Natural Language Interfaces for Semi-Structured Web Pages

University Oral Examination
Ice Pasupat
Motivation
Motivation
Motivation

utterance

Alexa, play Despacito
Motivation

utterance
Alexa, play Despacito

actions
play(track=Despacito)
Motivation

utterance
Alexa, play Despacito

actions
play(track=Despacito)

response

environment
Motivation

natural language interface to database

sales in 2018 and 2019

[Hendrix et al., 1978 / Androutsopoulos et al., 1995 / Popescu et al., 2003 / ...]
Motivation

natural language interface to database

sales in 2018 and 2019

smart search engines

[Hendrix et al., 1978 / Androutsopoulos et al., 1995 / Popescu et al., 2003 / ...]
Goal: Extend the capability of these systems along two axes:
Motivation

**Goal:** Extend the capability of these systems along two axes:

- Scope of the environment (breadth)

How much is a Steak Burrito at Ray's Grill?
Goal: Extend the capability of these systems along two axes:

- Scope of the environment (breadth)

How much is a Steak Burrito at Ray's Grill?
Motivation

- semantic parsing on databases
- most virtual assistants
- database / apps
  limited schema

Scope of the environment (breadth)

[Zelle + Mooney, 1996 / Zettlemoyer + Collins, 2007 / Liang et al., 2011 / …]
Motivation

- Semantic parsing on databases
- Most virtual assistants
- Database / apps with limited schema
- Semantic parsing on knowledge bases
- Knowledge base with still limited schema

[Cai + Yates, 2013 / Berant et al., 2013 / Kwiatkowski et al., 2013 / Bordes et al., 2015 / ...]
Motivation

Scope of the environment (breadth)

- semantic parsing on databases
  - most virtual assistants
  - database / apps
    - limited schema

- semantic parsing on knowledge bases
  - knowledge base
    - still limited schema

- web scraping
  - web pages
    - open schema

- question answering on paragraphs
  - any texts
    - open schema

[Robertson et al., 19xx / Hearst, 1992 / Richardson et al., 2013 / Rajpurkar et al., 2016 / ...]
**Goal:** Extend the capability of these systems along two axes:

- Scope of the environment (*breadth*)

**Motivation**

How much is a Steak Burrito at Ray's Grill?
Goal: Extend the capability of these systems along two axes:

- Scope of the environment (*breadth*)

Motivation

How much is a Steak Burrito at Ray's Grill?

$8.50
Goal: Extend the capability of these systems along two axes:

- Scope of the environment (breadth)
- Task complexity (depth)

How much is a Steak Burrito at Ray's Grill?

$8.50
**Goal:** Extend the capability of these systems along two axes:

- Scope of the environment (**breadth**)
- Task complexity (**depth**)
Goal: Extend the capability of these systems along two axes:

- Scope of the environment (breadth)
- Task complexity (depth)

Step 1: Find burritos
Step 2: Filter for chicken
Step 3: Find the cheapest one

What's the cheapest burrito with chicken?
Motivation

**Goal:** Extend the capability of these systems along two axes:

- Scope of the environment (breadth)
- Task complexity (depth)

Step 1: Find burritos
Step 2: Filter for chicken
Step 3: Find the cheapest one

What's the cheapest burrito with chicken?

Grilled Chicken Burrito ($6.95)
Motivation

Task Complexity (depth)

Find pages with "burrito"  surface form matching
Motivation

Task Complexity (depth)

- How much is a steak burrito? (1-2 step reasoning)
- Find pages with "burrito" (surface form matching)
Motivation

Task Complexity (depth)

- What's the cheapest burrito with chicken? multi-step reasoning
- How much is a steak burrito? 1-2 step reasoning
- Find pages with "burrito" surface form matching
Motivation

Task Complexity (depth)

Scope of the environment (breadth)

- web search
- web scraping
- question answering on paragraphs
Motivation

Task Complexity (depth)

- semantic parsing on databases
- most virtual assistants
- semantic parsing on knowledge bases

Scope of the environment (breadth)

- web scraping
- web search
- question answering on paragraphs
Motivation

Task Complexity (depth)

- semantic parsing on databases
- most virtual assistants
- semantic parsing on knowledge bases

Scope of the environment (breadth)

- web pages
- multi-step reasoning
- web scraping
- question answering on paragraphs
- web search
Motivation

Task Complexity (depth)

- semantic parsing on databases
- most virtual assistants
- semantic parsing on knowledge bases
- (simple) interaction with web pages
- web scraping
- question answering on paragraphs
- web search
- web pages
- multi-step reasoning

Scope of the environment (breadth)
Interacting with Web Pages

interact with HTML elements based on the queries

follow on facebook

[2018]
Interacting with Web Pages

interact with HTML elements based on the queries

follow on facebook

[P, Allan Jiang, Evan Liu, Kelvin Guu, Percy Liang, 2018]
Interacting with Web Pages

interact with HTML elements based on the queries

follow on facebook

extracting a list of entities from the web page

hiking trails in baltimore

[P, Allan Jiang, Evan Liu, Kelvin Guu, Percy Liang, 2018]

[P. and Liang, 2014]
Interacting with Web Pages

interact with HTML elements based on the queries

extracting a list of entities from the web page

follow on facebook

hiking trails in baltimore

[P. Allan Jiang, Evan Liu, Kelvin Guu, Percy Liang, 2018]  [P. and Liang, 2014]
Interacting with Web Pages

Structures such as tables have potential for complex reasoning!
Interacting with Web Pages

Structures such as tables have potential for complex reasoning!

### List of Hedge Funds in New York

<table>
<thead>
<tr>
<th>Rank</th>
<th>Firm</th>
<th>Headquarters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bridgewater Associates</td>
<td>Westport, CT</td>
</tr>
<tr>
<td>2</td>
<td>Man Group</td>
<td>London</td>
</tr>
<tr>
<td>3</td>
<td>J.P. Morgan Asset Management</td>
<td>New York</td>
</tr>
<tr>
<td>4</td>
<td>Brevan Howard Asset Management</td>
<td>London</td>
</tr>
<tr>
<td>5</td>
<td>Och-Ziff Capital Management Group</td>
<td>New York</td>
</tr>
<tr>
<td>6</td>
<td>Paulson &amp; Co.</td>
<td>New York</td>
</tr>
<tr>
<td>7</td>
<td>BlackRock Advisors</td>
<td>New York</td>
</tr>
</tbody>
</table>
Motivation

Task Complexity (depth)

- semantic parsing on databases
- most virtual assistants
- semantic parsing on knowledge bases

Scope of the environment (breadth)

- (simple) interaction with web pages
- question answering on paragraphs
- web scraping
- web search
Motivation

Task Complexity (depth)

- semantic parsing on databases
- most virtual assistants
- semantic parsing on knowledge bases
- (simple) interaction with web pages
- answering complex questions on web tables

Scope of the environment (breadth)

- web scraping
- question answering on paragraphs
- web search
Outline

▶ Motivation
▶ Task: Answering complex questions on web tables
▶ Approach: Semantic parsing with distant supervision
▶ Improvement 1: Make it faster
▶ Improvement 2: Convert to direct supervision
## Piotr Kędzia

### Competition record

<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>2005</td>
<td>European U23 Championships</td>
<td>Izmir, Turkey</td>
<td>7th</td>
<td>4x400 m relay</td>
<td>3:04.41</td>
</tr>
<tr>
<td></td>
<td>Universiade</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>Universiade</td>
<td>Bangkok, Thailand</td>
<td>7th</td>
<td>400 m</td>
<td>46.85</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>

[Source: P. and Liang, 2015]
Piotr Kędzia

Competition record

<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>

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

[P and Liang, 2015]
WikiTableQuestions Dataset

A new dataset of Wikipedia tables, complex questions, and answers

- 22K examples
- 2K tables

[P. and Liang, 2015]
### Task Complexity (depth):

<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, 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></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?]
WikiTableQuestions Dataset

Task Complexity (depth):

<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, 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></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?
## WikiTableQuestions Dataset

### Task Complexity (depth):

<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>

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?
Scope of the environment (breadth):

- Diverse data types: 4K different column headers among 2K tables

---

### WikiTableQuestions Dataset

#### Rationale

The WikiTableQuestions Dataset is large in scope as it includes a diverse set of data types across 2K tables. This diversity is crucial for testing natural language processing models as it simulates real-world data variability and complexity. The dataset is designed to challenge models on their ability to handle a wide range of data formats and structures, ensuring robust performance in practical applications.

---

#### Table Example

<table>
<thead>
<tr>
<th>Year</th>
<th>Division</th>
<th>League</th>
<th>Regular Season</th>
<th>Playoffs</th>
<th>Open Cup</th>
<th>Avg. Attendance</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>2</td>
<td>Uefa A-League</td>
<td>4th Western</td>
<td>Quarterfinals</td>
<td>Did not qualify</td>
<td>7,169</td>
</tr>
<tr>
<td>2002</td>
<td>2</td>
<td>Uefa A-League</td>
<td>2nd Pacific</td>
<td>1st Round</td>
<td>Did not qualify</td>
<td>6,260</td>
</tr>
<tr>
<td>2003</td>
<td>2</td>
<td>Uefa A-League</td>
<td>3rd Pacific</td>
<td>Did not qualify</td>
<td>Did not qualify</td>
<td>5,871</td>
</tr>
<tr>
<td>2004</td>
<td>2</td>
<td>Uefa A-League</td>
<td>1st Western</td>
<td>Quarterfinals</td>
<td>Quarterfinals</td>
<td>9,828</td>
</tr>
<tr>
<td>2005</td>
<td>2</td>
<td>Uefa A-League</td>
<td>5th</td>
<td>Quarterfinals</td>
<td>Quarterfinals</td>
<td>8,028</td>
</tr>
<tr>
<td>2006</td>
<td>2</td>
<td>Uefa A-League</td>
<td>11th</td>
<td>Did not qualify</td>
<td>3rd Round</td>
<td>5,575</td>
</tr>
<tr>
<td>2007</td>
<td>2</td>
<td>Uefa A-League</td>
<td>2nd</td>
<td>Semi-finals</td>
<td>2nd Round</td>
<td>6,851</td>
</tr>
<tr>
<td>2008</td>
<td>2</td>
<td>Uefa A-League</td>
<td>11th</td>
<td>Did not qualify</td>
<td>1st Round</td>
<td>8,987</td>
</tr>
<tr>
<td>2009</td>
<td>2</td>
<td>Uefa A-League</td>
<td>1st</td>
<td>Semi-finals</td>
<td>3rd Round</td>
<td>9,734</td>
</tr>
<tr>
<td>2010</td>
<td>2</td>
<td>Uefa A-League</td>
<td>3rd, Uefa (3rd)</td>
<td>Quarterfinals</td>
<td>3rd Round</td>
<td>10,727</td>
</tr>
</tbody>
</table>

---

**Notes:**

- The dataset includes a variety of sports, from football (soccer) to baseball, reflecting the breadth of its scope.
- The table illustrates the diversity in data types, with different years, leagues, and playoff systems.
- The average attendance figures provide insight into the popularity and engagement levels across different matches.

---

**References:**

- [P. and Liang, 2015]
WikiTableQuestions Dataset

Scope of the environment (breadth):

- Diverse data types: 4K different column headers among 2K tables
- Only ~20% can be answered by Freebase (a large knowledge base)
WikiTableQuestions Dataset

Scope of the environment (breadth):

- Diverse data types: 4K different column headers among 2K tables
- Only ~20% can be answered by Freebase (a large knowledge base)

We ensure that tables in the test set do not appear in the training data
- This tests the model's ability to generalize to unseen data schema

[P. and Liang, 2015]
Outline

- Motivation
- Task: Answering complex questions on web tables
  - **Approach:** Semantic parsing with distant supervision
  - Improvement 1: Make it faster
  - Improvement 2: Convert to direct supervision
**Semantic Parsing**

**Idea:** Parse the utterance into an executable *logical form*

Where did the last 1st place finish occur?

VenueOf.argmax(HasPosition.1st, Index)

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2005</td>
<td>Germany</td>
<td>5th</td>
<td>46.62</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>1st</td>
<td>53.13</td>
</tr>
</tbody>
</table>

[Zelle + Mooney, 1996 / Zettlemoyer + Collins, 2007 / Liang et al., 2011 / Berant et al., 2013 / ...]
Semantic Parsing

Idea: Parse the utterance into an executable logical form

Where did the last 1st place finish occur?

VenueOf.argmax(HasPosition.1st, Index)

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2005</td>
<td>Germany</td>
<td>5th</td>
<td>46.62</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>1st</td>
<td>53.13</td>
</tr>
</tbody>
</table>

[Zelle + Mooney, 1996 / Zettlemoyer + Collins, 2007 / Liang et al., 2011 / Berant et al., 2013 / ...]
Semantic Parsing

Idea: Parse the utterance into an executable logical form

Where did the last 1st place finish occur?

VenueOf.argmax(HasPosition.1st, Index)

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2005</td>
<td>Germany</td>
<td>5th</td>
<td>46.62</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>1st</td>
<td>53.13</td>
</tr>
</tbody>
</table>
Semantic Parsing

Idea: Parse the utterance into an executable logical form

Where did the last 1st place finish occur?

VenueOf.argmax(HasPosition.1st, Index)

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2005</td>
<td>Germany</td>
<td>5th</td>
<td>46.62</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>1st</td>
<td>53.13</td>
</tr>
</tbody>
</table>

[Zelle + Mooney, 1996 / Zettlemoyer + Collins, 2007 / Liang et al., 2011 / Berant et al., 2013 / ...]
Semantic Parsing

**Idea:** Parse the utterance into an executable **logical form**

Where did the last 1st place finish occur?

VenueOf.argmax(HasPosition.1st, Index)

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2005</td>
<td>Germany</td>
<td>5th</td>
<td>46.62</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>1st</td>
<td>53.13</td>
</tr>
</tbody>
</table>

[Zelle + Mooney, 1996 / Zettlemoyer + Collins, 2007 / Liang et al., 2011 / Berant et al., 2013 / ...]
Semantic Parsing

**Idea:** Parse the utterance into an executable logical form

Where did the last 1st place finish occur?

VenueOf.argmax(HasPosition.1st, Index)

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2005</td>
<td>Germany</td>
<td>5th</td>
<td>46.62</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>1st</td>
<td>53.13</td>
</tr>
</tbody>
</table>

"denotation"

[Zelle + Mooney, 1996 / Zettlemoyer + Collins, 2007 / Liang et al., 2011 / Berant et al., 2013 / ...]
Prediction

Where ... ?

[Zelle + Mooney, 1996 / Zettlemoyer + Collins, 2007 / Liang et al., 2011 / Berant et al., 2013 / ...]
Prediction

Where ... ?

Parser

logical forms

VenueOf ...
YearOf ...
count(…)

[Zelle + Mooney, 1996 / Zettlemoyer + Collins, 2007 / Liang et al., 2011 / Berant et al., 2013 / …]
Prediction

Parser

Where ... ?

logical forms:
- VenueOf ...
- YearOf ...
- count(...)

denotations:
- Thailand
- 2007
- 2

[Zelle + Mooney, 1996 / Zettlemoyer + Collins, 2007 / Liang et al., 2011 / Berant et al., 2013 / ...]
Prediction

Where ...

Parser

Logical forms:
- VenueOf ...
- YearOf ...
- count(...)

Denotations:
- Thailand
- 2007
- 2

Scorer

Parameter $\theta$

[Zelle + Mooney, 1996 / Zettlemoyer + Collins, 2007 / Liang et al., 2011 / Berant et al., 2013 / ...]
Where ...?

Prediction

Parser

logical forms

VenueOf ...

YearOf ...

count(...)

denotations

Thailand

2007

2

scores

3.5

4.3

1.1

Scorer
parameter θ

[Zelle + Mooney, 1996 / Zettlemoyer + Collins, 2007 / Liang et al., 2011 / Berant et al., 2013 / ...]
Learning from Denotations

Where ... ?

Parser

VenueOf ... Thailand 3.5
YearOf ... 2007 4.3
count(...) 2 1.1

Scorer
parameter $\theta$

During training:
correct answer
Thailand

[Clark et al., 2010 / Liang et al., 2011 / Berant et al., 2013 / ...]
During training:

correct answer
Thailand

Learner

Scorer
parameter \( \theta \)

VenueOf ...
Thailand 3.5
YearOf ...
2007 4.3
count(…)
2 1.1

During training:

[Clark et al., 2010 / Liang et al., 2011 / Berant et al., 2013 / ...]
Learning from Denotations

During training:

- **Parser**
  - **VenueOf**...
  - **YearOf**...
  - **count(...)**
  - Thailand
  - 2007
  - 2
  - 1.1

- **Scorer**
  - parameter $\theta$

- **Learner**

**Where...?**

Correct answer

**Thailand**

[Clark et al., 2010 / Liang et al., 2011 / Berant et al., 2013 / ...]
Where ... ?

Floating Parser

VenueOf ...  
Thailand  3.5
YearOf ...  
2007  4.3
count(...)  
2  1.1

Scorer
parameter $\theta$

During training:
correct answer

Thailand

Learner

[P. and Liang, 2015]
Parser

The parser **generates logical forms** from the utterance and table.

VenueOf.argmax(HasPosition.1st, Index)

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2005</td>
<td>Germany</td>
<td>5th</td>
<td>46.62</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>1st</td>
<td>53.13</td>
</tr>
</tbody>
</table>
Where did the last 1st place finish occur?

The parser generates logical forms from the utterance and table.

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2005</td>
<td>Germany</td>
<td>5th</td>
<td>46.62</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>1st</td>
<td>53.13</td>
</tr>
</tbody>
</table>
Where did the last 1st place finish occur?

The parser generates logical forms from the utterance and table

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2005</td>
<td>Germany</td>
<td>5th</td>
<td>46.62</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>1st</td>
<td>53.13</td>
</tr>
</tbody>
</table>

[P. and Liang, 2015]
Where did the last 1st place finish occur?

The parser generates logical forms from the utterance and table

Parser

VenueOf.argmax(HasPosition.1st, Index)

Cells

a + b → Hasb.a

1st

Position

Year | Venue | Position | Time
--- | --- | --- | ---
2003 | Finland | 1st | 47.12
2005 | Germany | 5th | 46.62
2007 | Thailand | 1st | 53.13

[P. and Liang, 2015]
Where did the last 1st place finish occur?

The parser **generates logical forms** from the utterance and table.
Floating Parser

Allows LF parts to be generated from other sources than the utterance

VenueOf.argmax(HasPosition.1st, Index)

Where did the last 1st place finish occur?

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2005</td>
<td>Germany</td>
<td>5th</td>
<td>46.62</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>1st</td>
<td>53.13</td>
</tr>
</tbody>
</table>
Floating Parser

Allows LF parts to be generated from other sources than the utterance. The scorer is responsible for capturing the relationship between "Venue" and words in the utterance.

Where did the last 1st place finish occur?

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2005</td>
<td>Germany</td>
<td>5th</td>
<td>46.62</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>1st</td>
<td>53.13</td>
</tr>
</tbody>
</table>
Most neural decoders today generate tokens in a floating fashion (with a soft guidance from attention)
## Results

<table>
<thead>
<tr>
<th>Rule Set</th>
<th>Test Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiny rule set (cell lookup + counting)</td>
<td>24.3</td>
</tr>
<tr>
<td>Small rule set (sum, argmax, next/prev row, subtraction, etc.)</td>
<td>37.1</td>
</tr>
<tr>
<td>Large rule set (fuzzy string matching, advanced argmax, etc.)</td>
<td>42.7</td>
</tr>
</tbody>
</table>

[P. and Liang, 2015]
## Results

<table>
<thead>
<tr>
<th>Rule Set Description</th>
<th>Test Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiny rule set (cell lookup + counting)</td>
<td>24.3</td>
</tr>
<tr>
<td>Small rule set (sum, argmax, next/prev row, subtraction, etc.)</td>
<td>37.1</td>
</tr>
<tr>
<td>Large rule set (fuzzy string matching, advanced argmax, etc.)</td>
<td>42.7</td>
</tr>
<tr>
<td>Upper bound (Ice writes a logical form)</td>
<td>84.0</td>
</tr>
</tbody>
</table>

[P. and Liang, 2015]
## Results

<table>
<thead>
<tr>
<th>Rule Set Description</th>
<th>Test Accuracy</th>
<th>Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiny rule set (cell lookup + counting)</td>
<td>24.3</td>
<td></td>
</tr>
<tr>
<td>Small rule set (sum, argmax, next/prev row, subtraction, etc.)</td>
<td>37.1</td>
<td>~ 4 hours</td>
</tr>
<tr>
<td>Large rule set (fuzzy string matching, advanced argmax, etc.)</td>
<td>42.7</td>
<td>~ 11 hours</td>
</tr>
<tr>
<td>Upper bound (Ice writes a logical form)</td>
<td>(84.0)</td>
<td></td>
</tr>
</tbody>
</table>

Slow runtime prevents us from increasing coverage!

[P. and Liang, 2015]
Outline

- Motivation
- Task: Answering complex questions on web tables
- Approach: Semantic parsing with distant supervision
  - Improvement 1: Make it faster
  - Improvement 2: Convert to direct supervision
Using macros to make search faster

For each training example:

Who took office right after Uriah Forrest?
Using macros to make search faster

For each training example:

- Fetch previously processed examples with similar utterances (Levenshtein distance)

Who took office right after Uriah Forrest?

Who ranked right after Turkey?

NationOf.NextOf.HasNation.Turkey

[Yuchen Zhang, P. and Liang, 2017]
Using macros to make search faster

For each training example:

- Fetch previously processed examples with *similar utterances* (Levenshtein distance)

Who took office right after Uriah Forrest?

macro


[Yuchen Zhang, P. and Liang, 2017]
Using macros to make search faster

For each training example:

- Fetch previously processed examples with similar utterances
- During search, try creating logical forms from macros found in those examples first

Who took office right after Uriah Forrest?

macro

<table>
<thead>
<tr>
<th>Column</th>
<th>Of.NextOf.Has</th>
<th>Column</th>
<th>Cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>NameOf.NextOf.HasName.UriahForrest</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Yuchen Zhang, P. and Liang, 2017]
Using macros to make search faster

<table>
<thead>
<tr>
<th></th>
<th>Dev Accuracy</th>
<th>Time (ms/example)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Train</td>
<td>Predict</td>
<td></td>
</tr>
<tr>
<td>Large rule set</td>
<td>40.6</td>
<td>1117</td>
<td>1150</td>
<td></td>
</tr>
<tr>
<td>+ Macros</td>
<td>40.4</td>
<td>99</td>
<td>70</td>
<td></td>
</tr>
</tbody>
</table>

11× speedup!  16× speedup!

[Yuchen Zhang, P. and Liang, 2017]
Using macros to make search faster

<table>
<thead>
<tr>
<th>Small rule set</th>
<th>37.1</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural Programmer</td>
<td>34.2</td>
<td>37.7</td>
</tr>
<tr>
<td>(Neelakantan et al., 2016)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neural Multi-Step Reasoning</td>
<td>34.8</td>
<td>38.7</td>
</tr>
<tr>
<td>(Haug et al., 2017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large rule set</td>
<td>42.7</td>
<td>-</td>
</tr>
<tr>
<td>+ Macros</td>
<td>43.7</td>
<td>-</td>
</tr>
<tr>
<td>Upper bound</td>
<td>(84.0)</td>
<td>-</td>
</tr>
<tr>
<td>(Ice writes a logical form)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Yuchen Zhang, P. and Liang, 2017]
Motivation

Task: Answering complex questions on web tables

Approach: Semantic parsing with distant supervision

Improvement 1: Make it faster

Improvement 2: Convert to direct supervision
Learning from Denotations

During training:

correct answer

Thailand
Learning from Denotations

Parser

VenueOf ... Thailand 3.5
YearOf ... 2007 4.3
count(...) 2 1.1

Scorer
parameter $\theta$

During training:
correct answer
Thailand
Learning from Denotations

During training:
- correct answer: Thailand

Parser
- VenueOf ...: Thailand
- YearOf ...: 2007
- count(...): 2
- count(...): 1

Scorer
- parameter \( \theta \)

Learner

Where ...?
Learning from Denotations

Pros:

- Collecting answers is easier than collecting LFs for each question
- The dataset is not tied to a specific LF formalism
Learning from Denotations

Pros:
- Collecting answers is easier than collecting LFs for each question
- The dataset is not tied to a specific LF formalism

Cons:
- Need to do search (both at training and test time)
  - Slow (macros helped a bit)
  - Two more problems during training: ...
Search is Hard

Where ... ?

Parser

VenueOf ...
Thailand 3.5

YearOf ...
2007 4.3

count(...)
2 1.1

Scorer
parameter $\theta$

Parser
Search is Hard

Parser uses **beam search** to make search tractible

- For each parsing state, only keep up to $B = 100$ highest-scoring partial LFs and **discard the rest**
Search is Hard

Where ...?

Parser

VenueOf ...  
Thailand  
2007  
4.3

count(...)

YearOf ...  
2007  
1.1

Scorer

parameter \( \theta \)
Search is Hard

**Problem 1:**

If the parser cannot find a **consistent** LF (consistent: denotation matches the correct answer) → cannot learn
Search is Hard

Some LFs are spurious (right for wrong reasons)

Where did the last 1st place finish occur?

VenueOf.argmax(HasPosition.1st, Index)

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2005</td>
<td>Germany</td>
<td>5th</td>
<td>46.62</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>1st</td>
<td>53.13</td>
</tr>
</tbody>
</table>
Search is Hard

Some LFs are spurious (right for wrong reasons)

Where did the last 1st place finish occur?

VenueOf.argmax(HasPosition.1st, Time)

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2005</td>
<td>Germany</td>
<td>5th</td>
<td>46.62</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>1st</td>
<td>53.13</td>
</tr>
</tbody>
</table>
Search is Hard

Problem 2:
Parser finds spurious LFs → bad parameter updates
Learning from Denotations

Pros:

- Collecting answers is easier than collecting LFs for each question
- The dataset is not tied to a specific LF formalism

Cons:

- Need to do search
  - Slow
  - Cannot find consistent LFs → cannot learn
  - Find spurious LFs → bad updates
Offline search

Let's avoid search during training altogether!
Let's avoid search during training altogether!

- For each training example (utterance, table, answer), augment it with LFs

- Use (utterance, table, LFs) for supervised training

[P. and Liang, 2016]
Offline search

Let's avoid search during training altogether!

- For each training example (utterance, table, answer), augment it with LFs
  - 
  - 
- Use (utterance, table, LFs) for supervised training
Let's avoid search during training altogether!

- For each training example (utterance, table, answer), augment it with LFs
  - **Enumerate** all LFs consistent with the correct answer
  - ...

- Use (utterance, table, LFs) for supervised training
Let's avoid search during training altogether!

- For each training example (utterance, table, answer), augment it with LFs
  - Enumerate all LFs consistent with the correct answer
- Use (utterance, table, LFs) for supervised training
Offline search

Let's avoid search during training altogether!

- For each training example (utterance, table, answer), augment it with **semantically correct** LFs
  - **Enumerate** all LFs consistent with the correct answer
  - **Filter out** spurious LFs
- Use (utterance, table, LFs) for supervised training

[P. and Liang, 2016]
Let's avoid search during training altogether!

- For each training example (utterance, table, answer), augment it with semantically correct LFs
  - **Enumerate** all LFs consistent with the correct answer
  - **Filter out** spurious LFs
- Use (utterance, table, LFs) for supervised training
Enumerating consistent LFs

Beam search controls the search space, but can discard crucial LF parts → low coverage!

If we have access to the correct answer, is there a better way to control the search space?

[P. and Liang, 2016]
## Dynamic Programming on Denotations

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2005</td>
<td>Germany</td>
<td>5th</td>
<td>46.62</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>1st</td>
<td>53.13</td>
</tr>
</tbody>
</table>
Dynamic Programming on Denotations

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2005</td>
<td>Germany</td>
<td>5th</td>
<td>46.62</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>1st</td>
<td>53.13</td>
</tr>
</tbody>
</table>

[P. and Liang, 2016]
Dynamic Programming on Denotations

TimeOf. HasIndex.2
TimeOf. HasPosition.5th
TimeOf. HasVenue.Germany

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2005</td>
<td>Germany</td>
<td>5th</td>
<td>46.62</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>1st</td>
<td>53.13</td>
</tr>
</tbody>
</table>

[P. and Liang, 2016]
Dynamic Programming on Denotations

Group LFIs with the same denotation together during search

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2005</td>
<td>Germany</td>
<td>5th</td>
<td>46.62</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>1st</td>
<td>53.13</td>
</tr>
</tbody>
</table>

[P. and Liang, 2016]
Phase 1: Group LF\(s\) with the same denotation together during search

\[\{r_2\}\]

[Dynamic Programming on Denotations](#)

**Phase 1:** Group LF\(s\) with the same denotation together during search

TimeOf. \(\{r_2\}\)

\[46.62\]

[Time]

[Phase 1: Group LF\(s\) with the same denotation together during search](#)
Dynamic Programming on Denotations

**Phase 1:** Group LFs with the same denotation together during search
Dynamic Programming on Denotations

Phase 1: Group LFs with the same denotation together during search

- Remove paths that do not give the correct answer

[Diagram showing Thailand and Germany with numbers indicating 153,000 → 2,000 cells (99% reduction!).]

[P. and Liang, 2016]
Dynamic Programming on Denotations

Phase 2: Search over LFs, but only on the paths found in Phase 1
Results

Testing coverage:

- Annotate each example with a semantically correct LF
- Test whether the algorithm can generate the annotated LF

Note: Maximum success rate = 84.0% (% of examples with annotated LFs)

[P. and Liang, 2016]
Results

Testing coverage:

- Annotate each example with a semantically correct LF
- Test whether the algorithm can generate the annotated LF

Success rate: 53.7%

Note: Maximum success rate = 84.0% (% of examples with annotated LFs)

[P. and Liang, 2016]
Results

Testing coverage:

▸ Annotate each example with a semantically correct LF
▸ Test whether the algorithm can generate the annotated LF

Success rate: 53.7%

Note: Maximum success rate = 84.0% (% of examples with annotated LFs)

[P. and Liang, 2016]
Results

Testing coverage:

- Annotate each example with a semantically correct LF
- Test whether the algorithm can generate the annotated LF

Success rate: 53.7%
Success rate: 76.0%

Note: Maximum success rate = 84.0% (% of examples with annotated LFs)

[P. and Liang, 2016]
Let's avoid search during training altogether!

- For each training example (utterance, table, answer), augment it with semantically correct LFs
  - Enumerate all LFs consistent with the correct answer
  - Filter out spurious LFs
- Use (utterance, table, LFs) for supervised training

[Offline search](P. and Liang, 2016)
Let's avoid search during training altogether!

- For each training example (utterance, table, answer), augment it with semantically correct LFs
  - Enumerate all LFs consistent with the correct answer
  - Filter out spurious LFs
- Use (utterance, table, LFs) for supervised training

[Online search: all logical forms]

- semantically correct
- spurious (right for wrong reasons)

[P. and Liang, 2016]
Offline search

Let's avoid search during training altogether!

- For each training example (utterance, table, answer), augment it with semantically correct LFs
  - Enumerate all LFs consistent with the correct answer
  - **Filter out** spurious LFs
- Use (utterance, table, LFs) for supervised training

[P. and Liang, 2016]
Filtering out Spurious LFs

Where did the last 1st place finish occur?

Human: Thailand
Correct LF: Thailand
Spurious LF: Thailand

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2005</td>
<td>Germany</td>
<td>5th</td>
<td>46.62</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>1st</td>
<td>53.13</td>
</tr>
</tbody>
</table>

[P. and Liang, 2016]
Filtering out Spurious LFs

Where did the last 1st place finish occur?

- **Human**: Finland
- **Correct LF**: Finland
- **Spurious LF**: Germany

### Table

<table>
<thead>
<tr>
<th>Year</th>
<th>Venue</th>
<th>Position</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Germany</td>
<td>1st</td>
<td>53.13</td>
</tr>
<tr>
<td>2005</td>
<td>Finland</td>
<td>1st</td>
<td>47.12</td>
</tr>
<tr>
<td>2007</td>
<td>Thailand</td>
<td>5th</td>
<td>46.62</td>
</tr>
</tbody>
</table>

Randomize cells in each column.

[P. and Liang, 2016]
Fictitious Tables

Generate *fictitious tables* and execute the logical forms on them

Original Table

<table>
<thead>
<tr>
<th>LF 1</th>
<th>Thailand</th>
</tr>
</thead>
<tbody>
<tr>
<td>LF 2</td>
<td>Thailand</td>
</tr>
<tr>
<td>LF 3</td>
<td>Thailand</td>
</tr>
<tr>
<td>LF 4</td>
<td>Thailand</td>
</tr>
<tr>
<td>LF 5</td>
<td>Thailand</td>
</tr>
<tr>
<td>Human</td>
<td>Thailand</td>
</tr>
</tbody>
</table>

[P. and Liang, 2016]
## Fictitious Tables

Generate **fictitious tables** and execute the logical forms on them

<table>
<thead>
<tr>
<th>Original Table</th>
<th>Fictitious Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>LF 1</td>
<td>Thailand</td>
</tr>
<tr>
<td></td>
<td>Finland</td>
</tr>
<tr>
<td>LF 2</td>
<td>Thailand</td>
</tr>
<tr>
<td></td>
<td>Thailand</td>
</tr>
<tr>
<td>LF 3</td>
<td>Thailand</td>
</tr>
<tr>
<td></td>
<td>Finland</td>
</tr>
<tr>
<td>LF 4</td>
<td>Thailand</td>
</tr>
<tr>
<td></td>
<td>Finland</td>
</tr>
<tr>
<td>LF 5</td>
<td>Thailand</td>
</tr>
<tr>
<td></td>
<td>Germany</td>
</tr>
<tr>
<td>Human</td>
<td>Thailand</td>
</tr>
</tbody>
</table>
|                | Finland             

[P. and Liang, 2016]
Fictitious Tables

Generate **fictitious tables** and execute the logical forms on them

<table>
<thead>
<tr>
<th>Original Table</th>
<th>Fictitious Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>LF 1</td>
<td>Thailand</td>
</tr>
<tr>
<td>LF 2</td>
<td>Thailand</td>
</tr>
<tr>
<td>LF 3</td>
<td>Thailand</td>
</tr>
<tr>
<td>LF 4</td>
<td>Thailand</td>
</tr>
<tr>
<td>LF 5</td>
<td>Thailand</td>
</tr>
<tr>
<td>Human</td>
<td>Thailand</td>
</tr>
</tbody>
</table>

[P. and Liang, 2016]
Generate **fictitious tables** and execute the logical forms on them.

We also propose a way to select the **most informative** fictitious tables.

---

<table>
<thead>
<tr>
<th>Original Table</th>
<th>Fictitious Table 1</th>
<th>Fictitious Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>LF 1</td>
<td>Thailand</td>
<td>Finland</td>
</tr>
<tr>
<td>LF 2</td>
<td>Thailand</td>
<td>Thailand</td>
</tr>
<tr>
<td>LF 3</td>
<td>Thailand</td>
<td>Finland</td>
</tr>
<tr>
<td>LF 4</td>
<td>Thailand</td>
<td>Finland</td>
</tr>
<tr>
<td>Human</td>
<td>Thailand</td>
<td>Germany</td>
</tr>
</tbody>
</table>

[122]

[P. and Liang, 2016]
Fictitious Tables

Results:

- Accidentally pruned correct LFs in 20% of the examples
  - because randomizing cells can create nonsensical tables
- But for the remaining examples, pruned out 92.1% of spurious LFs

[P. and Liang, 2016]
Fictitious Tables

Relax assumption: Only filter LFs disagreeing with humans $> 1$ once

- Accidentally pruned correct LFs in 20% of the examples
  - because randomizing cells can create nonsensical tables
- But for the remaining examples, pruned out 92.1% of spurious LFs
  - 78%

[P. and Liang, 2016]
Offline search

Let's avoid search during training altogether!

- For each training example (utterance, table, answer), augment it with **semantically correct** LFs
  - **Enumerate** all LFs consistent with the correct answer
  - **Filter out** spurious LFs
- Use (utterance, table, LFs) for supervised training

[125] all logical forms

[P. and Liang, 2016]
Offline search

Let's avoid search during training altogether!

▸ For each training example (utterance, table, answer), augment it with semantically correct LFs
  ▹ Enumerate all LFs consistent with the correct answer
  ▹ Filter out spurious LFs

▸ Use (utterance, table, LFs) for supervised training

[P. and Liang, 2016]
Using macros to make search faster

<table>
<thead>
<tr>
<th></th>
<th>Test Acc.</th>
<th>+Ensemble</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small rule set</td>
<td>37.1</td>
<td>-</td>
</tr>
<tr>
<td>Neural Programmer</td>
<td>34.2</td>
<td>37.7</td>
</tr>
<tr>
<td>(Neelakantan et al., 2016)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neural Multi-Step Reasoning</td>
<td>34.8</td>
<td>38.7</td>
</tr>
<tr>
<td>(Haug et al., 2017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large rule set</td>
<td>42.7</td>
<td>-</td>
</tr>
<tr>
<td>+ Macros</td>
<td>43.7</td>
<td>-</td>
</tr>
</tbody>
</table>

Krishnamurthy et al., 2017 uses 100 shortest LFs we generated to train a neural parser (top-down tree generation)
Using macros to make search faster

<table>
<thead>
<tr>
<th></th>
<th>Test Acc.</th>
<th>+Ensemble</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small rule set</td>
<td>37.1</td>
<td>-</td>
</tr>
<tr>
<td>Neural Programmer (Neelakantan et al., 2016)</td>
<td>34.2</td>
<td>37.7</td>
</tr>
<tr>
<td>Neural Multi-Step Reasoning (Haug et al., 2017)</td>
<td>34.8</td>
<td>38.7</td>
</tr>
<tr>
<td>Large rule set</td>
<td>42.7</td>
<td>-</td>
</tr>
<tr>
<td>+ Macros</td>
<td>43.7</td>
<td>-</td>
</tr>
<tr>
<td>Neural Parser trained on consistent LFs (DPD; no filtering)</td>
<td>36.3</td>
<td></td>
</tr>
</tbody>
</table>

Krishnamurthy et al., 2017 uses 100 shortest LFs we generated to train a neural parser (top-down tree generation)
### Using macros to make search faster

<table>
<thead>
<tr>
<th>Method</th>
<th>Test Acc.</th>
<th>+Ensemble</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small rule set</td>
<td>37.1</td>
<td>-</td>
</tr>
<tr>
<td>Neural Programmer</td>
<td>34.2</td>
<td>37.7</td>
</tr>
<tr>
<td>(Neelakantan et al., 2016)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neural Multi-Step Reasoning</td>
<td>34.8</td>
<td>38.7</td>
</tr>
<tr>
<td>(Haug et al., 2017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large rule set</td>
<td>42.7</td>
<td>-</td>
</tr>
<tr>
<td>+ Macros</td>
<td>43.7</td>
<td>-</td>
</tr>
<tr>
<td>Neural Parser trained on <strong>consistent</strong> LFs</td>
<td>36.3</td>
<td></td>
</tr>
<tr>
<td>(DPD; no filtering)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neural Parser trained on <strong>correct</strong> LFs</td>
<td>43.3</td>
<td>45.9</td>
</tr>
<tr>
<td>(DPD + filtering)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Krishnamurthy et al., 2017* uses 100 shortest LFs we generated to train a neural parser (top-down tree generation)
Summary

Task Complexity (depth)

- database / apps
- knowledge base
- web pages
- any texts

Scope of the environment (breadth)

The Lacey Act of 1900 was the first federal law that regulated commercial animal markets. It prohibited interstate commerce of animals killed in violation of state game laws, and covered all fish and wildlife and their parts or products, as well as plants. Other legislation followed, including the Migratory Bird Conservation Act of 1929, a 1937 treaty prohibiting the hunting of right and gray whales, and the Bald Eagle Protection Act of 1940. These later laws had a low cost to society—the species were relatively rare—and little opposition was raised.\(^1\)
Summary

Task Complexity (depth)

What's the cheapest burrito with chicken? multi-step reasoning

How much is a steak burrito? 1-2 step reasoning

Find pages with "burrito" surface form matching

Scope of the environment (breadth)
Summary

Task Complexity (depth)

- semantic parsing on databases
- most virtual assistants
- semantic parsing on knowledge bases
- (simple) interaction with web pages
- answering complex questions on web tables

Scope of the environment (breadth)

- web search
- web scraping
- question answering on paragraphs
Future Directions

Task Complexity (depth)

Scope of the environment (breadth)

answering complex questions on web tables
Future Directions

Task Complexity (depth)

- Learn implicit columns
  ("What is ....?" → which column?)
- Detect if the question cannot be answered by the table
- More complex reasoning
  (e.g., "consecutive")
- Better table understanding
  (╯°□°）╯︵ ┻━┻
- ...

Scope of the environment (breadth)

[WikiSQL (Zhong et al., 2017) / SQUAD 2.0 (Rajpurkar et al., 2018) / Dong and Lapata, 2018 / ...]
Future Direct

Task Complexity (depth)

- multi-step web interaction
- answering complex questions on web tables

Scope of the environment (breadth)

[PLow (Allen et al., 2007) / Branavan et al., 2010 / World of Bits (Shi et al., 2017 / Liu et al., 2018 / Gur et al., 2019) / ...]
Future Directions

Task Complexity (depth)

- multi-step web interaction
- combine information from multiple sources
- answering complex questions on web tables

Scope of the environment (breadth)

- gas station near Jack's office
- which Jack?
- office?
- gas station?

[QALD shared tasks / Spider (Yu et al., 2018) / HotpotQA (Yang et al., 2018) / DROP (Dua et al., 2019) / ...]
Thank you!