

Stanford CS224v Course

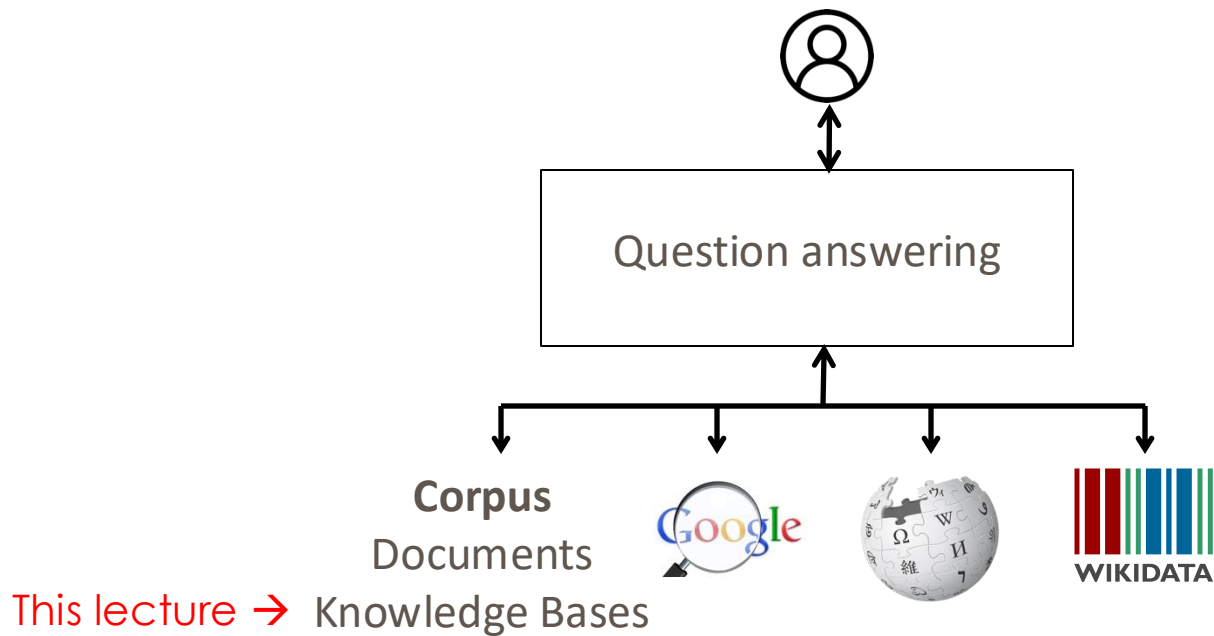
Conversational Virtual Assistants with Deep Learning

Lecture 6

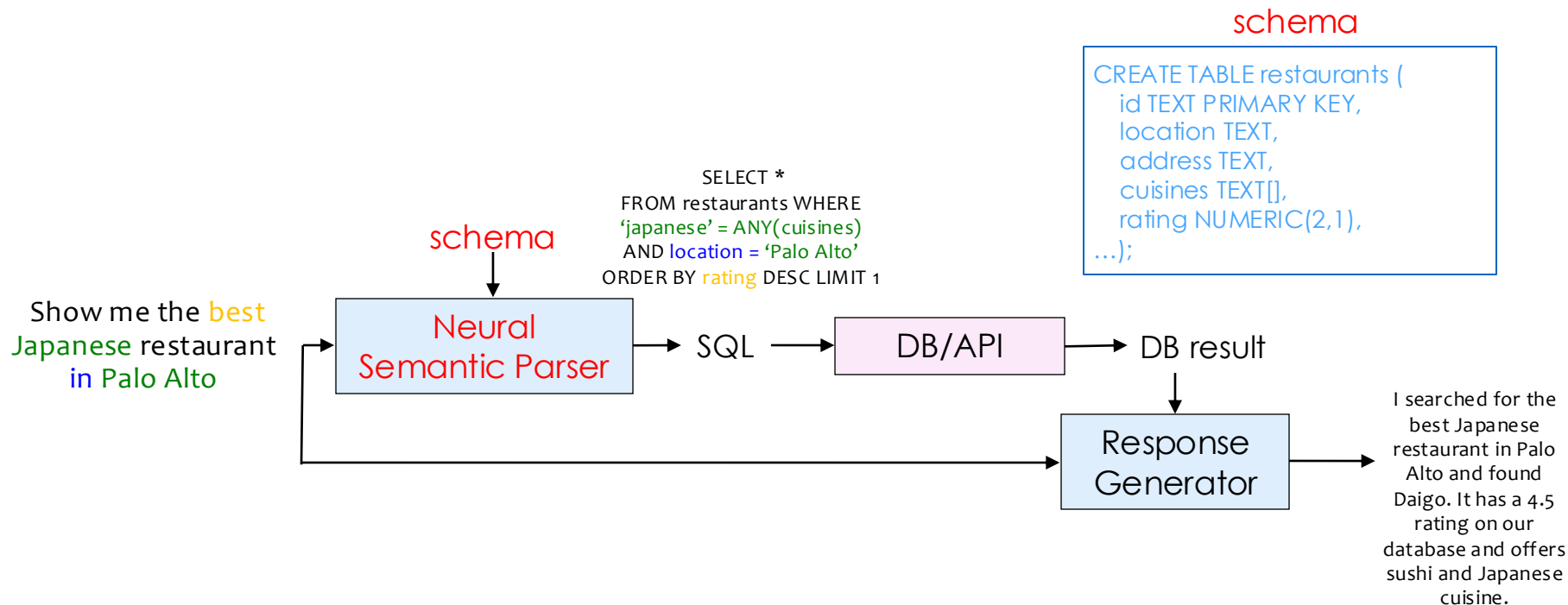
Introduction to Agents for Structured Data

Monica Lam & Shicheng Liu

Architecture



Semantic Parsing: NL $\rightarrow$ Formal Query



Outline

- **Our casual questions are formal queries**
- A basic LLM-based agent that consults a database – Yelp
- Evaluation issues

Power of Query Languages

- Domain agnostic
- All queries of any domain in a table are compositions of a FEW relational algebra operations
 - Basic: Selection, Projection, Cartesian product (Join), Union, Set Difference
 - Extended: Sort, Aggregate Operators (Sum, Max, Avg, ...)
- Expressive, succinct, well-defined

Amazing
CS Idea

WHERE clause: filter

```
CREATE TABLE restaurants (  
  id TEXT PRIMARY KEY,  
  location TEXT,  
  address TEXT,  
  cuisines TEXT[],  
  rating NUMERIC(2,1),  
  ...);
```

```
SELECT * FROM  
table [WHERE filter]?
```

```
SELECT * FROM restaurants  
WHERE location == "Palo Alto"  
AND 'chinese' = ANY (cuisines)
```

Show me Chinese restaurants in Palo Alto

WHERE clause: filter

```
CREATE TABLE restaurants (  
  id TEXT PRIMARY KEY,  
  location TEXT,  
  address TEXT,  
  cuisines TEXT[],  
  rating NUMERIC(2,1),  
  ...);
```

```
SELECT * FROM  
table [WHERE filter]?
```

```
SELECT * FROM restaurants  
WHERE location = "Palo Alto"  
AND "Chinese" = ANY(cuisines)  
AND rating >= 4.5
```

Show me Chinese restaurants in Palo Alto
with at least 4.5 stars

SELECT clause: Projection

```
CREATE TABLE restaurants (  
  id TEXT PRIMARY KEY,  
  location TEXT,  
  address TEXT,  
  cuisines TEXT[],  
  rating NUMERIC(2,1),  
  ...);
```

```
SELECT field+ FROM  
table [WHERE filter];
```

```
SELECT address FROM restaurants  
WHERE location = "Palo Alto"  
AND "Chinese" = ANY(cuisines)  
AND rating >= 4.5
```

Show me **the address of** Chinese restaurants **in** Palo Alto
with at least 4.5 stars

Subquery

```
CREATE TABLE restaurants (  
  id TEXT PRIMARY KEY,  
  location TEXT,  
  address TEXT,  
  cuisines TEXT[],  
  rating NUMERIC(2,1),  
  ...);
```

```
CREATE TABLE reviews (  
  id TEXT PRIMARY KEY,  
  restaurant_id TEXT  
  REFERENCES restaurants(id),  
  author TEXT,  
  ...);
```

SELECT **field**+
FROM **table** [WHERE **filter**];

Atom filter:
param op value

Subquery filter:
param op
*(SELECT * FROM ...)*

SELECT **address** FROM restaurants

WHERE location = "Palo Alto"

AND "Chinese" = ANY(cuisines)

AND id IN

(SELECT **restaurant_id** FROM reviews WHERE author =~ "Bob")

Show me the **address** of **Chinese** restaurants in **Palo Alto**
reviewed by **Bob**

Sorting

```
CREATE TABLE restaurants (  
  id TEXT PRIMARY KEY,  
  location TEXT,  
  address TEXT,  
  cuisines TEXT[],  
  rating NUMERIC(2,1),  
  ...);
```

```
CREATE TABLE reviews (  
  id TEXT PRIMARY KEY,  
  restaurant_id TEXT  
  REFERENCES restaurants(id),  
  author TEXT,  
  ...);
```

```
SELECT field+ FROM table [WHERE filter];  
ORDER BY field DESC/ASC
```

```
SELECT address FROM restaurants  
WHERE location = "Palo Alto"  
AND "Chinese" = ANY(cuisines)  
AND id IN  
(SELECT restaurant_id FROM reviews WHERE author =~ "Bob")  
ORDER BY rating DESC
```

Show me the address of top-rated Chinese restaurants in Palo Alto
reviewed by Bob

Joins

```
CREATE TABLE restaurants (  
  id TEXT PRIMARY KEY,  
  location TEXT,  
  address TEXT,  
  cuisines TEXT[],  
  rating NUMERIC(2,1),  
  ...);
```

```
CREATE TABLE reviews (  
  id TEXT PRIMARY KEY,  
  restaurant_id TEXT  
  REFERENCES restaurants(id),  
  author TEXT,  
  ...);
```

```
SELECT * FROM  
table join table [WHERE filter]? [on filter]?
```

```
SELECT * FROM restaurants JOIN reviews ON  
  restaurants.id = reviews.restaurant_id  
  WHERE location = "Palo Alto"  
  AND "Chinese" = ANY(cuisines)  
  ORDER BY rating DESC
```

Show me top-rated Chinese restaurants in Palo Alto
with their reviews

Can Assistants Understand Compound Queries?

Examples of Compound Questions	Alexa	Google	Siri	Genie
Show me restaurants rated at least 4 stars with at least 100 reviews				✓
Show restaurants in San Francisco rated higher than 4.5		✓		✓
What is the highest rated Chinese restaurant near Stanford?			✓	✓
How far is the closest 4 star restaurant?				✓
Find a W3C employee that went to Oxford				✓
Who worked for Google and lives in Palo Alto?				✓
Who graduated from Stanford and won a Nobel prize?	✓	✓		✓
Who worked for at least 3 companies?				✓
Show me hotels with checkout time later than 12PM				✓
Which hotel has a pool in this area?		✓	✓	✓

<https://arxiv.org/abs/2001.05609>

Restaurant Assistant



```
CREATE TABLE restaurants (  
    name TEXT,  
    address TEXT,  
    location TEXT,  
    phone_number TEXT,  
    opening_hours TEXT,  
    cuisines TEXT[],  
    price ENUM ('cheap', 'moderate', 'expensive', 'luxury'),  
    rating NUMERIC(2,1),  
    num_reviews NUMBER,  
    reviews FREE_TEXT,  
    popular_dishes FREE_TEXT,  
);
```

NUMERIC(2,1) means 2-digit precision, 1 digit after decimal

Queries in a Conversation (Yelp)

My dad is visiting me in Mountain View. What do you think about dinner choices?

I found Cascal, a moderate Spanish restaurant with great tapas platters, appetizers, and drinks. It has a great ambience with live music on the weekends and both indoor and outdoor seating with heat lamps to keep guests warm.

Generated SQL:

*SELECT * FROM restaurants
WHERE location = "Mountain View"*

How about Italian instead? I really like the food there when I visited Florence last year.

I found Vaso Azzurro Ristorante, a moderate Italian restaurant near Mountain View. It has a lively atmosphere and offers excellent customer service. Reviews say their pasta is authentic and perfectly seasoned.

Generated SQL:

*SELECT * FROM restaurants
WHERE location = "Mountain View"
AND cuisines = ANY("Italian")*

Oh, what is the ham called in Italian?

Prosciutto is the Italian word for ham.

GPT responds directly

Thanks. We'll be visiting Stanford. How about in Palo Alto?

I found Terún, a 4 star Italian and Pizza restaurant near Stanford in Palo Alto. Reviews say the pizza is great and they can seat customers outside without a reservation.

Generated SQL:

*SELECT * FROM restaurants
WHERE location = "Palo Alto"
AND cuisines = ANY("Italian")*

NL → FORMAL DATABASE QUERIES

NATURAL LANGUAGE IS COMPOSITIONAL

A CASUAL NL STATEMENT CAN CORRESPOND TO
FORMAL QUERIES WITH MANY CLAUSES

QUIZ: FOR YELP,
HOW MANY POSSIBLE QUERIES ARE THERE?

Outline

- Our casual questions are formal queries
- **A basic LLM-based agent that consults a database – Yelp**
- Evaluation issues

LLM: Few-Shot Prompt

You are a semantic parser. Generate a query for a restaurant database with the following signature:

```
CREATE TABLE restaurants (  
  name TEXT,  
  address TEXT,  
  location TEXT,  
  phone_number TEXT,  
  opening hours TEXT,  
  cuisines TEXT[],  
  price ENUM ('cheap', 'moderate', 'expensive', 'luxury'),  
  rating NUMERIC(2,1),  
  num_review NUMBER,  
);
```

+ examples of semantic parses

Casual Assessment

- Show that LLM works surprising well
 - For domains such as restaurants because of its familiarity with the domain
- LLMs already know the basic SQL syntax
- Observed issue: Enumerated values

Issue: Enumerated Types

```
1 CREATE TABLE restaurants ( ...  
2     price ENUM ('cheap', 'moderate', '  
3         expensive', 'luxury'),  
    ...)
```

- Small # values (e.g. 10): included in schema description
- LLMs know how to use these enumerated values

What are some good **inexpensive** chinese restaurants in Palo Alto?

↓
SELECT * FROM restaurants WHERE
'chinese' = ANY (cuisines) AND location = 'Palo Alto'
AND **price** = 'cheap' AND rating >= 4.0
ORDER BY num_reviews DESC LIMIT 3;

Large Enumerated Types

- Example: Yelp cuisines (text) has over 200 values, e.g.
 - "brazilian", "southern", "soul food", "modern european", "champagne bars", "latin american", "health & medical", "singaporean", ...
- Problem: Not feasible to put all choices in the schema
- Database search expects an exact match

show me a **café**
↓
SELECT * FROM restaurants WHERE 'coffee' = ANY (cuisines)

Yelp: no coffee cuisine
Available cuisines:
coffee & tea,
cafe

Large Enumerated Types

show me a **café**

↓ Semantic parser

SELECT * FROM restaurants WHERE '**coffee**' = ANY (cuisines)

↓ Rewrite '=' operator by calling
classify ("coffee", cuisines)

SELECT * FROM restaurants WHERE
'**coffee & tea**' = ANY (cuisines) OR '**cafe**' = ANY (cuisines)

Yelp: no coffee cuisine
Available cuisines:
coffee & tea,
cafe

Solution

- Change '=' operator for a text field into a value classification operation
- $\text{classify}(x, V)$ finds the closest values of x in V ,
returns $\{\}$ if no close values found
- Given a text field f , collect all possible values V

$$x = \text{ANY}(f) \Rightarrow c_1 = \text{ANY}(f) \text{ OR } c_2 = \text{ANY}(f) \dots,$$

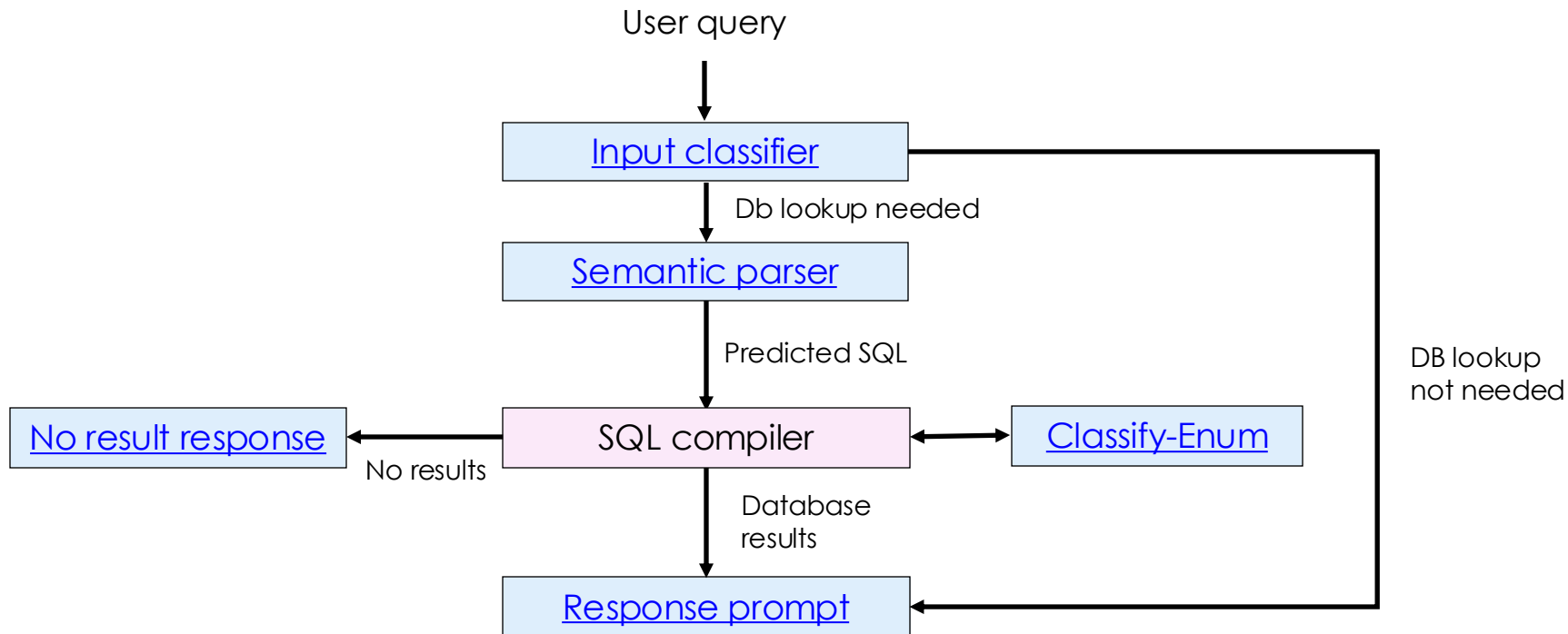
where $c_i \in \text{classify}(x, V)$

Quiz: How to implement Classify?

The Full Agent Design

Some turns are not queries, let LLMs answer those
Few-shot prompts based on the domain

Agent Design



Click the links to see the prompts (written in [jinja syntax](#))

FEW-SHOT LLM-BASED DATABASE AGENTS
ARE EASY TO BUILD

DO THEY PERFORM WELL?

Outline

- Our casual questions are formal queries
- A basic LLM-based agent that consults a database – Yelp
- **Evaluation issues**

Evaluation

- Measure query accuracy of our annotated data
 - Poor accuracy
- Quiz: what should we do next?

Example

What's some great Mexican food around Bernal Heights?

- Gold Target:
 - SELECT FROM restaurants
WHERE “mexican” = ANY (cuisines) AND location = “bernal heights”;
- Predicted Target:
 - SELECT FROM restaurants
WHERE “mexican” = ANY (cuisines) AND location = “bernal heights”
AND rating >= 4;

Quiz: Which is correct?

Evaluation example (SQL, BIRD dataset)

Name all cards with 2015 frame style ranking below 100 on EDHRec.

- Gold Target:
 - `SELECT id FROM cards WHERE edhrecRank < 100 AND frameVersion = 2015`
- Predicted Target:
 - `SELECT name FROM cards WHERE edhrecRank < 100 AND frameVersion = 2015;`

Quiz: which is correct?

Our Discovery Process

- Tried few-shot ChatGPT to make it conform to our convention
- Tried to fix the annotations in the benchmark



ChatGPT is better than annotations by PhD students!

We have landed on a new planet!



Quiz: What should we do next?

Conclusion

- Few-shot Chat-GPT parses SQL queries for Yelp
 - Restaurants: well-known domain to ChatGPT
 - Small table: 11 fields
(incl. 2 Free-text, 1 small, 1 large ENUM)
 - Well-understood field names
- Cannot evaluate based using old benchmarks!
- Cannot address the full problem without free-text retrieval!
 - Must handle hybrid data before rigorous evaluation!