Question Answering on Web Data

Silei Xu
CS294S April 9, 2020

Joint work with Giovanni Campagna, Sina Semnani, Jian Li, and Monica S. Lam
Commercial Assistants

Alexa
User hand-codes question/code 1 by 1

*get me an upscale restaurants*

*What are the restaurants around here?*

*What is the best restaurant?*

*search for Chinese restaurants*
Commercial Assistants

Alexa
User hand-codes question/code 1 by 1

get me an upscale restaurants
What are the restaurants around here?
What is the best restaurant?
search for Chinese restaurants

100K Alexa skills
Sep 2019
Commercial Assistants

Alexa
User hand-codes question/code 1 by 1

*get me an upscale restaurants*

*What are the restaurants around here?*

*What is the best restaurant?*

*search for Chinese restaurants*

100K Alexa skills
Sep 2019

1.8 billion websites
Genie: Synthesize Question/Code from a Schema

User

Schema

<table>
<thead>
<tr>
<th>Name</th>
<th>Price</th>
<th>Cuisine</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Property Annotations

Genie

500 Domain-Independent Templates

What is the <prop> of <subject>?
What is the <subject>'s <prop>?
Genie: Synthesize Question/Code from a Schema

User

Schema

| Name | Price | Cuisine | ...
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Property Annotations

Genie

500 Domain-Independent Templates

What is the <prop> of <subject>?
What is the <subject>’s <prop>?

get me an upscale restaurants
What are the restaurants around here?
What is the best restaurant?
search for Chinese restaurants
What is the best restaurant within 10 miles?
Find restaurants that serve Chinese or Japanese food
What is the best non-Chinese restaurant near here?
Show me a cheap restaurant with 5-star review.
Are there any restaurant with at least 4.5 stars?
What is the phone number of Wendy’s?
I’m looking for an Italian fine dining restaurant.
Give me the best Italian restaurant.
Find me the best restaurant with 500 or more reviews
Show me some restaurant with less than 10 reviews

Stanford University
The Web Has a Schema!
The Web Has a Schema!

• Schema.org
  • Structure data to mark up web pages
  • Mainly used by search engines
  • It covers many domains, including restaurants, hotels, people, recipes, products, news ...
The Web Has a Schema!

- Schema.org
- Structure data to mark up web pages
- Mainly used by search engines
- It covers many domains, including restaurants, hotels, people, recipes, products, news ...

```json
<script type="application/ld+json">
{
  @type: "restaurant",
  name: "The French Laundry",
  servesCuisine: "French",
  aggregateRating: {
    @type: "AggregateRating",
    reviewCount: 2527,
    ratingValue: 4.5
  }
}
...

Schema.org markup on Yelp
```
The Web Has a Schema!

• Schema.org
  • Structure data to mark up web pages
  • Mainly used by search engines
  • It covers many domains, including restaurants, hotels, people, recipes, products, news ...

40% of the websites use it!

```json
<script type="application/ld+json">
{
  @type: "restaurant",
  name: "The French Laundry",
  servesCuisine: "French",
  aggregateRating: {
    @type: "AggregateRating",
    reviewCount: 2527,
    ratingValue: 4.5
  }

  ...
}

Schema.org markup on Yelp
```
Outline

• Introduction to Schema.org

• Represent Questions in ThingTalk

• LUINet: NL to ThingTalk
  • Training data generation
  • Experimental results

• Work in progress: automate everything!
Introduction to Schema.org
Graph Data Model of Schema.org
Graph Data Model of Schema.org

Organization
legalName: Text
slogan: Text
aggregateRating: AggregateRating
...
Graph Data Model of Schema.org

class

Organization
legalName: Text
slogan: Text
aggregateRating: AggregateRating
...
Graph Data Model of Schema.org

class

properties

Organization
  legalName: Text
  slogan: Text
  aggregateRating: AggregateRating
  ...

Stanford University
Graph Data Model of Schema.org

- **class**
- **properties**
  - `Organization`
    - `legalName`: Text
    - `slogan`: Text
    - `aggregateRating`: `AggregateRating`
    - ...
- **types – primitive or class**
Graph Data Model of Schema.org

Organization
legalName: Text
slogan: Text
aggregateRating: AggregateRating
...

AggregateRating
ratingCount: Integer
ratingValue: Integer
...

Stanford University
Schema.org Hierarchy

**Thing**
- name: Text
- url: URL
...

**Organization (Thing)**
- legalName: Text
- slogan: Text
- aggregateRating: AggregateRating
...

**AggregateRating**
- ratingCount: Integer
- ratingValue: Integer
...

Stanford University
Some useful tools

- Google Structured Data Testing Tool
  - Show schema.org markups in a web page

- Google Custom Search
  - Search for pages that contain certain schema.org domains
ThingTalk for Questions
ThingTalk for QA

\text{table\ }[.,\text{filter}]^2
Show me restaurants in Stanford

```
@QA.restaurant(), geo == makeLocation("Stanford")
now => table [, filter]
=> notify
```
Show me Chinese restaurants in Stanford

```plaintext
@QA.restaurant(), geo == makeLocation("Stanford")
&& servesCuisine =~ "Chinese"

now => => notify
```

Table \[, \text{filter}\]?
Show me Chinese restaurants in Stanford

@QA.restaurant(), geo == makeLocation("Stanford")
&& servesCuisine =~ "Chinese"

=> notify

sort fn asc|desc of table [, filter]?
Show me top-rated Chinese restaurants in Stanford

```
now =>

sort fn asc|desc of table [, filter]?

sort aggregateRating.ratingValue desc of ( QA.restaurant(), geo == makeLocation("Stanford") && servesCuisine =~ "Chinese" ) => notify
```
Show me top-rated Chinese restaurants in Stanford

Now =>

sort aggregateRating.ratingValue desc of ( @QA.restaurant(), geo == makeLocation("Stanford") && servesCuisine ~= "Chinese" ) => notify
Show me top-rated Chinese restaurants in Stanford reviewed by Bob
ThingTalk for QA

sort \( fn \) asc|desc of \( table [, filter] \)? \( [join table [, filter]?] \)^*  

\[ fn^+ \text{ of} \]? \( table [, filter] \)?
ThingTalk for QA

\[
\text{sort } fn \text{ asc|desc of } table [, filter]? \left[ \text{join } table [, filter]? \right]^* \\
\left[ fn^+ \text{ of} \right]? table [, filter]?
\]

\[
\text{aggregate min|max|sum|avg|count } fn \text{ of } table [, filter]?
\]
ThingTalk for QA

\[
\text{sort } fn \text{ asc|desc of } table [,.filter]? \left[ \text{join } table [,.filter]? \right]^* \\
\left[ fn^+ \text{ of} \right]? \text{ table [,.filter]?} \\
\text{aggregate min|max|sum|avg|count fn of } table [,.filter]? \\
\ldots
\]
LUINet: NL to ThingTalk
What is the top-rated Chinese restaurant in Palo Alto?

sort aggregateRating.ratingValue desc of (@QA.restaurant(),
geo == new MakeLocation(“Stanford”)
&& servesCuisine =~ “Chinese” )
Genie Pipeline

Natural language ➔ LUINet ➔ ThingTalk
Genie Pipeline

Thingpedia Manifest

Schema

<table>
<thead>
<tr>
<th>Name</th>
<th>Price</th>
<th>Cuisine</th>
<th>...</th>
</tr>
</thead>
</table>

Natural Language Annotations

- cuisine of the restaurant
- restaurant’s cuisine
- cuisine served by the restaurant
Genie Pipeline

Thingpedia Manifest

Schema

<table>
<thead>
<tr>
<th>Name</th>
<th>Price</th>
<th>Cuisine</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Natural Language Annotations

cuisine of the restaurant
restaurant’s cuisine
cuisine served by the restaurant

ThingTalk Grammar

Domain-independent

Templates

What is the <prop> of <table>?
What is the <table>’s <prop>?

Natural language → LUINet → ThingTalk
Genie Pipeline

Thingpedia Manifest

Schema

<table>
<thead>
<tr>
<th>Name</th>
<th>Price</th>
<th>Cuisine</th>
<th>…</th>
</tr>
</thead>
</table>

Natural Language Annotations

- cuisine of the restaurant
- restaurant’s cuisine
- cuisine served by the restaurant

Synthesize sentence/code pairs

Domain-independent Templates

What is the <prop> of <table>?
What is the <table>’s <prop>?

ThingTalk Grammar

Stanford University
Genie Pipeline

ThingTalk Grammar
Domain-independent Templates
What is the <prop> of <table>?
What is the <table>'s <prop>?

Schema

<table>
<thead>
<tr>
<th>Name</th>
<th>Price</th>
<th>Cuisine</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Natural Language Annotations

- cuisine of the restaurant
- restaurant’s cuisine
- cuisine served by the restaurant

Synthesize sentence/code pairs
Paraphrase

LUINet

Stanford University
Genie Pipeline

- **ThingTalk Grammar**
- **Domain-independent Templates**
  - What is the `<prop>` of `<table>`?
  - What is the `<table>`’s `<prop>`?

The diagram shows a flow from **Natural Language Annotations** to **Synthesize sentence/code pairs**, then to **Paraphrase**, and finally to **Parameter & data augmentation**. The annotations include:

- **Table**
  - **Name**
  - **Price**
  - **Cuisine**

Example annotations:
- Cuisine of the restaurant
- Restaurant’s cuisine
- Cuisine served by the restaurant

The diagram also shows the integration with **LUINet** and **ThingTalk** from Stanford University.
Genie Pipeline

Natural Language Annotations

Thingpedia Manifest

Schema

<table>
<thead>
<tr>
<th>Name</th>
<th>Price</th>
<th>Cuisine</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>cuisine of the restaurant</td>
<td>cuisine served by the restaurant</td>
</tr>
</tbody>
</table>

Natural Language Annotations

What is the <prop> of <table>?
What is the <table>’s <prop>?

ThingTalk Grammar

Domain-independent Templates

Synthesize sentence/code pairs

Paraphrase

Parameter & data augmentation

Training Data

Natural language

LUINet

ThingTalk

Stanford University
Genie Pipeline

Thingpedia Manifest

Natural Language Annotations
- Cuisine of the restaurant
- Restaurant’s cuisine
- Cuisine served by the restaurant

Schema

| Name | Price | Cuisine | ...
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Natural Language Annotations
- What is the <prop> of <table>?
- What is the <table>’s <prop>?

Training Data
- Synthesize sentence/code pairs
- Paraphrase
- Parameter & data augmentation

Domain-independent Templates
- ThingTalk Grammar
- Evaluation & Test: Real User Input

LUINet
- Natural language

ThingTalk
- Stanford University
Automatically Turn Schema.org into Thingpedia Manifest

- Tables with hierarchy
  - properties are inherited from parent tables
- only keep classes & properties with data
- decide types based on schema.org types and data

```json
@org.schema {
  Restaurant extends FoodEstablishment {}
  FoodEstablishment extends LocalBusiness {
    acceptsReservation: Boolean,
    servesCuisine: String,
  }
  LocalBusiness extends Place, Organizations {
    priceRange: String,
    openingHours: String,
  }
  Organizations extends Thing {
    aggregateRating: {
      ratingCount: Number,
      ratingValue: Number,
    },
    review: Array(Review),
  }
  Thing {
    name: String,
  }
}
```
Map properties to natural language
Map properties to natural language

- Long, non-word, property names
  - E.g., ratingValue, servesCuisine
- Variety in natural language usage
Map properties to natural language

- Long, non-word, property names
  - E.g., ratingValue, servesCuisine
- Variety in natural language usage

<table>
<thead>
<tr>
<th>servesCuisine</th>
<th>ratingValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese restaurant ✓</td>
<td>4.5 restaurant ✗</td>
</tr>
<tr>
<td>Restaurant with Chinese cuisine ✓</td>
<td>Restaurant with 4.5 rating ✓</td>
</tr>
<tr>
<td>Restaurant served Chinese cuisine ✗</td>
<td>Restaurant rated 4.5 ✓</td>
</tr>
<tr>
<td>Restaurant that serves Chinese cuisine ✓</td>
<td>Restaurant rates 4.5 ✗</td>
</tr>
<tr>
<td>Restaurant with Chinese ✗</td>
<td>Restaurant with 4.5 ✗</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
NL Annotations by Part-Of-Speech Categories

- “servesCuisine”
  - Noun phrase
    - “cuisine”: e.g., “the cuisine of the restaurant”, “restaurants with Chinese cuisine”
  - Verb phrase
    - “serves # cuisine”, “serves #”: e.g., “restaurant that serves Chinese cuisine”, “what does the restaurant serve”
  - Adjective-phrase value (with no property name)
    - E.g., “Chinese restaurants”
NL Annotation Generation

servesCuisine
NL Annotation Generation

servesCuisine → serves → cuisine
servesCuisine serves cuisine

VBP

NN

Stanford University
NL Annotation Generation

servesCuisine: Verb: “serves # cuisine”
Noun: “# cuisine”
NL Annotation Generation

• Automatic: Heuristics based on POS (Part-Of-Speech) tags

  servesCuisine -> serves (VBP) -> cuisine (NN)

  NL Annotations
  servesCuisine:
  Verb: “serves # cuisine”
  Noun: “# cuisine”

• Manual:
  • Provides additional synonyms, and annotations in different POS categories
NL Annotation Generation

- Automatic: Heuristics based on POS (Part-Of-Speech) tags

  servesCuisine → serves → cuisine
  
  Verb: "serves # cuisine"
  Noun: "# cuisine"

- Manual:
  - Provides additional synonyms, and annotations in different POS categories

  servesCuisine:
  Verb: "serves # cuisine"
  "offers # food"
  Noun: "# cuisine"
  "# food"
  Adjective: "#"
Domain-independent Templates

servesCuisine:
Verb: “serves # cuisine”
      “offers # food”
Noun: “# cuisine”
      “# food”
Adjective: “#”
Domain-independent Templates

servesCuisine:
Verb: “serves # cuisine”
“offers # food”
Noun: “# cuisine”
“# food”
Adjective: “#”

Show me <table> that <verb>.
Show me <table> with <noun>.
Show me <adjective> <table>.
Domain-independent Templates

servesCuisine:

Verb: “serves # cuisine”
    “offers # food”

Noun: “# cuisine”
    “# food”

Adjective: “#”

now => @QA.restaurant(),
servesCuisine =~ “Chinese” => notify;

Show me <table> that <verb>.
Show me <table> with <noun>.
Show me <adjective> <table>.

Show me restaurants that serve Chinese cuisine.
Show me restaurants with Chinese food.
Show me Chinese restaurants.
Domain-independent Templates

servesCuisine:
Verb: “serves # cuisine”
   “offers # food”
Noun: “# cuisine”
   “# food”
Adjective: “#”

Show me <table> that <verb>.  
Show me <table> with <noun>.  
Show me <adjective> <table>.

Show me <table> with <noun:NUMBER> greater than <value>.

now => @QA.restaurant(),
servesCuisine =~ “Chinese” => notify;

Show me restaurants that serve Chinese cuisine. 
Show me restaurants with Chinese food. 
Show me Chinese restaurants.

Show me restaurants with rating greater than 4
Domain-independent Templates

```
servesCuisine:
Verb: “serves # cuisine”
     “offers # food”
Noun: “# cuisine”
     “# food”
Adjective: “#”
```

now => @QA.restaurant(),
servesCuisine =~ “Chinese” => notify;

Show me <table> that <verb>.
Show me <table> with <noun>.
Show me <adjective> <table>.

Show me <table> with <noun:NUMBER> greater than <value>.

Show me <table> with <noun:MEASURE(m)> longer than <value>.

Show me restaurants that serve Chinese cuisine.
Show me restaurants with Chinese food.
Show me Chinese restaurants.

Show me restaurants with rating greater than 4
Show me surfboard with length longer than 3m
Domain-dependent Templates
Domain-dependent Templates

- Some natural sentences cannot be generated by domain-independent templates:
  - “the top-rated restaurant”, “the best restaurant”
  - We allow developers to improve the accuracy by providing domain-dependent templates
Domain-dependent Templates

• Some natural sentences cannot be generated by domain-independent templates:

<table>
<thead>
<tr>
<th>ThingTalk</th>
<th>Sentence by domain-independent templates</th>
</tr>
</thead>
<tbody>
<tr>
<td>sort aggregateRating.ratingValue desc of @QA.restaurant()</td>
<td>restaurant with the highest rating</td>
</tr>
<tr>
<td></td>
<td>restaurant that have the highest rating</td>
</tr>
<tr>
<td></td>
<td>...</td>
</tr>
</tbody>
</table>

• “the top-rated restaurant”, “the best restaurant”
• We allow developers to improve the accuracy by providing domain-dependent templates
Domain-dependent Templates

- Some natural sentences cannot be generated by domain-independent templates:

<table>
<thead>
<tr>
<th>ThingTalk</th>
<th>Sentence by domain-independent templates</th>
</tr>
</thead>
<tbody>
<tr>
<td>sort aggregateRating.ratingValue desc of @QA.restaurant()</td>
<td>restaurant with the highest rating</td>
</tr>
<tr>
<td></td>
<td>restaurant that have the highest rating</td>
</tr>
<tr>
<td></td>
<td>...</td>
</tr>
</tbody>
</table>

- “the top-rated restaurant”, “the best restaurant”

- We allow developers to improve the accuracy by providing domain-dependent templates:

<table>
<thead>
<tr>
<th>ThingTalk</th>
<th>Domain-dependent templates</th>
</tr>
</thead>
<tbody>
<tr>
<td>sort aggregateRating.ratingValue desc of @QA.restaurant()</td>
<td>the top-rated restaurant</td>
</tr>
<tr>
<td></td>
<td>the best restaurant</td>
</tr>
<tr>
<td></td>
<td>...</td>
</tr>
</tbody>
</table>
Experiments
Experimental Results

- **Domains**
  - Restaurants: data from Yelp
  - Person: data from LinkedIn

- **Training set**

<table>
<thead>
<tr>
<th></th>
<th>Restaurant</th>
<th>Person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synthetic</td>
<td>1,294,278</td>
<td>553,067</td>
</tr>
<tr>
<td>Paraphrase</td>
<td>6,288</td>
<td>6,000</td>
</tr>
<tr>
<td><strong>Total (augmented)</strong></td>
<td><strong>1,809,109</strong></td>
<td><strong>930,564</strong></td>
</tr>
</tbody>
</table>

- **Realistic evaluation set**

<table>
<thead>
<tr>
<th></th>
<th>Restaurant</th>
<th>Person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dev 1 property</td>
<td>134</td>
<td>6</td>
</tr>
<tr>
<td>2 properties</td>
<td>47</td>
<td>144</td>
</tr>
<tr>
<td>3+ properties</td>
<td>59</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>240</strong></td>
<td><strong>160</strong></td>
</tr>
<tr>
<td>Test 1 property</td>
<td>96</td>
<td>127</td>
</tr>
<tr>
<td>2 properties</td>
<td>79</td>
<td>106</td>
</tr>
<tr>
<td>3+ properties</td>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>215</strong></td>
<td><strong>233</strong></td>
</tr>
</tbody>
</table>
Comparison with Commercial Virtual Assistants

Answer Accuracy on Restaurant Queries

- Alexa
- Google
- Siri
- Almond
Comparison with Commercial Virtual Assistants

Answer Accuracy on Restaurant Queries

Trained with no real data!
Experimental Results

![Bar chart showing results for different property counts and types]

- **1 property**
  - Restaurant: 80%
  - Person: 80%

- **2 properties**
  - Restaurant: 70%
  - Person: 70%

- **3+ properties**
  - Restaurant: 60%
  - Person: 60%

- **Overall**
  - Restaurant: 70%
  - Person: 70%

Stanford University
Can We Do Better?
Manual Effort in Genie Pipeline
Manual Effort in Genie Pipeline

- Natural language annotations
  - The heuristics based on part-of-speech doesn’t provide good variety, and sometimes unnatural

- Paraphrase
  - We ask crowdworkers to manually paraphrase synthetic sentences
  - We can only do this for a small sample of synthetic because of cost

- Can we replace them with something automatic?
Automatic NL Annotation Generation

• Generate context-aware synonyms by a language model
Automatic NL Annotation Generation

- Generate context-aware synonyms by a language model

A Sample Sentence
Automatically Constructed based on POS

Show me restaurants with Italian cuisine.
Automatic NL Annotation Generation

- Generate context-aware synonyms by a language model

A Sample Sentence
Automatically Constructed based on POS

Show me restaurants with Italian cuisine.

BERT (pretrained)
Automatic NL Annotation Generation

- Generate context-aware synonyms by a language model

A Sample Sentence
Automatically Constructed based on POS

Show me restaurants with Italian cuisine.

BERT (pretrained)

Show me restaurants with Italian dishes.
Show me restaurants with Italian food.
Show me restaurants with Italian menu.
...

Generate Context-aware Synonyms
Automatic NL Annotation Generation

- Generate context-aware synonyms by a language model

A Sample Sentence
Automatically Constructed based on POS

Show me restaurants with "Italian cuisine."

BERT (pretrained)

Show me restaurants with "Italian dishes."
Show me restaurants with "Italian food."
Show me restaurants with "Italian menu."
...

Generate Context-aware Synonyms

Templatize

noun: "# cuisine | dishes | menu ... "

Stanford University
Automatic NL Annotation Generation (cont.)

• Predict adjective properties by a language model
Automatic NL Annotation Generation (cont.)

• Predict adjective properties by a language model

Construct a sample sentence with mask

Show me a [MASK] restaurant.
Automatic NL Annotation Generation (cont.)

- Predict adjective properties by a language model

Construct a sample sentence with mask

Show me a [MASK] restaurant.

BERT (pretrained)

Show me a good restaurant. Show me a Chinese restaurant. ...

Predict [MASK]
Automatic NL Annotation Generation (cont.)

- Predict adjective properties by a language model

  - Construct a sample sentence with mask
    - Show me a [MASK] restaurant.

  - BERT (pretrained)
    - Show me a good restaurant.
    - Show me a Chinese restaurant.
    - ...

  - Predict [MASK]

  - Look up predicted words in property value sets
Automatic NL Annotation Generation (cont.)

- Predict adjective properties by a language model

  - Construct a sample sentence with mask
  - Show me a [MASK] restaurant.
  - BERT (pretrained)
  - Show me a good restaurant.
  - Show me a Chinese restaurant.
  - Look up predicted words in property value sets
  - servesCuisine – adjective: “#”
  - Add adjective annotation to found properties
Preliminary Experimental Result

Accuracy on Restaurant Queries

- POS-based Heuristics
- Automatic
- Manual

Auto NL Annotation Generation
Automatic Paraphrasing
Automatic Paraphrasing

- Paraphrase dataset
- GPT-2 (Pretrained)
Automatic Paraphrasing

Paraphrase dataset

GPT-2 (Pretrained)

Fine-tune

GPT-2 Paraphraser

Stanford University
Automatic Paraphrasing

- Synthetic Training Examples
  - Show me restaurants with Chinese cuisine.

- Fine-tune
  - GPT-2 Paraphraser
  - What is a restaurant that is Chinese?
    - Give me Chinese dining places.
    - Show me top-rated Chinese restaurants.
    - ...

- Paraphrase dataset
- GPT-2 (Pretrained)
Automatic Paraphrasing

- Synthetic Training Examples
  - Show me restaurants with Chinese cuisine.
  - What is a restaurant that is Chinese? Give me Chinese dining places. Show me top-rated Chinese restaurants.
  - Fine-tune
    - GPT-2 Paraphraser
  - Inference
    - LUINet Trained w/ Synthetic data
      - Filter paraphrases that do not preserve meaning

Paraphrase dataset
GPT-2 (Pretrained)
Automatic Paraphrasing

- **Paraphrase dataset**
- **GPT-2 (Pretrained)**
- **Synthetic Training Examples**
  - Show me restaurants with Chinese cuisine.
  - What is a restaurant that is Chinese? Give me Chinese dining places.
  - Show me top rated Chinese restaurants.
  - ...
- **GPT-2 Paraphraser**
- **Fine-tune**
- **Inference**
  - LUINet Trained w/ Synthetic data
  - Filter paraphrases that do not preserve meaning
- **Paraphrased Examples**
  - What is a restaurant that is Chinese? Give me Chinese dining places.
  - ...

Stanford University
Preliminary Experimental Result

Accuracy on Restaurant Queries

- **Synthetic only**
- **Auto Paraphrase**
- **Humann Paraphrase**

Auto Paraphrasing
Thank you!

Hope you will enjoy your homework 😊