Toward Conversational Query Reformulation

DESIRES 2021


Johannes Kiesel[1]


Xiaoni Cai[1]


Roxanne El Baff[2]


Benno Stein[1]


Matthias Hagen[3]

Bauhaus-Universität Weimar [1]

DLR [2]

MARTIN-LUTHER-UNIVERSITÄT HALLE-WITTENBERG [3]

Webis [1,2,3]

What are Complex Queries?

From all the rivers in Mexico, but not New Mexico, starting with the letter E, which one passes closest to a city that has at least 100,000 inhabitants and is at least 200 km from the American border at the closest point and does not have a name that exists in an American or European city as well?

— Nadja and Benjamin, Yesterday

What are Conversational Query Reformulations?

News

COVID-19

COVID-19 surge slows travel industry's recovery

PBS NewsHour · 23 minutes ago

With cases surging, Biden to launch six-point plan against COVID-19

Reuters · 23 minutes ago

COVID-19 surge in the US: The summer of hope ends in gloom

Associated Press · 17 hours ago

What are Conversational Query Reformulations?

News COVID-19

COVID-19 surge slows travel industry's recovery
PBS NewsHour · 23 minutes ago

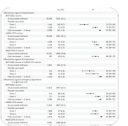
With cases surging, Biden to launch six-point plan against COVID-19
Reuters · 23 minutes ago

COVID-19 surge in the US: The summer of hope ends in gloom
Associated Press · 17 hours ago

→

News COVID-19 vaccination

Effectiveness of Covid-19 Vaccines in Ambulatory and Inpatient Care Settings
nejm.org · 17 hours ago

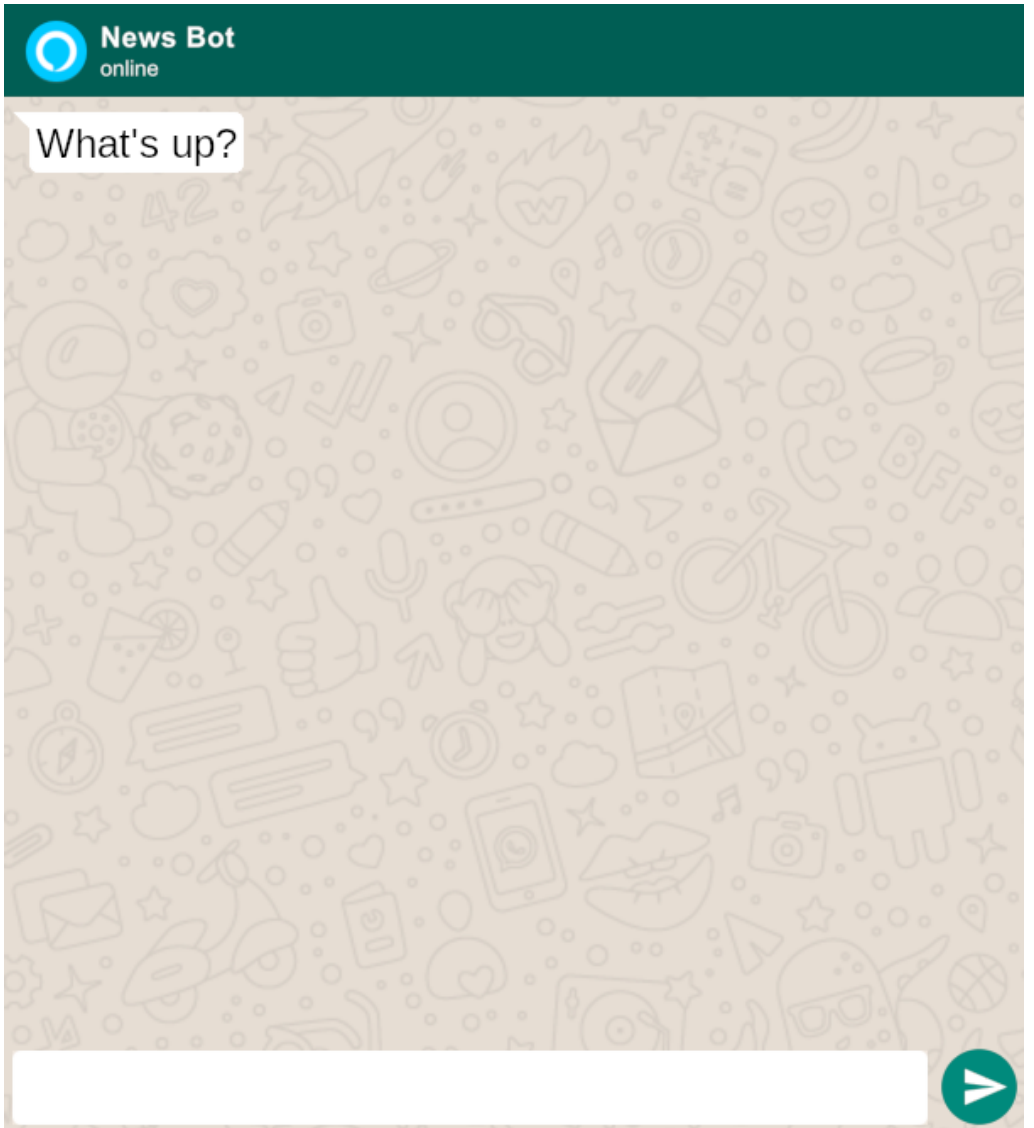


How, When and Where Older Adults Can Get COVID-19 Vaccines
AARP · 16 hours ago

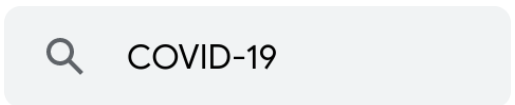
- Studies offer reassuring data on the safety of COVID-19 vaccination for breastfeeding women
  News-Medical.Net · 1 hour ago

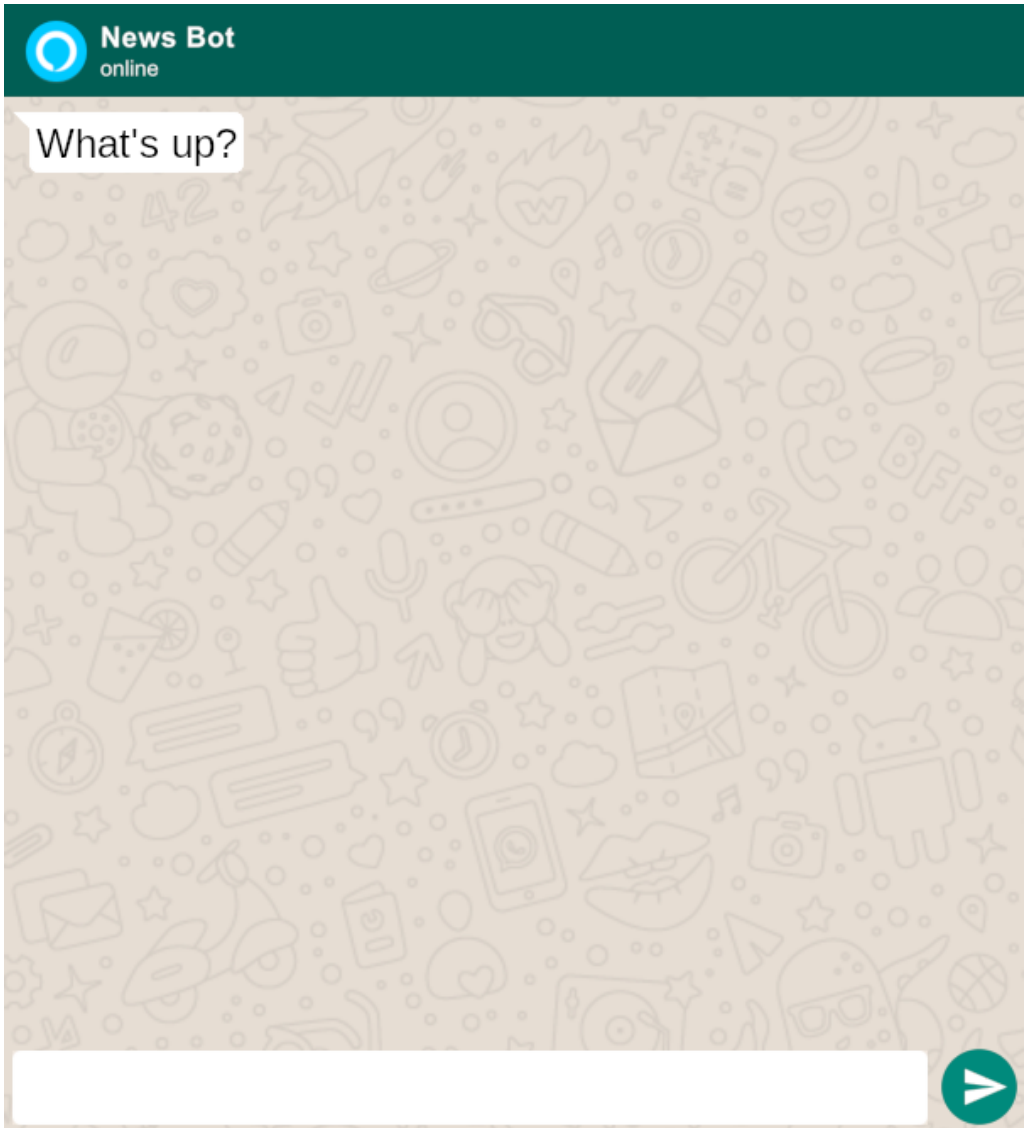
View Full Coverage

What are Conversational Query Reformulations?

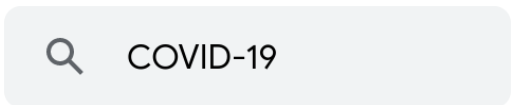


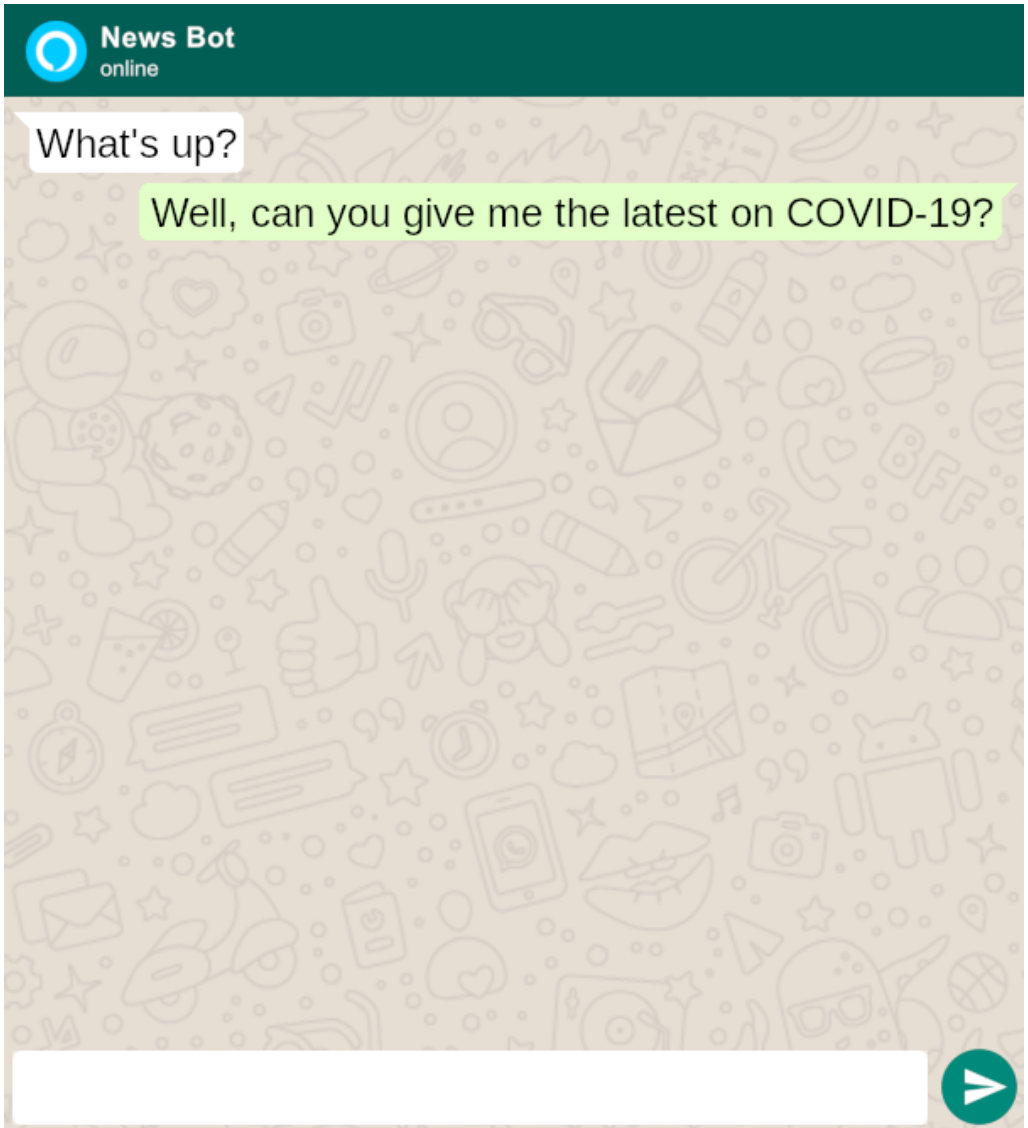
What are Conversational Query Reformulations?





What are Conversational Query Reformulations?

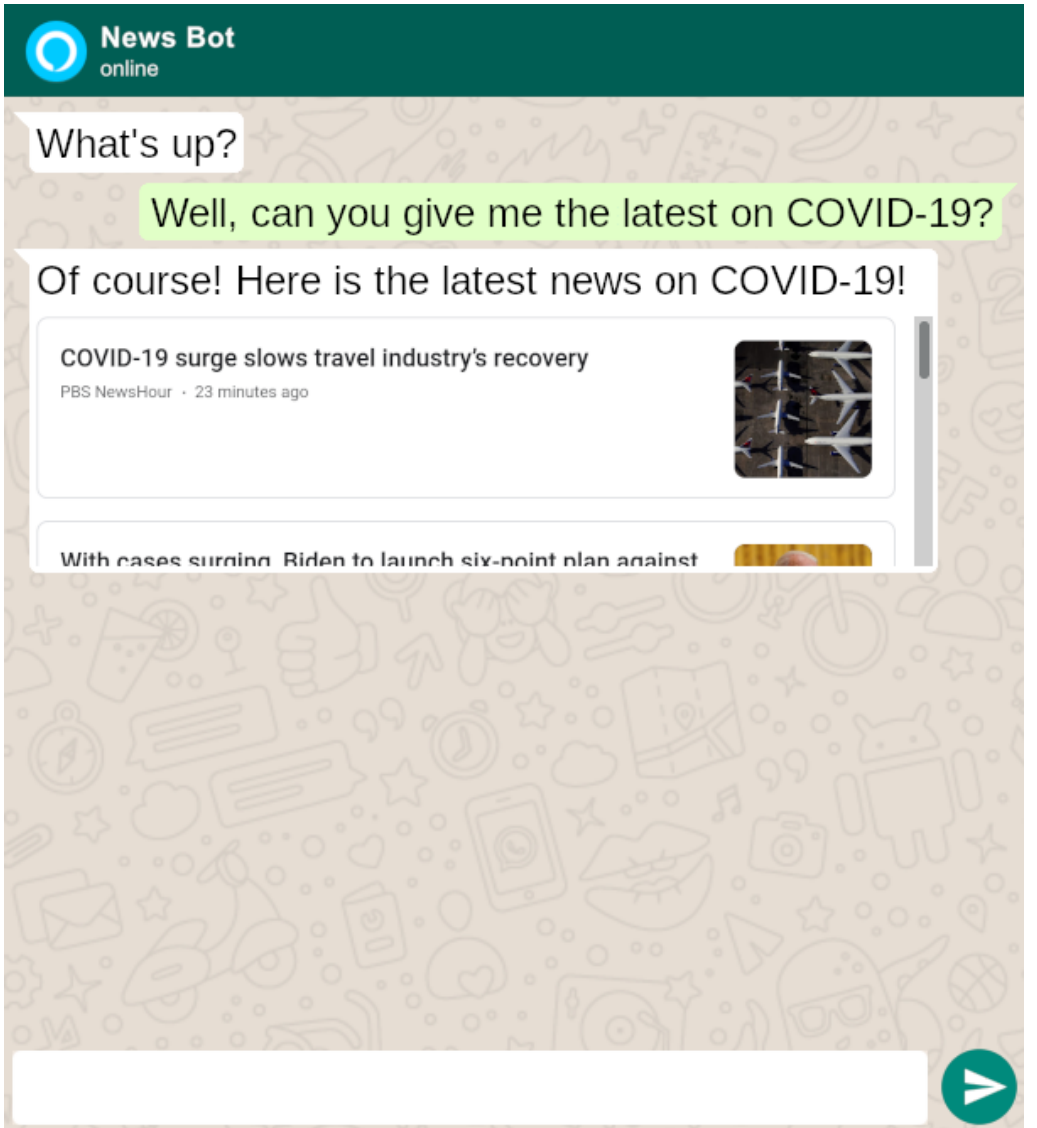




What are Conversational Query Reformulations?

COVID-19

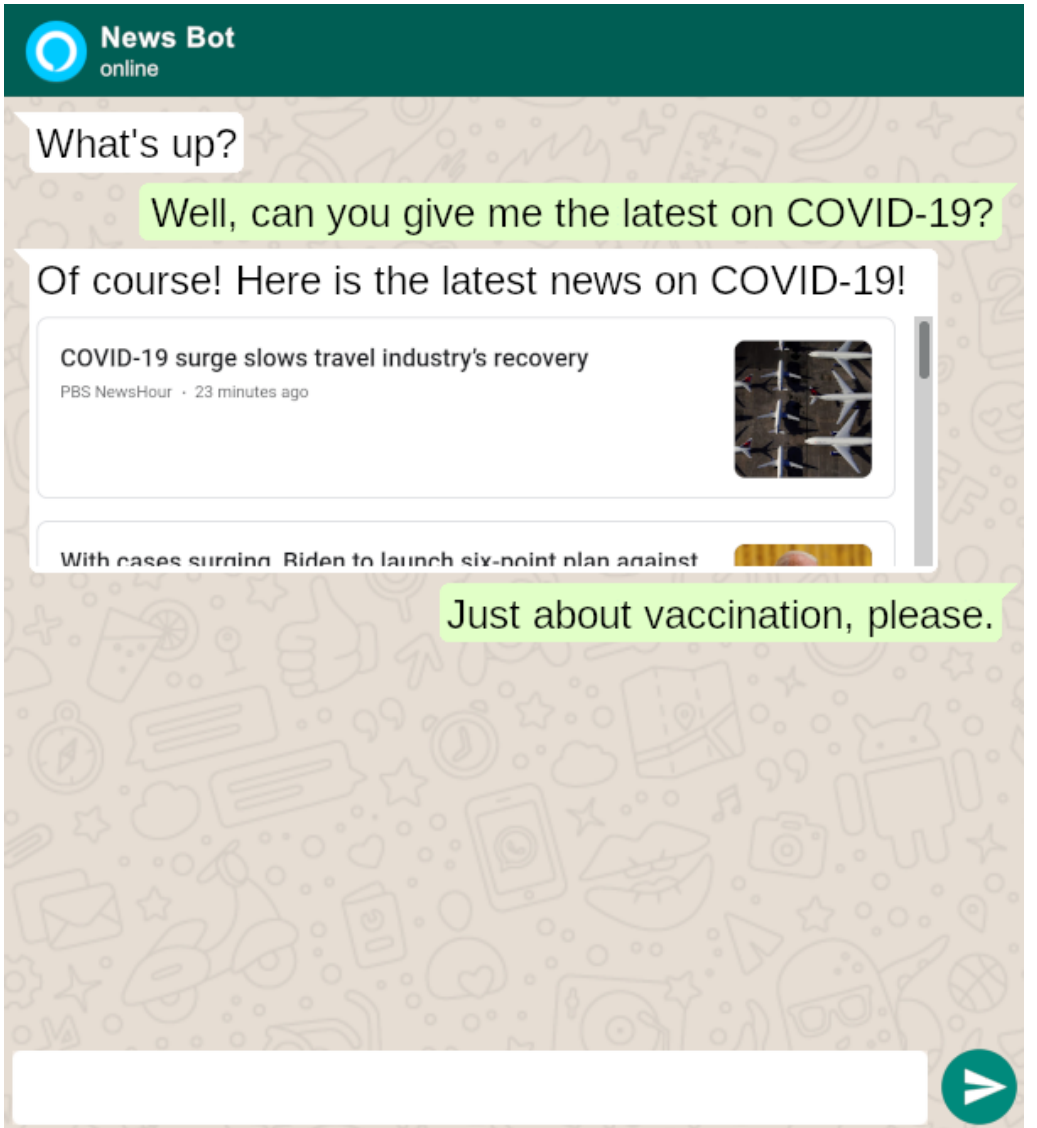
COVID-19 vaccination



What are Conversational Query Reformulations?

COVID-19

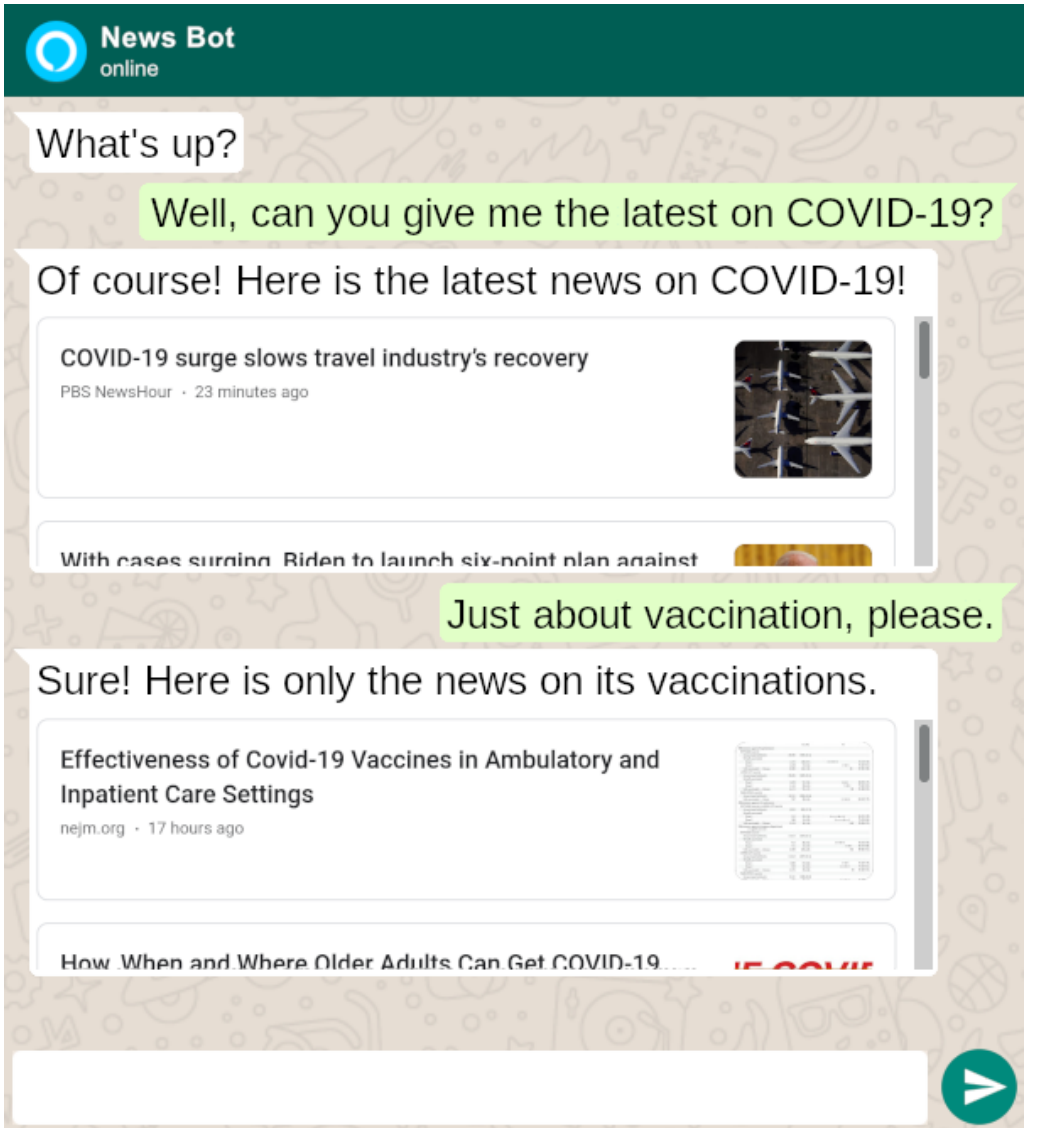
COVID-19 vaccination



What are Conversational Query Reformulations?

COVID-19

COVID-19 vaccination



What are Conversational Query Reformulations?

From all the rivers in Mexico, but not New Mexico, starting with the letter E, which one passes closest to a city that has at least 100,000 inhabitants and is at least 200 km from the American border at the closest point and does not have a name that exists in an American or European city as well?

— Nadja and Benjamin, Yesterday

What are Conversational Query Reformulations?

From all the rivers in Mexico, but not New Mexico, starting with the letter E, which one passes closest to a city that has at least 100,000 inhabitants and is at least 200 km from the American border at the closest point and does not have a name that exists in an American or European city as well?

— Nadja and Benjamin, Yesterday

I'm looking for a river in Mexico.

But not in New Mexico.

Its name starts with the letter E.

It is the one that passes closest to some city.

Only cities with at least 100,000 inhabitants count.

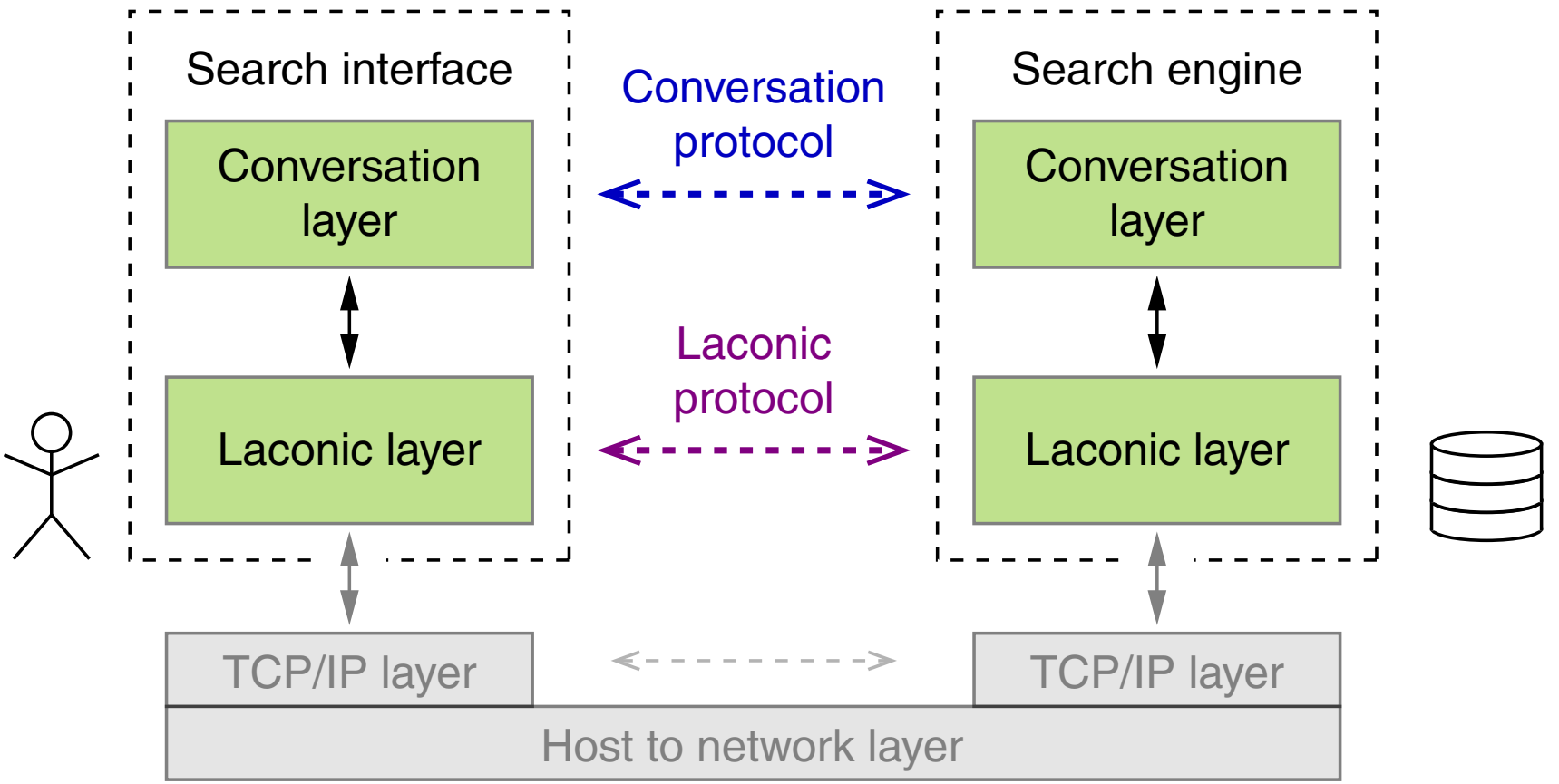
The river's closes point to the American border is at least 200 km away from it.

The river's name exists in no American or European city.

How to Deal with Conversational Query Reformulations?

Just about vaccination, please

How to interpret this message as a search engine?



How to Deal with Conversational Query Reformulations?

Just about vaccination, please

How to interpret this message as a search engine?

Approach 1: Query rewriting.

Make the query context explicit by modifying the query.

Well, can you give me the latest on COVID-19 just about vaccination, please

Approach 2: Interpreting as conversational query reformulation.

Identify the intended operation on a latent "context query."

Specification: add term "vaccination"

Basic query reformulation operations:

- Add a term (specification)
- Remove a term (generalization)
- Change/replace a term

→ **CRUD** (**C**reate, **R**ead, **U**pdate, **D**elete), the basic operations of data systems

→ "Meta-queries"

How to Deal with Conversational Query Reformulations?

Just about vaccination, please

How to interpret this message as a search engine?

Approach 1: Query rewriting.

Make the query context explicit by modifying the query.

Well, can you give me the latest on COVID-19 just about vaccination, please

Approach 2: Interpreting as conversational query reformulation.

Identify the intended operation on a latent "context query."

Specification: add term "vaccination"

Basic query reformulation operations:

- Add a term (specification)
- Remove a term (generalization)
- Change/replace a term

→ **CRUD** (**C**reate, **R**ead, **U**pdate, **D**elete), the basic operations of data systems

→ "Meta-queries"

How to Deal with Conversational Query Reformulations?

Just about vaccination, please

How to interpret this message as a search engine?

Approach 1: Query rewriting.
Make the query context explicit by modifying the query.

Well, can you give me the latest on COVID-19 just about vaccination, please

Approach 2: Interpreting as conversational query reformulation.
Identify the intended operation on a latent "context query."

Specification: add term "vaccination"

Basic query reformulation operations:

- Add a term (specification)
- Remove a term (generalization)
- Change/replace a term

→ **CRUD** (**C**reate, **R**ead, **U**pdate, **D**elete), the basic operations of data systems
→ "Meta-queries"

How do People Perform Conversational Query Reformulations?

MTurk study

- 284 participants (AU/CA/UK/IN/US; arguments/books/news/trip)
- 12 tasks each (CRUD, same between domains)

⇒ 2694 messages (after quality control, publicly available)

How do People Perform Conversational Query Reformulations?

Task 2: $l_1 \rightarrow l_1 \wedge l_2$

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?

- ❑ Filter list to just about vaccination.
- ❑ Show me articles that mention vaccination.
- ❑ Please filter out all articles that are not about vaccination.
- ❑ Remove all articles which do not cover the subject of vaccination.
- ❑ Only show me articles about vaccination.
- ❑ Tell me only about the vaccination related.
- ❑ Which of these include vaccination?

How do People Perform Conversational Query Reformulations?

Task 2: $l_1 \rightarrow l_1 \wedge l_2$

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?

- Filter list to just about vaccination.
- Show me articles that mention vaccination. ($\rightarrow l_2$)?
- Please filter out all articles that are not about vaccination.
- Remove all articles which do not cover the subject of vaccination.
- Only show me articles about vaccination. ($\rightarrow l_2$)?
- Tell me only about the vaccination related.
- Which of these include vaccination?

64% could also ask for l_2 (starting a new query)

How do People Perform Conversational Query Reformulations?

Task 2: $l_1 \rightarrow l_1 \wedge l_2$

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?

- Filter list to just about vaccination.
- Show me articles that mention vaccination. ($\rightarrow l_2$)? ($\rightarrow l_1 \vee l_2$)?
- Please filter out all articles that are not about vaccination.
- Remove all articles which do not cover the subject of vaccination.
- Only show me articles about vaccination. ($\rightarrow l_2$)?
- Tell me only about the vaccination related.
- Which of these include vaccination?

64% could also ask for l_2 (starting a new query)
35% could also ask for $l_1 \vee l_2$ (OR instead of AND)

How do People Perform Conversational Query Reformulations?

Task 3: $l_1 \wedge l_2 \rightarrow l_1 \wedge (l_2 \vee l_3)$

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?

- Or treatment.
- Vaccination or treatment articles.
- I also want to see treatments.
- Can you get me more news on vaccination or treatment?
- Alter previous filter and keep articles containing the word treatment.
- List can also include treatment options for the disease.

How do People Perform Conversational Query Reformulations?

Task 3: $l_1 \wedge l_2 \rightarrow l_1 \wedge (l_2 \vee l_3)$

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?

- Or treatment. ($\rightarrow (l_1 \wedge l_2) \vee l_3$)?
- Vaccination or treatment articles.
- I also want to see treatments. ($\rightarrow (l_1 \wedge l_2) \vee l_3$)?
- Can you get me more news on vaccination or treatment?
- Alter previous filter and keep articles containing the word treatment.
- List can also include treatment options for the disease. ($\rightarrow (l_1 \wedge l_2) \vee l_3$)?

72% could also ask for $(l_1 \wedge l_2) \vee l_3$ (unclear precedence)

To Take Away

- How to specify complex queries? Problem solved!
- Conversational query reformulations (you heard it here first)
- But understanding reformulations is hard
  - The main problem may not be the large variety of formulations
  - But rather their ambiguity (solutions: clarification questions? teaching?)
  - Likely feasible: domain-independent reformulation module

Further implications:

- More search operators (e.g., negation, phrases, boosting)
- Personalization (query feeds)
- . . . ?