

# What's The Latest?

## A Question Driven News Chatbot

Philippe Laban - [phillab@berkeley.edu](mailto:phillab@berkeley.edu)  
John Canny and Marti Hearst



**Berkeley**  
UNIVERSITY OF CALIFORNIA

# Talk organization

Part 1

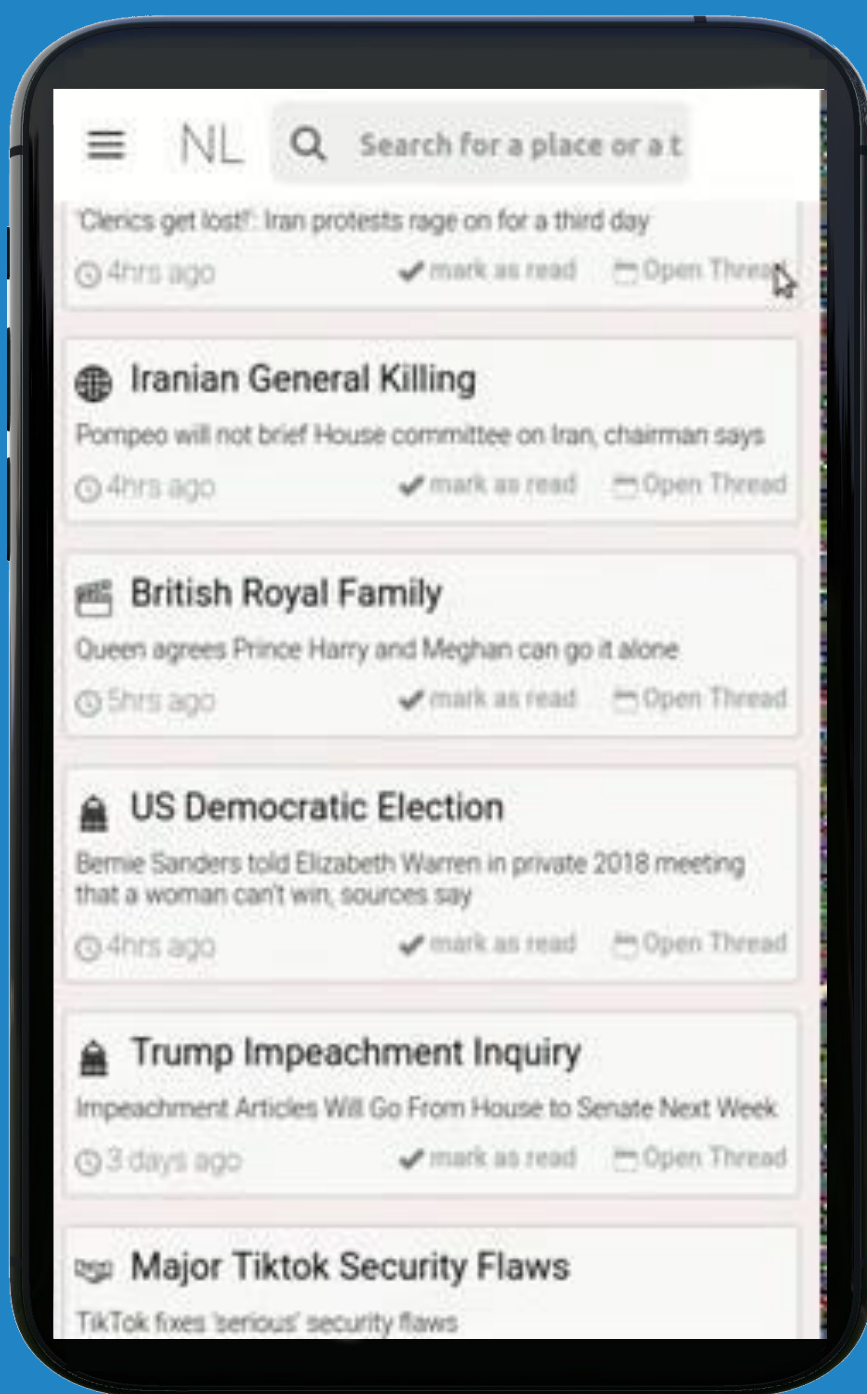
Demo

Part 2

Conversation State  
Tracking Method

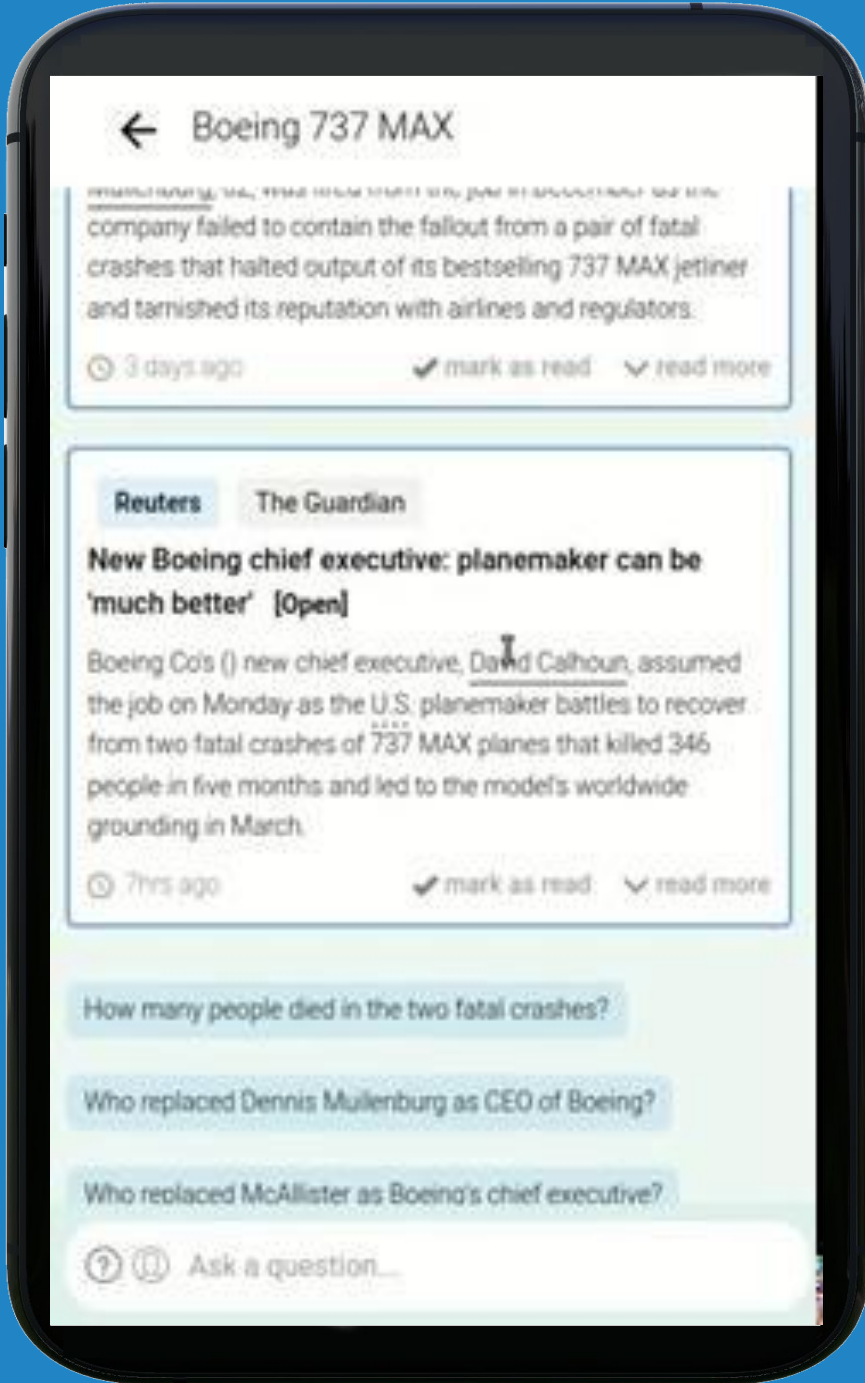
Part 3

User Study  
And Results



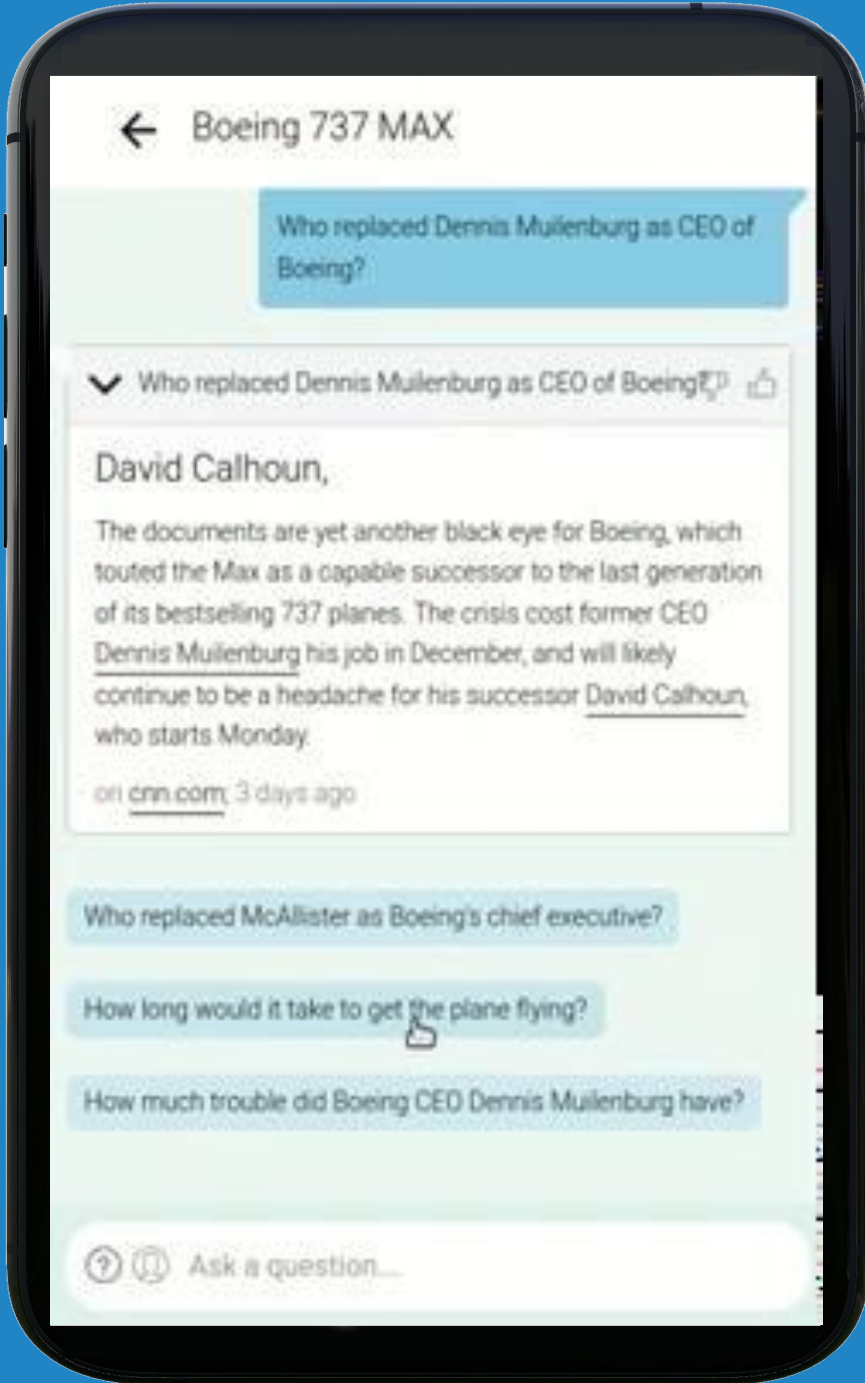
News is organized into stories, each forming a chatroom.

Example stories:  
Iran Plane Crash,  
Boeing 737 MAX...



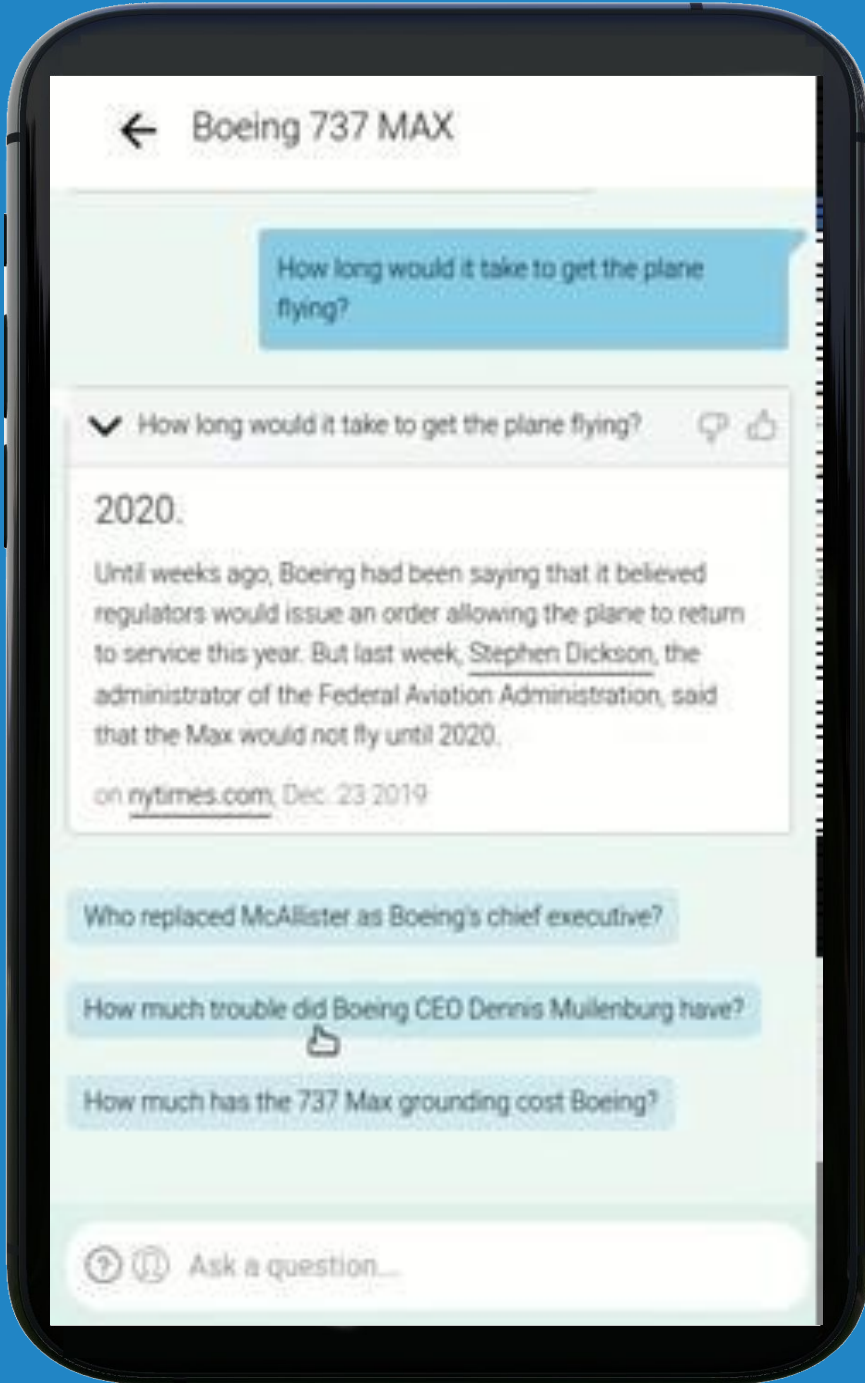
A chatroom starts with a timeline of events.

Each event is composed of multiple sources.

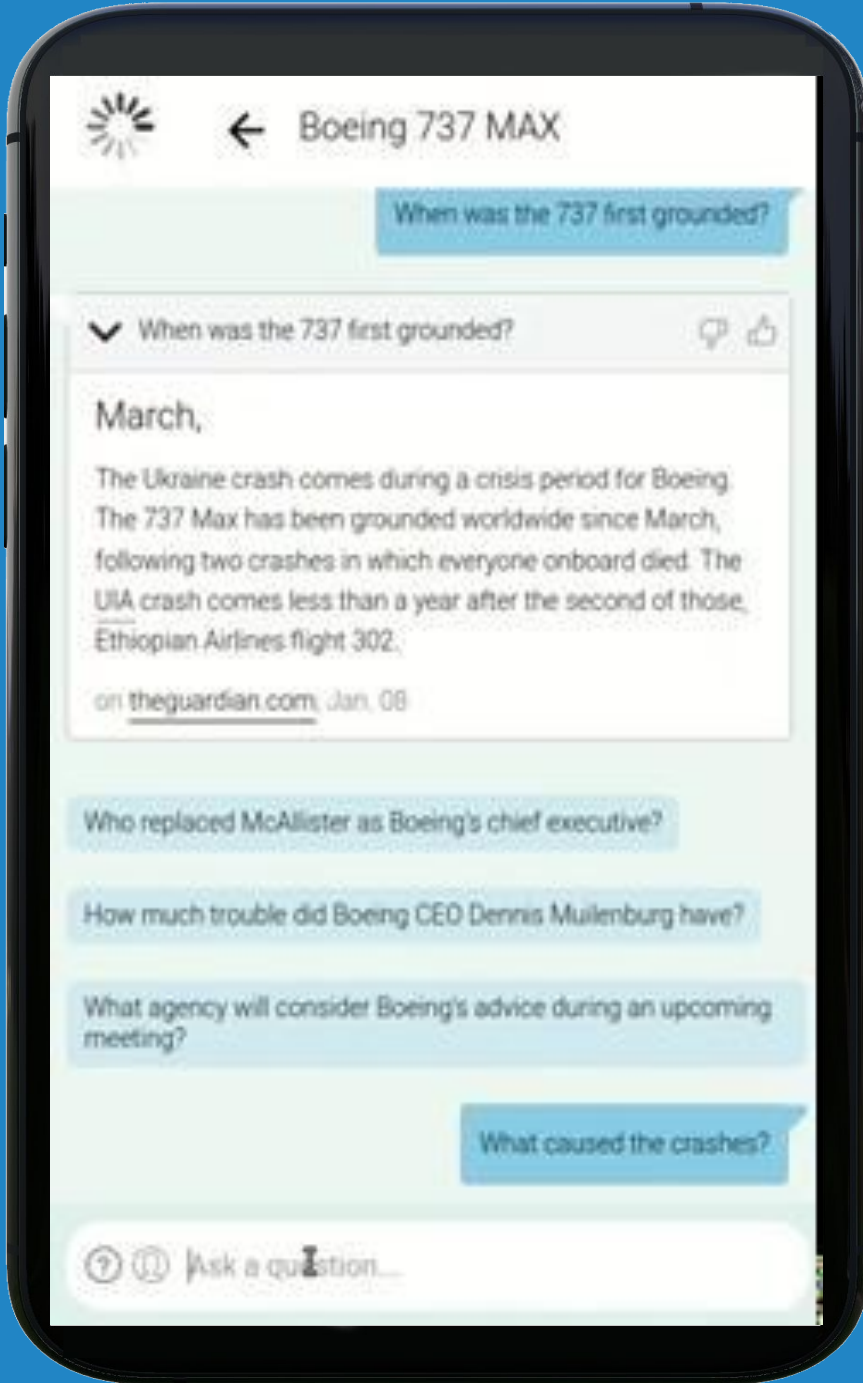


Questions are recommended to the reader.

Questions are updated as information is revealed.



**Information is gathered from news and Wikipedia sources.**



**The reader can ask their own questions.**

**An extractive Q&A system finds a likely answer.**



# Method

Leveraging question answering to keep track of conversation state



# You said that already...

Since October in Australia, fires scorched more than 10.3 million hectares and 27 people have been killed

what else should I know



The fires, which have been raging since October, have killed at least 24 people and burned 10 million hectares

you said that already...



**For a given story, there is a wealth of redundant text content.**  
Throughout the conversation, new information must be provided.

# You said that already...

The two paragraphs answer the same questions

Since October in Australia, fires scorched more than 10.3 million hectares and 27 people have been killed

what else should I know



The fires, which have been raging since October, have killed at least 24 people and burned 10 million hectares

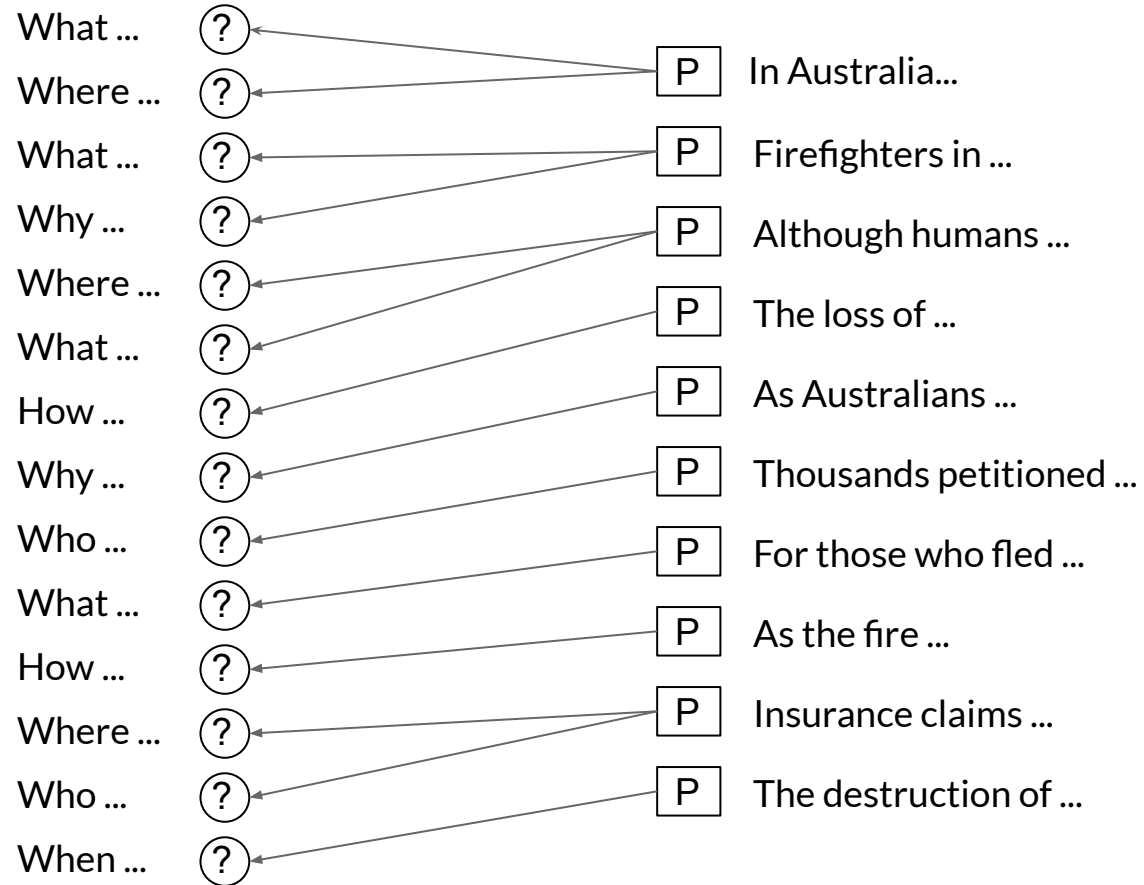
you said that already...



**For a given story, there is a wealth of redundant text content.**  
Throughout the conversation, new information must be provided.

- P In Australia...
- P Firefighters in ...
- P Although humans ...
- P The loss of ...
- P As Australians ...
- P Thousands petitioned ...
- P For those who fled ...
- P As the fire ...
- P Insurance claims ...
- P The destruction of ...

Start with all the paragraphs in a story ...

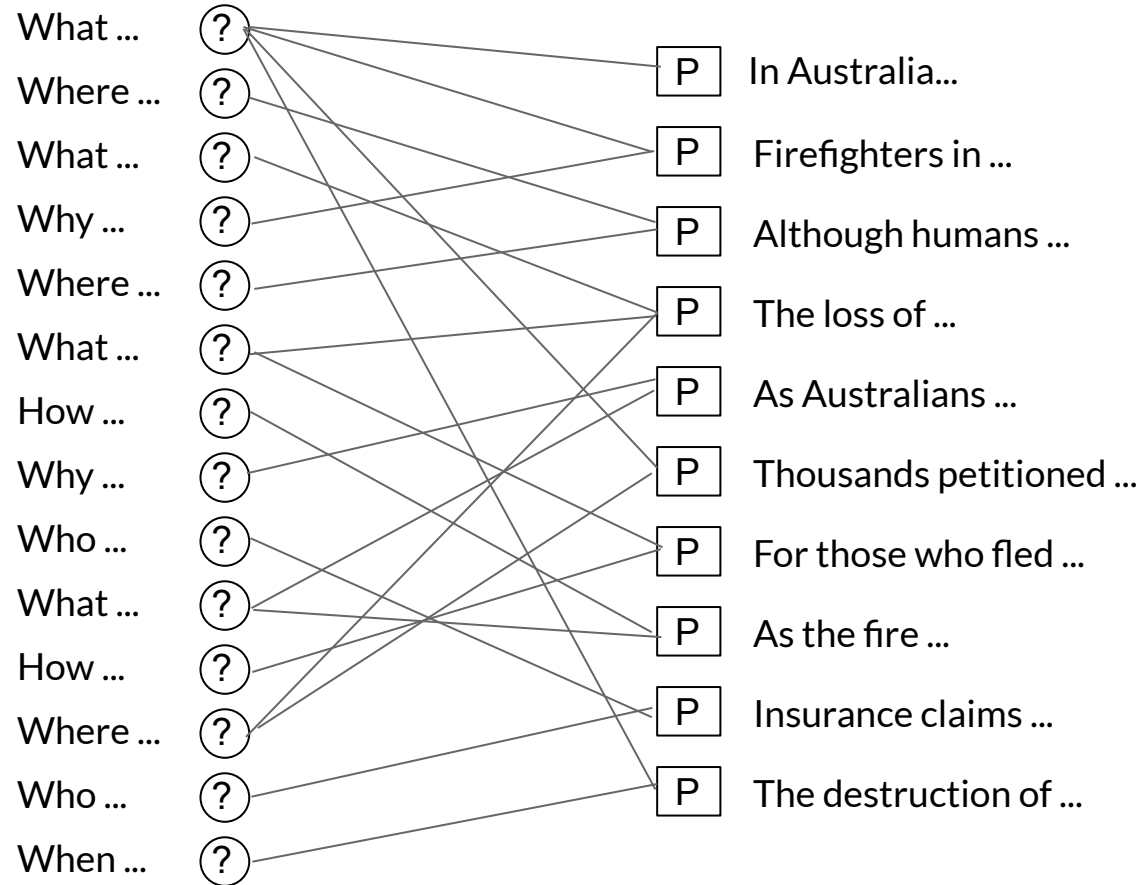


Use a standard Question Generator to generate many questions

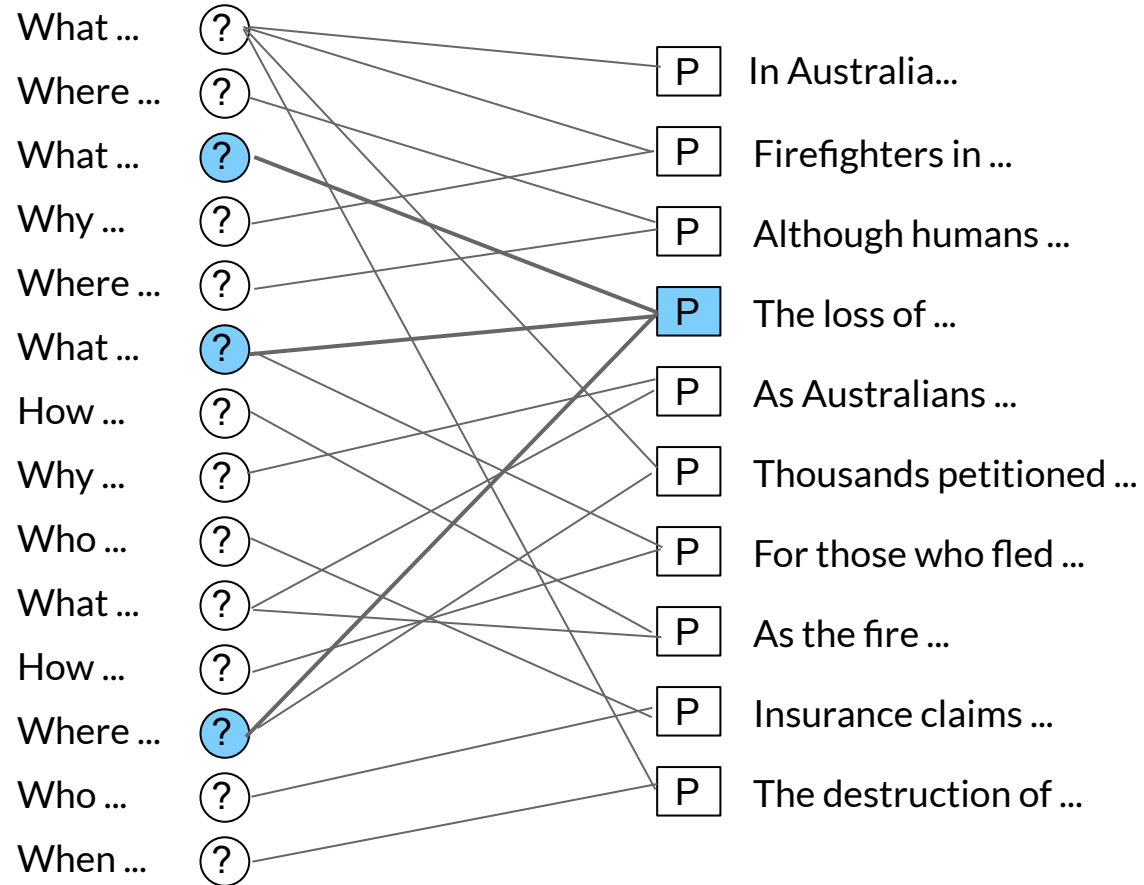
What ... (?)  
Where ... (?)  
What ... (?)  
Why ... (?)  
Where ... (?)  
What ... (?)  
How ... (?)  
Why ... (?)  
Who ... (?)  
What ... (?)  
How ... (?)  
Where ... (?)  
Who ... (?)  
When ... (?)

P In Australia...  
P Firefighters in ...  
P Although humans ...  
P The loss of ...  
P As Australians ...  
P Thousands petitioned ...  
P For those who fled ...  
P As the fire ...  
P Insurance claims ...  
P The destruction of ...

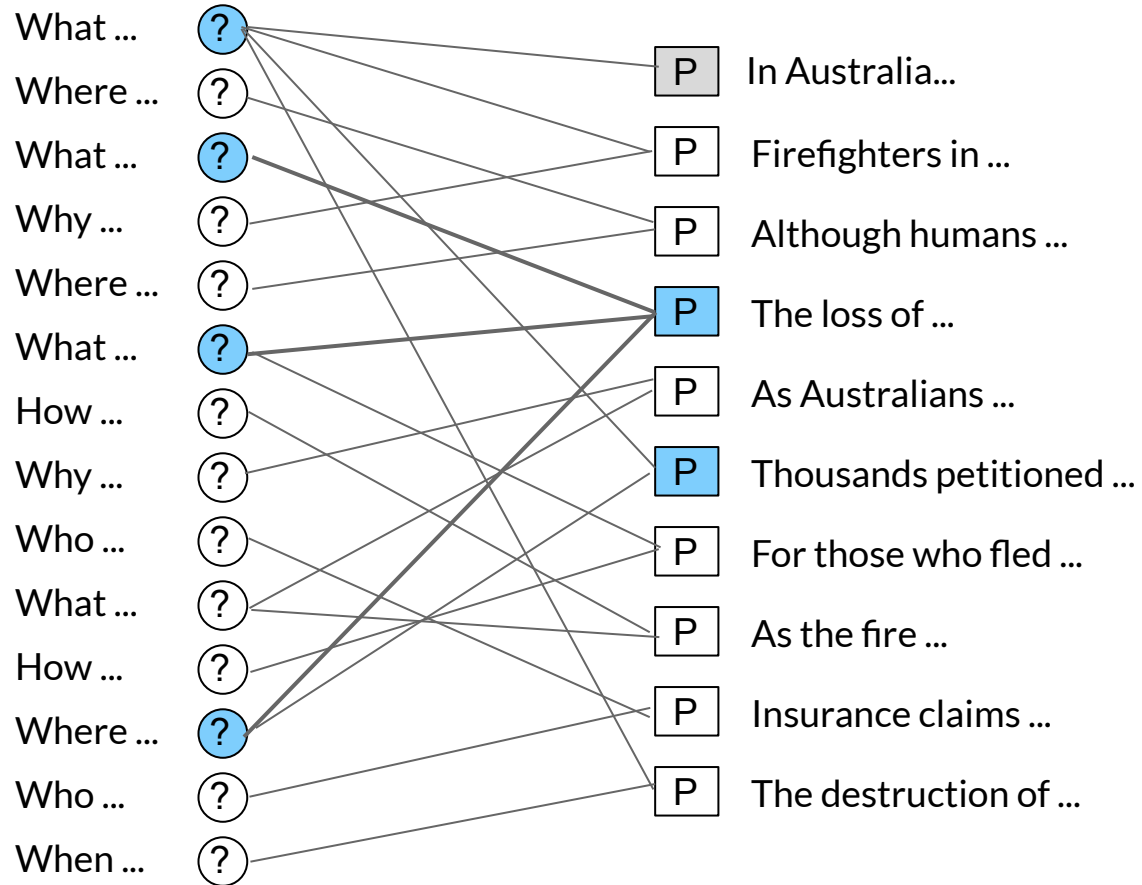
Now we have many paragraphs, and many questions.



Use a Q&A system to see what paragraphs answer what questions  
*Two paragraphs are redundant if they cover the same questions.*



As content is given to the user, keep track of the *answerable questions*.



Some content becomes redundant: all questions it answers are already covered. We can also recommend unanswered questions.





# User Study

Can recommending questions help users have longer conversations?

# Study Setting

Three groups of study:



**NOQR:** No Question Recommendation but the users can ask their own questions



**RANDQR:** Questions recommended are randomly sampled from the set of available questions



**TOPQR:** Questions Recommended are chosen in a greedy order, selecting questions that answers most unanswered paragraphs first.

# Study Settings

## Research Questions:

- **Q1:** Does recommending any questions help users have longer conversations?
- **Q2:** Does the order of questions recommendation help users have longer conversations?
- **Q3:** When users see recommended questions, do they still ask their own question?

# Study Results - Statistics

**Q1:** confirmed. **TOPQR** users have the longest conversations.

**Q3:** confirmed. **TOPQR** and **RANDQR** still ask their own questions.

Measured Value	TOPQR	RANDQR	NOQR
# participants	18	16	22
# chatrooms opened	3.2	2.9	3.1
# turns / chatroom	<b>24.9*</b>	<b>15.3*</b>	8.1
# rec. questions asked	<b>11.9*</b>	<b>8.2*</b>	-
# own questions asked	1.5	1.1	2.2
Latency (seconds)	<b>1.84*</b>	<b>1.88*</b>	4.51

\* is statistically significant with  $p < 0.05$  (comparison with NOQR)

# Study Results - User Satisfaction

Users preferred the TOPQR interface on **4 out of 10 metrics** asked in QUIS survey. No statistical difference on the other 6.

User Survey (7-pt Likert)	TOPQR	RANDQR	NOQR
<b>1</b> dull ... stimulating <b>7</b>	<b>5.28</b> *	5.06	4.20
<b>1</b> frustrating ... satisfying <b>7</b>	<b>5.00</b> *	4.43	4.00
rec. questions are clear	<b>5.78</b> *	4.87	4.28
answers are informative	<b>5.07</b> *	4.44	3.64

\* is statistically significant with  $p < 0.05$  (comparison with NOQR)

# Thanks!

## Any questions?

Come to the Live Q&A.

phillab@berkeley.edu

<https://newslens.berkeley.edu>

NewsLens is also on



&



We thank our sponsors:

**Bloomberg**

