>Dan Boneh</td>
<td>3-1131</td>
<td>GATES 439</td>
<td>cheriton</td>
</tr>
<tr>
<td>Moses Charikar</td>
<td>723-9798</td>
<td>Gates 190</td>
<td>coopers</td>
</tr>
<tr>
<td>David Cheriton</td>
<td>5-9145</td>
<td>GATES 344</td>
<td>dill</td>
</tr>
<tr>
<td>Steve Cooper</td>
<td>497-8586</td>
<td>Gates 204</td>
<td>rondror</td>
</tr>
<tr>
<td>Bill Dally</td>
<td>3-0762</td>
<td>GATES 314</td>
<td>engler</td>
</tr>
<tr>
<td>David Dill</td>
<td>3-0685</td>
<td>GATES 434</td>
<td>hector</td>
</tr>
<tr>
<td>Ron Dror</td>
<td>3-0934</td>
<td>GATES 220</td>
<td></td>
</tr>
</tbody>
</table>
Motivation

Web text in general is too unstructured

However, the Web also contains semi-structured data (tables, lists, repeated headings, ...)

- Open-domain: lots of information with arbitrary data schema [Cafarella et al., 2008 (WebTables)]
Motivation

Web text in general is too unstructured

However, the Web also contains semi-structured data (tables, lists, repeated headings, ...)

- **Open-domain**: lots of information with arbitrary data schema [Cafarella et al., 2008 (WebTables)]
- **Structured enough** to allow complex logical operations (~ mini knowledge base)

How many Stanford CS professors do not have offices in the Gates building?
Motivation

Web text in general is too unstructured

However, the Web also contains semi-structured data (tables, lists, repeated headings, ...)

- **Open-domain**: lots of information with arbitrary data schema [Cafarella et al., 2008 (WebTables)]
- **Structured enough** to allow complex logical operations (~ mini knowledge base)

**Task**: Answer compositional questions based on semi-structured tables from the Web
Motivation

Depth
(compositionality)

Early Systems

Scale to KBs

Web Search

Breadth
(domain size)

Semantic Parsing on Semi-Structured Data
Outline

- Background and Related Work
- Task and Dataset
- Approach
- Experiments
Task Description

Input: utterance \( x \) and HTML table \( t \)

Output: answer \( y \)

<table>
<thead>
<tr>
<th>Year</th>
<th>City</th>
<th>Country</th>
<th>Nations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1896</td>
<td>Athens</td>
<td>Greece</td>
<td>14</td>
</tr>
<tr>
<td>1900</td>
<td>Paris</td>
<td>France</td>
<td>24</td>
</tr>
<tr>
<td>1904</td>
<td>St. Louis</td>
<td>USA</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>2004</td>
<td>Athens</td>
<td>Greece</td>
<td>201</td>
</tr>
<tr>
<td>2008</td>
<td>Beijing</td>
<td>China</td>
<td>204</td>
</tr>
<tr>
<td>2012</td>
<td>London</td>
<td>UK</td>
<td>204</td>
</tr>
</tbody>
</table>

\( x = \) Greece held its last Summer Olympics in which year? 

\( y = 2004 \)
Task Description

Input: utterance $x$ and HTML table $t$

Output: answer $y$

Training data: list of $(x, t, y)$ — no logical form

Tables in test data are **not seen** during training

- The model must generalize to unseen table schemas!
Dataset

WikiTableQuestions dataset:

- Tables $t$ are from Wikipedia
<table>
<thead>
<tr>
<th>Year</th>
<th>Competition</th>
<th>Venue</th>
<th>Position</th>
<th>Event</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>World Youth Championships</td>
<td>Debrecen, Hungary</td>
<td>2nd</td>
<td>400 m</td>
<td>47.12</td>
</tr>
<tr>
<td></td>
<td>European Junior Championships</td>
<td>Grosseto, Italy</td>
<td>1st</td>
<td>Medley relay</td>
<td>1:50.46</td>
</tr>
<tr>
<td>2003</td>
<td>European Junior Championships</td>
<td>Tampere, Finland</td>
<td>3rd</td>
<td>400 m</td>
<td>46.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2nd</td>
<td>4x400 m relay</td>
<td>3:08.62</td>
</tr>
<tr>
<td>2005</td>
<td>European U23 Championships</td>
<td>Erfurt, Germany</td>
<td>11th (sf)</td>
<td>400 m</td>
<td>46.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1st</td>
<td>4x400 m relay</td>
<td>3:04.41</td>
</tr>
<tr>
<td></td>
<td>Universiade</td>
<td>Izmir, Turkey</td>
<td>7th</td>
<td>400 m</td>
<td>46.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1st</td>
<td>4x400 m relay</td>
<td>3:02.57</td>
</tr>
<tr>
<td>2006</td>
<td>World Indoor Championships</td>
<td>Moscow, Russia</td>
<td>2nd (h)</td>
<td>4x400 m relay</td>
<td>3:06.10</td>
</tr>
<tr>
<td></td>
<td>European Championships</td>
<td>Gothenburg, Sweden</td>
<td>3rd</td>
<td>4x400 m relay</td>
<td>3:01.73</td>
</tr>
<tr>
<td>2007</td>
<td>European Indoor Championships</td>
<td>Birmingham, United Kingdom</td>
<td>3rd</td>
<td>4x400 m relay</td>
<td>3:08.14</td>
</tr>
<tr>
<td></td>
<td>Universiade</td>
<td>Bangkok, Thailand</td>
<td>7th</td>
<td>400 m</td>
<td>46.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1st</td>
<td>4x400 m relay</td>
<td>3:02.05</td>
</tr>
<tr>
<td>2008</td>
<td>World Indoor Championships</td>
<td>Valencia, Spain</td>
<td>4th</td>
<td>4x400 m relay</td>
<td>3:08.76</td>
</tr>
<tr>
<td></td>
<td>Olympic Games</td>
<td>Beijing, China</td>
<td>7th</td>
<td>4x400 m relay</td>
<td>3:00.32</td>
</tr>
<tr>
<td>2009</td>
<td>Universiade</td>
<td>Belgrade, Serbia</td>
<td>2nd</td>
<td>4x400 m relay</td>
<td>3:05.69</td>
</tr>
</tbody>
</table>

[31](https://en.wikipedia.org/wiki/Piotr_Kędzia)
Dataset

WikiTableQuestions dataset:

- Tables $t$ are from Wikipedia
- Questions $x$ and answers $y$ are from Mechanical Turk — Prompts are given to encourage compositionality

How many ...     ... ___est ...     ... last ...
... above ...    ... same ... as ...    ... difference ...
... or ...       ... his ...           Requires counting

etc.
Prompt: The question must contains "last" (or a synonym)

In what city did Piotr's last 1st place finish occur?
<table>
<thead>
<tr>
<th>Year</th>
<th>Competition</th>
<th>Venue</th>
<th>Position</th>
<th>Event</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>World Youth Championships</td>
<td>Debrecen, Hungary</td>
<td>2nd</td>
<td>400 m</td>
<td>47.12</td>
</tr>
<tr>
<td></td>
<td>European Junior Championships</td>
<td>Grosseto, Italy</td>
<td>1st</td>
<td>Medley relay</td>
<td>1:50.46</td>
</tr>
<tr>
<td>2003</td>
<td>European Junior Championships</td>
<td>Tampere, Finland</td>
<td>3rd</td>
<td>400 m</td>
<td>46.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2nd</td>
<td>4x400 m relay</td>
<td>3:08.62</td>
</tr>
<tr>
<td></td>
<td>European U23 Championships</td>
<td>Erfurt, Germany</td>
<td>11th (sf)</td>
<td>400 m</td>
<td>46.62</td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td></td>
<td>1st</td>
<td>4x400 m relay</td>
<td>3:04.41</td>
</tr>
<tr>
<td></td>
<td>Universiade</td>
<td>Izmir, Turkey</td>
<td>7th</td>
<td>400 m</td>
<td>46.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1st</td>
<td>4x400 m relay</td>
<td>3:02.57</td>
</tr>
<tr>
<td>2006</td>
<td>World Indoor Championships</td>
<td>Moscow, Russia</td>
<td>2nd (h)</td>
<td>4x400 m relay</td>
<td>3:06.10</td>
</tr>
<tr>
<td></td>
<td>European Championships</td>
<td>Gothenburg, Sweden</td>
<td>3rd</td>
<td>4x400 m relay</td>
<td>3:01.73</td>
</tr>
<tr>
<td>2007</td>
<td>European Indoor Championships</td>
<td>Birmingham, United Kingdom</td>
<td>3rd</td>
<td>4x400 m relay</td>
<td>3:08.14</td>
</tr>
<tr>
<td></td>
<td>Universiade</td>
<td>Bangkok, Thailand</td>
<td>7th</td>
<td>400 m</td>
<td>46.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1st</td>
<td>4x400 m relay</td>
<td>3:02.05</td>
</tr>
<tr>
<td>2008</td>
<td>World Indoor Championships</td>
<td>Valencia, Spain</td>
<td>4th</td>
<td>4x400 m relay</td>
<td>3:08.76</td>
</tr>
<tr>
<td></td>
<td>Olympic Games</td>
<td>Beijing, China</td>
<td>7th</td>
<td>4x400 m relay</td>
<td>3:00.32</td>
</tr>
<tr>
<td>2009</td>
<td>Universiade</td>
<td>Belgrade, Serbia</td>
<td>2nd</td>
<td>4x400 m relay</td>
<td>3:05.69</td>
</tr>
</tbody>
</table>

In what city did Piotr's last 1st place finish occur?
<table>
<thead>
<tr>
<th>Year</th>
<th>Competition</th>
<th>Venue</th>
<th>Position</th>
<th>Event</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>World Youth Championships</td>
<td>Debrecen, Hungary</td>
<td>2nd</td>
<td>Medley relay</td>
<td>47.12</td>
</tr>
<tr>
<td></td>
<td>European Junior Championships</td>
<td>Grosseto, Italy</td>
<td>1st</td>
<td>4x400 m relay</td>
<td>3:06.12</td>
</tr>
<tr>
<td>2003</td>
<td>European Junior Championships</td>
<td>Tampere, Finland</td>
<td>3rd</td>
<td>400 m</td>
<td>46.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2nd</td>
<td>4x400 m relay</td>
<td>3:08.62</td>
</tr>
<tr>
<td></td>
<td>European U23 Championships</td>
<td>Erfurt, Germany</td>
<td>11th (sf)</td>
<td>400 m</td>
<td>46.62</td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td></td>
<td>1st</td>
<td>4x400 m relay</td>
<td>3:04.41</td>
</tr>
<tr>
<td></td>
<td>Universiade</td>
<td>Izmir, Turkey</td>
<td>7th</td>
<td>400 m</td>
<td>46.89</td>
</tr>
<tr>
<td></td>
<td>World Indoor Championships</td>
<td>Moscow, Russia</td>
<td>2nd (h)</td>
<td>4x400 m relay</td>
<td>3:06.10</td>
</tr>
<tr>
<td></td>
<td>European Championships</td>
<td>Gothenburg, Sweden</td>
<td>3rd</td>
<td>4x400 m relay</td>
<td>3:01.73</td>
</tr>
<tr>
<td>2007</td>
<td>European Indoor Championships</td>
<td>Birmingham, UK</td>
<td>3rd</td>
<td>4x400 m relay</td>
<td>3:08.14</td>
</tr>
<tr>
<td></td>
<td>Universiade</td>
<td>Bangkok, Thailand</td>
<td>7th</td>
<td>400 m</td>
<td>46.85</td>
</tr>
<tr>
<td></td>
<td>World Indoor Championships</td>
<td>Valencia, Spain</td>
<td>4th</td>
<td>4x400 m relay</td>
<td>3:08.76</td>
</tr>
<tr>
<td>2008</td>
<td>Olympic Games</td>
<td>Beijing, China</td>
<td>7th</td>
<td>4x400 m relay</td>
<td>3:00.32</td>
</tr>
<tr>
<td></td>
<td>Universiade</td>
<td>Belgrade, Serbia</td>
<td>2nd</td>
<td>4x400 m relay</td>
<td>3:05.69</td>
</tr>
</tbody>
</table>

In what city did Piotr's last 1st place finish occur?
In what city did Piotr's last 1st place finish occur?
In what city did Piotr's last 1st place finish occur?

<table>
<thead>
<tr>
<th>Year</th>
<th>Competition</th>
<th>Venue</th>
<th>Position</th>
<th>Event</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>World Youth Championships</td>
<td>Debrecen, Hungary</td>
<td>2nd</td>
<td>400 m</td>
<td>47.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1st</td>
<td>Medley relay</td>
<td>1:50.46</td>
</tr>
<tr>
<td>2003</td>
<td>European Junior Championships</td>
<td>Grosseto, Italy</td>
<td>1st</td>
<td>4x400 m relay</td>
<td>3:06.12</td>
</tr>
<tr>
<td>2005</td>
<td>European Junior Championships</td>
<td>Tampere, Finland</td>
<td>3rd</td>
<td>400 m</td>
<td>46.69</td>
</tr>
<tr>
<td></td>
<td>European U23 Championships</td>
<td>Erfurt, Germany</td>
<td>11th (sf)</td>
<td>400 m</td>
<td>46.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1st</td>
<td>4x400 m relay</td>
<td>3:04.41</td>
</tr>
<tr>
<td>2006</td>
<td>Universiade</td>
<td>Izmir, Turkey</td>
<td>7th</td>
<td>400 m</td>
<td>46.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1st</td>
<td>4x400 m relay</td>
<td>3:02.57</td>
</tr>
<tr>
<td>2007</td>
<td>World Indoor Championships</td>
<td>Moscow, Russia</td>
<td>2nd (h)</td>
<td>4x400 m relay</td>
<td>3:06.10</td>
</tr>
<tr>
<td></td>
<td>European Championships</td>
<td>Gothenburg, Sweden</td>
<td>3rd</td>
<td>4x400 m relay</td>
<td>3:01.73</td>
</tr>
<tr>
<td></td>
<td>European Indoor Championships</td>
<td>Birmingham, UK</td>
<td>3rd</td>
<td>4x400 m relay</td>
<td>3:08.14</td>
</tr>
<tr>
<td>2008</td>
<td>Universiade</td>
<td>Bangkok, Thailand</td>
<td>7th</td>
<td>400 m</td>
<td>46.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1st</td>
<td>4x400 m relay</td>
<td>3:02.05</td>
</tr>
<tr>
<td>2009</td>
<td>World Indoor Championships</td>
<td>Valencia, Spain</td>
<td>4th</td>
<td>4x400 m relay</td>
<td>3:08.76</td>
</tr>
<tr>
<td></td>
<td>Olympic Games</td>
<td>Beijing, China</td>
<td>7th</td>
<td>4x400 m relay</td>
<td>3:00.32</td>
</tr>
<tr>
<td></td>
<td>Universiade</td>
<td>Belgrade, Serbia</td>
<td>2nd</td>
<td>4x400 m relay</td>
<td>3:05.69</td>
</tr>
<tr>
<td>Year</td>
<td>Competition</td>
<td>Venue</td>
<td>Position</td>
<td>Event</td>
<td>Notes</td>
</tr>
<tr>
<td>------</td>
<td>------------------------------------</td>
<td>---------------------</td>
<td>----------</td>
<td>------------------</td>
<td>---------</td>
</tr>
<tr>
<td>2001</td>
<td>World Youth Championships</td>
<td>Debrecen, Hungary</td>
<td>2nd</td>
<td>400 m</td>
<td>47.12</td>
</tr>
<tr>
<td></td>
<td>European Junior Championships</td>
<td>Grosseto, Italy</td>
<td>1st</td>
<td>Medley relay</td>
<td>1:50.46</td>
</tr>
<tr>
<td>2003</td>
<td>European Junior Championships</td>
<td>Tampere, Finland</td>
<td>3rd</td>
<td>400 m</td>
<td>46.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2nd</td>
<td>4x400 m relay</td>
<td>3:08.62</td>
</tr>
<tr>
<td>2005</td>
<td>European U23 Championships</td>
<td>Erfurt, Germany</td>
<td>11th (sf)</td>
<td>400 m</td>
<td>46.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1st</td>
<td>4x400 m relay</td>
<td>3:04.41</td>
</tr>
<tr>
<td></td>
<td>Universiade</td>
<td>Izmir, Turkey</td>
<td>7th</td>
<td>400 m</td>
<td>46.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1st</td>
<td>4x400 m relay</td>
<td>3:02.57</td>
</tr>
<tr>
<td>2006</td>
<td>World Indoor Championships</td>
<td>Moscow, Russia</td>
<td>2nd (h)</td>
<td>4x400 m relay</td>
<td>3:06.10</td>
</tr>
<tr>
<td></td>
<td>European Championships</td>
<td>Gothenburg, Sweden</td>
<td>3rd</td>
<td>4x400 m relay</td>
<td>3:01.73</td>
</tr>
<tr>
<td>2007</td>
<td>European Indoor Championships</td>
<td>Birmingham, UK</td>
<td>3rd</td>
<td>4x400 m relay</td>
<td>3:08.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7th</td>
<td>400 m</td>
<td>46.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1st</td>
<td>4x400 m relay</td>
<td>3:02.05</td>
</tr>
<tr>
<td>2008</td>
<td>World Indoor Championships</td>
<td>Valencia, Spain</td>
<td>4th</td>
<td>4x400 m relay</td>
<td>3:08.76</td>
</tr>
<tr>
<td></td>
<td>Olympic Games</td>
<td>Beijing, China</td>
<td>7th</td>
<td>4x400 m relay</td>
<td>3:00.32</td>
</tr>
<tr>
<td>2009</td>
<td>Universiade</td>
<td>Belgrade, Serbia</td>
<td>2nd</td>
<td>4x400 m relay</td>
<td>3:05.69</td>
</tr>
</tbody>
</table>

How long did it take this competitor to finish the 4x400 meter relay at Universiade in 2005?
Where was the competition held immediately before the one in Turkey?
How many times has this competitor placed 5th or better in competition?

<table>
<thead>
<tr>
<th>Year</th>
<th>Competition</th>
<th>Venue</th>
<th>Position</th>
<th>Event</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>World Youth Championships</td>
<td>Debrecen, Hungary</td>
<td>2nd</td>
<td>400 m</td>
<td>47.12</td>
</tr>
<tr>
<td></td>
<td>European Junior Championships</td>
<td>Grosseto, Italy</td>
<td>1st</td>
<td>Medley relay</td>
<td>1:50.46</td>
</tr>
<tr>
<td>2003</td>
<td>European Junior Championships</td>
<td>Tampere, Finland</td>
<td>3rd</td>
<td>400 m</td>
<td>46.69</td>
</tr>
<tr>
<td></td>
<td>European U23 Championships</td>
<td>Erfurt, Germany</td>
<td>11th (sf)</td>
<td>400 m</td>
<td>46.62</td>
</tr>
<tr>
<td></td>
<td>Universiade</td>
<td>Izmir, Turkey</td>
<td>7th</td>
<td>400 m</td>
<td>46.62</td>
</tr>
<tr>
<td></td>
<td>World Indoor Championships</td>
<td>Moscow, Russia</td>
<td>2nd (h)</td>
<td>4x400 m relay</td>
<td>3:06.10</td>
</tr>
<tr>
<td></td>
<td>European Championships</td>
<td>Gothenburg, Sweden</td>
<td>3rd</td>
<td>4x400 m relay</td>
<td>3:01.73</td>
</tr>
<tr>
<td>2007</td>
<td>European Indoor Championships</td>
<td>Birmingham, United Kingdom</td>
<td>3rd</td>
<td>4x400 m relay</td>
<td>3:08.14</td>
</tr>
<tr>
<td></td>
<td>Universiade</td>
<td>Bangkok, Thailand</td>
<td>7th</td>
<td>400 m</td>
<td>46.85</td>
</tr>
<tr>
<td></td>
<td>World Indoor Championships</td>
<td>Valencia, Spain</td>
<td>4th</td>
<td>4x400 m relay</td>
<td>3:08.76</td>
</tr>
<tr>
<td>2009</td>
<td>Olympic Games</td>
<td>Beijing, China</td>
<td>7th</td>
<td>4x400 m relay</td>
<td>3:00.32</td>
</tr>
<tr>
<td></td>
<td>Universiade</td>
<td>Belgrade, Serbia</td>
<td>2nd</td>
<td>4x400 m relay</td>
<td>3:05.69</td>
</tr>
</tbody>
</table>
Dataset

WikiTableQuestions dataset:

- 2100 tables
  - Average: 6.3 columns / 27.5 rows
- 22000 examples
Challenges

With increased breadth (semi-structured data):

- Must generalize to arbitrary table schemas (as opposed to the fixed database schema)
- Test tables are unseen → Cannot precompute a lexicon mapping phrases to table relations

Table headers (Year, Competition, Venue, ...)

Challenges

With increased breadth (semi-structured data):

▸ Must generalize to arbitrary table schemas (as opposed to the fixed database schema)
▸ Test tables are unseen → Cannot precompute a lexicon mapping phrases to table relations

With increased depth (compositional questions):

▸ More operations and deeper recursion → Number of possible parses grows exponentially
Outline

▸ Background and Related Work
▸ Task and Dataset
▸ Approach
▸ Experiments
Approach

Greece held its last Summer Olympics in which year?

Greece held its last Summer Olympics in 2004.
Greece held its last Summer Olympics in which year?

R[λx[Year.Date.x]]. argmax(..., Index)

(3) Execution

2004
Greece held its last Summer Olympics in which year?

\( \text{Set of candidates} \rightarrow Z \)

\( R[\lambda x[\text{Year.Date}.x]]. \arg\max(..., \text{Index}) \)

\( (3) \text{ Execution} \)

\( y \rightarrow 2004 \)

\( (1) \text{ Generation} \)

\( x \rightarrow t \)
Greece held its last Summer Olympics in which year?

1. Generation
2. Ranking
3. Execution

\[ \text{argmax}(..., \text{Index}) \]

\[ R[\lambda x[\text{Year.Date.x}]]. \text{argmax}(..., \text{Index}) \]

2004
Greece held its last Summer Olympics in which year?

(1) Generation

(2) Ranking

\[ \text{R[λx[Year.Date.x]]. argmax(..., Index)} \]

(3) Execution

\[ 2004 \]
**Representation**

Convert table $t$ to knowledge graph $w$

<table>
<thead>
<tr>
<th>Year</th>
<th>City</th>
<th>Country</th>
<th>Nations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1896</td>
<td>Athens</td>
<td>Greece</td>
<td>14</td>
</tr>
<tr>
<td>1900</td>
<td>Paris</td>
<td>France</td>
<td>24</td>
</tr>
<tr>
<td>1904</td>
<td>St. Louis</td>
<td>USA</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>2004</td>
<td>Athens</td>
<td>Greece</td>
<td>201</td>
</tr>
<tr>
<td>2008</td>
<td>Beijing</td>
<td>China</td>
<td>204</td>
</tr>
<tr>
<td>2012</td>
<td>London</td>
<td>UK</td>
<td>204</td>
</tr>
</tbody>
</table>

![Knowledge graph diagram](image-url)
### Representation

Convert **table $t$** to **knowledge graph $w$**

<table>
<thead>
<tr>
<th>Year</th>
<th>City</th>
<th>Country</th>
<th>Nations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1896</td>
<td>Athens</td>
<td>Greece</td>
<td>14</td>
</tr>
<tr>
<td>1900</td>
<td>Paris</td>
<td>France</td>
<td>24</td>
</tr>
<tr>
<td>1904</td>
<td>St. Louis</td>
<td>USA</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>Athens</td>
<td>Greece</td>
<td>201</td>
</tr>
<tr>
<td>2008</td>
<td>Beijing</td>
<td>China</td>
<td>204</td>
</tr>
<tr>
<td>2012</td>
<td>London</td>
<td>UK</td>
<td>204</td>
</tr>
</tbody>
</table>
Greece held its last Summer Olympics in which year?

(1) Generation

(2) Ranking

R[\lambda x[Year.Date.x]]. argmax(..., Index)

(3) Execution

2004
Approach

Greece held its last Summer Olympics in which year?

(1) Generation

(2) Ranking

(3) Execution

\[ \text{argmax}(..., \text{Index}) \]

\[ 2004 \]
Greece held its last Summer Olympics in which year?

Approach

1. Generation

\[ R[\lambda x[Year\cdot Date\cdot x]]. \text{argmax}(\ldots, \text{Index}) \]

2. Ranking

\[ Z \]

3. Execution

\[ y \]

\[ \text{2004} \]
Generation

Build formulas **bottom-up** according to a set of deduction rules

\[
\text{R}[^\lambda_x[\text{Year.Date}.x]].\text{argmax(Country.Greece, Index)}
\]

Greece held its last Summer Olympics in which year?
Generation

Build formulas **bottom-up** according to a set of deduction rules

**Index**: 0

**Next**:
- Year: 1896
- City: Athens
- Country: Greece

Greece **held its last Summer Olympics in which year?**
Generation

Build formulas **bottom-up** according to a set of deduction rules

"anchored" to the word Greece

TokenSpan → Entity

Greece held its last Summer Olympics in which year?
Generation

Complication: Some logical predicates (e.g., relation Country) don't map to any phrase

\[ R[\lambda x[Year.Date.x]].\arg\max(Country.Greece, Index) \]

Greece held its last Summer Olympics in which year?
**Generation**

**Complication:** Some logical predicates (e.g., relation `Country`) don't map to any phrase.

$$R[\lambda x[\text{Year}.\text{Date}.x]].\text{argmax}(\text{Country}.\text{Greece}, \text{Index})$$

Even when there is such a phrase, we may still don't know the mapping if we have not seen the relation in any table in the training data.

Greece held its last Summer Olympics in which year?
Generation

Idea: Allow formulas to be created from nothing ("floating")

- Inspired by "bridging" [Berant et al., 2013]
Generation

Idea: Allow formulas to be created from nothing ("floating")

- Inspired by "bridging" [Berant et al., 2013]
Generation

Idea: Allow formulas to be created from nothing ("floating")

- Inspired by "bridging" [Berant et al., 2013]
Generation

- **Entities** are anchored to token spans
- **Relations** and **Operations** are not

The last row with country Greece

Greece held its last Summer Olympics in which year?
Connection between floating predicates and phrases in the question are made during ranking.
Problem: Over-generation due to high recursion

Handled by beam search and pruning heuristics
Greece held its last Summer Olympics in which year?

(1) Generation

\( R[\lambda x[\text{Year.Date}.x]]. \arg\max(..., \text{Index}) \)

(2) Ranking

(3) Execution

\( 2004 \)
Greece held its last Summer Olympics in which year?

1. Generation

2. Ranking

3. Execution

\[ \text{R}[\lambda x[\text{Year.Date.x}]]. \text{argmax}(\ldots, \text{Index}) \]

\[ \text{2004} \]
Ranking

Given a set $Z$ of candidate formulas $z$, define a log-linear distribution:

$$p_\theta(z \mid x, w) \propto \exp \{\theta^T \phi(x, w, z)\}$$

where

- $\theta =$ parameter vector
- $\phi(x, w, z) =$ feature vector
Ranking

Features:

- Relate phrases in $x$ to predicates in $z$
  
  (phrase = last, predicate = argmax)

  (phrase = year, predicate = Year)

  phrase == predicate
Ranking

Features:

- Relate phrases in $x$ to predicates in $z$
  - (phrase = last, predicate = $\text{argmax}$)
  - (phrase = year, predicate = $\text{Year}$)
    - phrase == predicate

- Relate phrases in $x$ to properties of $y = \lfloor [z] \rfloor_w$
  - (headword = year, answer's type = $\text{NUMBER}$)
    - headword == answer's column
Learning

Given training example \((x, w, y)\), define

\[
p_\theta(y | x, w) = \sum_{z \in Z} p_\theta(z | x, w) \mathbf{I}(y = [z]_w)
\]

As usual, we choose \(\theta\) to maximize the (L1 regularized) expectation of \(\log p_\theta(y | x, w)\) over training data.
Outline

▸ Background and Related Work
▸ Task and Dataset
▸ Approach
▸ Experiments
Results

- **Oracle**: Able to generate a candidate formula \( z \in Z \) that executes to \( y \)

- **Accuracy**: The highest-ranked \( z \) executes to \( y \)
Results

- **Oracle**: Able to generate a candidate formula $z \in Z$ that executes to $y$
- **Accuracy**: The highest-ranked $z$ executes to $y$

Two baselines:

- **IR-inspired**: Pick an answer among table cells by putting softmax over table cells
- **WQ**: Restrict the generation rules to the ones from Berant and Liang (2014)
## Results on Test Set

<table>
<thead>
<tr>
<th></th>
<th>accuracy</th>
<th>oracle</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR-inspired</td>
<td>12.7</td>
<td>70.6</td>
</tr>
<tr>
<td>WQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>This work</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In all settings, tables in test data are not seen during training.
## Results on Test Set

<table>
<thead>
<tr>
<th></th>
<th>accuracy</th>
<th>oracle</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR-inspired</td>
<td>12.7</td>
<td>70.6</td>
</tr>
<tr>
<td>WQ</td>
<td>24.3</td>
<td>35.6</td>
</tr>
<tr>
<td>This work</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In all settings, tables in test data are not seen during training.
Results on Test Set

<table>
<thead>
<tr>
<th></th>
<th>accuracy</th>
<th>oracle</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR-inspired</td>
<td>12.7</td>
<td>70.6</td>
</tr>
<tr>
<td>WQ</td>
<td>24.3</td>
<td>35.6</td>
</tr>
<tr>
<td>This work</td>
<td>37.1</td>
<td>76.6</td>
</tr>
</tbody>
</table>

In all settings, tables in test data are not seen during training.
Positive Examples

What is the last title that spicy horse produced?

How many districts have a population density of at least 1000?

Who finished directly after the driver who finished in 1:28.745?

(Information retrieval alone can't answer these questions)
(1) Anchoring [18%]

How many **mexican** swimmers ranked in the top 10?

<table>
<thead>
<tr>
<th>Rank</th>
<th>Swimmer</th>
<th>Country</th>
<th>Time</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mexico</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(2) Normalization [29%]

how long does the show defcon 3 last?

<table>
<thead>
<tr>
<th>ET</th>
<th>Days</th>
<th>Program</th>
<th>Hosts</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2pm–3pm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Error Analysis

(3) Unhandled Operations [19%]

- was there more gold medals won than silver? (boolean answer)
- which movies were number 1 for at least two consecutive weeks? (consecutive count)
- how many titles had the same author listed as the illustrator? (count rows with arbitrary conditions)
Error Analysis

(4) Ranking Errors [24%]

how many buildings on the list are taller than 200 feet?

<table>
<thead>
<tr>
<th>Name</th>
<th>Street Address</th>
<th>Years as Tallest</th>
<th>Height ft (m)</th>
<th>Floors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>792 (241)</td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

Depth
(compositionality)

Breadth
(domain size)

Early Systems

Scale to KBs

Semantic Parsing on Semi-Structured Data

Web Search
Conclusion

Dataset and reproducible experiments are available on CodaLab:

[Link to dataset and experiments]

nlp.stanford.edu/software/sempre/wikitible

Thank you