

Bachelor Thesis Defense:

Building Complex Queries in Conversational Search

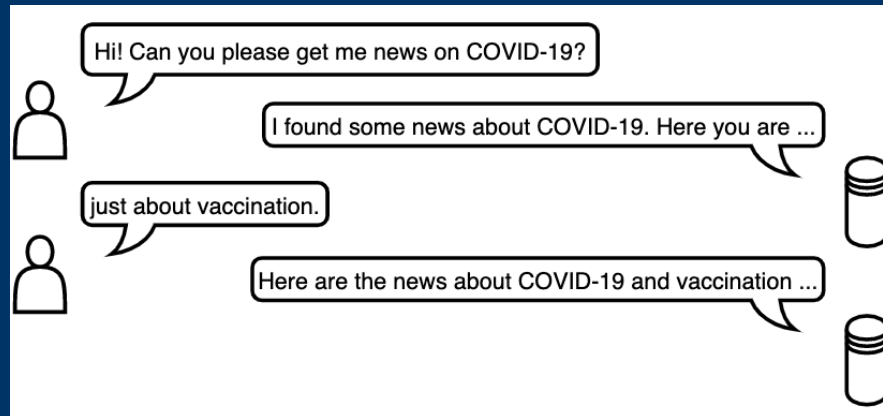
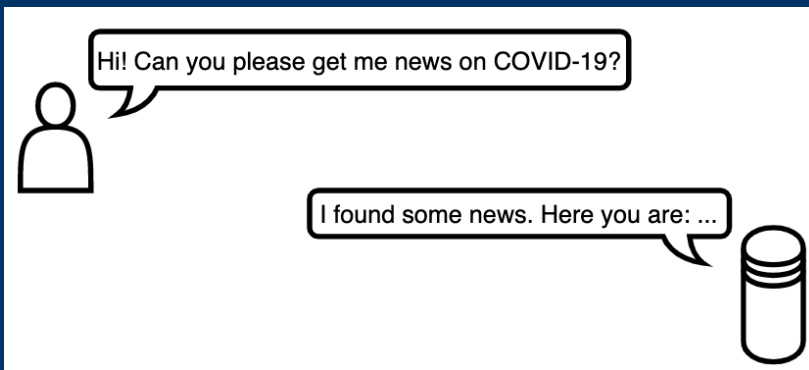
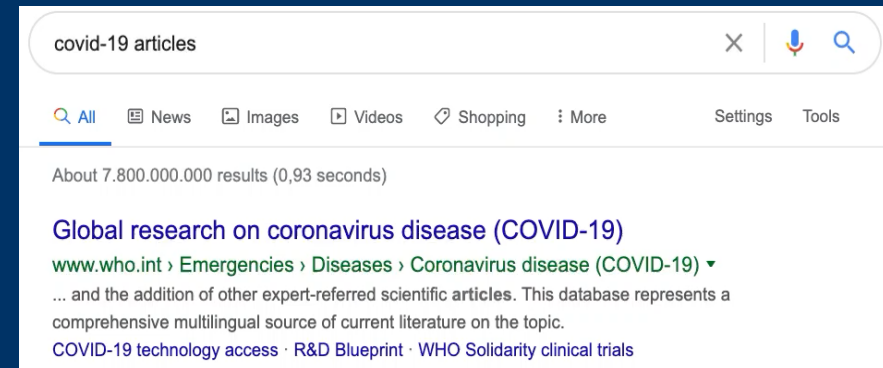
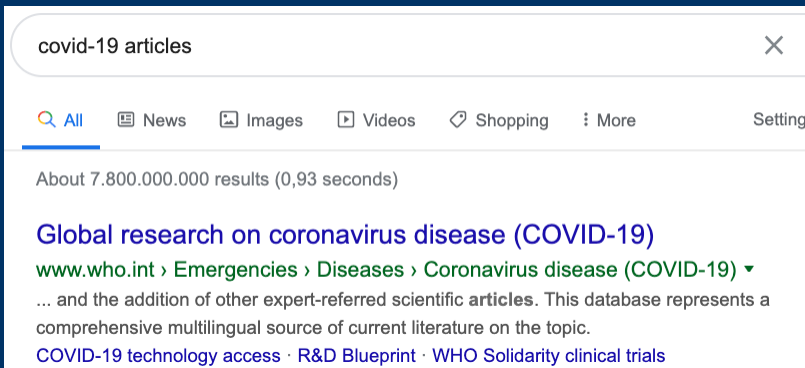
Xiaoni Cai

Advisor: Johannes Kiesel

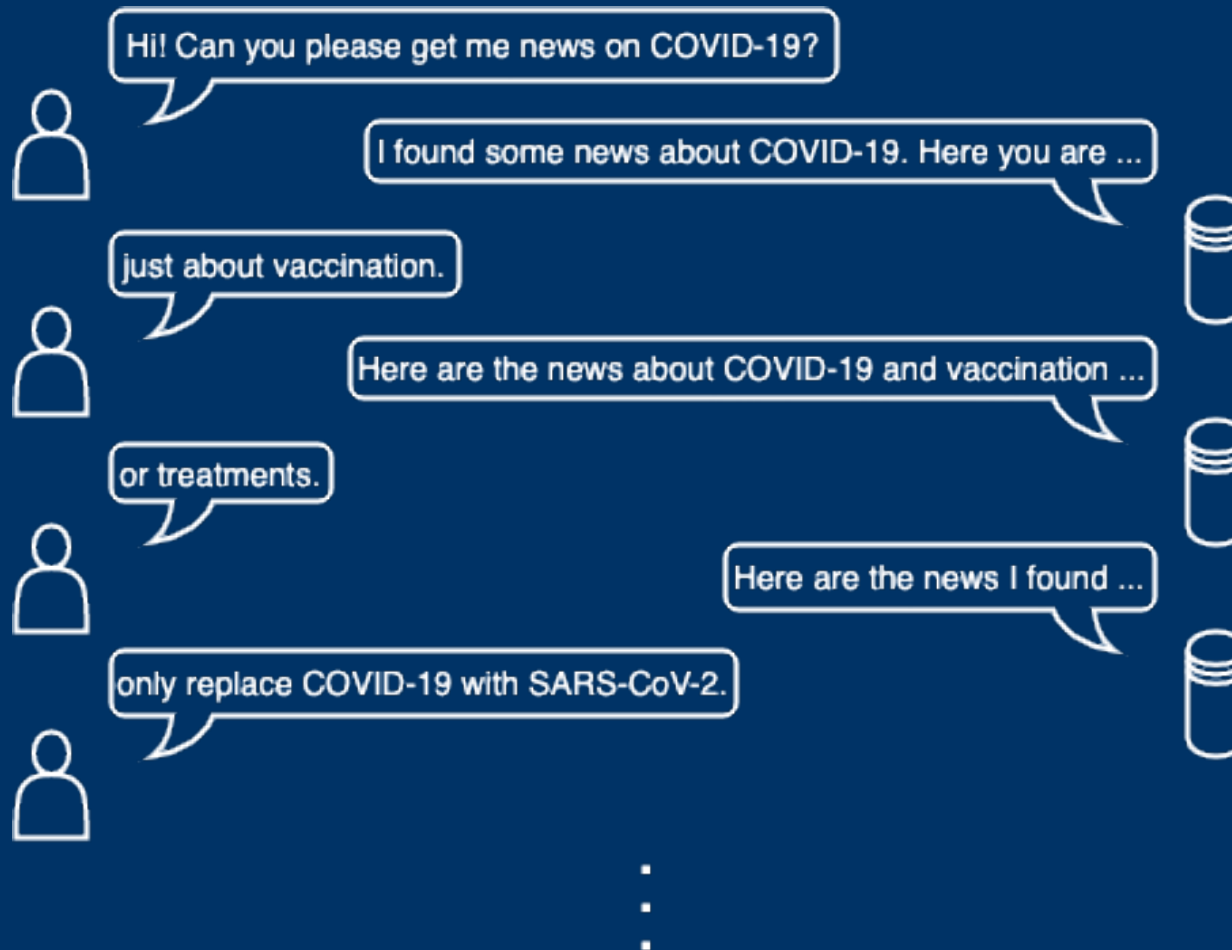
Referees: Prof. Benno Stein, Jr. Prof. Jan Ehlers

Build Queries

Traditional Search v.s. Conversational Search



Complex Queries



Complex Queries

content=SARS-CoV-2 AND (content=vaccination OR content=treatment)

```
{
  "query": {
    "bool": {
      "must": [
        "match": { "content": "SARS-CoV-2" }
      ],
      {
        "bool": {
          "should" : [
            { "match" : { "content" : "vaccination" } },
            { "match" : { "content" : "treatment" } }
          ]
        }
      }
    }
  }
}
```

**How will seekers formulate their queries
while interacting with a system
in a multi-turn conversational search?**

1. Conduct a study to collect human utterances
2. Analyze the collected utterances and recognize patterns
3. Build the interaction model as front-end of a prototype

Crowdsourcing Study

News Bot
online

17:03

infections' or 'the death rate' or other more specific information about the disease. Just ask for news articles about the disease.

But otherwise formulate your own messages as you would for a real chatbot!

You must have good English skills to work on this HIT: If we can not understand your messages or they are not what we asked for, we can not accept your work.

If you understood, tell me that you are ready.

Task 1 of 13

ready ✓

How would you ask me to get you all news articles on the COVID-19 disease (also known as Coronavirus)?

Task 2 of 13

Can you please get me news on COVID-19? ✓

Imagine I would show you the list of all news articles about COVID-19 as you asked for.

You now want the list to contain **fewer** articles: It should contain just the news articles

- about COVID-19 (you already told me)
- that are about a vaccination

How would you ask me to change the list?

Task 3 of 13

just about vaccination|

➤

- Mechanical Turk
- 5 countries (Australia, Canada, India, the United Kingdom, the United States)
- 4 scenarios (argument, book, news, trip)
- 1 'ready' task + 12 query reformulation tasks
- $20 * 4 * 5 = 400$ participants
- $400 * 12 = 4800$ human natural language utterances
- 8 pilot studies for news scenario

Tasks of Crowdsourcing Study

CRUD operations

Operation	Target		
	Query	Part	Literal
Create	<i>Show me news about COVID-19.</i>	<i>Show me news that contain NCD, NI or WHO in the headline.</i>	<i>Is there any news for its treatment?</i>
Read	<i>What do I have so far?</i>	<i>What filters did I set for the headline?</i>	<i>What was the last filter?</i>
Update	<i>Please start a new search on flu.</i>	<i>Remove my criteria for headline and search any news about economy.</i>	<i>I don't mean North Ireland by NI.</i>
Delete	<i>No, let's start again.</i>	<i>Remove the word filters vaccination and treatment.</i>	<i>Remove news about treatment.</i>

Query Reformulation Intent: Operation + Target

- ReadQuery: Memorization & Navigation
- UpdateLiteral: e.g., Negative Feedback
- DeleteQuery: start v.s. restart

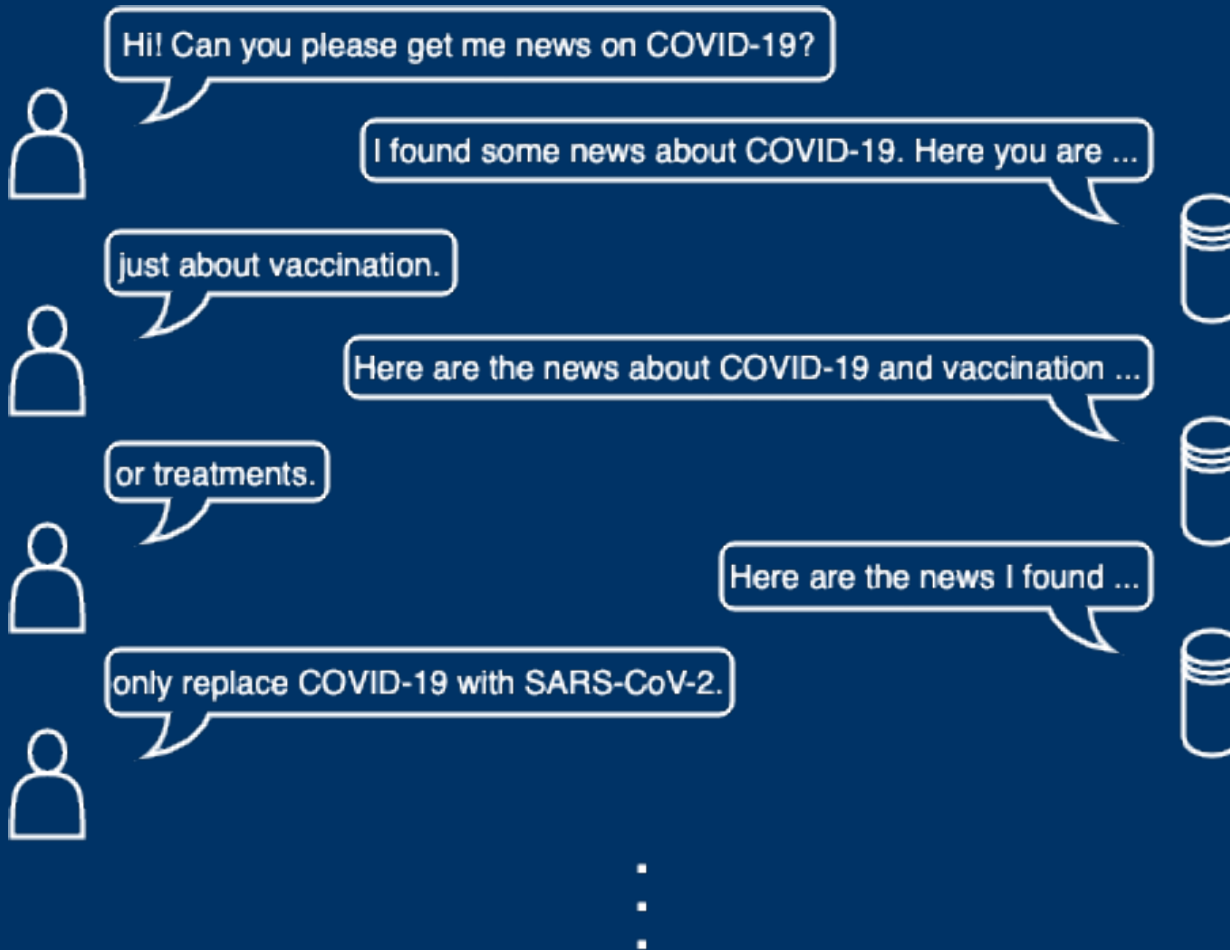
Curation of Crowdsourcing Study

Choose File		news-in-2.csv												Comment
Status	Time	ask for COVID-19	add vaccination	add 'or treatment'	delete two filters(so results should include all articles about virus whether they are related to both filters or not)	specify headline with NCD, NI, WHO	say not North Ireland(indicate the error made by chatbot)	say National Insurance	say not COVID-19 but SARS-CoV-2	ask chatbot to repeat your query	replace NCD, NI, WHO(headline filters) with economy	change to flu	ask flu but not about COVID-19	
reject <input checked="" type="checkbox"/>	11:11	Do you know about all of Covid-19 disease? good <input type="radio"/> bad <input checked="" type="radio"/> very bad <input type="radio"/>	I want to know all about vaccination good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	Can you tell me about treatment of it? good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	tell me about the vaccination and treatment of this disease good <input type="radio"/> bad <input checked="" type="radio"/> very bad <input type="radio"/>	What is suggested by WHO IN HEADLINES and NCD in headline and NI in headline? good <input type="radio"/> bad <input checked="" type="radio"/> very bad <input type="radio"/>	FULL Form of NI IS NORTH IRELAND? good <input type="radio"/> bad <input checked="" type="radio"/> very bad <input type="radio"/>	What do you know about national insurance? good <input type="radio"/> bad <input checked="" type="radio"/> very bad <input type="radio"/>	How COVID-19 is related to SARS-CoV-2 disease? good <input type="radio"/> bad <input checked="" type="radio"/> very bad <input type="radio"/>	I want to know about it good <input type="radio"/> bad <input checked="" type="radio"/> very bad <input type="radio"/>	how this affect economy and what comes in headlines good <input type="radio"/> bad <input checked="" type="radio"/> very bad <input type="radio"/>	You know about flu good <input type="radio"/> bad <input checked="" type="radio"/> very bad <input type="radio"/>	how THIS is related to COVID-19 good <input type="radio"/> bad <input checked="" type="radio"/> very bad <input type="radio"/>	THANKS
reject <input type="checkbox"/>	36:56	Show the news about COVID-19 good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	show me the news about vaccination good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	show me the treatment available good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	show me the news articles that are not on vaccination or treatment. good <input type="radio"/> bad <input checked="" type="radio"/> very bad <input type="radio"/>	show me the articles that contain NCD or NI or Who as headline. good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	do not show me the articles that contain North Ireland instead it should contain NI only. good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	show me the articles that discusses about National Insurance in the disease. good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	show me the articles about SARS-Cov-2 instaed of COVID-19 which should contain at least NCD or NI means National Insurance or WHO as headline. good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	show me the details about my previous search. good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	Show me articles about the economy that does not have NCD or NI or WHO as headline. good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	show me the articles about flu. good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	show me only those articles that are not about COVID-19 good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	na
reject <input type="checkbox"/>	12:35	Hey, show me news articles on COVID-19 pandemic good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	show me article having details about vaccination good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	filter those article having detail about treatment good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	now filter article which doesn't have vaccination or treatment details good <input type="radio"/> bad <input checked="" type="radio"/> very bad <input type="radio"/>	display article which must have NCD or NI or WHO in the headline good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	hey you misunderstood NI doesn't means North Ireland good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	so correct it's National Insurance good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	Now show me articles about SARS-CoV2 rather than COVID-19 good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	show me all previous filters good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	Filter me articles about the economy as headline good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	show me news articles about flu good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	filter that don't have COVID-19 good <input checked="" type="radio"/> bad <input type="radio"/> very bad <input type="radio"/>	{}

- News scenario: 5 countries (AU, CA, GB, IN, US)
- Argument, news and book scenarios: 3 countries (CA, GB, US)
- Number of approved participants: 284
- Three categories: “good”, “bad”, “very bad”
- 2919 “good” utterances (85.65%), 1434 patterns

Analysis of Crowdsourcing Study

Ambiguity



Ambiguity

A. mix up with other tasks due to the existence of overlapping patterns

- Task 2 (createQuery) v.s. Task 3 (createLiteral) v.s. Task 4 (updatePart)
- Task 7 (rejectLiteral) v.s. Task 13 (createNegLiteral)
- Task 2 (createQuery) v.s. Task 12 (updateQuery)

B. misunderstood by participants

- Task 5 (deletePart)
- Task 11 (updatePart)

Ambiguity

A. mix up with other tasks due to the existence of overlapping patterns

How would you ask me to get you all news articles on the COVID-19 disease (also known as Coronavirus)?

Imagine I would show you the list of all news articles about COVID-19 as you asked for.

You now want the list to contain **fewer** articles: It should contain just the news articles

- about COVID-19 (you already told me)
- that are about a vaccination

How would you ask me to change the list?

Assume I changed the list accordingly.

You now want the list to contain **more** articles: It should contain the news articles

- about COVID-19 (you already told me)
- that are about either:
 - vaccination (you already told me), or
 - treatment (as a new alternative to vaccination).

How would you ask me to change the list?

Task 2 Create(*COVID-19*): start a search for main-topic *COVID-19*

Task 3 Create(*vaccination*): add subtopic *vaccination* to filter previously obtained results

- *just show me news about vaccination.*

⇒ Create(*vaccination*) OR Update(*COVID-19, vaccination*)

- *show me news that are only about vaccination.*

⇒ Update(*COVID-19, vaccination*) OR Update (*COVID-19, COVID-19 V vaccination*)

Task 4 Update(*vaccination, vaccination V treatment*): add another subtopic *treatment* as an alternative to *vaccination*

- *please also add news about treatment to the list.*

⇒ Update(*vaccination, vaccination V treatment*) OR Create(*treatment*)

- *show me news about treatment.*

⇒ Update(*vaccination, vaccination V treatment*) OR Update(*COVID-19 ∧ vaccination, treatment*)

Unambiguous patterns for Task 3

1. Use pronouns (co-reference):
 - can you [only|just] [show|give] [me|] [ones|those] [about|on] vaccination?
 - which [of|] [these|ones] [include|relate to] vaccination?
2. Use verbs like filter, trim down, narrow down, reduce, shorten
 - can you [filter|shorten|trim down|reduce] list to [those|ones|] about vaccination?
 - filter [these|this list] [with|for] vaccination [only|].
3. Elimination
 - [please|] [remove|filter out] {collection} that are not [relate to|about|] vaccination.

Ambiguity

B. misunderstood by participants

- Task 5 (deletePart)

Assume I changed the list accordingly.

You now want the list to contain **more** articles: It should again contain all news articles

- about COVID-19 (you already told me, but now again with those news articles that are not about vaccination or treatment)

How would you ask me to change the list?

Task 5a delete(*vaccination* \vee *treatment*): remove part of filters

- *add back all news that are not about vaccination or treatment to the list.*
- *can you show me all news not just the ones about vaccination or treatment?*
- *please remove filters vaccination and treatment.*

Task 5b update(*vaccination* \vee *treatment*, \neg *vaccination* \vee \neg *treatment*):
update part of filters to negation of filters

- *now please show me news that are not about vaccination or treatment.*
- *please remove all news about vaccination or treatment.*

Interaction Model

Task	Intent	Utterances	Slots	Descriptions
2	CreateQueryIntent	274	2	Seekers start a search session for collection about a main-topic.
3	CreateLiteralIntent	82	2	Seekers add a subtopic to filter previously obtained result list.
4	UpdatePartIntent	55	3	Seekers add an alternative subtopic.
5	DeletePartIntent	70	3	Seekers remove part of filters.
6	CreatePartIntent	262	5	Seekers specify certain field with part of filters.
7	RejectLiteralIntent	155	4	Seekers indicate a error made by the system and reject the unexpected filter.
9	UpdateLiteralIntent	258	4	Seekers replace a existing filter with a new filter.
10	ReadQueryIntent	189	1	Seekers ask for system to recall the search history.
11	UpdatePartFieldIntent	133	6	Seekers replace part of filters for certain field with another new filter.
12	UpdateQueryIntent	140	2	Seekers replace the entire query with a new filter.
13	CreateNegLiteralIntent	263	2	Seekers add a negation of existing filter.

- 11 custom Intents
- Max. 6 custom Slot Types
- Replace patterns with human annotations as sample utterances
- Bad generalizability for different scenarios

Front-end of Prototype

Evaluation

[< Go Back](#)

NLU Evaluation / 37140832-3f... Source: TASK2

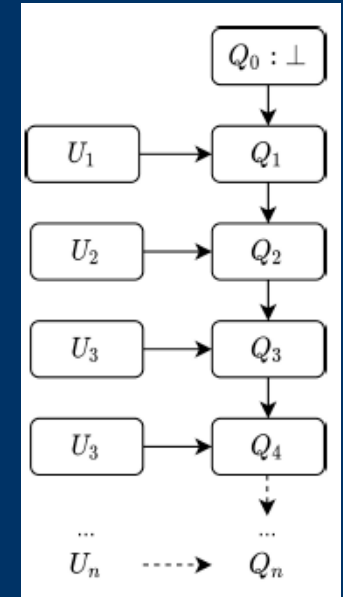
0% Passed () Failed (275) Total Test Cases: 275

ASK_SAA_EVALUATION_RESULTS_ENDED_LABEL
17 minutes ago in **DEVELOPMENT**
Sep 12, 2020 7:48 PM

#	UTTERANCE	EXPECTED [E] VS ACTUAL [A] INTENT
256	Show me all trips to San Jose	E: createQuery, A: updateQuery
257	Show me trips to San Jose	E: createQuery, A: updateQuery
258	list trips to San Jose	✓ createQuery
259	Can you open up all trips to San Jose	E: createQuery, A: updateQuery
260	I want you to get all trips to San Jose	E: createQuery, A: updateQuery
261	Can you find me all trips to San Jose please.	✓ createQuery
262	Trips to San Jose.	E: createQuery, A: confirmLiteral
263	Show me trips to San Jose.	E: createQuery, A: updateQuery

- Prove the existence of ambiguous patterns (explicit)
- Figure out the implicit ambiguity

- Implement back-end of the prototype
 - Detect ambiguity of human utterances
 - Minimize ambiguity
 - Check correlation between different filters in the documents (semantic, occurrence etc.)
 - Reasoning & Memorization (e.g. Negative Feedback)
- Follow-up studies
 - Test for resolving ambiguity
 - Test for prototype



Thank you for your attention!

Question Time