Stanford CS224v Course

Conversational Virtual Assistants with Deep Learning

Lecture 5

SUQL: Structured/Unstructured Query Language

Monica Lam & Shicheng Liu

Summary from Lecture 4

- **Many questions require information from hybrid data sources**
- **Structure OR Text: is inadequate**
  - Binary classifier up front (SK-TOD, 2023)
  - Pick afterwards (Stanford Chirpy Cardinal, 2021)
- **Different approaches to combine structures and free-text**
  - Structures → Text: Linearization (one hop)
  - Text → Structure: Semantic parser (Hard to represent free text in KB)
  - Hybrid: Retrieve from both and combine (one hop each)
- **Hybrid questions are multi-hop questions**
  - Break-it-Down: the only truly multi-hop solution for text only

Quiz: What should we do given prior results?

Desired Solution

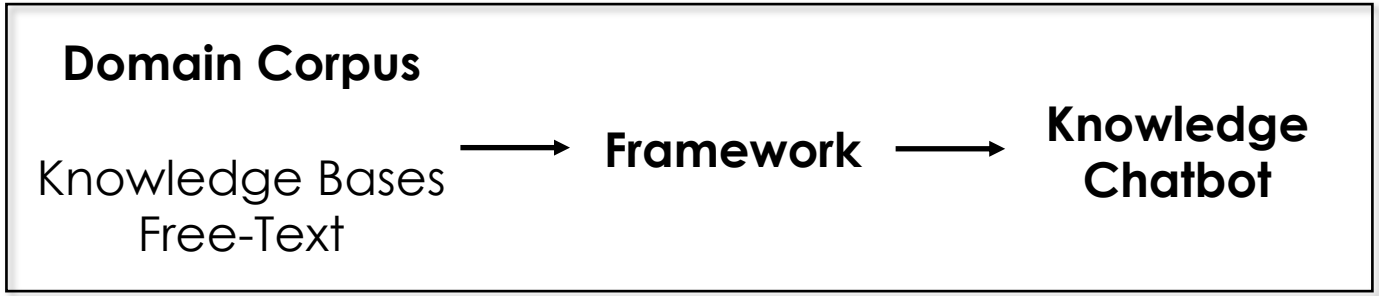
1. Keep hybrid data sources
   - Don't convert KB → Text, or Text → KB
2. Hybrid multi-hop: Compose IR and KB accesses arbitrarily
3. Break-it-down decomposes questions into "English-like" SQL operations
   - Has issues with executing "English" operations
   - Can't support DBes

Since LLMs understand SQL

- SUQL (Structured & Unstructured Query Language)
- Extends SQL to include free-text operation with structured access
- Create an optimizing compiler to support full SUQL (all compositions)

NEW!

Lecture Goals



- SUQL Language
- SUQL Semantic Parser
- SUQL Performance

Many practical hints for some of your projects!

SUQL Free-Text Support

- Add free-text primitives into SQL, implemented with IR & LLM
- Two functions: summary, answer

"I want a family-friendly restaurant in Palo Alto"

```
SELECT *, summary(reviews) FROM restaurants
WHERE location = 'Palo Alto'
AND answer(reviews, 'is it a family friendly restaurant') = 'Yes' LIMIT 1;
```

Quiz: How do you implement summary and answer?

HybridQA Dataset

Wikipedia Tables *hyperlinked* Wikipedia Pages

The 2016 Summer Olympics officially known as the Games of the XXXI Olympiad (Portuguese : Jogos da XXXI Olimpíada) and commonly known as **Rio** 2016 , was an international multi-sport event ……

Flag bearers of Myanmar at the Olympics

Name	Year	Season	Flag bearer
XXXI	2016	Summer	Yan Naing Soe
XXX	2012	Summer	Zaw Win Thet
XXIX	2008	Summer	Phone Myint Tayzar
XXVIII	2004	Summer	Hla Win U
XXVII	2000	Summer	Maung Maung Nge
XX	1972	Summer	Win Maung

Yan Naing Soe (born **31 January 1979**) is a Burmese judoka . He competed at the 2016 Summer Olympics in the **men 's 100 kg event** , …… He was the flag bearer for Myanmar at the **Parade of Nations** .

Zaw Win Thet (born **1 March 1991** in Kyonpyaw , Pathein District , Ayeyarwady Division , Myanmar) is a Burmese runner who ……

Myint Tayzar Phone (Burmese : မြင့်တေဇာဖုန်း) born **July 2 , 1978**) is a sprint canoer from Myanmar who competed in the late 2000s .

……

Win Maung (born **12 May 1949**) is a Burmese footballer . He competed in the men 's tournament at the 1972 Summer Olympics …

Hardness ↓

Q: In which year did the judoka bearer participate in the Olympic opening ceremony?	A: 2016
Q: Which event does the does the XXXI Olympic flag bearer participate in?	A: men's 100 kg event
Q: Where does the Burmesse jodoka participate in the Olympic opening ceremony as a flag bearer?	A: Rio
Q: For the Olympic event happening after 2014, what session does the Flag bearer participate?	A: Parade of Nations
Q: For the XXXI and XXX Olympic event, which has an older flag bearer?	A: XXXI
Q: When does the oldest flag Burmese bearer participate in the Olympic ceremony?	A: 1972

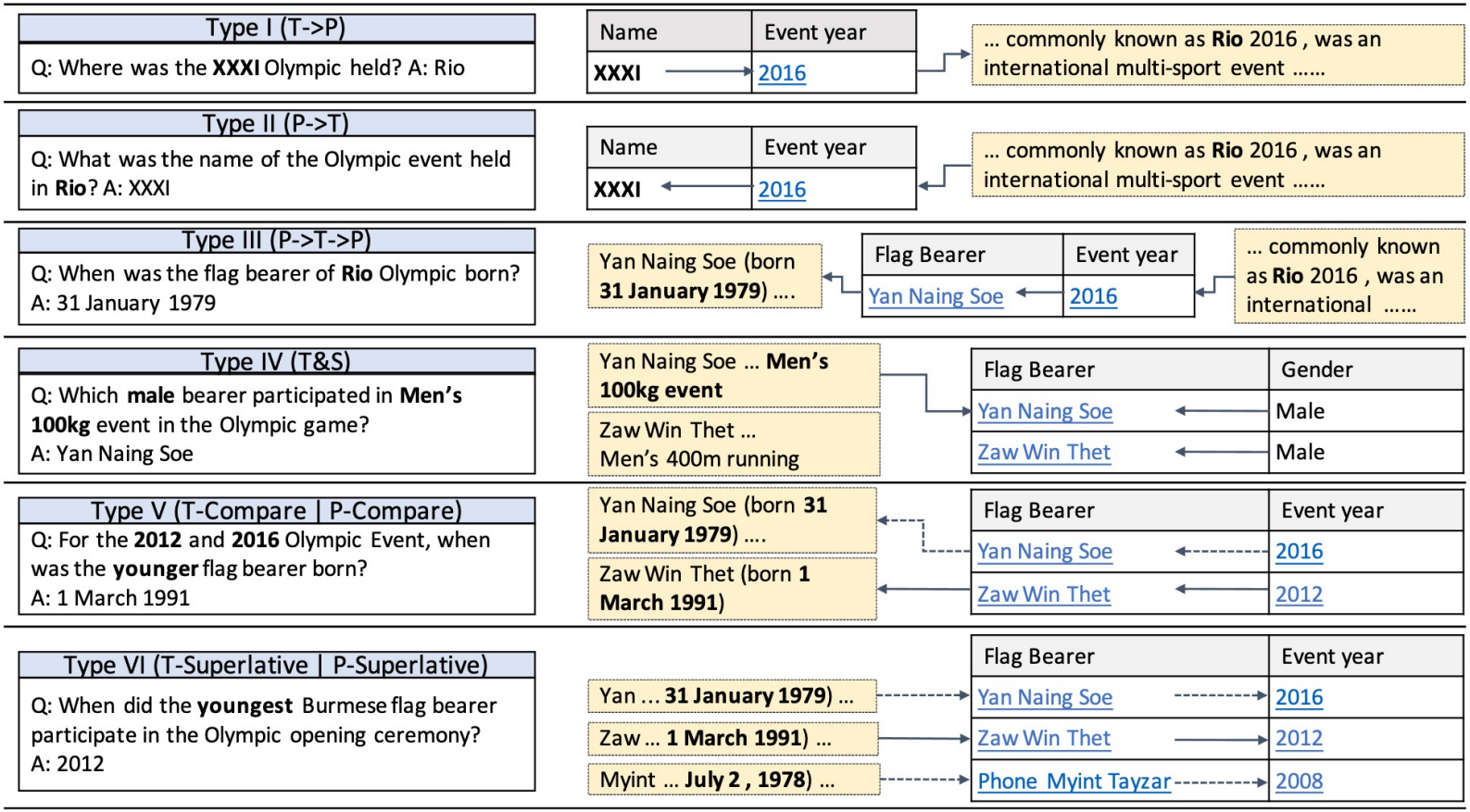


Figure 3: Illustration of different types of multi-hop questions.

T: Table
P: Passage

HybridQA Questions in SUQL

```
CREATE TABLE Flag (
"Name" TEXT,
"Flag Bearer" TEXT,
"Flag Bearer_Info" TEXT[],
"Gender" TEXT,
"Event year" TEXT,
"Event year_Info" TEXT[]);
```

T: Table
P: Paragraph

Type I (T->P)

Where was the XXXI Olympic held?

```
SELECT answer("Event year_Info",
                    'where is this event held?')
        FROM "Flag" WHERE "Name" = 'XXXI'
```

Type II (P->T) What was the name of the Olympic event held in Rio?

```
SELECT "Name" FROM "Flag"
WHERE answer("Event year_Info",
        'where is this event held?') = 'Rio'
```

HybridQA Questions in SUQL

```
CREATE TABLE Flag (
"Name" TEXT,
"Flag Bearer" TEXT,
"Flag Bearer_Info" TEXT[],
"Gender" TEXT,
"Event year" TEXT,
"Event year_Info" TEXT[]);
```

T: Table
P: Paragraph

Type III (P->T->P) When was the flag bearer of Rio Olympic born?

```
SELECT answer("Flag Bearer_Info", 'when is this person born?')
FROM "Flag"
WHERE answer("Event year_Info", 'where is this event held?') = 'Rio'
```

Type IV (T&P) Which **male** bearer participated in **Men's 100kg** event in the Olympic game?

```
SELECT "Flag Bearer" FROM "Flag" WHERE "Gender" = 'Male' AND
answer("Flag Bearer_Info",
        'what event did this person participate in?')
        = "Men's 100kg event"
```

HybridQA Questions in SUQL

```
CREATE TABLE Flag (
"Name" TEXT,
"Flag Bearer" TEXT,
"Flag Bearer_Info" TEXT[],
"Gender" TEXT,
"Event year" TEXT,
"Event year_Info" TEXT[]);
```

T: Table
P: Paragraph

Type V (T-Compare | P-Compare)

For the **2012** and 2016 Olympic Event, when was the younger flag bearer born?

```
SELECT MAX
(answer("Flag Bearer_Info",  'when is this person born?')::date)
FROM "Flag" WHERE "Event year" IN ('2016', '2012')
```

Type VI (T-Superlative | P-Superlative) When did the **youngest** Burmese flag bearer participate in the Olympic opening ceremony?

```
SELECT "Event year" FROM "Flag" ORDER BY
answer("Flag Bearer_Info", 'when is this person born?')::date
DESC LIMIT 1;
```

At-A-Glance: HybridQA Questions in SUQL

```
CREATE TABLE Flag (
"Name" TEXT,
"Flag Bearer" TEXT,
"Flag Bearer_Info" TEXT[],
"Gender" TEXT,
"Event year" TEXT,
"Event year_Info" TEXT[]);
```

Type I (T->P)
Where was the XXXI Olympic held?

```
SELECT
answer("Event year_Info",
        'where is this event held?')
FROM "Flag" WHERE "Name" = 'XXXI'
```

Type II (P->T) What was the name of the Olympic event held in Rio?

```
SELECT "Name" FROM "Flag"
WHERE answer("Event year_Info",
'where is this event held?') = 'Rio'
```

Type III (P->T->P) When was the flag bearer of Rio Olympic born?

```
SELECT answer("Flag Bearer_Info",
'when is this person born?')
FROM "Flag"
WHERE answer("Event year_Info",
'where is this event held?') = 'Rio'
```

Type IV (T&P) Which **male** bearer participated in **Men's 100kg** event in the Olympic game?

```
SELECT "Flag Bearer" FROM "Flag"
WHERE "Gender" = 'Male' AND
answer("Flag Bearer_Info",
'what event did this person
participate in?')
= "Men's 100kg event"
```

Type V (T-Compare | P-Compare) For the **2012** and 2016 Olympic Event, when was the younger flag bearer born?

```
SELECT MAX
(answer("Flag Bearer_Info",
  'when is this person born?')::date)
FROM "Flag"
WHERE "Event year" IN ('2016', '2012')
```

Type VI (T-Superlative | P-Superlative) When did the **youngest** Burmese flag bearer participate in the Olympic opening ceremony?

```
SELECT "Event year" FROM "Flag"
ORDER BY answer("Flag Bearer_Info",
'when is this person born?')::date
DESC LIMIT 1;
```

Pros and Cons of SUQL

- **Pros: Formal representation & semantic parsing**
  - Compositionality
  - Domain independence
  - Interpretability
  - Allows query optimization over the whole expression (better than Break-it-down)

- **Cons: It is new – many unknowns**
  - Can LLMs generate the right formal representation?
    - Choosing between the different fields
    - Can it generate complex queries
  - What is the speed?

Conversational Examples

Restaurants

Do you have a recommendation for a first date restaurant in Palo Alto?
We're thinking sushi but not sure what's good around here.

↓

```
SELECT *, summary(reviews) FROM restaurants
WHERE 'sushi' = ANY (cuisines) AND location = 'Palo Alto' AND rating >= 4.0
AND answer(reviews, 'is this restaurant good for a first date?') = 'Yes'
ORDER BY num_reviews DESC LIMIT 1;
```

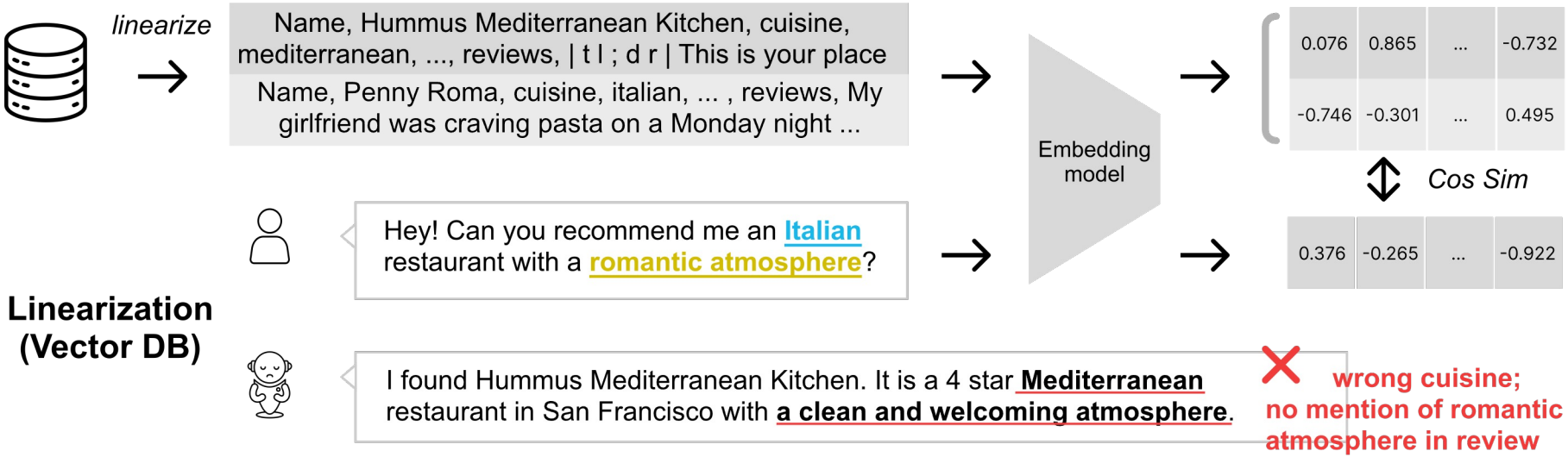
Laptops

I need a laptop with a Thunderbolt 3 port and at least 16GB RAM for my workstation setup.

↓

```
SELECT *, summary(reviews) FROM laptops
WHERE  ram >= 16 AND
(answer(about, 'does this laptop have Thunderbolt 3?') = 'Yes'
OR answer(description, 'does this laptop have Thunderbolt 3?') = 'Yes')
LIMIT 3;
```

Previous Example



Using SUQL

Hey! Can you recommend me an **Italian** restaurant with a **romantic atmosphere**?

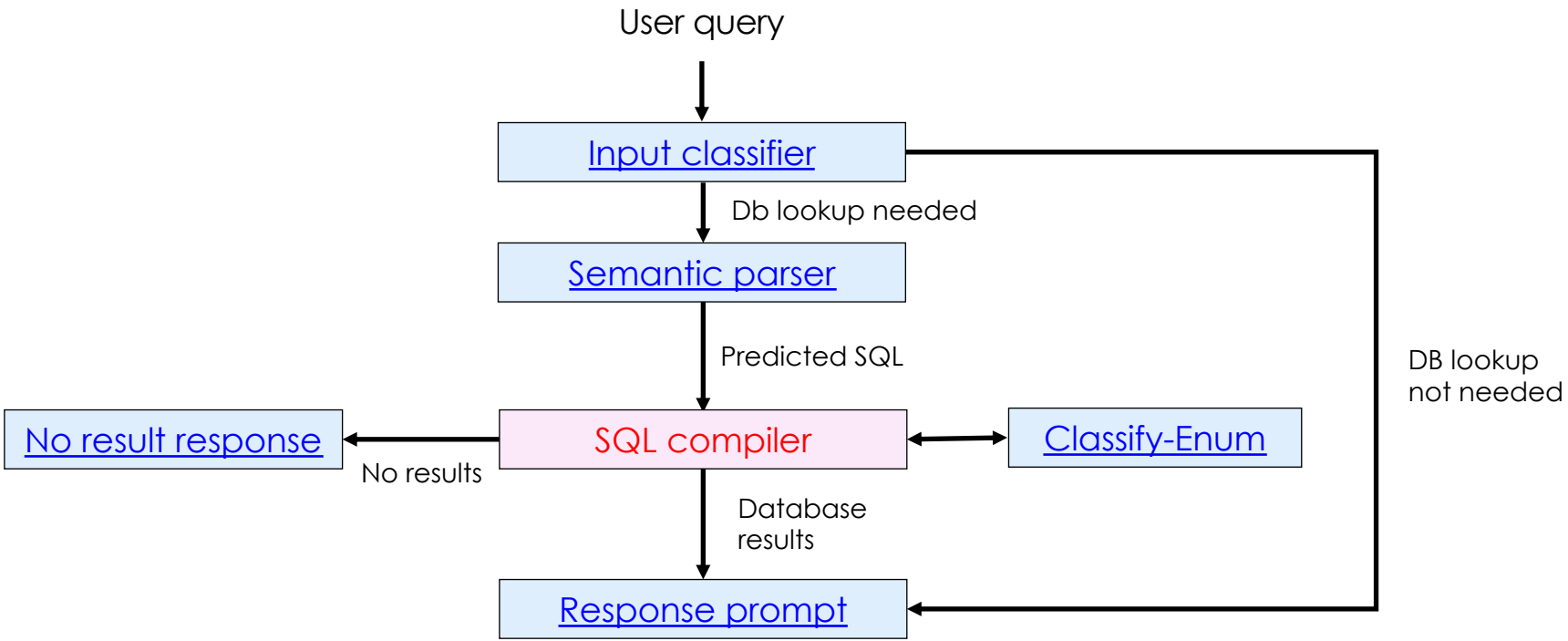

```
SELECT *, summary(reviews) FROM restaurants
WHERE 'italian' = ANY (cuisines) AND
answer(reviews, 'is this restaurant romantic?') = 'Yes'  LIMIT 1;
```

I found Penny Roma, which has a 4.0 rating on our database and offers a variety of **Italian** dishes. Overall, the atmosphere is described as delightful, authentic, and **perfect for a date spot**.

DISCUSSION

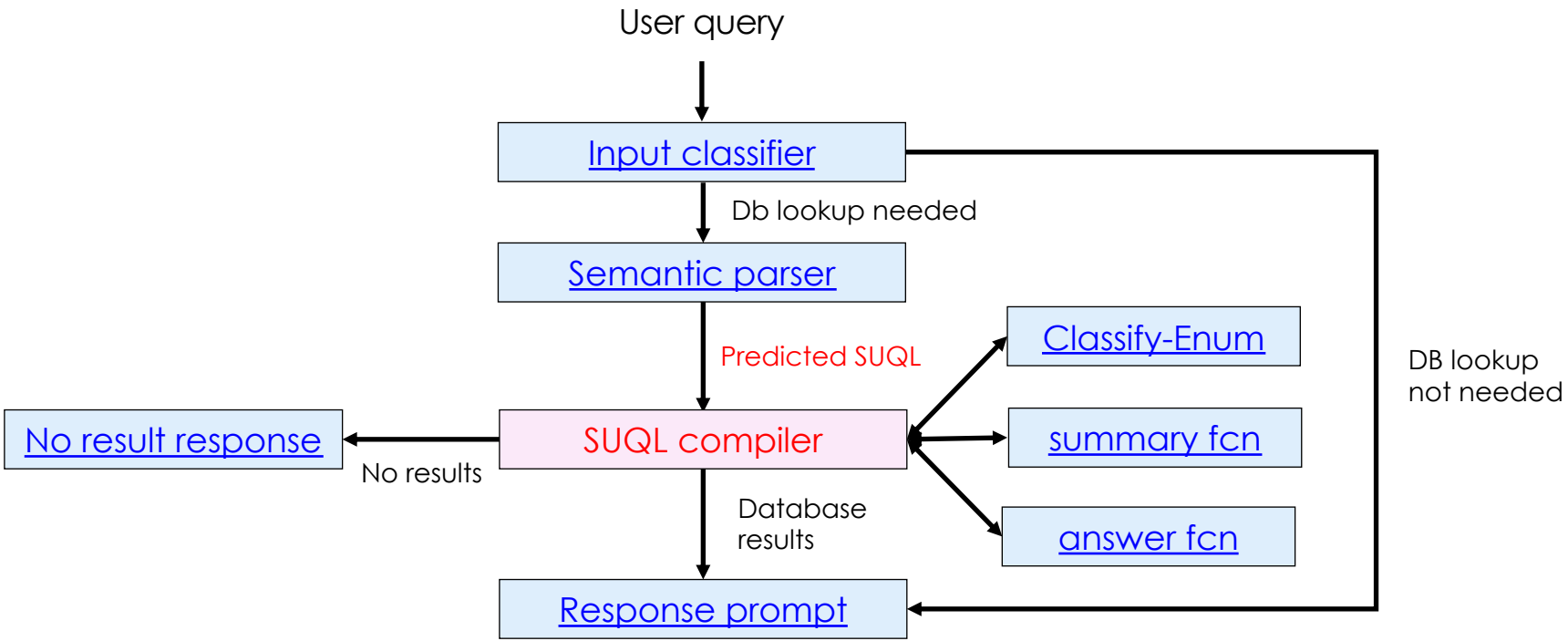
WHAT CAN WE USE SUQL FOR?

Agent Design for SQL



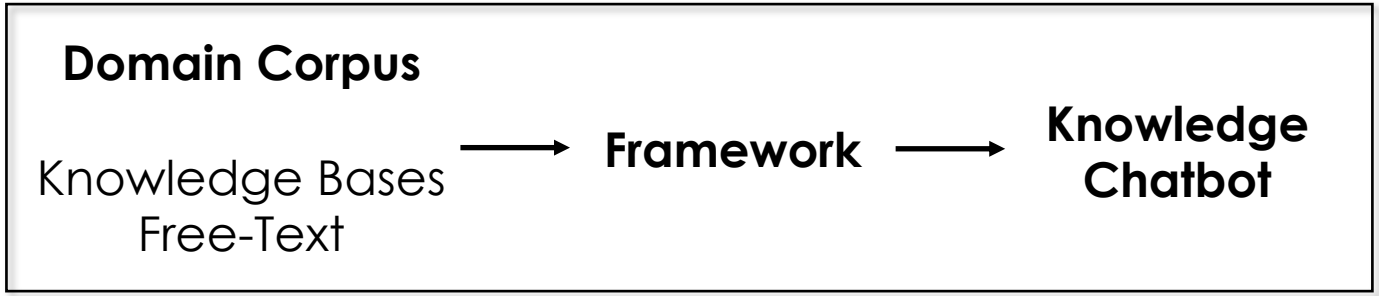
Click the links to see the prompts (written in jinja syntax)

Agent Design: SQL Updated with SUQL



7 prompts with few-shot examples. Templates written in jinja syntax

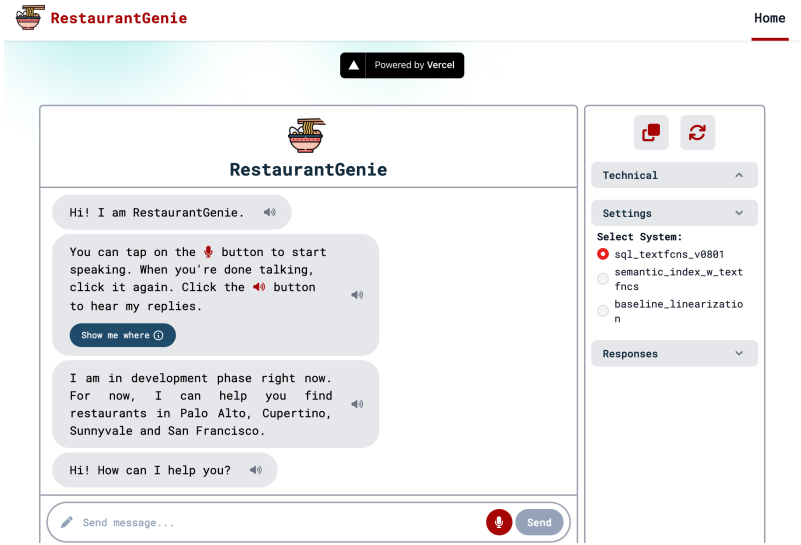
Lecture Goals



- SUQL Language
- SUQL Semantic Parser
- SUQL Performance

Many practical hints for some of your projects!

Experiment



A real, large dataset

- Yelp on San Francisco, Palo Alto, Cupertino, Sunnyvale
- Scraped reviews and popular dishes information

Components

- Open-source agent interface ← You can use this too for your project!
- LLM: gpt-3.5-turbo
- Information retrieval on text fields
  - indexed with Coco-DR
- A new optimizing SUQL compiler
  - SQL can run SUQL programs, but it is too slow

COCO-DR reference: https://arxiv.org/abs/2210.15212

Preliminary Evaluation (in Restaurants)

- Dataset
  - Development: iterate on 50 simulated questions (dev set)
  - Evaluation: 100 real-user questions (test set)
    - 55 of them need SUQL (combo of structured & unstructured)
- Parsing accuracy: 96.0% on dev and 94.0% on test.
- Query *precision*: #correct results / #results

	Dev set (N = 50)	Test set (N = 100)
Linearization (return 3 results)	58.0 %	52.2%
SUQL	**95.6%**	**93.8%**

Quiz: how can you get incorrect results with semantic parsing on SUQL?

What About Recall?

- Are all the answers found?
  - Incomplete
  - Not at all

← Quiz: Is this OK?

Defensive Programming:
What if the Semantic Parser is Wrong?

Assistant fails to find an answer → verbalizes the user query

I searched for 5-star restaurants in Sunnyvale that serve kids food.
Unfortunately, I couldn't find any search results.
Is there anything else I can help you with?

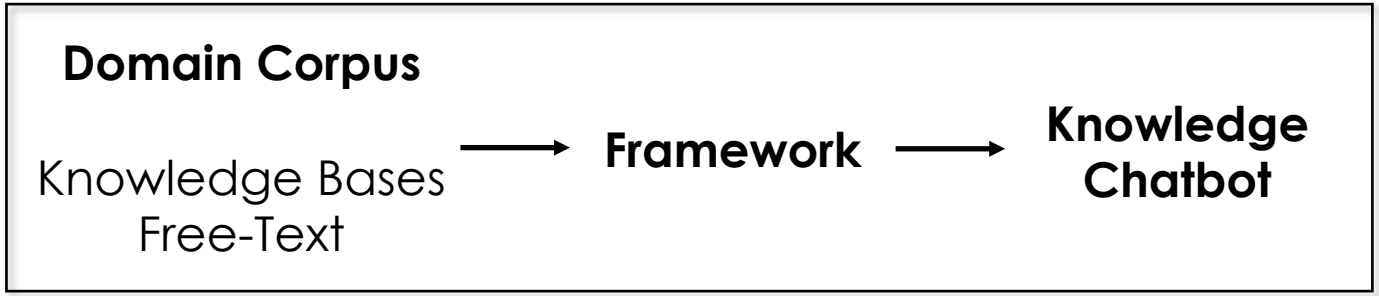
Allows the user to spot the answer

Error Analysis:
What Happened When No Answers Were Returned?

14 false negatives on our test set (100 questions)

- Parsing: 2 errors (one syntactic, one enum field confusion)
- Query evaluation
  - Structured: ← Can be improved
    - 4 Unsupported location service
    - 2 Opening hours errors
    - 2 dishes: searched in popular dishes, but results are in reviews
  - Free-text: 4 false negatives from 'answer' (ChatGPT)

Lecture Goals



- SUQL Language
- SUQL Semantic Parser
- SUQL Performance

Many practical hints for some of your projects!

Performance

Text Search over Entire Corpus is Too Slow

Text Preprocessing

Large Free-Text Corpus

- Speed of query resolution

answer(reviews, ‘is it a family friendly restaurant’) = ‘Yes’

Problem: Takes too long to query over every review

Solution: Create a new field “if_family_friendly” in the database

answer(reviews, ‘is it a family friendly restaurant’) = ‘Yes’
→ if_family_friendly = True

Auto-Creating Databases from Free-Text

1. Use LLM to simulate users' common questions
2. Encourage LLM semantic parser on user questions to generate new fields (with few-shot examples)
3. Stuff the database
   - Give LLM the free-text and the DB form (JSON)
   - Ask LLM to fill in the form

It works surprisingly well!
LLMs are good at hypothesizing fields
LLMs fill in a long form in one LLM call

Performance

Text Search over Entire Corpus is Too Slow

Query Optimization

SUQL Compiler

- SQL can run SUQL programs without modification
  - summary / answer are just external functions

- But it is slow

Query Execution Optimization

Developed an optimizing SUQL compiler
to optimize the execution for all queries!

1. Return only necessary results
2. Order filtering to reduce slow operations
3. Lazy evaluation: produce results only when needed

1. Return Only Necessary Results

answer(reviews, ‘is it a family friendly restaurant’) = ‘Yes’

- For applications such as recommendation, it is not necessary to return all the answers
- IR uses embedding model (vector similarity) to return top candidates
- Return only top results to LLM-based answer functions

2. Order Filtering

answer(reviews, ‘is it a family friendly restaurant’) = ‘Yes’
AND
‘french’ = ANY(cuisines)

- Execution of structured predicates is much cheaper
- Always execute structured predicates first

3. Lazy Evaluation

answer(reviews, ‘is it a family friendly restaurant’) = ‘Yes’
AND ‘french’ = ANY(cuisines) **LIMIT 1**

- Lazy evaluation: Evaluate only when the result is needed
- No need to keep calling answer as soon as LIMIT 1 is reached

SUQL Compiler Overview

- For each SELECT statement with answer in it (begin with bottom node with no sub-queries)
  - Apply optimizations to the SELECT statement
  - Store the processed results in a temporary table *temp*
  - Substitute this statement with *SELECT from temp*
- SQL compiler handles the final processing

Conclusion

SUQL: Extends semantic parsing to hybrid data sources

- Expressiveness: Unifies the hybrid data sources
  - Arbitrary composition (including multi-hop questions on free-text)
  - Automatic extracting DB columns for efficiency
- Uniquely enables query optimization
- In-context learning with LLMs works well for natural queries on small domains in real life

Next challenges

- HybridQA (text+SQL)
  - Complex SQL queries used in research need fine-tuning with synthesized data
- Compmix (Wikipedia+Wikidata)
  - Requires adding tables to free-text and knowledge bases
  - Handing large knowledge bases needs fine-tuning with synthesized data

Consider using SUQL for your project if you have structured/unstructured/multihop problems.