

# Toolkits for Creating Conversational Interfaces

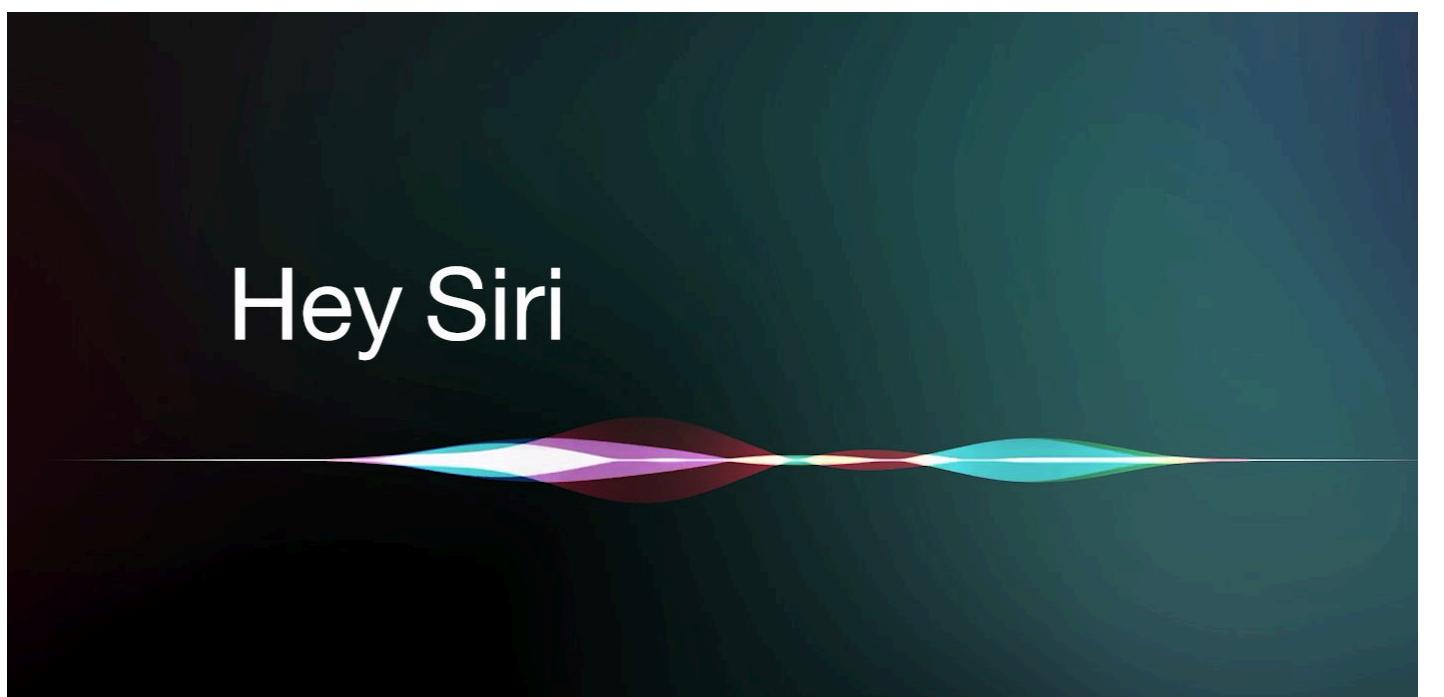
Toby Jia-Jun Li  
<http://toby.li/>

04/20/2020

05-830  
Advanced User Interface Software  
Spring, 2020

# Conversational Interfaces

- Intelligent personal assistants  
Alexa, Siri, Google Assistant, Cortana...



# Conversational Interfaces

- Intelligent personal assistants  
Alexa, Siri, Google Assistant, Cortana...
- Voice command support for specific task domains  
e.g., Talking to your car



# Conversational Interfaces

- Intelligent personal assistants  
Alexa, Siri, Google Assistant, Cortana...
- Voice command support for specific task domains  
e.g., Talking to your car
- Automated phone systems for customer service



Hello! Welcome to XYZ  
Press 1 for Sales  
Press 2 for Support  
Press 3 for Billing

# Conversational Interfaces

- Intelligent personal assistants  
Alexa, Siri, Google Assistant, Cortana...
- Voice command support for specific task domains  
e.g., Talking to your car
- Automated phone systems for customer service
- Chatbots for fun

# History



Turing Test (1950)



# History

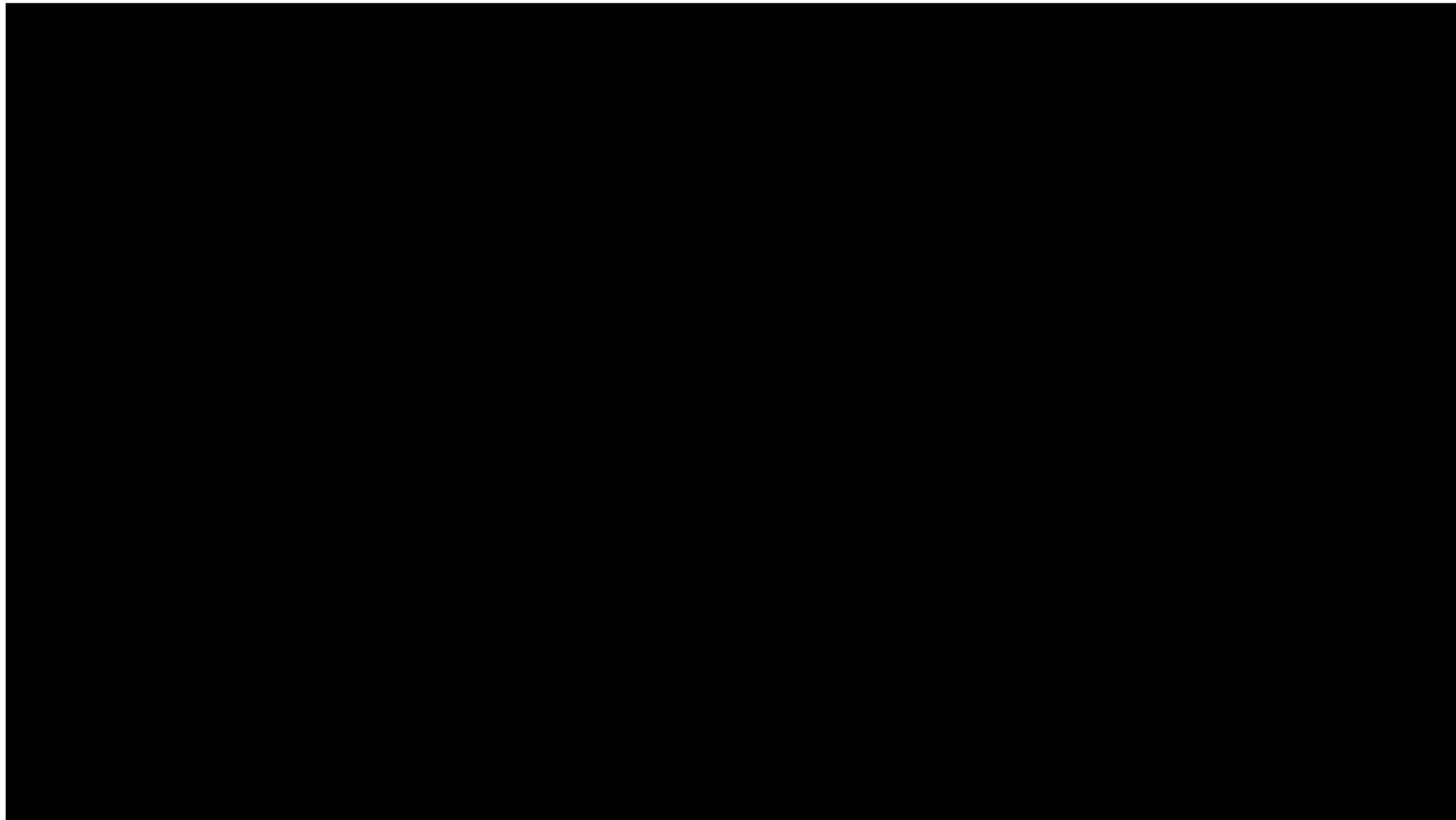
Man-Computer Symbiosis (1960): Cooperative interaction between men and electronic computers

- Let computers facilitate formulative thinking as they now facilitate the solution of formulated problems
- Enable men and computers to cooperate in making decisions and controlling complex situations without inflexible dependence on predetermined programs.



J. C. R. Licklider

# History



Automated Phone System

# History

```
=====
EEEEEEEEE L      IIIIIII  ZZZZZZZZ      AAA
E          L      I          Z      A  A
E          L      I          Z      A  A
EEEEEE    L      I          Z      A  A
E          L      I          Z      AAAAAAAA
E          L      I          Z      A  A
EEEEEEEEE LLLLLLLL IIIIIII  ZZZZZZZZ      A  A

=====
ELIZA > (PLEASE TYPE IN ALL CAPS) WHAT'S YOUR NAME DEAR ?
AMIT
ELIZA > HI, I'M ELIZA. WHAT DO YOU WANT TO TALK ABOUT ?
AMIT > CAN I TALK ABOUT MY PROBLEM ?
ELIZA > SURE... !
AMIT > DUE TO ON GOING EXAMS IN COLLEGE, I HAVE A LOT OF STRESS.
ELIZA > PLEASE RELAX A LITTLE. SLEEP WELL.
AMIT > THANKS FOR YOUR ADVICE.
ELIZA > NO MENTION.
AMIT > BYE
ELIZA > BYE AND KEEP IN TOUCH...
```

**ELIZA (1965)**

# History

## **Spoken Dialog Systems (SDS)**

- ATIS (Air Travel Information Service) (Hempill 1990)
- SUNDIAL (McGlashan 1992)



# History

## AI in Conversational Interfaces

- TRAINS and TRIPS  
(Allen et al. 1995, 2001)

**0:17 H:** And where are the transports?

**0:19 S:** There are two trucks and one helicopter at Delta. *Displays a chart showing vehicle status and contents (also updated through the conversation)*

**0:24 H:** Use a truck to get the people from Calypso to Delta.

**0:31 S:** Your wish is my command. *Displays a construction window with a timeline view of the plan under development, with the vehicles on the vertical axis and time horizontally. The map is also updated to show the route taken to Calypso and back.*

**0:34 H:** How long will that take?

**0:37 S:** It will take six hours and forty-two minutes.

**0:43 H:** What if we went along the coast instead?

**0:47 S:** That option would take ten hours and twenty-seven minutes. *Displays another construction window showing the option alongside the original plan.*

**0:53 H:** Forget it.

**0:55 S:** Ok. *Deletes second construction window from the display.*

**1:05 H:** Use the other truck to get the people from Exodus to Delta.

**1:13 S:** Ok. *Adds the new task to the original construction window, using a different color, and updates the map. In this case, the task is clearly seen to take more than ten hours.*

**1:22 H:** Let's use the helicopter instead.

**1:34 S:** Ok. *Changes the construction window view of this task and updates the map display.*

# Advantages of conversational interfaces

- **Hands-free:** can handle situations where direct manipulation is not possible or convenient (e.g., far away, driving, users with accessibility needs)
- **Screen size independence:** can operate on devices with small screens (e.g., wearable) and no screen.
- **Intuitive to use:** well-designed conversational interfaces should have low learning barriers to users.
- **Efficient:** takes less time and effort for **\*some tasks\*** that require a lot of text entry, or navigating complex menus.
  - Can be inefficient and hard-to-use in some situations too! E.g., when the prompts are too verbose, when the affordances are unclear (discoverability), or when the error handling mechanism is lacking.

# Two classes of conversational systems

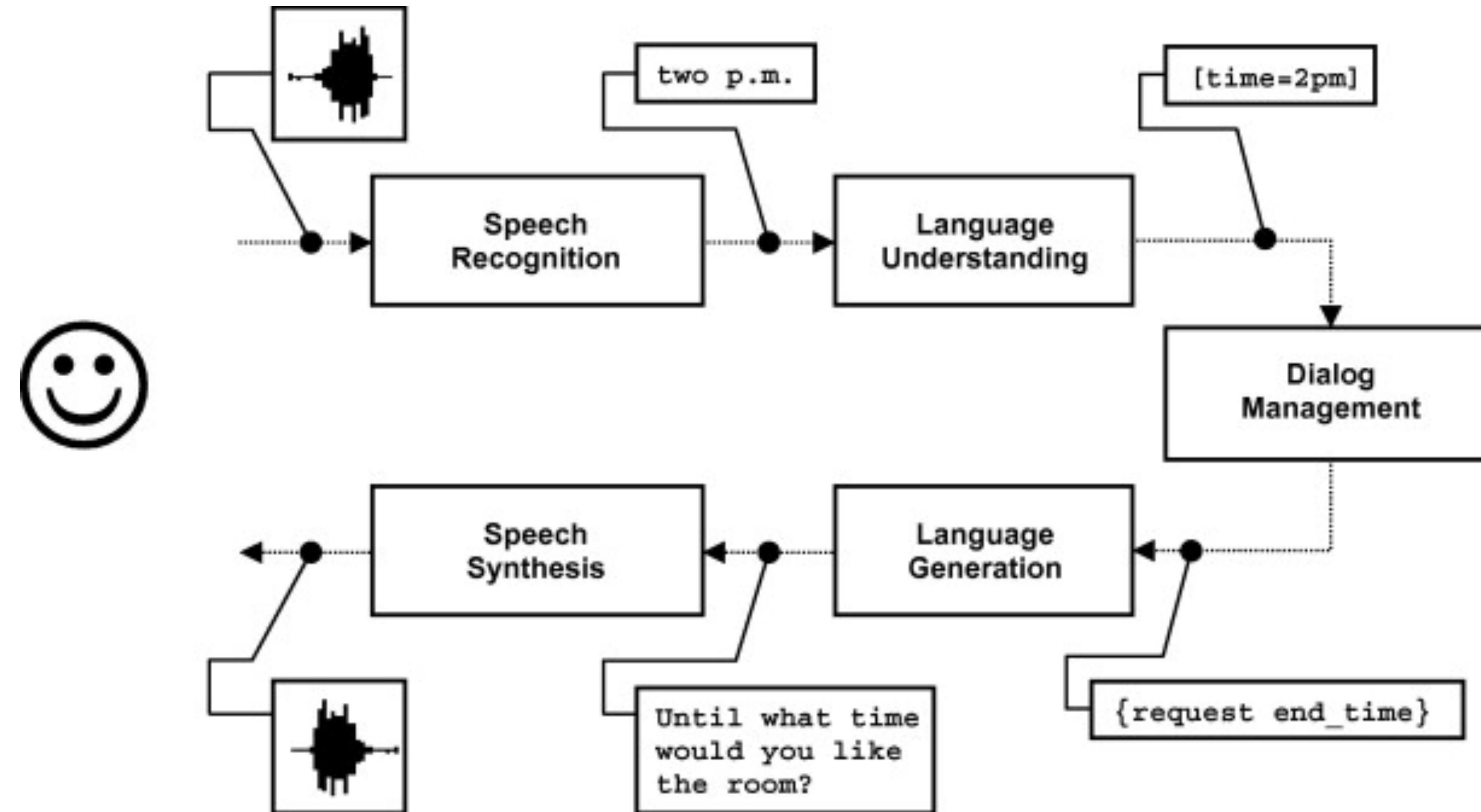
## 1. Task-oriented conversational agents

Purpose: help the user perform some specific tasks

## 2. Social chatbots (“chit-chat” bots)

Purpose: maintain realistic conversations with humans

# Practical architectures for task-oriented dialog systems



**RavenClaw (Bohus and Rudnicky, 2003)**

Bohus, Dan, and Alexander I. Rudnicky. "RavenClaw: Dialog management using hierarchical task decomposition and an expectation agenda." Eighth European Conference on Speech Communication and Technology. 2003.

# Practical architectures for task-oriented dialog systems

## 1. Finite-state

The developer manually defines all the conversation states in the system, and the transitions between the states.

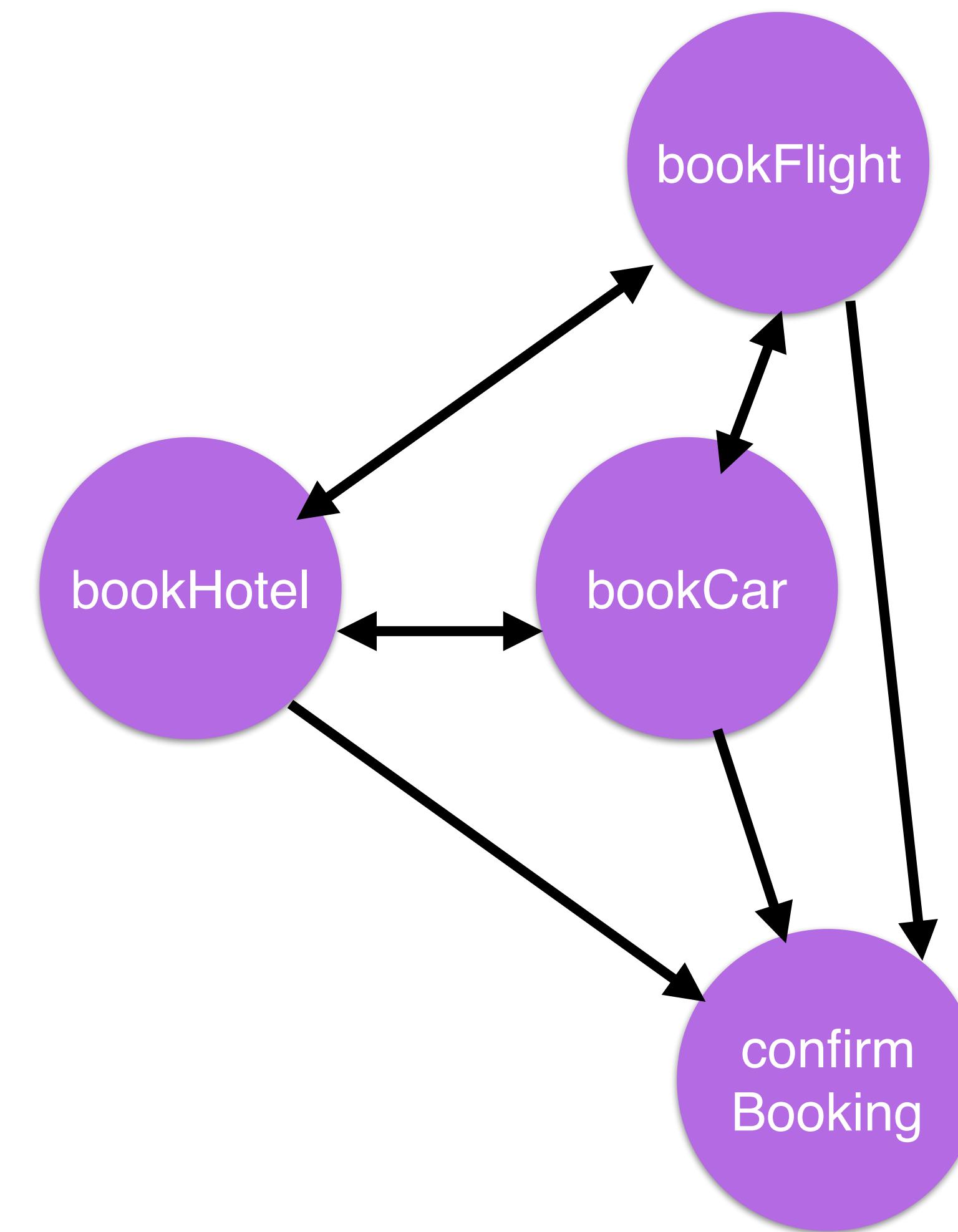
## 2. Frame-based

**frame** (intent): the user's intention for one conversation turn (e.g., book\_flight)

**slot**: the information that the system needs to know to fulfill an intent (e.g., departure\_date, destination\_city)

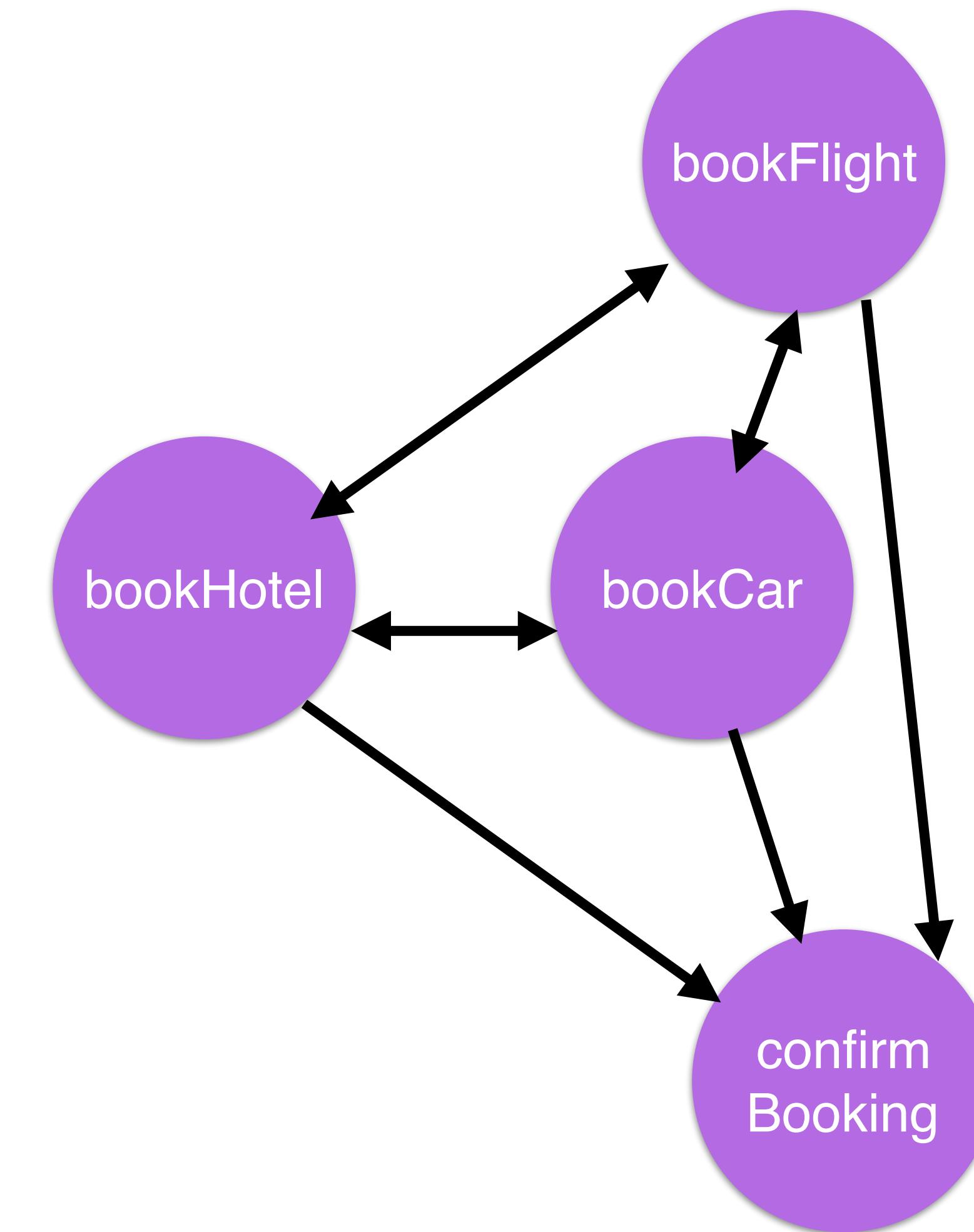
**slot values**: the values that each slot can take

**User:** I want to book a flight for 2 to Munich.



**User:** I want to **book a flight** for 2 to Munich.

### *Intent recognition*

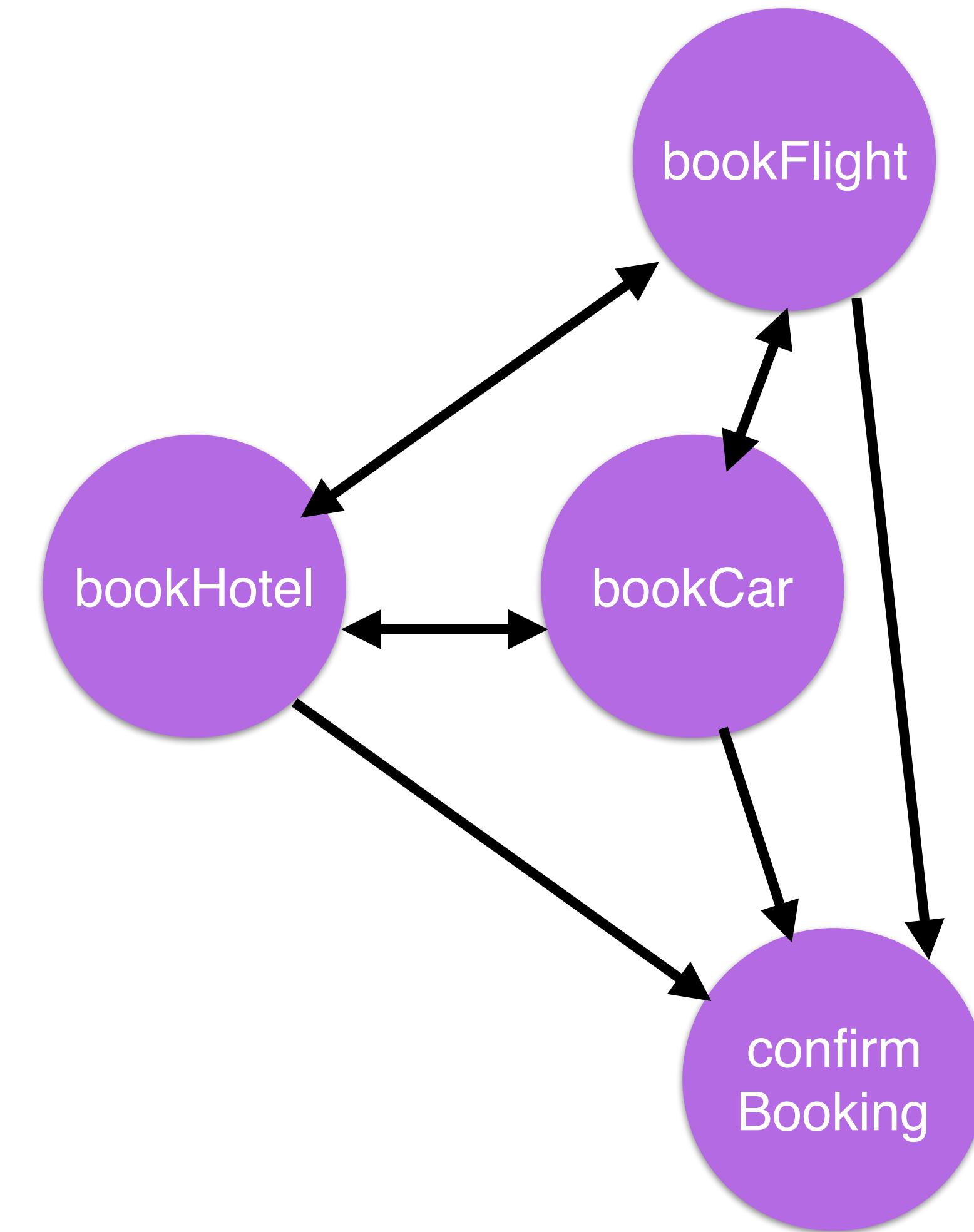


**Intent:** bookFlight

**Slots:** departureCity, arrivalCity, personCount, date

**User:** I want to book a flight for **2** to **Munich**.

### ***Entity extraction / slot filling***



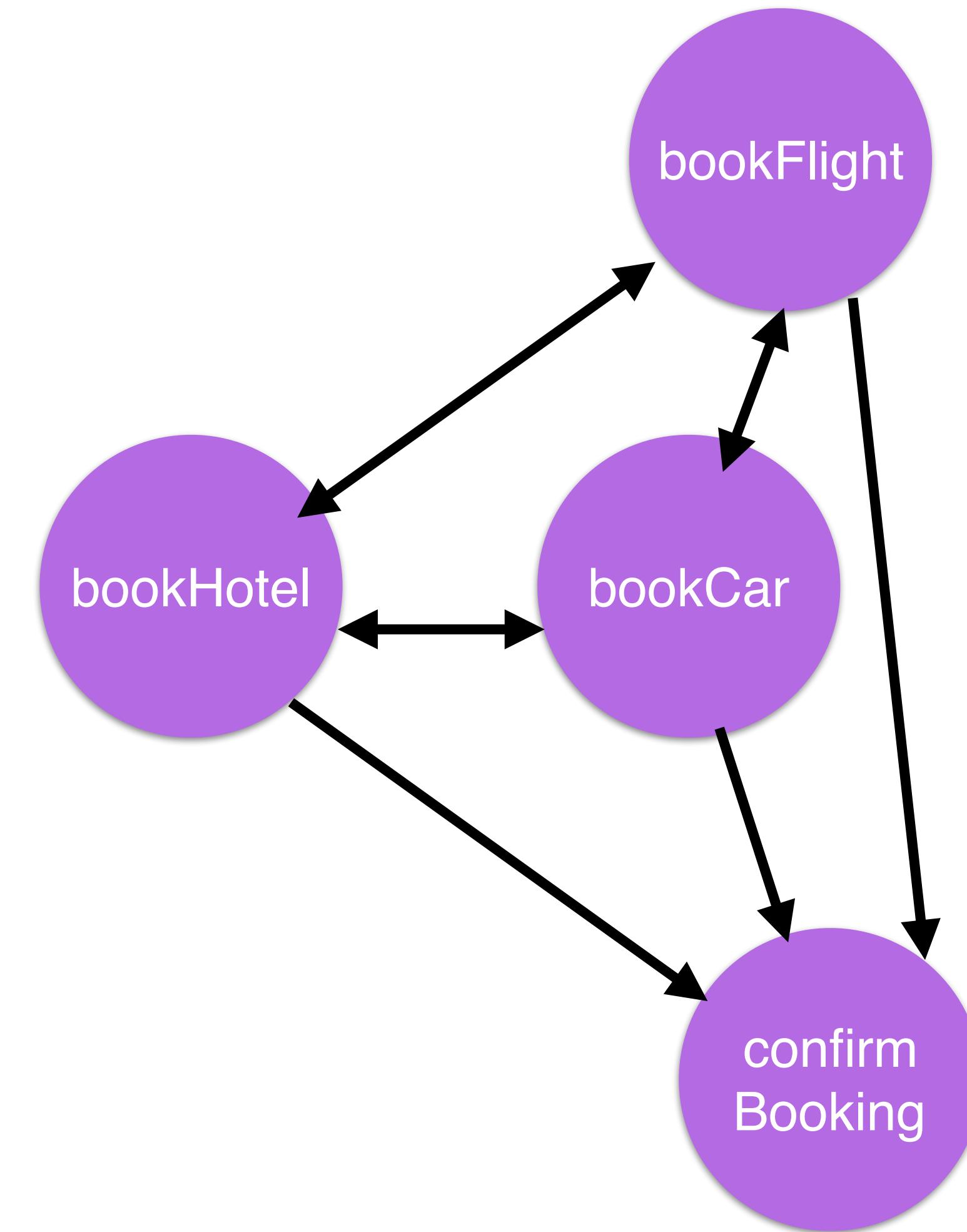
**Intent:** bookFlight

**Slots:** departureCity, ~~arrivalCity, personCount, date~~

**User:** I want to **book a flight** for **2** to **Munich**.

**Bot:** What city are you flying from?

**User:** Pittsburgh.



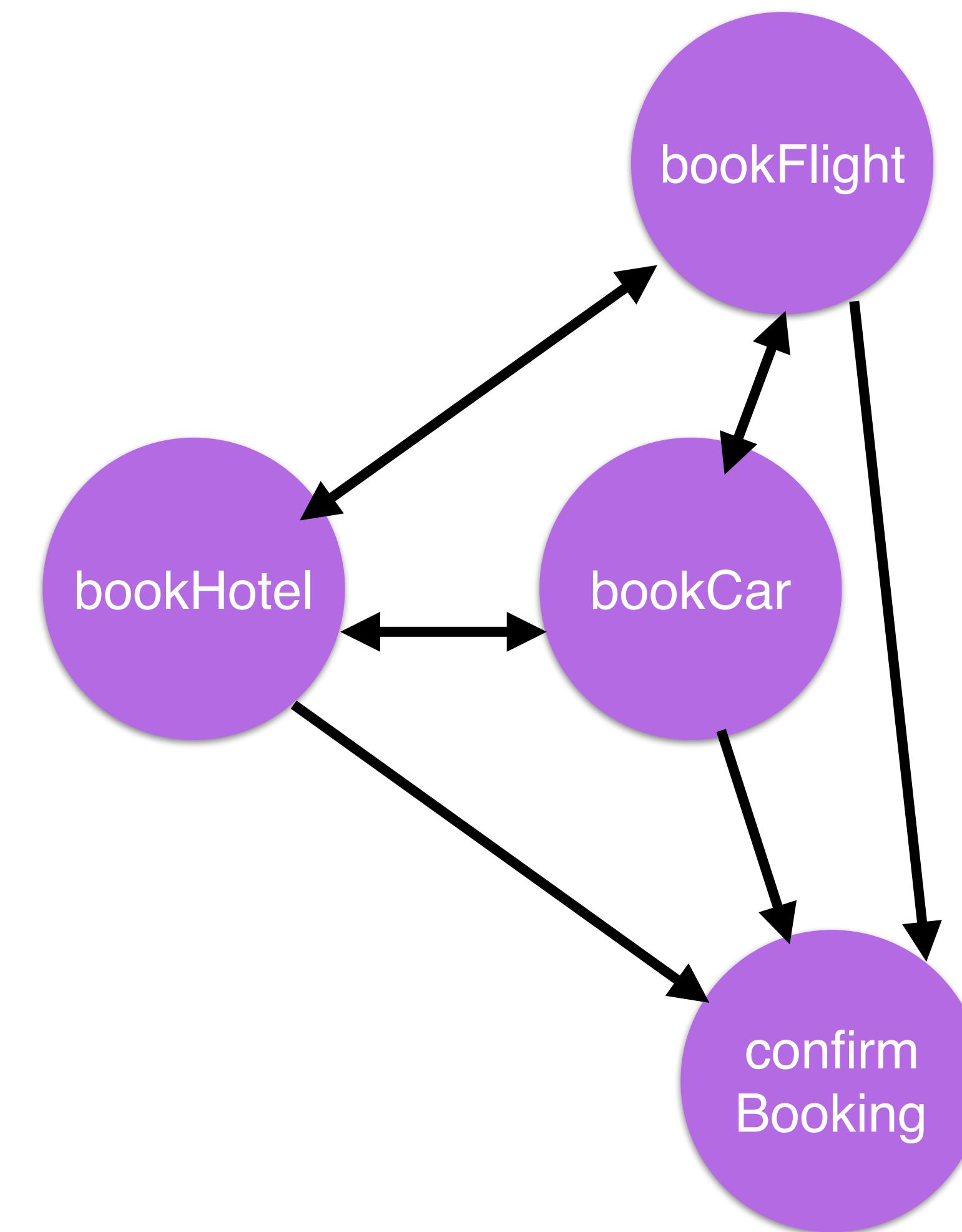
**Intent:** bookFlight

**Slots:** departureCity, ~~arrivalCity, personCount, date~~

**User:** I want to **book a flight** for **2** to **Munich**.

**Bot:** What city are you flying from?

**User:** **Pittsburgh**.



**Intent:** bookFlight

**Slots:** ~~departureCity, arrivalCity, personCount, date~~

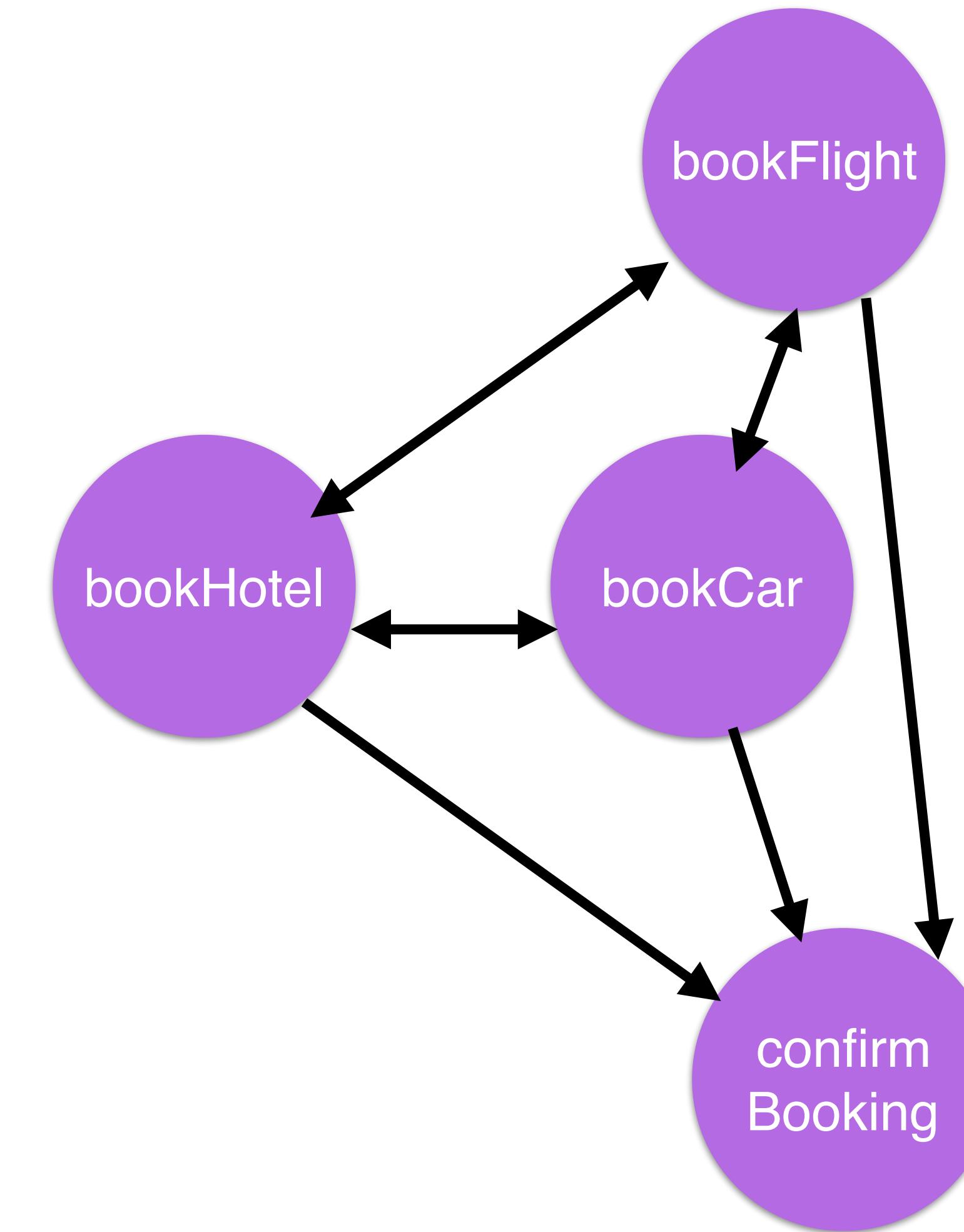
**User:** I want to **book a flight** for **2** to **Munich**.

**Bot:** What city are you flying from?

**User:** **Pittsburgh**.

**Bot:** What's the departure date for the flight?

**User:** **Tomorrow**.



**Intent:** bookFlight

**Slots:** ~~departureCity, arrivalCity, personCount, date~~

**User:** I want to **book a flight** for **2** to **Munich**.

**Bot:** What city are you flying from?

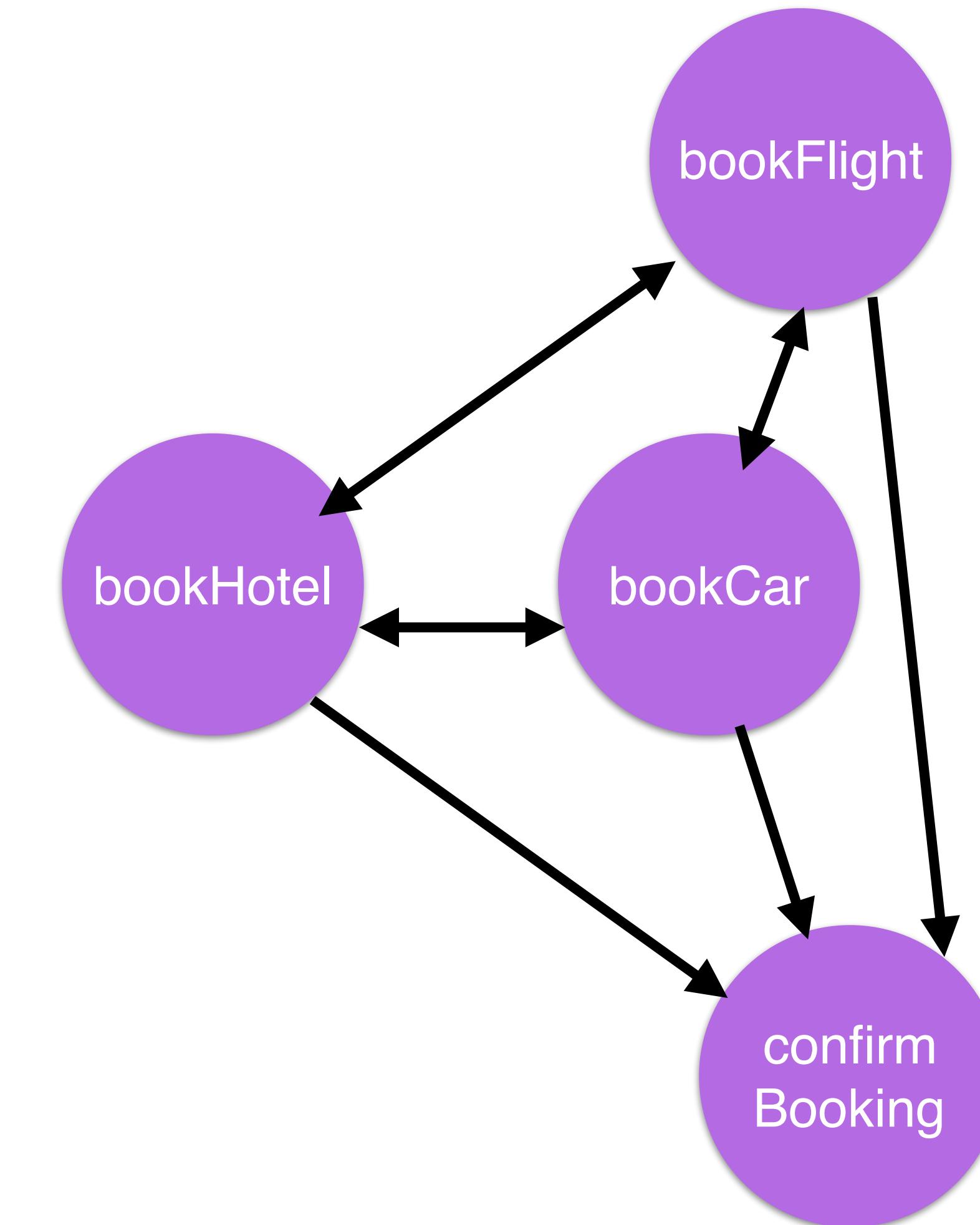
**User:** **Pittsburgh**.

**Bot:** What's the departure date for the flight?

**User:** **Tomorrow**.

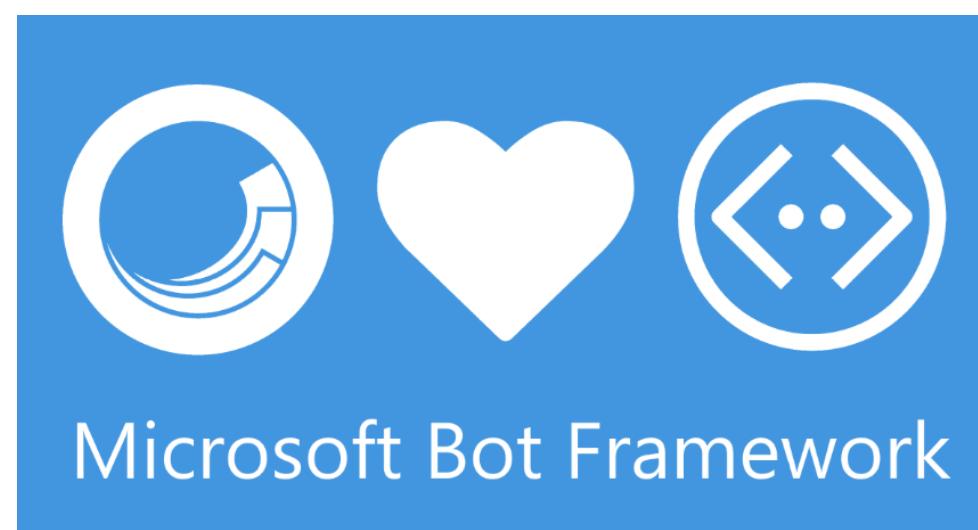
**Bot:** Do you want to also book a hotel or a car?

**User:** I'd like to **get a place to stay** too.

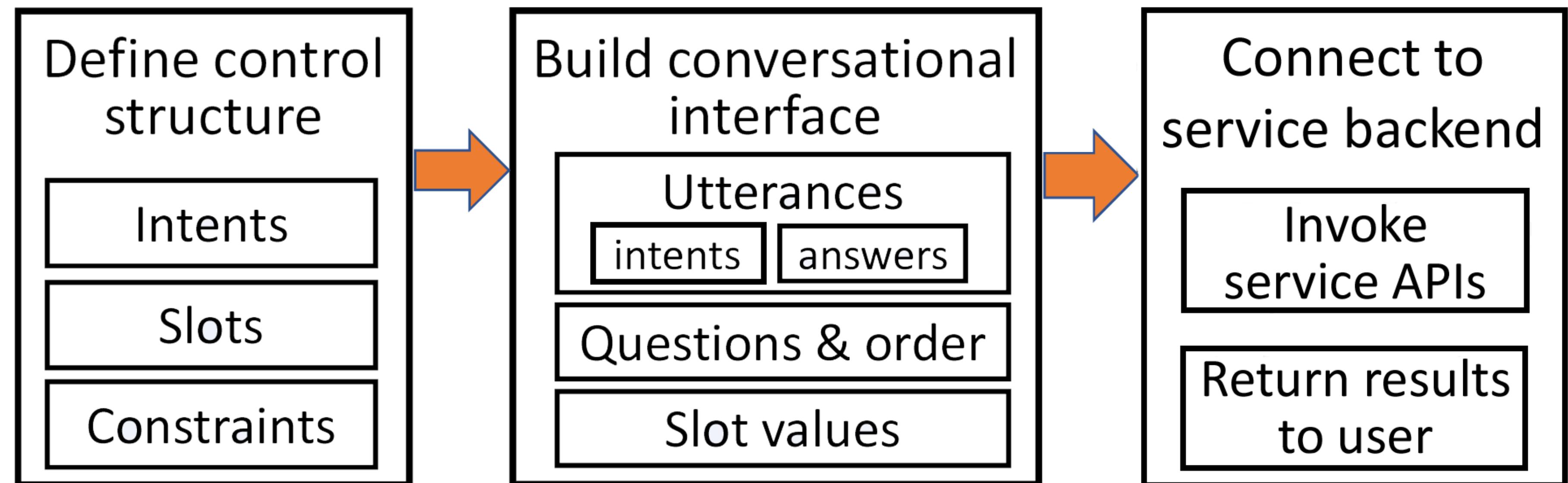


**Intent:** bookHotel    **Slots:** .....

# Existing tools for building slot-filling bots



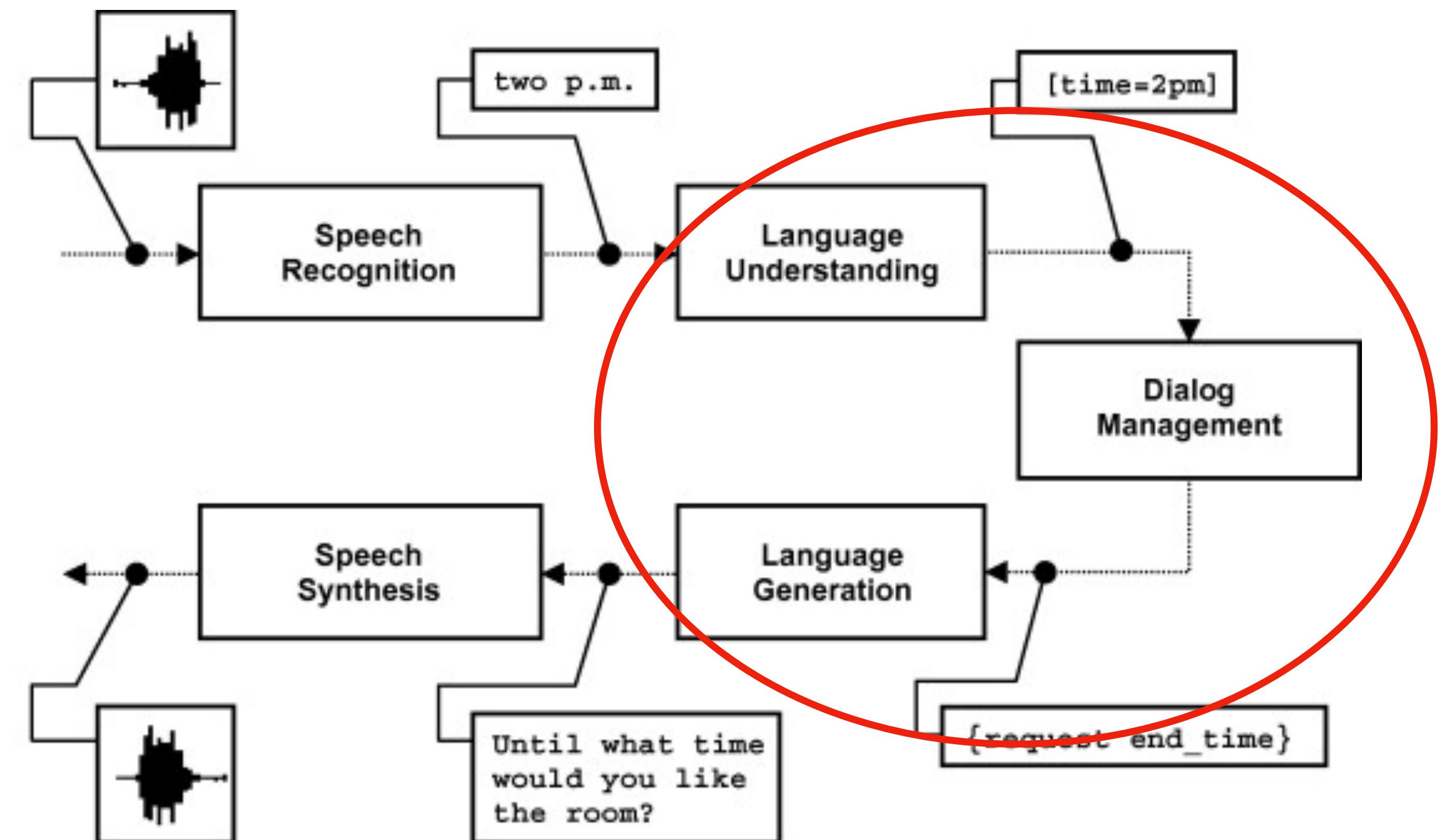
Amazon Lex



# Dialogflow



- One of the more popular toolkits
- Can easily connect to other Google components (e.g., speech recognition, speech synthesis, knowledge graph...)



# Dialogflow demo

# Practical architectures for task-oriented dialog systems

## 1. Finite-state

The developer manually defines all the conversation states in the system, and the transitions between the states.

## 2. Frame-based

**frame** (intent): the user's intention for one conversation turn (e.g., book\_flight)

**slot**: the information that the system needs to know to fulfill an intent (e.g., departure\_date, destination\_city)

**slot values**: the values that each slot can take

# Other architectures for dialog systems

## 1. Rule-based

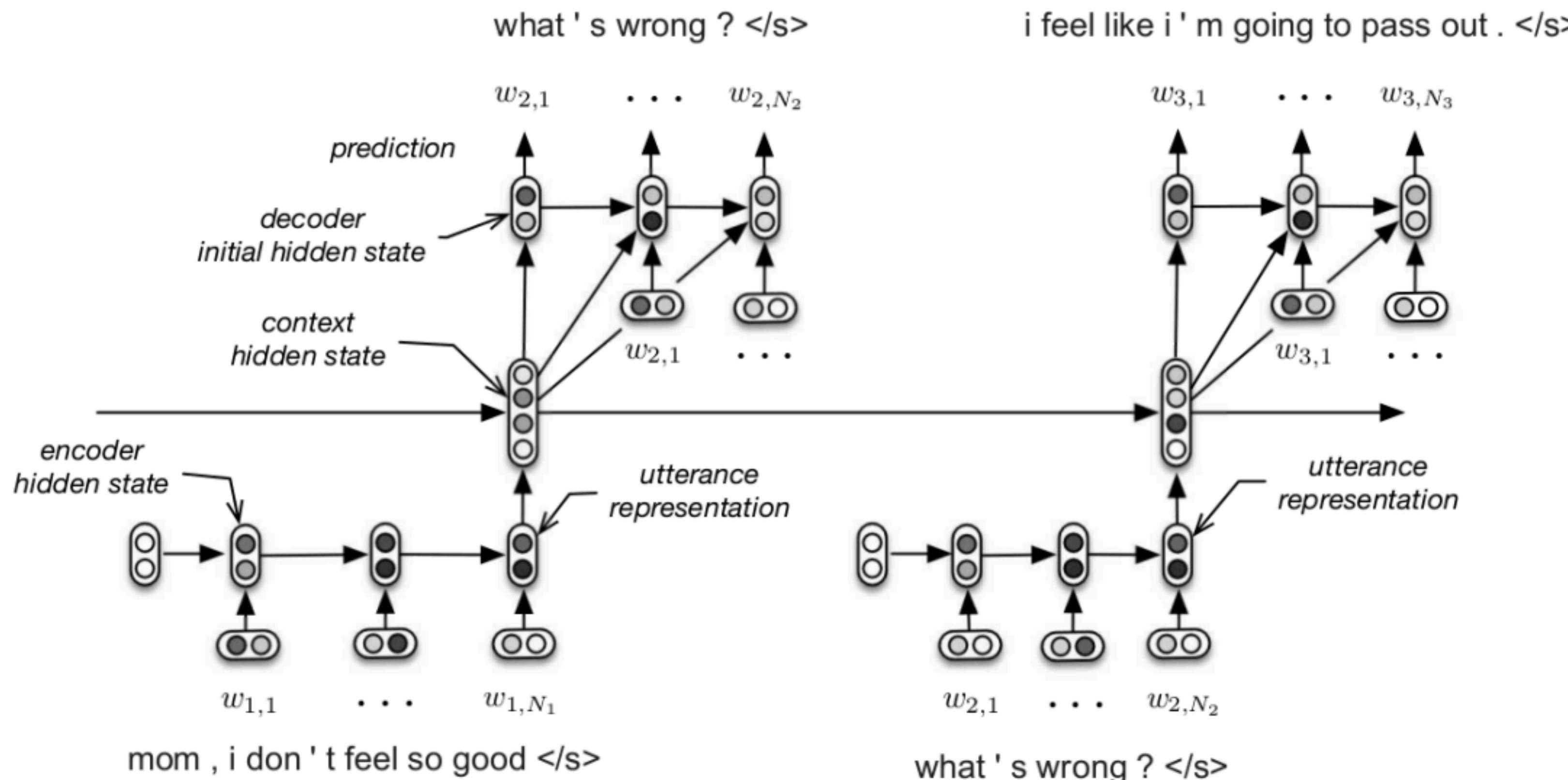
```
(if (contains (or "hi" "hello")) (output "hello"))
(if (and (= detect_comm_type SELF_DISCLOSURE) (= detect_emotion SAD))
    (output "I'm sorry to hear [$USER_DISCLOSURE]"))
```

## 2. Corpus-based: use a very large corpus of human-human or human-machine conversations

- **Information retrieval (IR) based approach:** find the best-matched prior utterance for the user's input in the corpus, and use the prior response for that utterance
- **Sequence-to-sequence dialog generation:** model conversation as a sequence transduction problem -> generate a response from a user input (and probably with some other contexts encoded in)

# Seq2seq model architecture

Serban, Iulian V., Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau. 2015. "Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models."



# Sample output from a neural net chatbot (2015)

Sordoni A, Galley M, Auli M, Brockett C, Ji Y, Mitchell M, Nie JY, Gao J, Dolan B. A neural network approach to context-sensitive generation of conversational responses. NAACL

trained on 127M Twitter context-message-response triples

#	Context <i>c</i>	Message <i>m</i>	Generated Response <i>r</i>
1	noone can help this time its just i need a support system	well im here to support you. whenever you need me.	thank you, i appreciate it thank you
2	nothin much, and how's the book?!	its good but i'm only like halfway through cuz i don't feel like reading. i'm so bored ...	that's good! i have the book but i'm bored too.
3	i know! what have you been up to?	nothing really. just enjoying summer and working. what about you?	that's good! nothing much just listening to music and you?
4	thanks, your way better though	i'm just good at expressing it that's but trust me you are freakin amazing and waay better	i know, i don't trust you, i love you too! thank you
5	thanks. yeah i am good thanks! good, but angry!	sorry you're angry, been there, i hope it dies down.	i hope so too, i'm not angry!!
6	figures. bass. never ate bass.	bass is extremely mild with very few bones. my favorite after catfish	that's one of the bones, chuck bass is mine too.
7	today i will walk for miles. clear sky clear mind.	love walking. it's too hot to walk this month. i haven't moved. frozen. heat freeze.	it's too hot for you to get up to honey if you have frozen yogurt to you.
8	arresting people over "nasty tweets"? please. i'm not condoning what the lad did but arresting him is a tad silly.	it appears he may have sent death threats to others while all this was kicking off too.	i did! he was armed with nuclear threats? that's what's happening to you.

<b>Baseline mutual information model (Li et al. 2015)</b>	<b>Proposed reinforcement learning model</b>
A: Where are you going? (1)	A: Where are you going? (1)
B: I'm going to the restroom. (2)	B: I'm going to the police station. (2)
A: See you later. (3)	A: I'll come with you. (3)
B: See you later. (4)	B: No, no, no, no, you're not going anywhere. (4)
A: See you later. (5)	A: Why? (5)
B: See you later. (6)	B: I need you to stay here. (6)
...	A: I don't know what you are talking about. (7)
...	...
A: how old are you? (1)	A: How old are you? (1)
B: I'm 16. (2)	B: I'm 16. Why are you asking? (2)
A: 16? (3)	A: I thought you were 12. (3)
B: I don't know what you are talking about. (4)	B: What made you think so? (4)
A: You don't know what you are saying. (5)	A: I don't know what you are talking about. (5)
B: I don't know what you are talking about. (6)	B: You don't know what you are saying. (6)
A: You don't know what you are saying. (7)	...
...	...

## Deep reinforcement learning for dialog generation (2016)

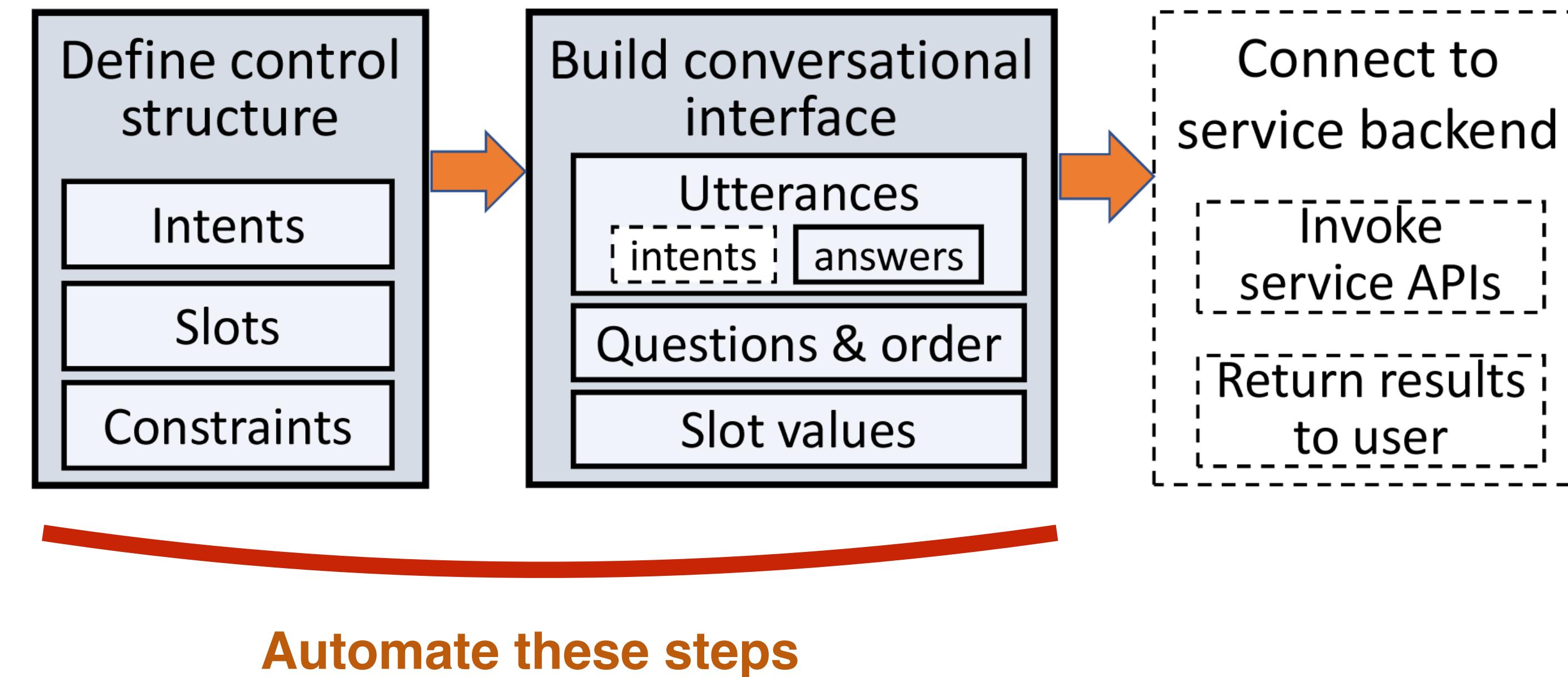
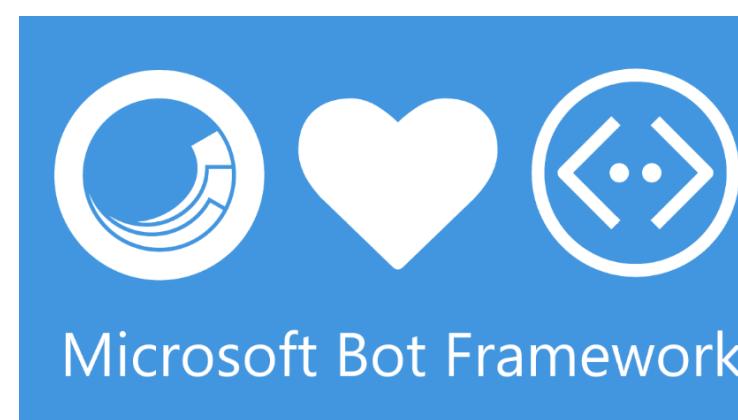
Li, Jiwei, et al. "Deep reinforcement learning for dialogue generation." arXiv preprint arXiv:1606.01541 (2016).

<b>Input</b>	tell me ... how long have you had this falling sickness ?
<b>Vanilla-MLE</b>	i 'm not a doctor .
<b>Vanilla-Sample</b>	well everything you did was totally untrue .
<b>REINFORCE</b>	i don 't know how long it 's been .
<b>REGS Monte Carlo</b>	A few months, I guess .
<b>Input</b>	so i had the doctors test sammy 's response to conditioning .
<b>Vanilla-MLE</b>	sammy wrote the test sammy wrote the test .
<b>Vanilla-Sample</b>	objects pick objects up objects objects objects objects objects
<b>REINFORCE</b>	i 'm not sure that 's relevant .
<b>REGS Monte Carlo</b>	so he took the pills .
<b>Input</b>	what are you going to do ?
<b>Vanilla-MLE</b>	i 'm going to the movies .
<b>Vanilla-Sample</b>	get him outta here first !
<b>REINFORCE</b>	i 'm going to get you .
<b>REGS Monte Carlo</b>	i 'm going to get you a new car .
<b>Input</b>	they fear your power your intellect .
<b>Vanilla-MLE</b>	you 're the only one who knows what 's going on .
<b>Vanilla-Sample</b>	when they are conquered and you surrender they will control all of us .
<b>REINFORCE</b>	i 'm afraid i 'm not ready yet .
<b>REGS Monte Carlo</b>	i 'm not afraid of your power .

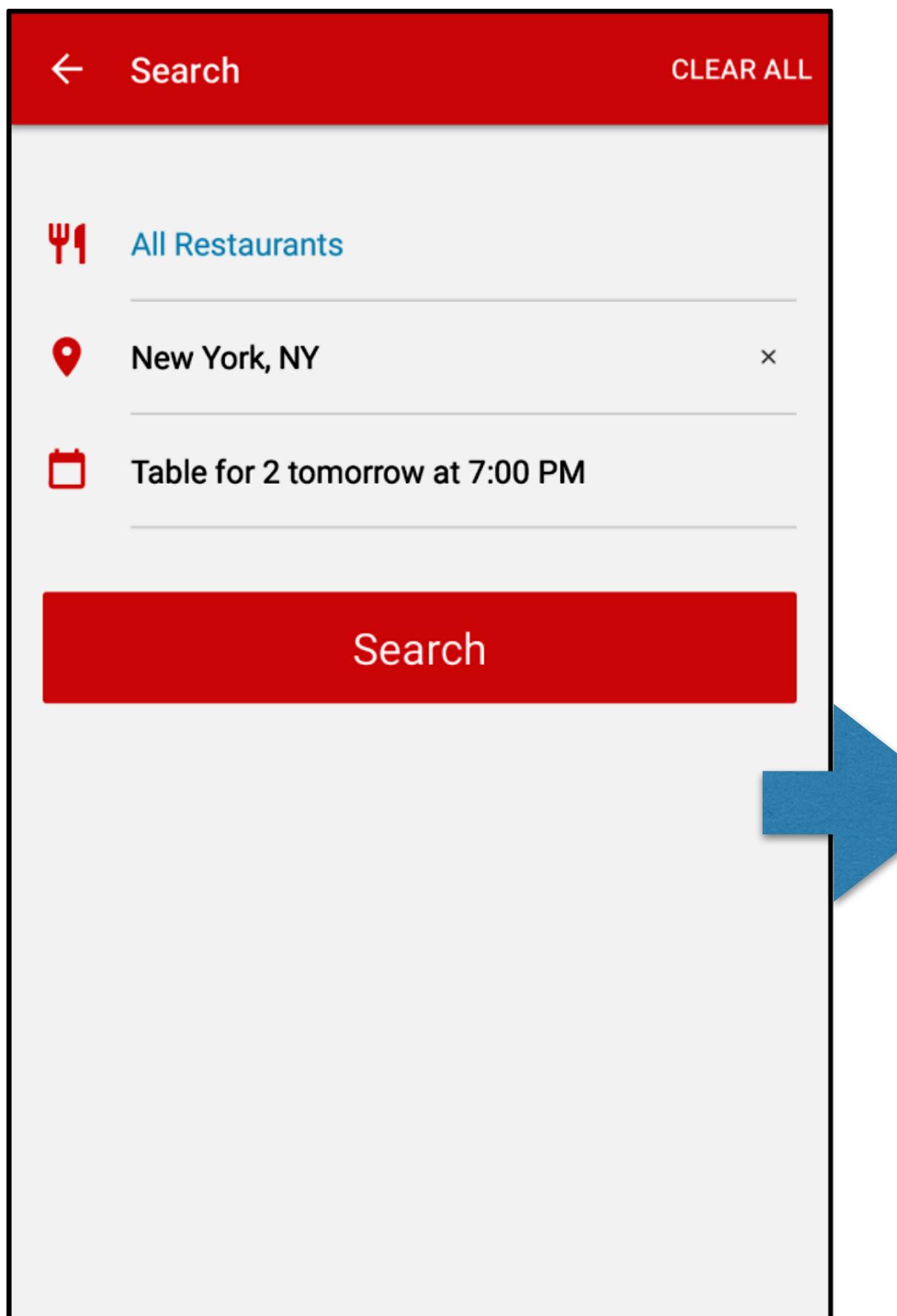
## Adversarial learning for dialog generation (2017)

Li, Jiwei, et al. "Adversarial learning for neural dialogue generation." arXiv preprint arXiv:1701.06547 (2017).

# KITE: generating frame-based interfaces from GUIs (2018)



Toby Jia-Jun Li and Oriana Riva. KITE: Building conversational bots from mobile apps. *MobiSys 2018*.



a)

**Intent**

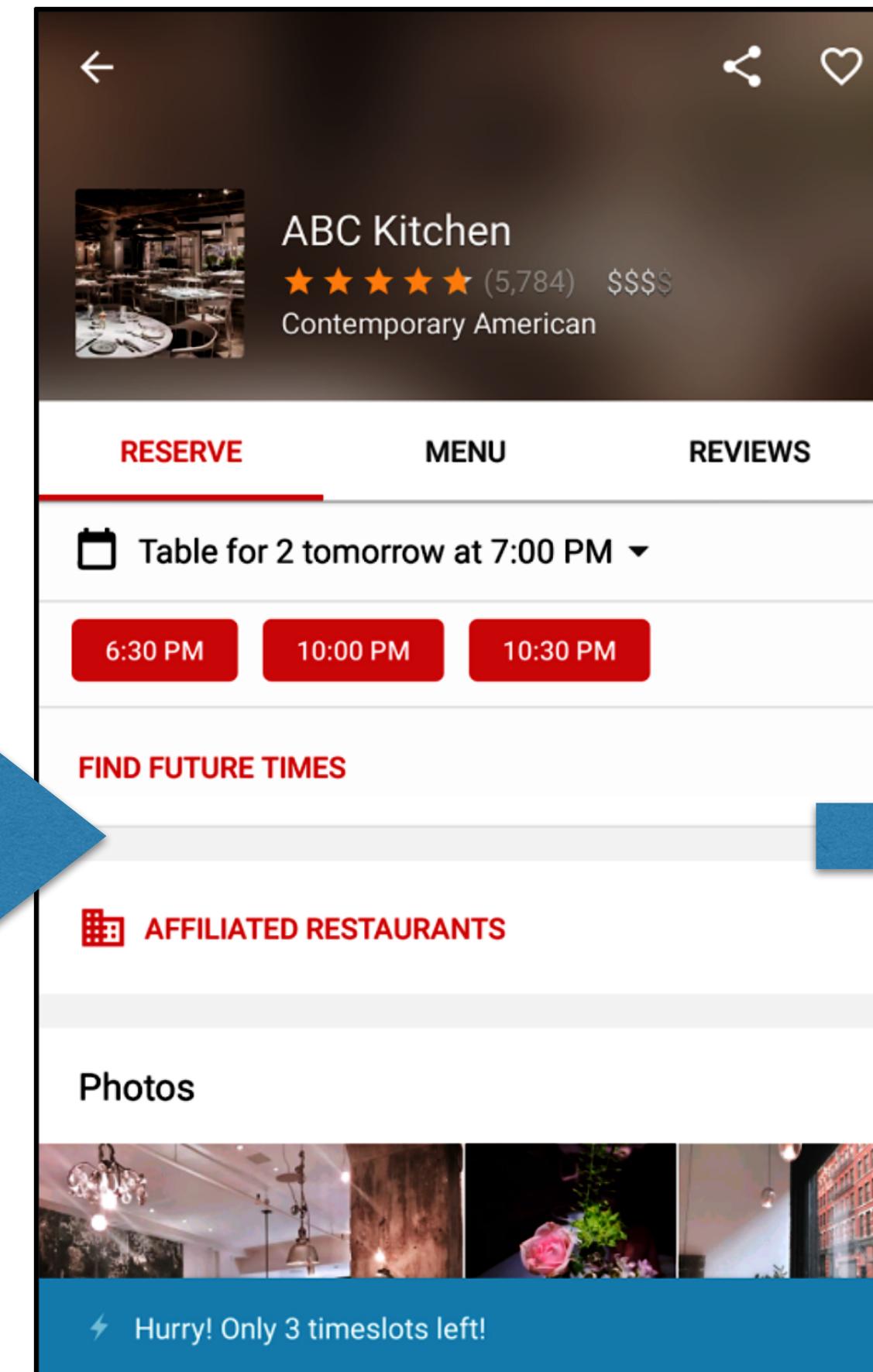
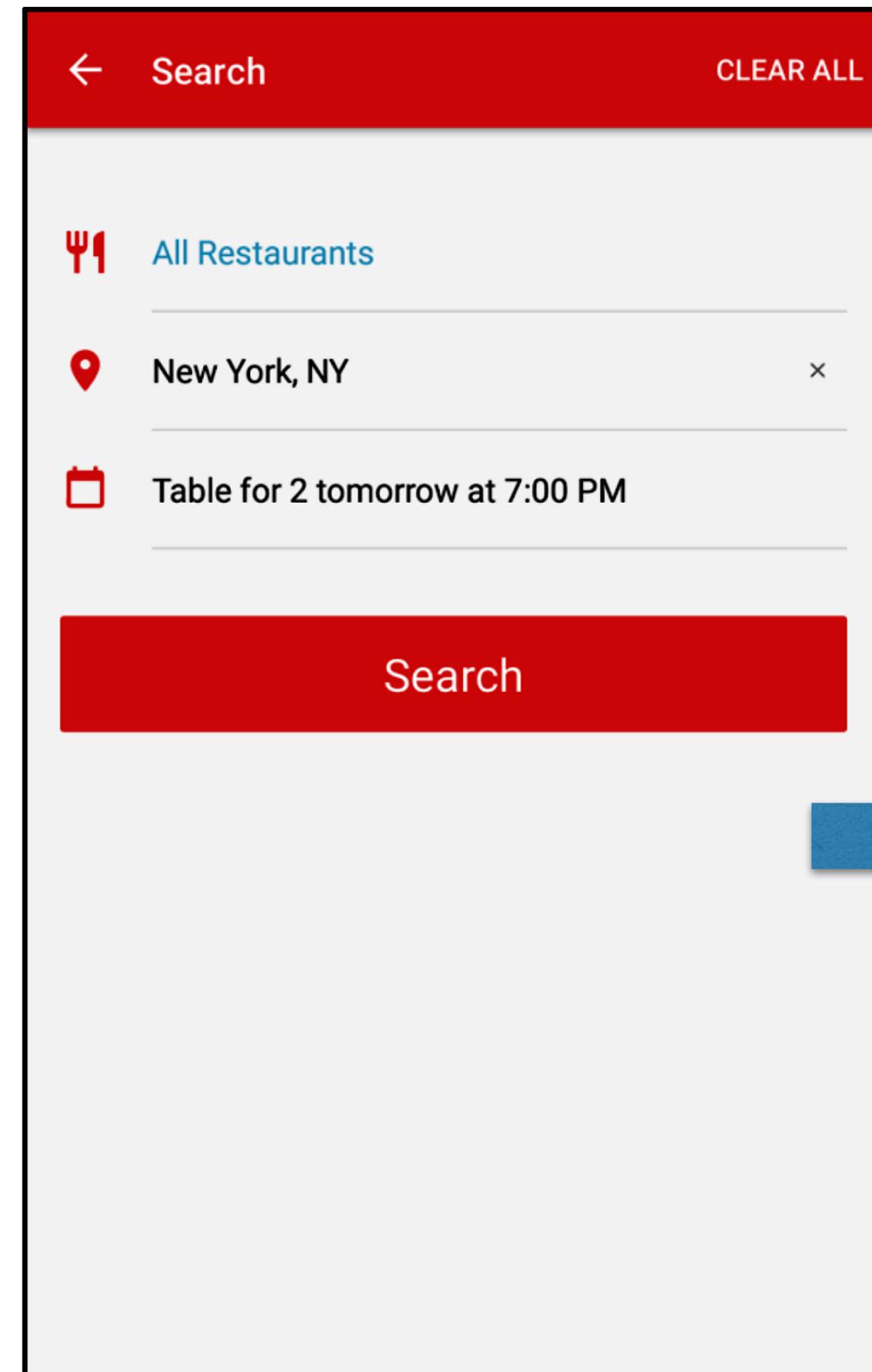
**SearchCriteria**

**Slot**

CuisineType

City

PersonCount...



a)

b)

**Intent**

**SearchCriteria**

**RestaurantProfile**

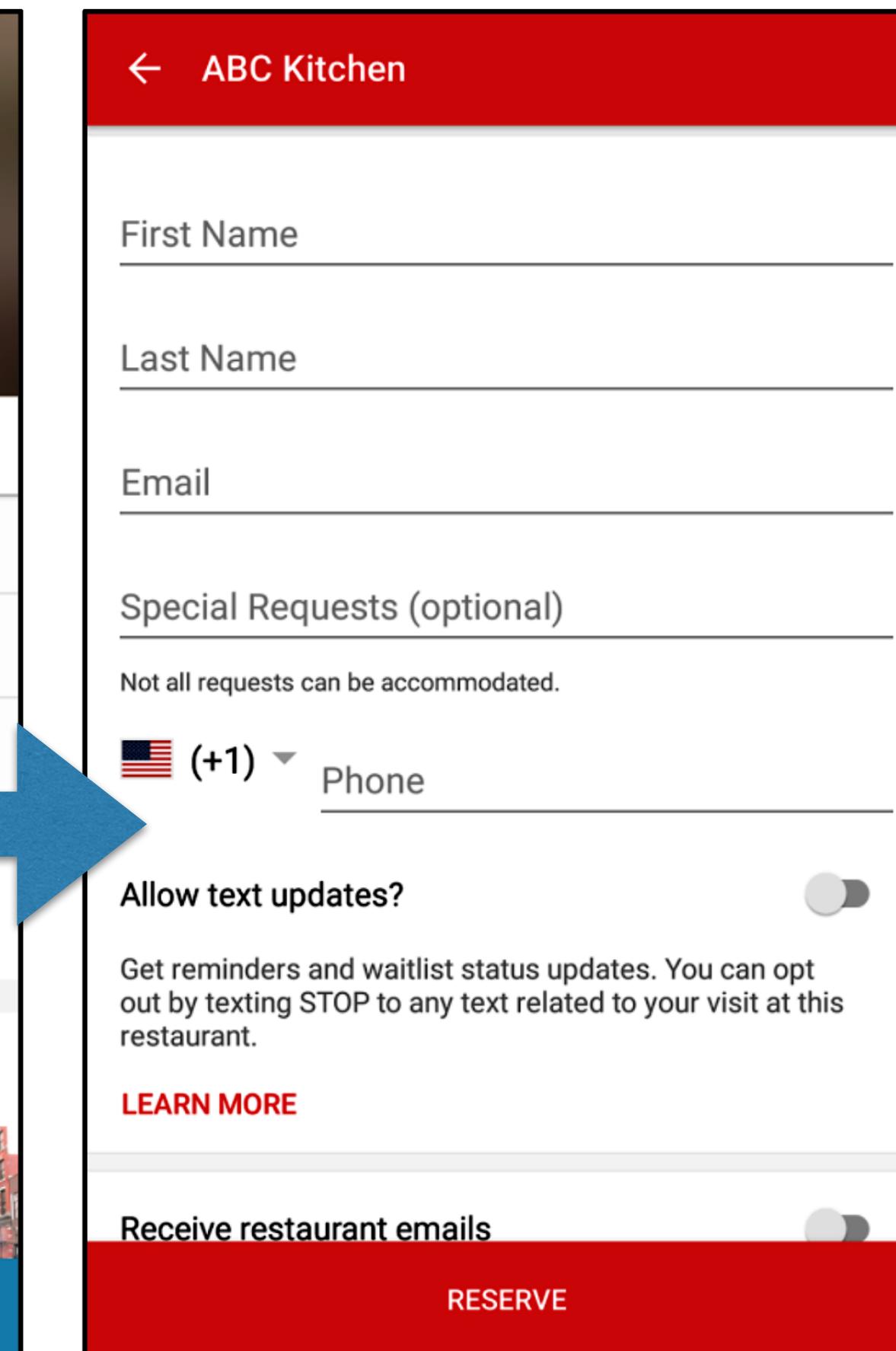
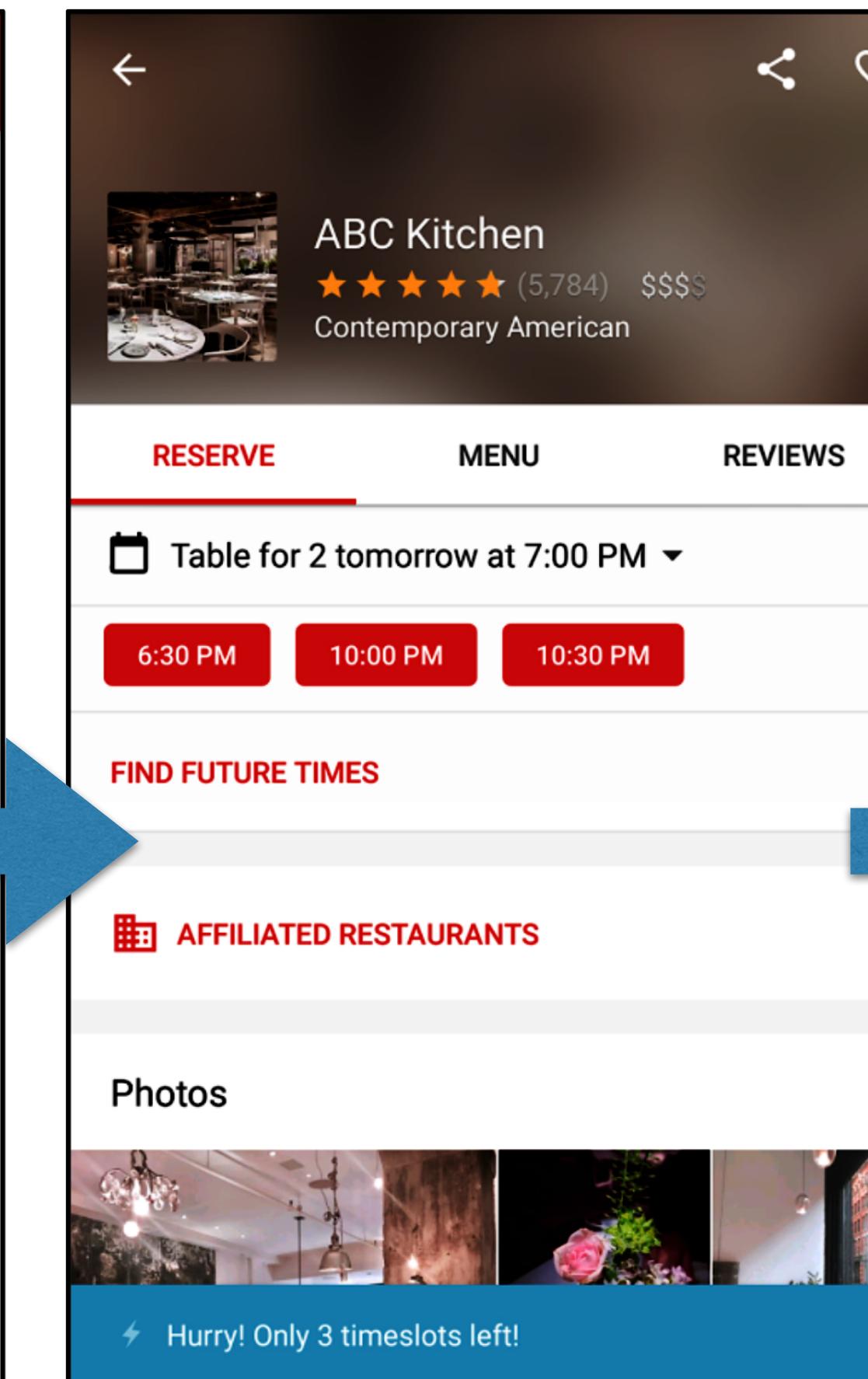
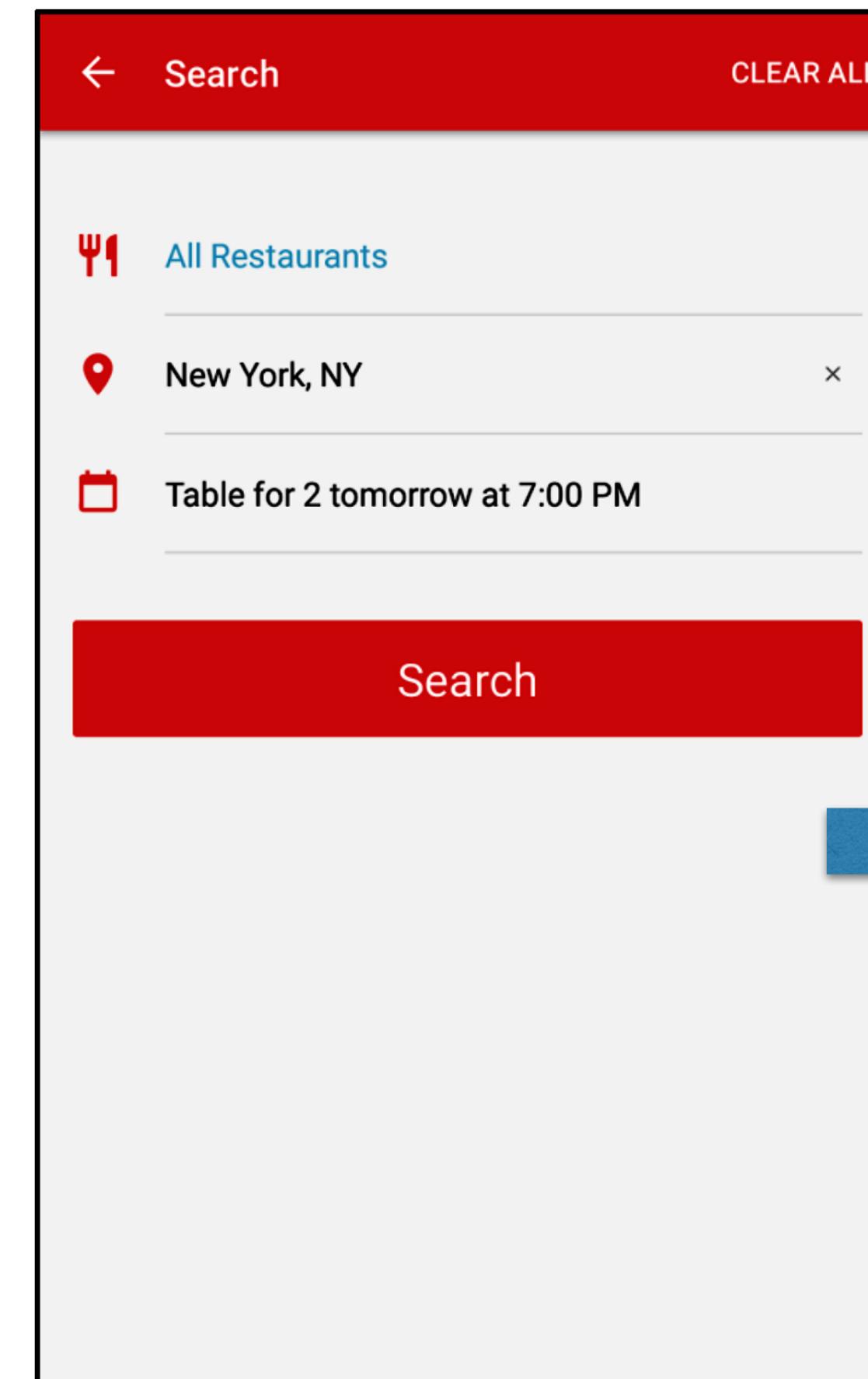
**Slot**

**CuisineType**

**Time**

**City**

**PersonCount...**



**Intent**

**SearchCriteria**

**RestaurantProfile**

**ConfirmBooking**

**Slot**

CuisineType

Time

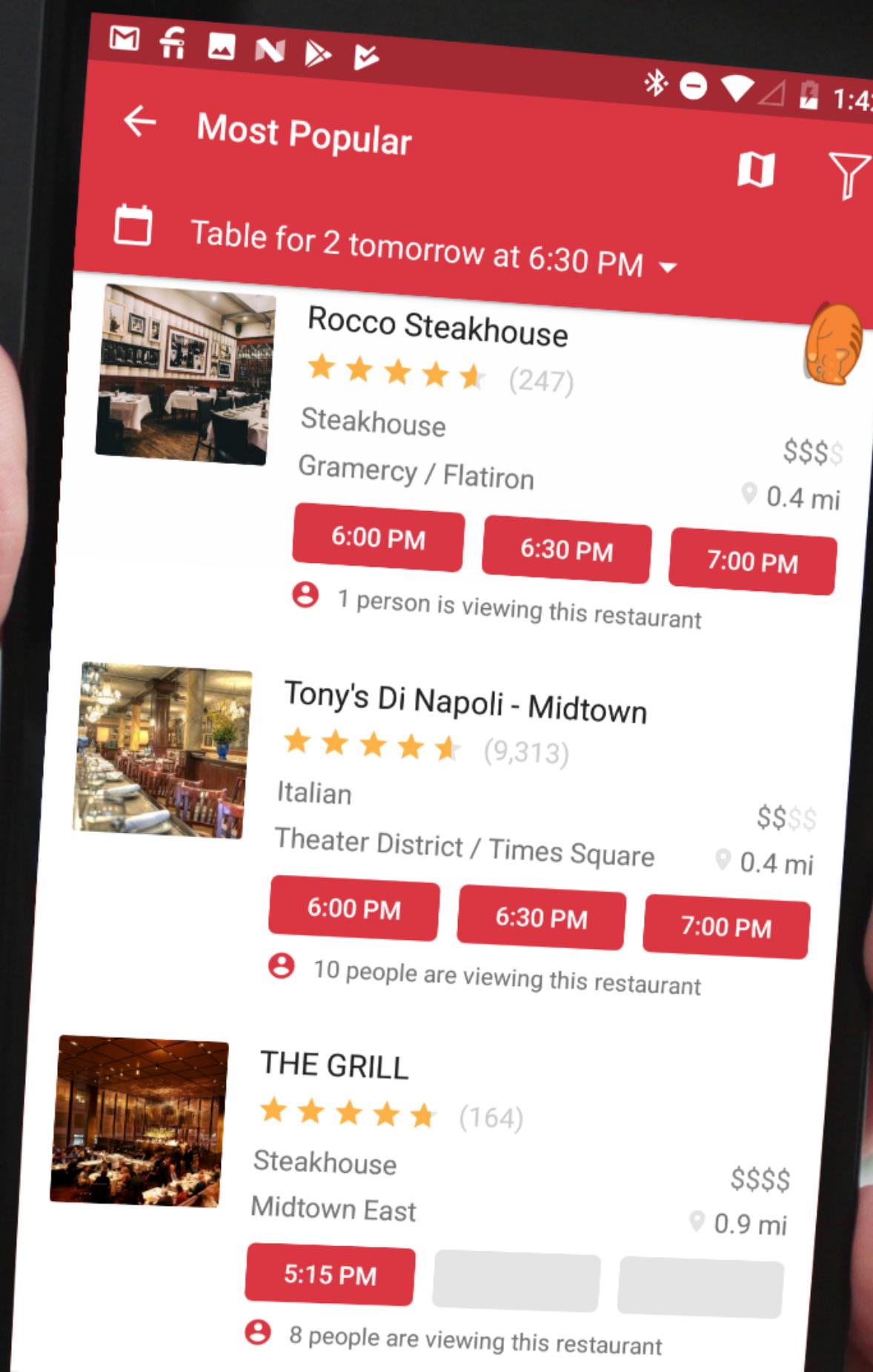
FirstName

City

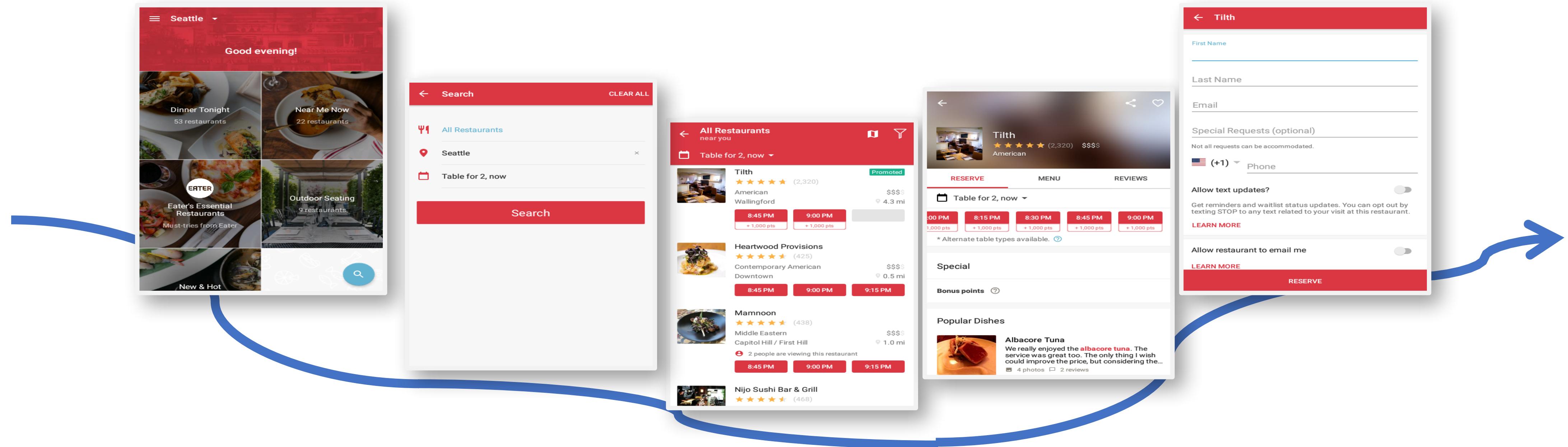
LastName

PersonCount...

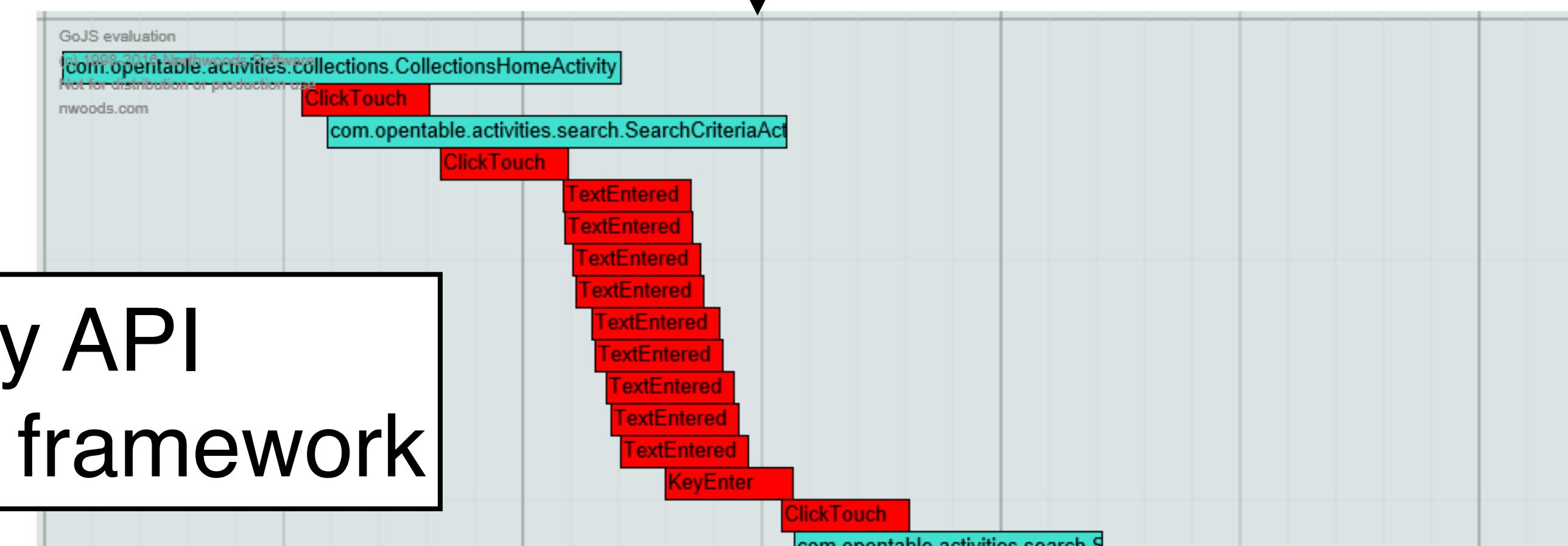
Email...



# Trace collection and aggregation



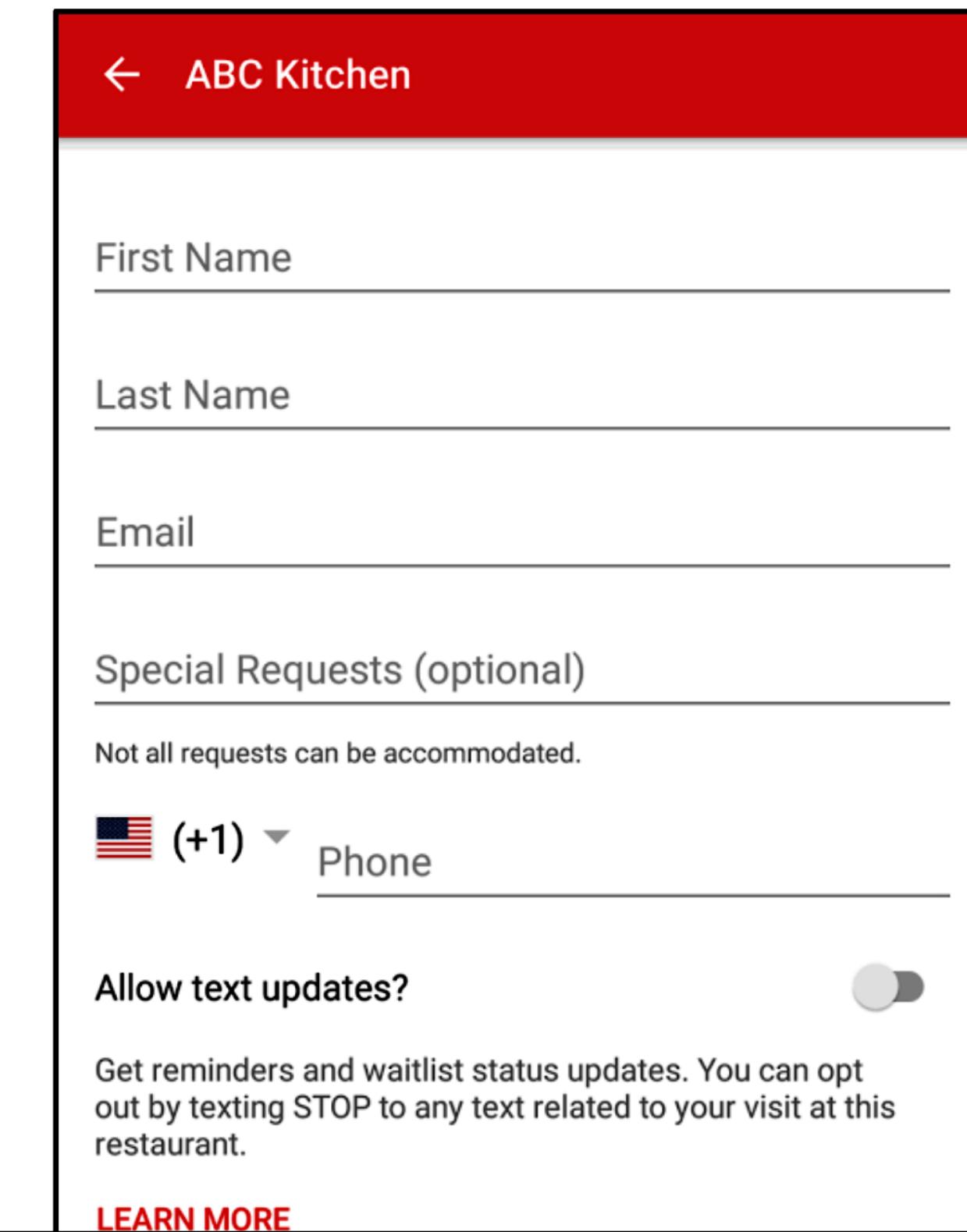
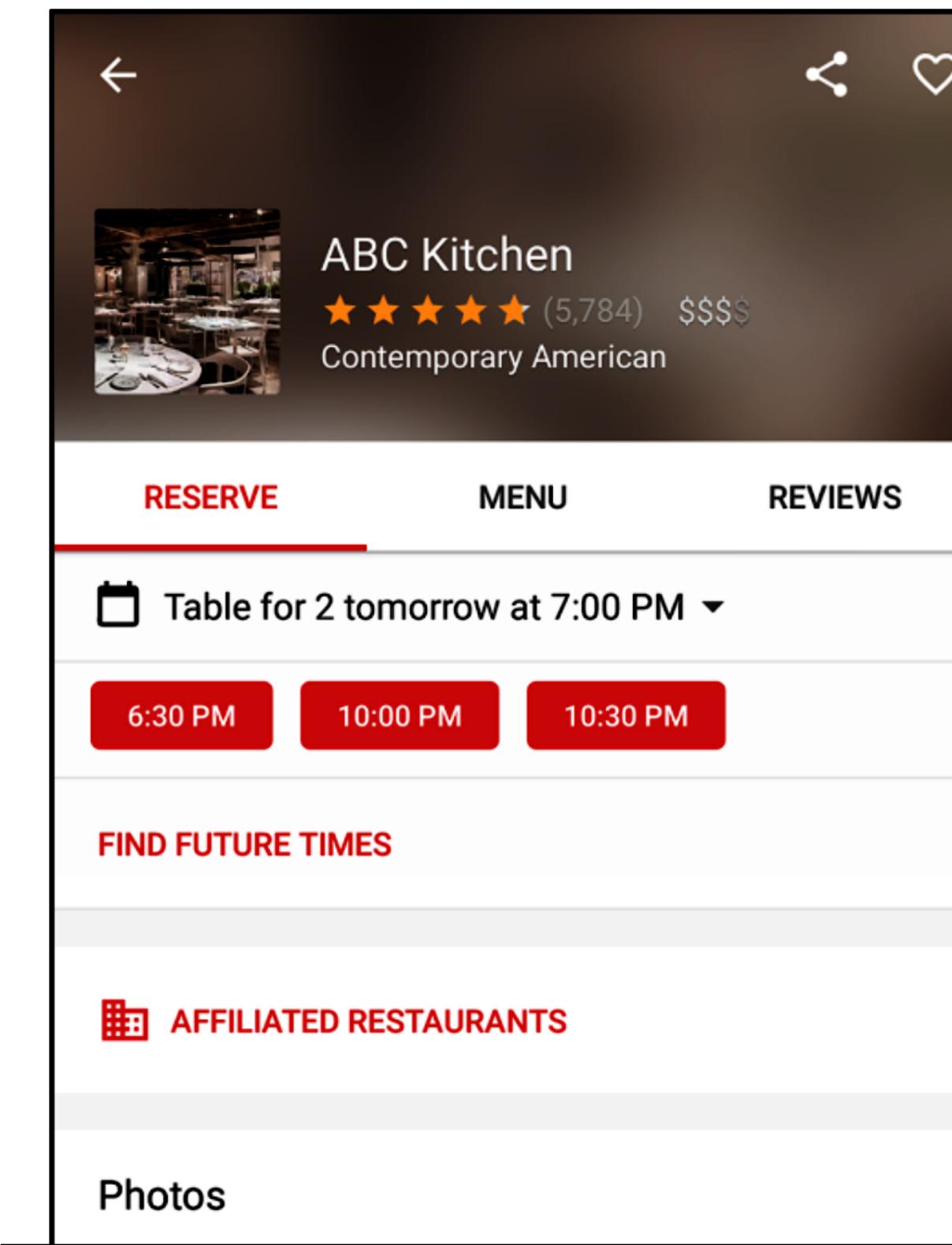
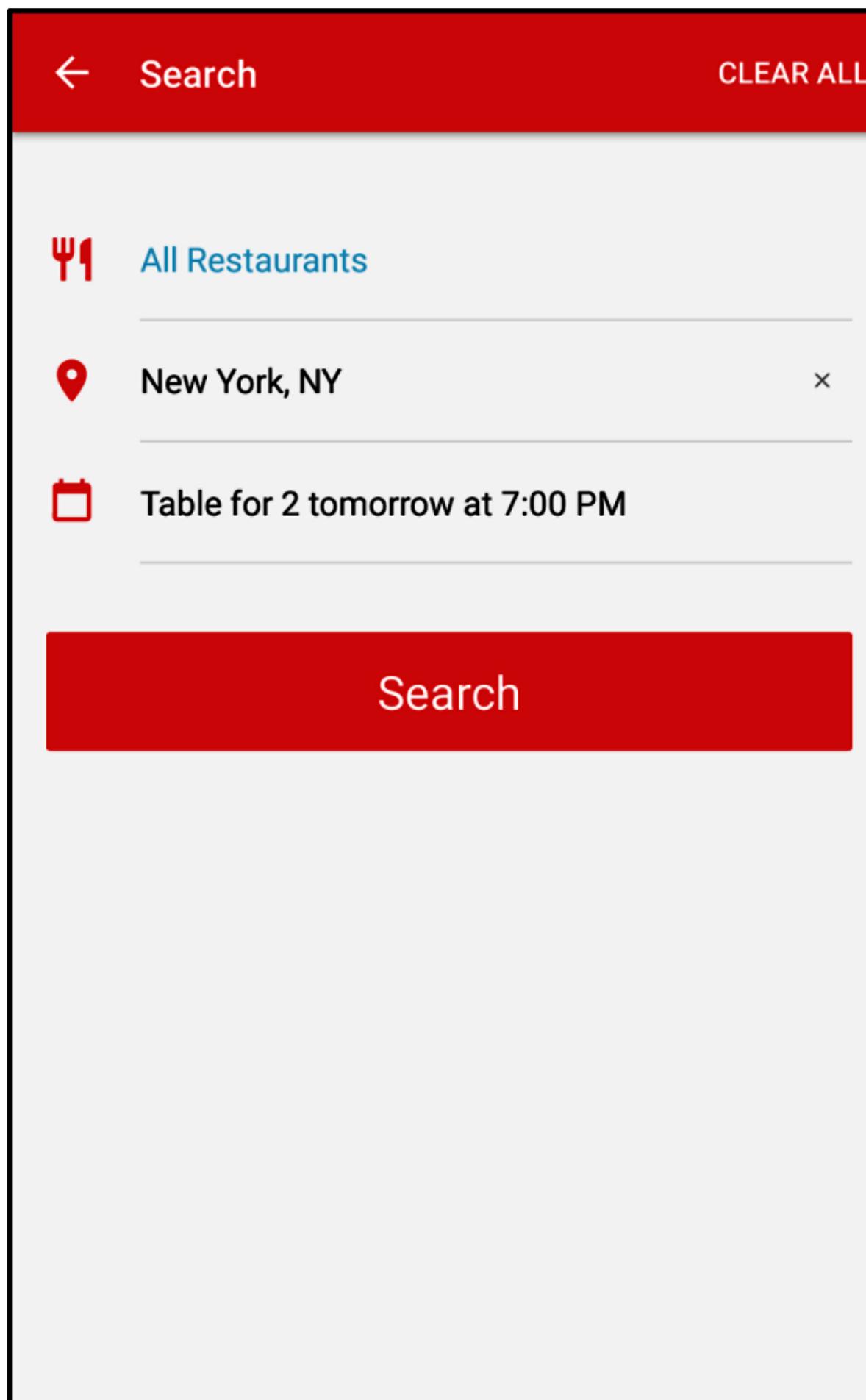
Trace Timeline



Android accessibility API  
Android application framework

UI event

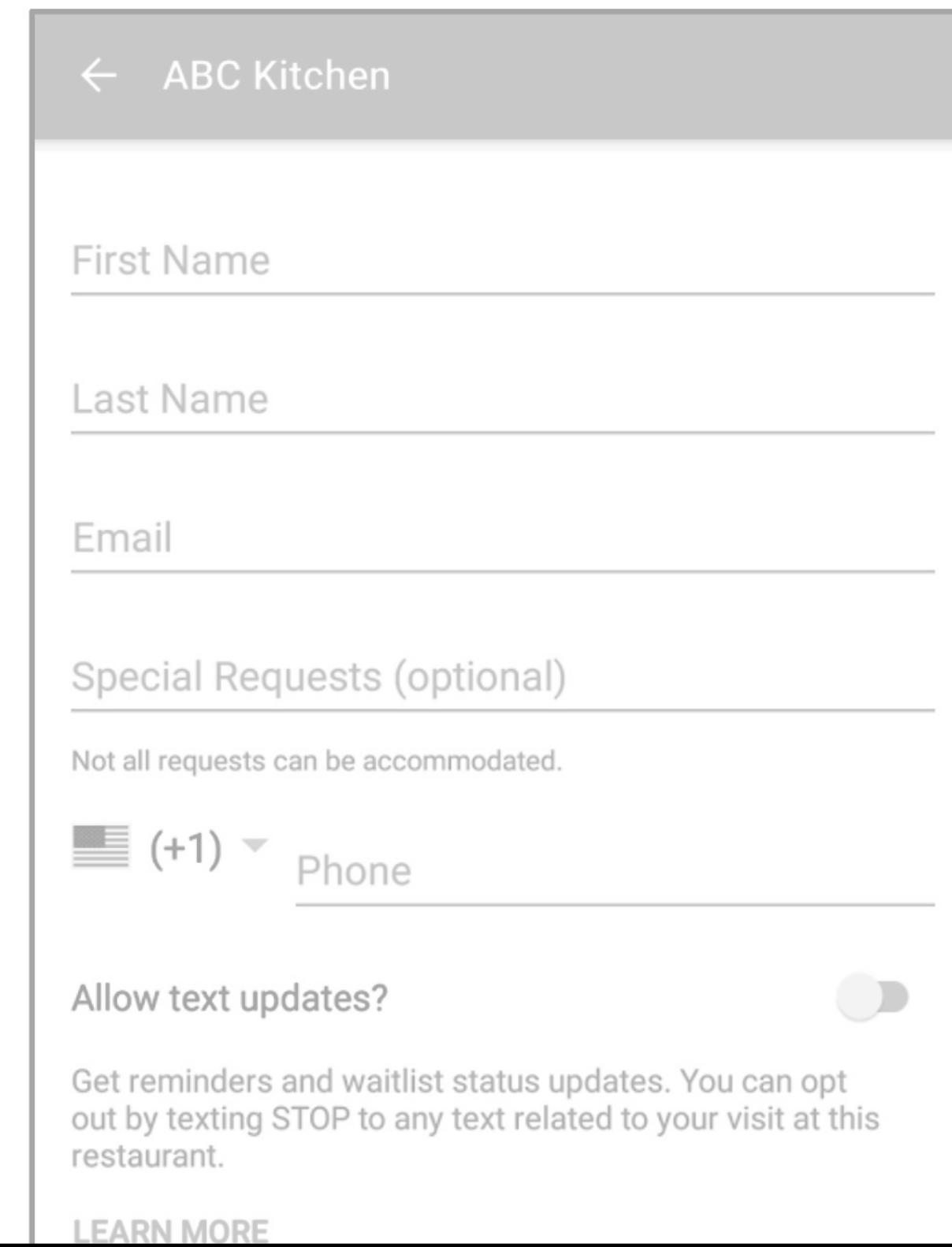
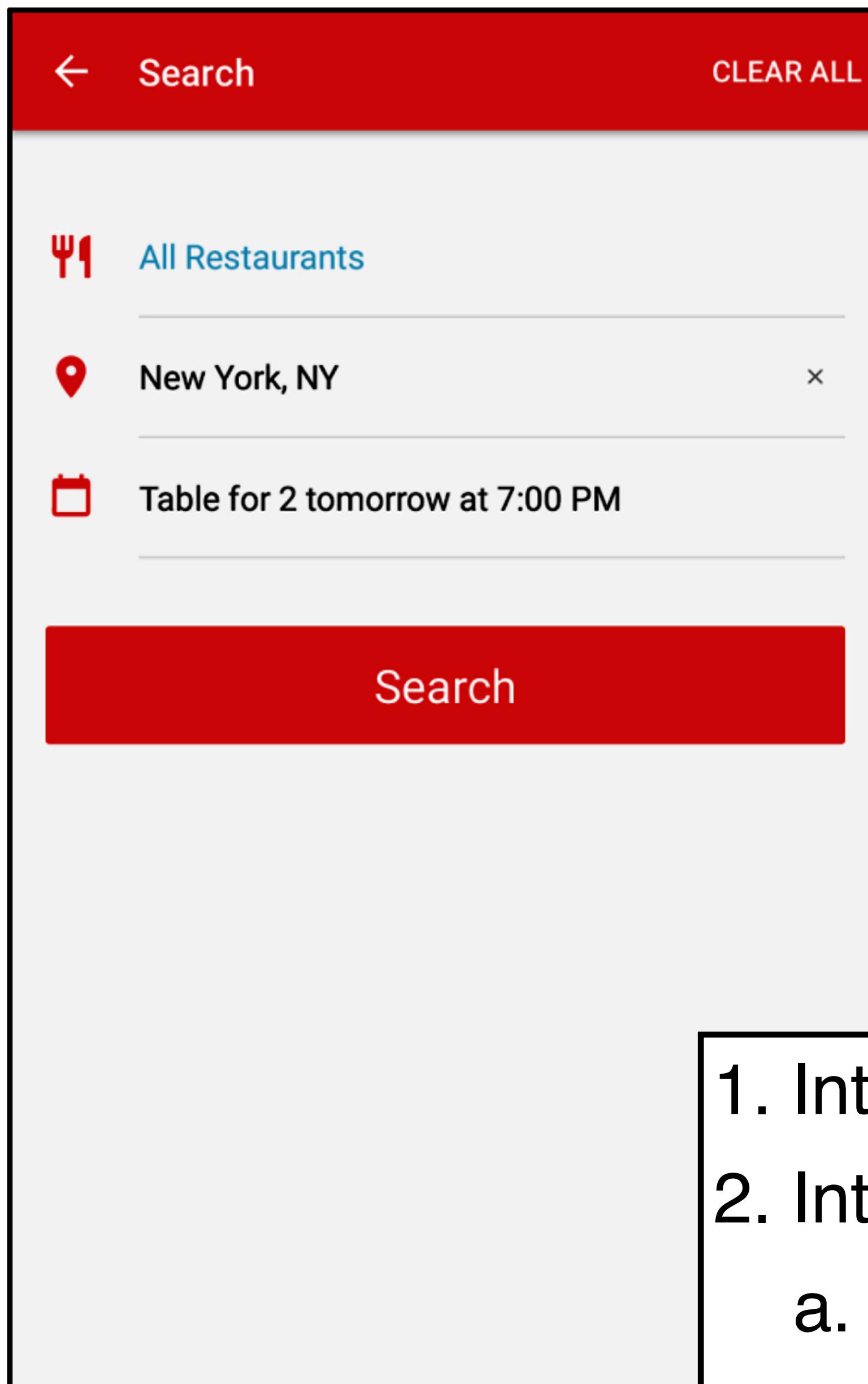
# Intent extraction



a)

1. Intents from activity (pages in an app) transitions
2. Intents within an activity
  - a. Sub-pages: tabs and fragments
  - b. Immutable buttons in a GUI that represents an action

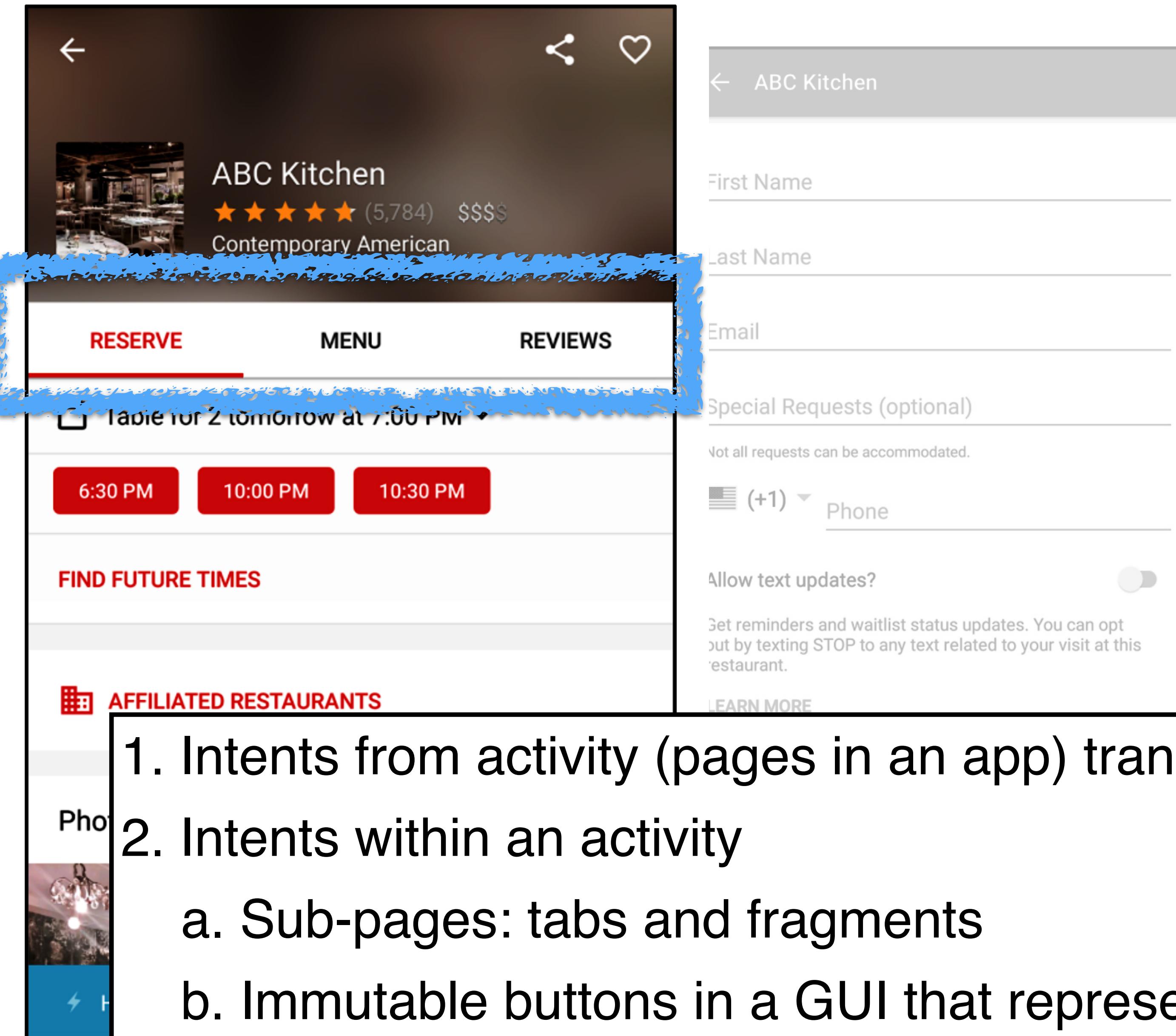
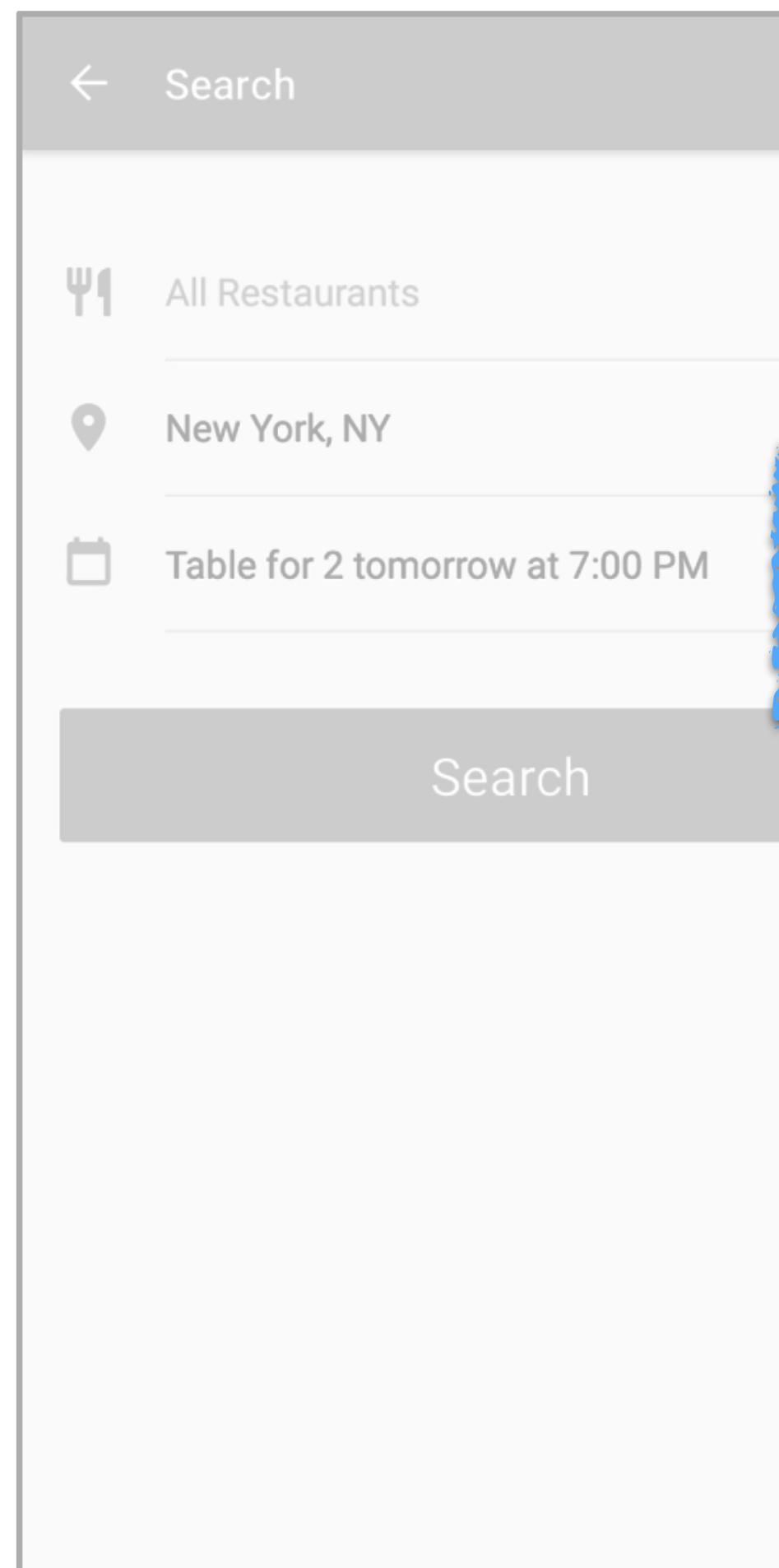
# Intent extraction



1. Intents from activity (pages in an app) transitions
2. Intents within an activity
  - a. Sub-pages: tabs and fragments
  - b. Immutable buttons in a GUI that represents an action

a)

# Intent extraction



a)

1. Intents from activity (pages in an app) transitions
2. Intents within an activity
  - a. Sub-pages: tabs and fragments
  - b. Immutable buttons in a GUI that represents an action

# Slot extraction

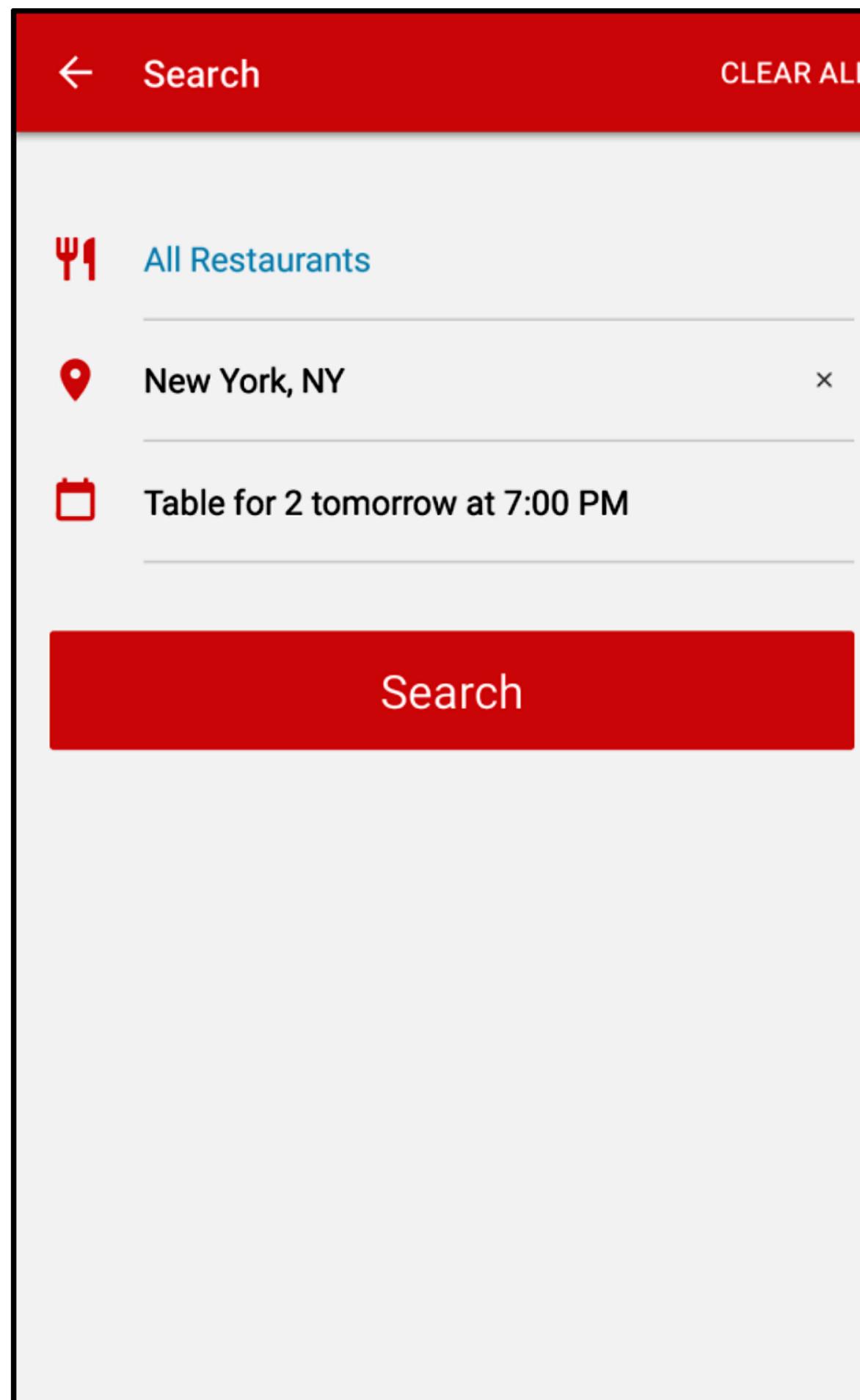
← Search      CLEAR ALL

All Restaurants

New York, NY

Table for 2 tomorrow at 7:00 PM

**Search**



← ABC Kitchen

ABC Kitchen  
★★★★★ (5,784) \$\$\$  
Contemporary American

**RESERVE**   **MENU**   **REVIEWS**

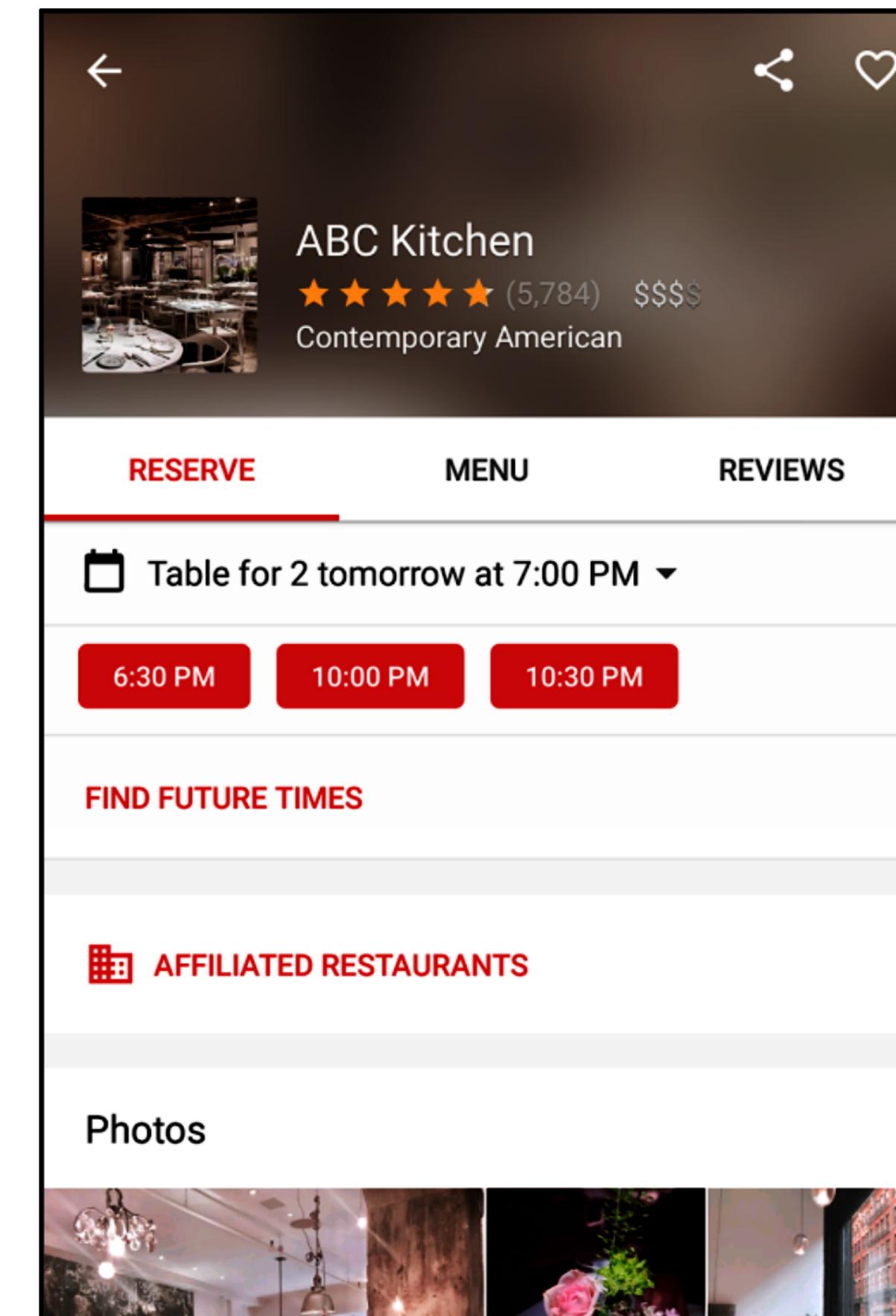
Table for 2 tomorrow at 7:00 PM ▾

6:30 PM   10:00 PM   10:30 PM

**FIND FUTURE TIMES**

**AFFILIATED RESTAURANTS**

**Photos**



← ABC Kitchen

First Name

Last Name

Email

Special Requests (optional)

Not all requests can be accommodated.

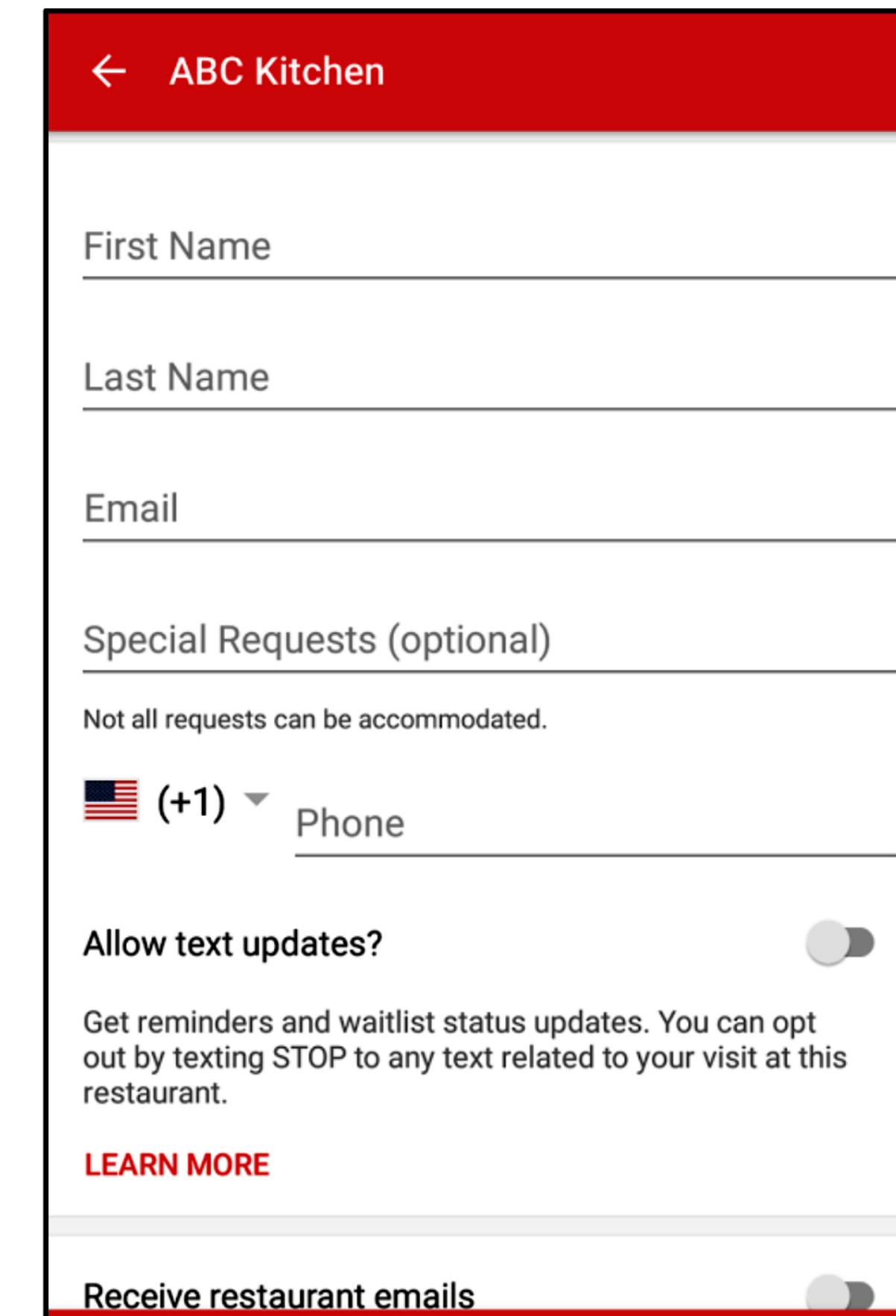
(+1) ▾ Phone

Allow text updates?

Get reminders and waitlist status updates. You can opt out by texting STOP to any text related to your visit at this restaurant.

**LEARN MORE**

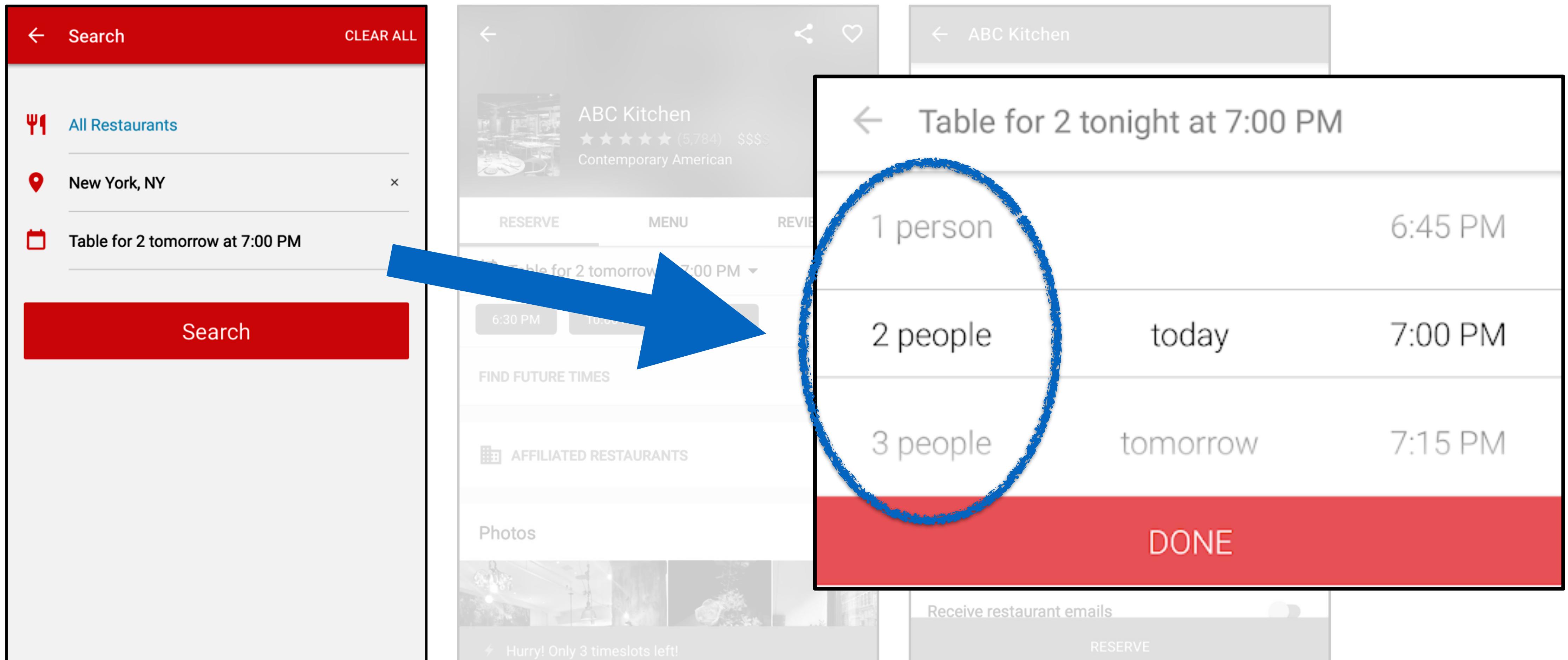
Receive restaurant emails



a)

1. Slots are determined by the UI element type and the structure of the UI layout

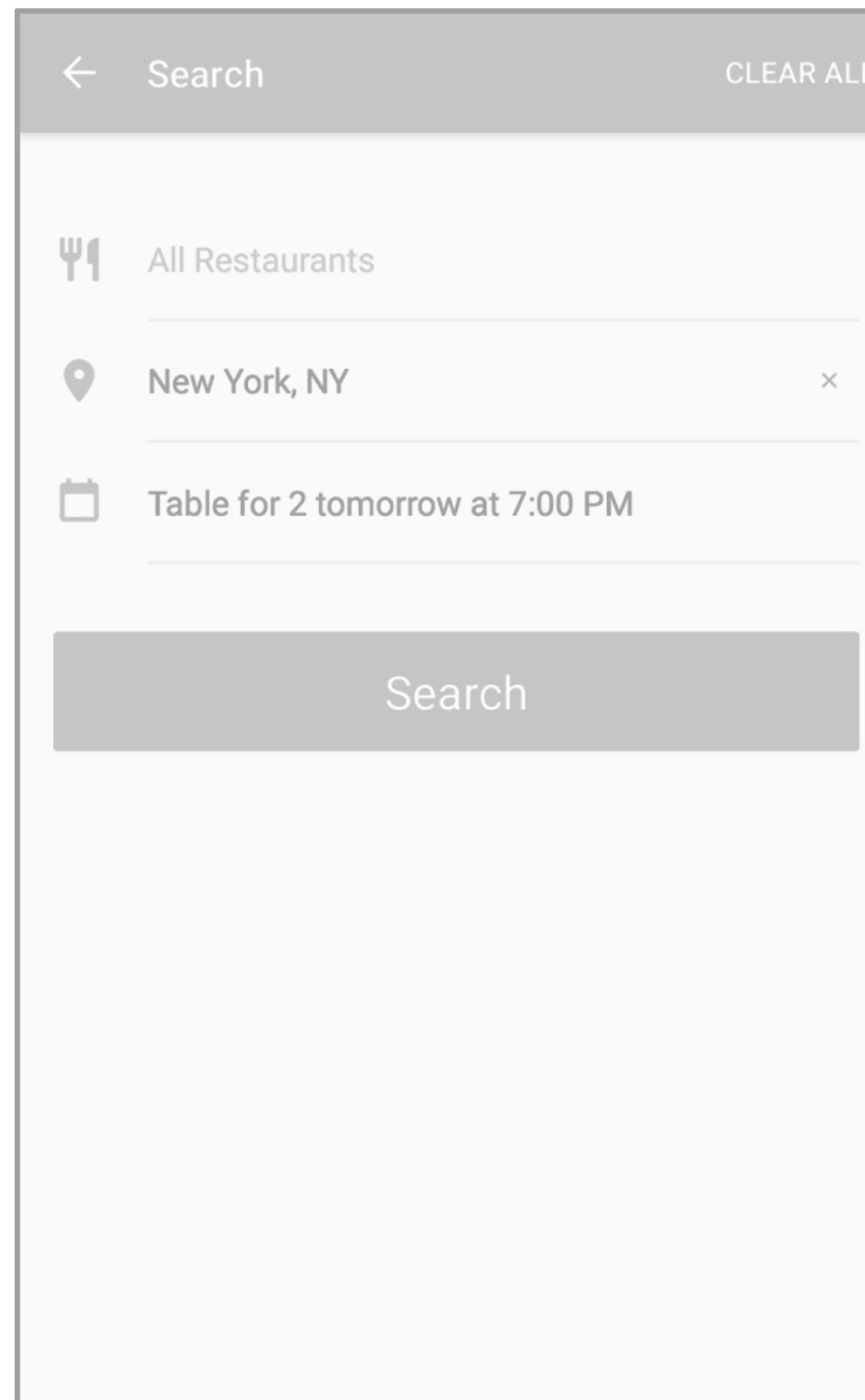
# Slot extraction



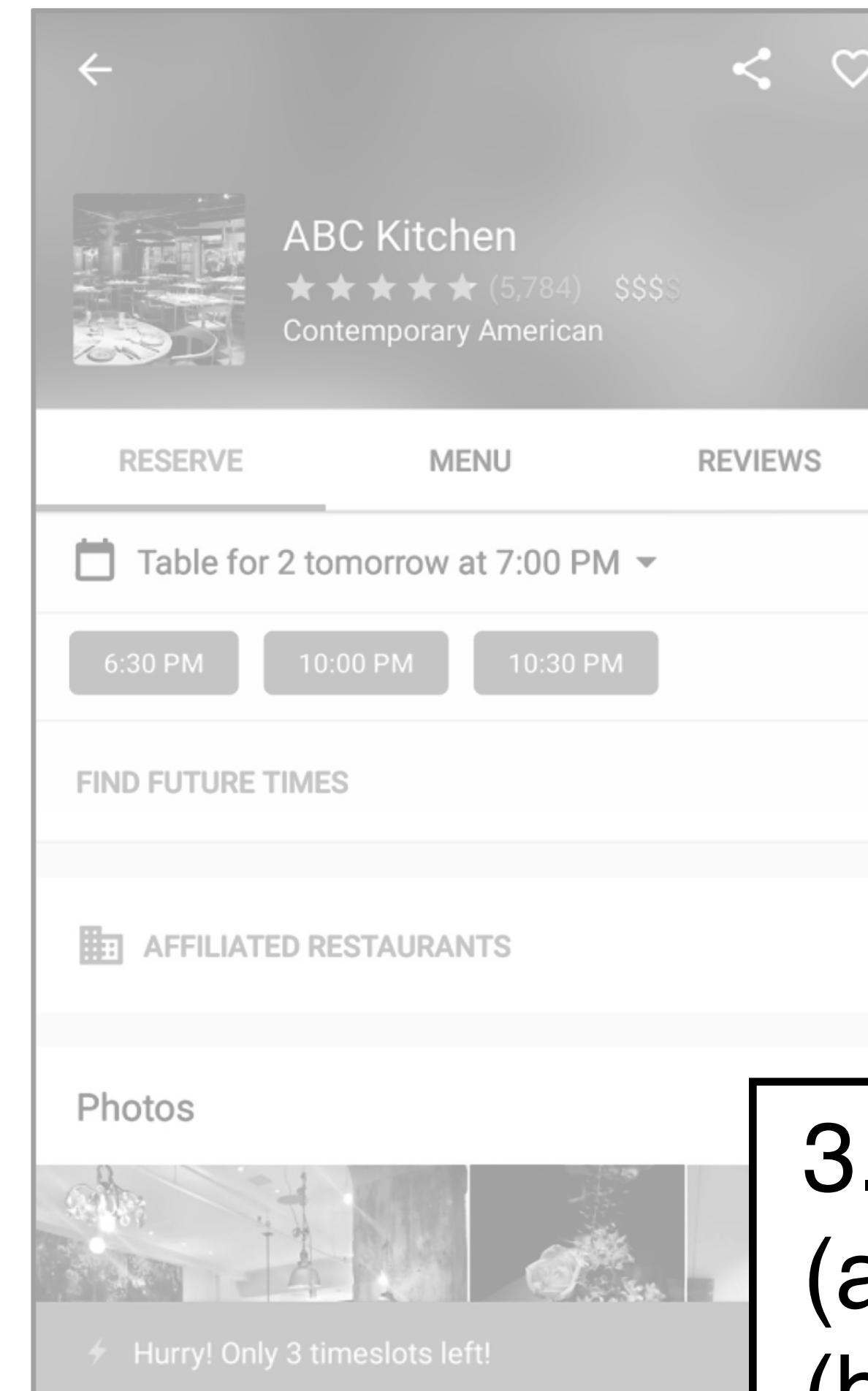
a)

2. Possible values are extracted from the GUI

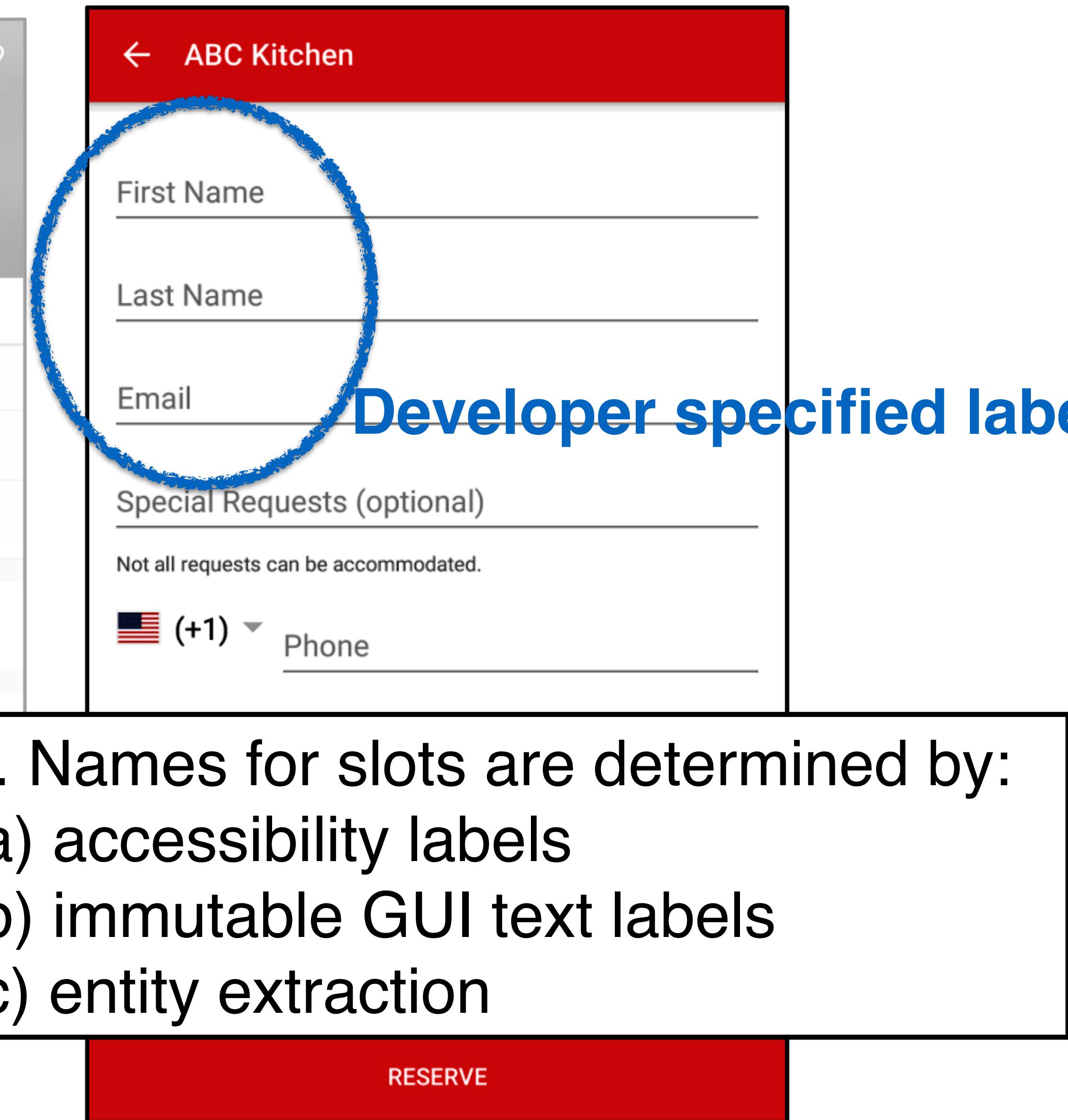
# Slot extraction



a)



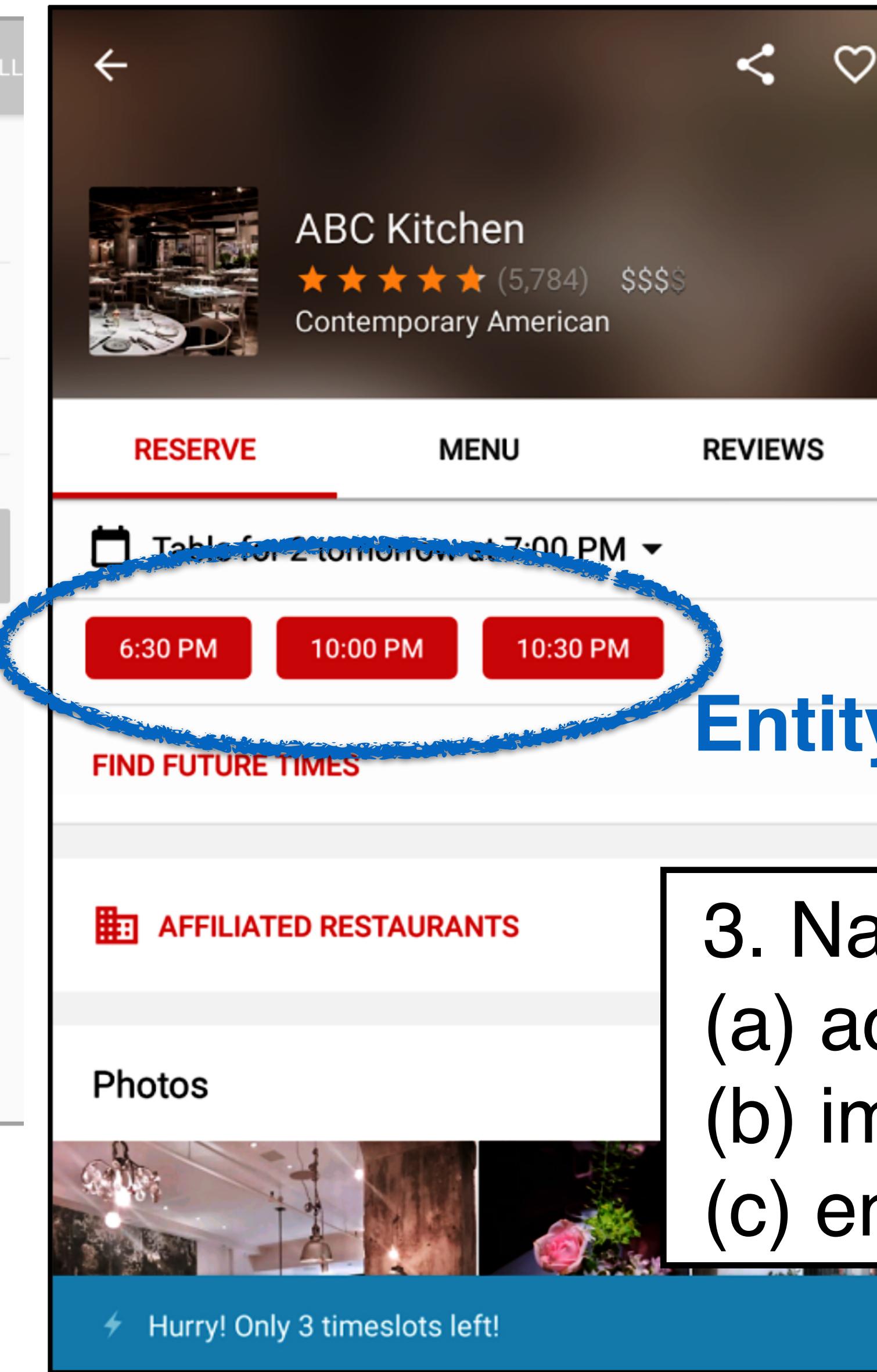
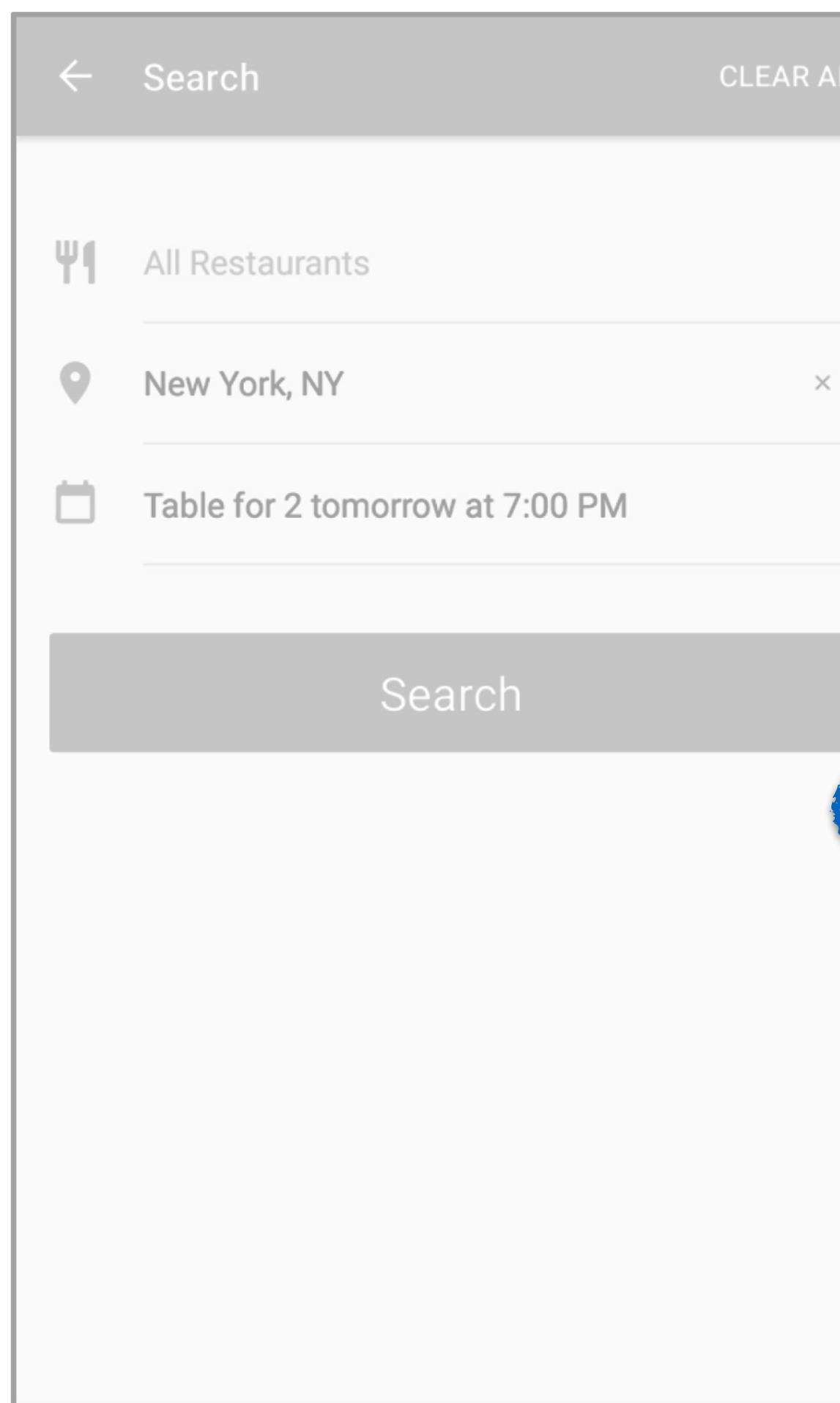
b)



3. Names for slots are determined by:

- (a) accessibility labels
- (b) immutable GUI text labels
- (c) entity extraction

# Slot extraction



ABC Kitchen

Name \_\_\_\_\_

Name \_\_\_\_\_

ial Requests (optional)

requests can be accommodated.

+1) ▾ Phone

minders and waitlist status updates. You can opt

texting STOP to any text related to your visit at this

rant.

**Entity extraction: type=TIME**

3. Names for slots are determined by:

- accessibility labels
- immutable GUI text labels
- entity extraction



# Bot Generator

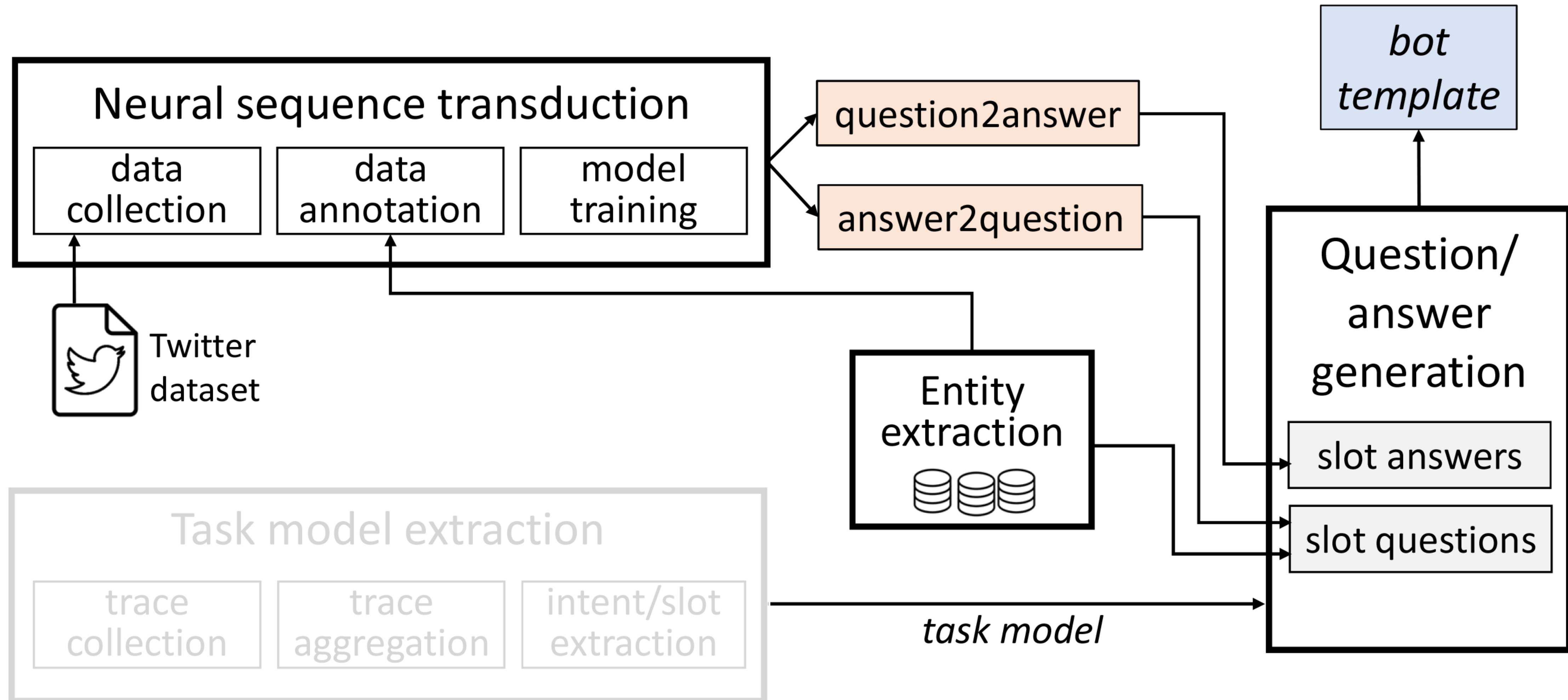
**Upload JSON File:**

no file selected

# Overview of developer experience

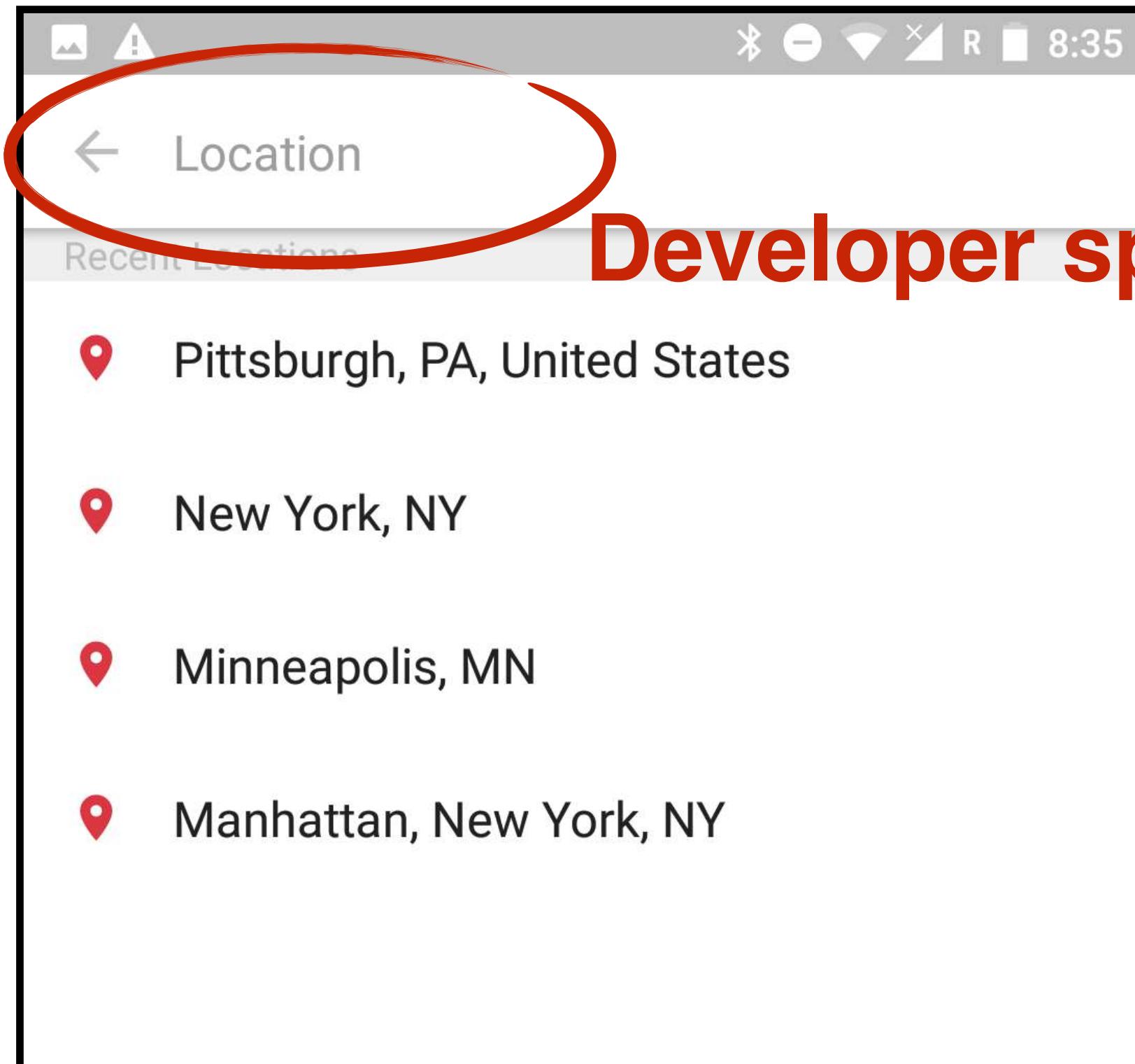


# KITE's architecture

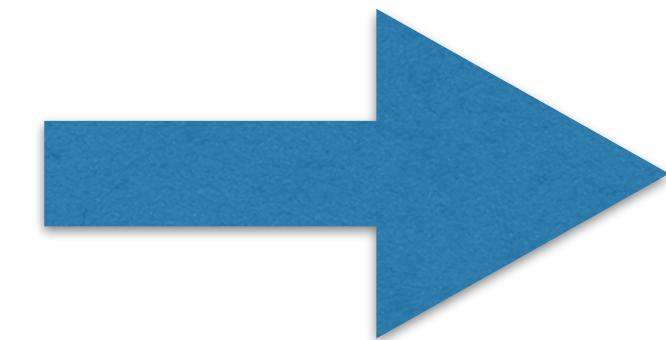


# Question generation

## Syntactic rule based approach



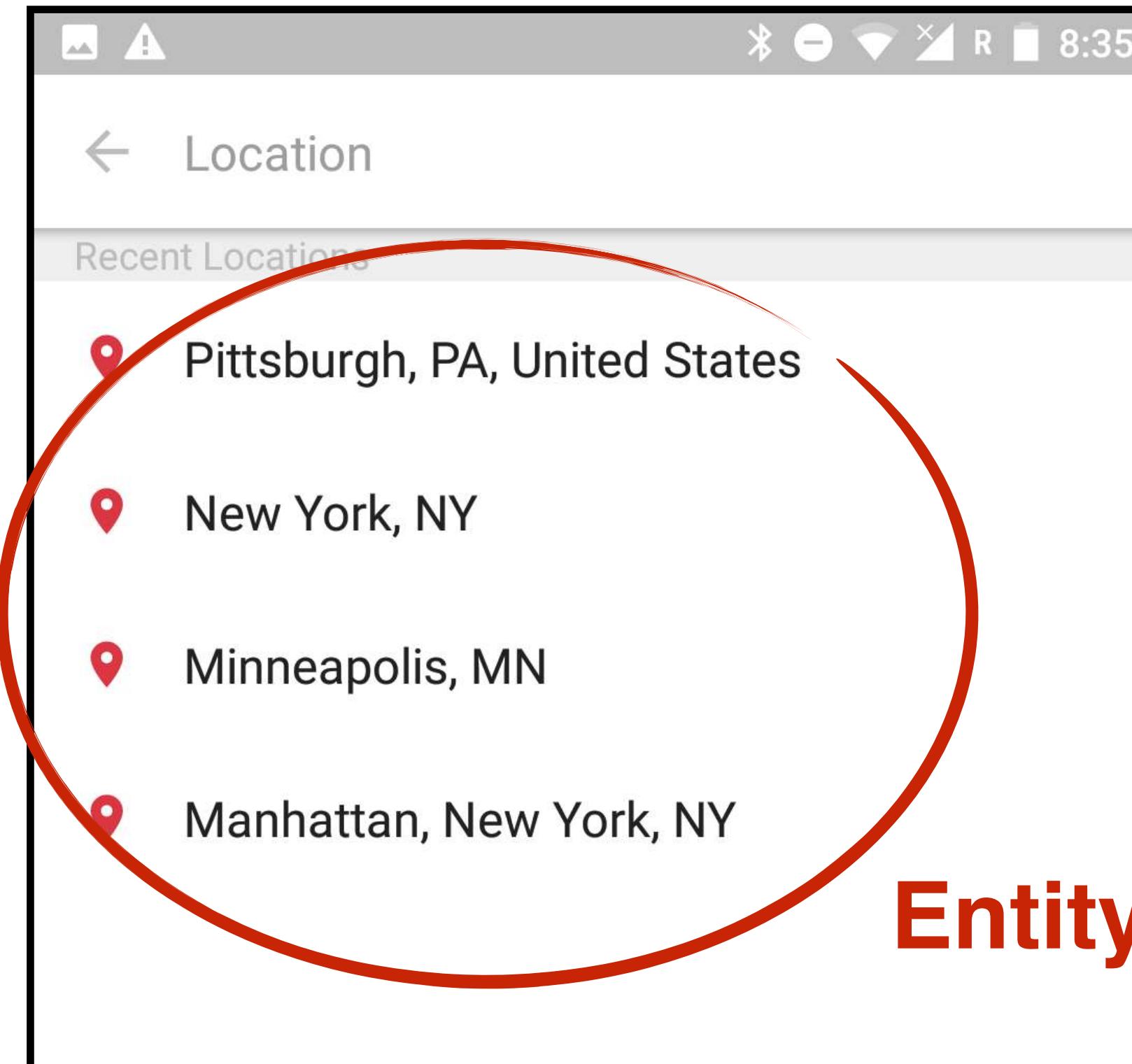
**Developer specified label: Location**



**What location do you want to choose?**

# Question generation

## Syntactic rule based approach



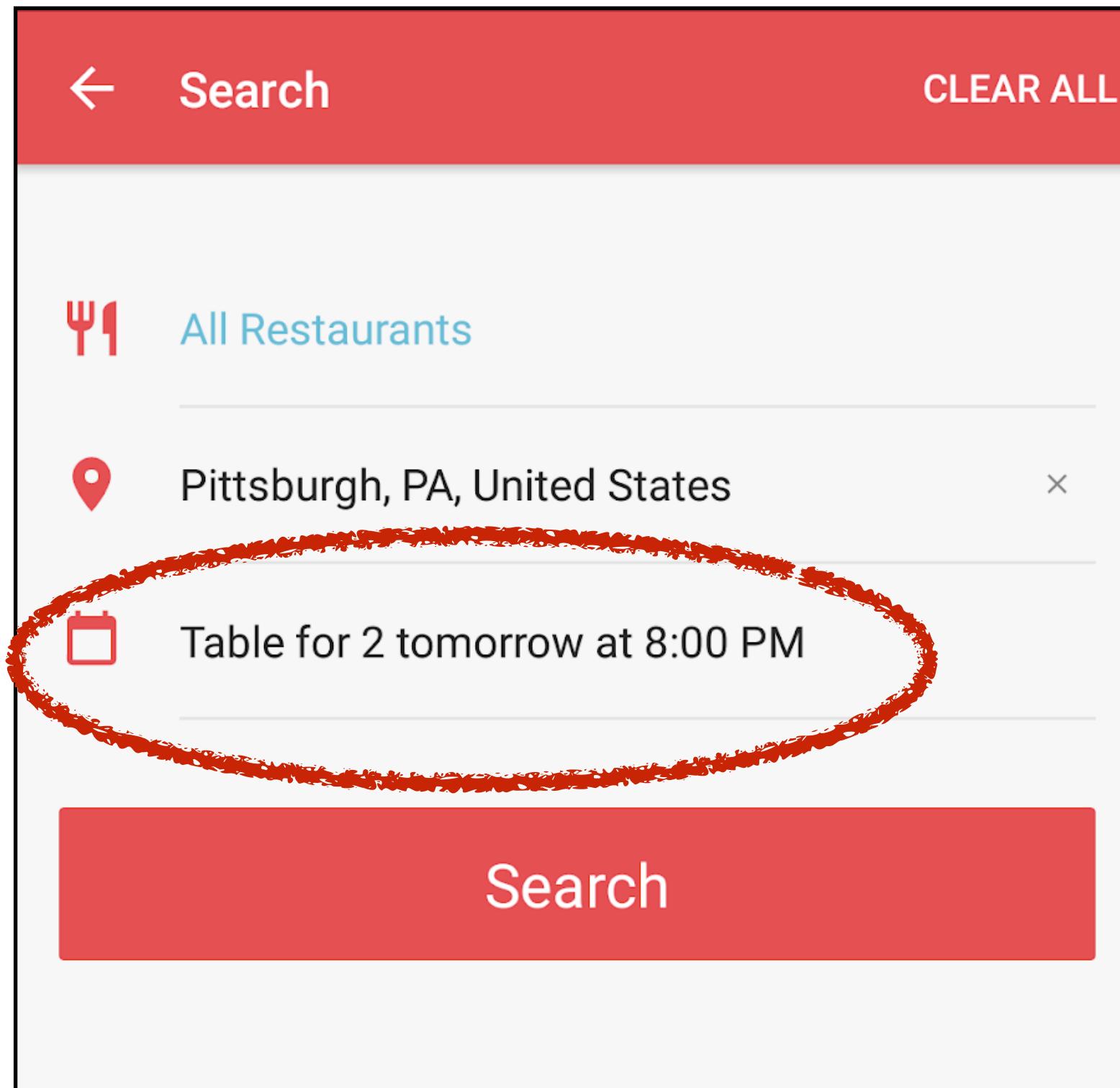
**Entity extraction: type=CITY**

What location do you want to choose?

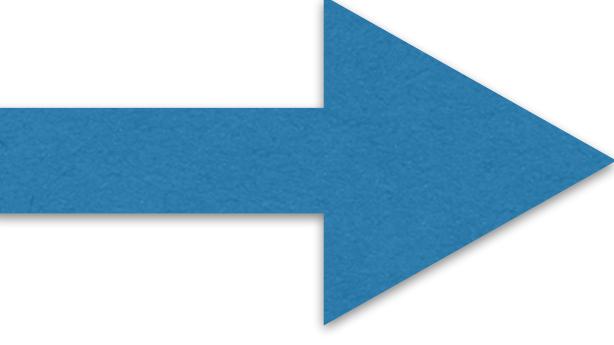
**What city do you want to choose?**

# Question generation

## Neural sequence transduction



**Domain:**  
Restaurant

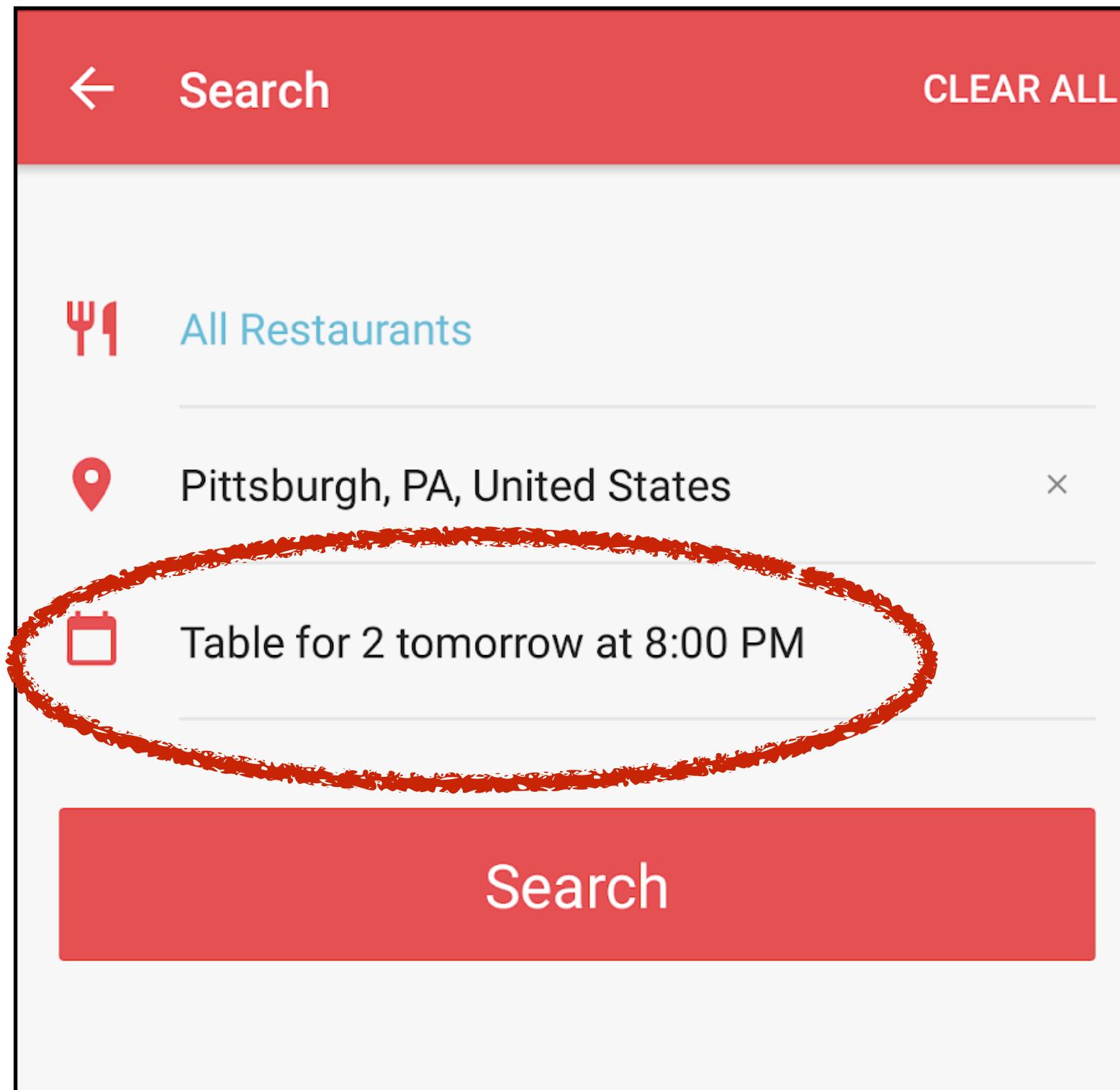


**Entity:**  
DATE, TIME

how many are in your party? what time?  
how many are you and what time are you  
planning on coming in?  
y all got a table for two in ten minutes?  
what time do you close?  
how many people are in line?  
...

# Question generation

## Neural sequence transduction



**Domain:**

Restaurant

**Entity:**

DATE, TIME



how many are in your party? what time?

✓ **how many are you and what time are you planning on coming in?**

y all got a table for two in ten minutes?

what time do you close?

how many people are in line?

...

# Overview of developer experience



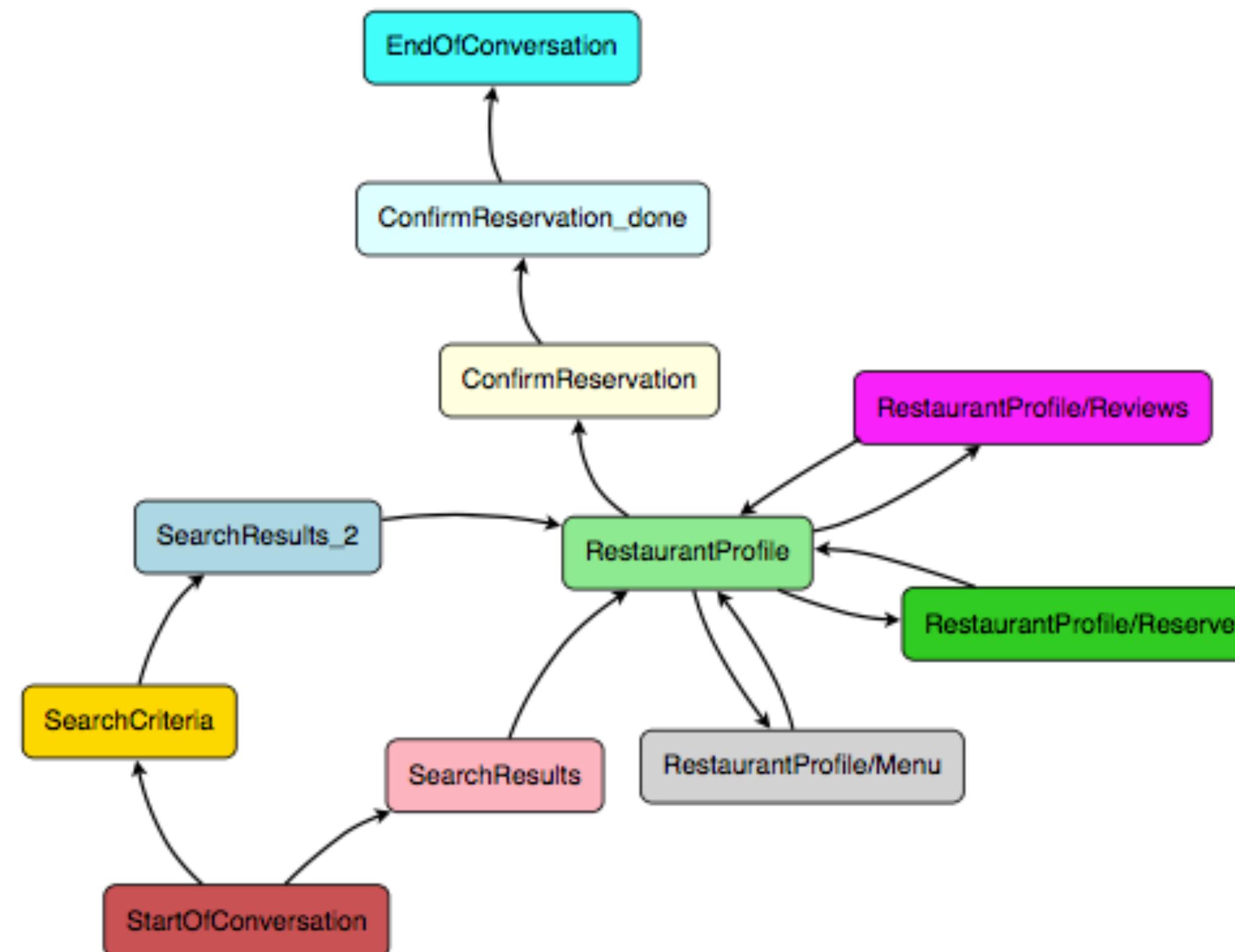


# Bot Generator

**Upload JSON File:**

**bot (1).json**

GoJS 1.7 evaluation  
(c) 1998-2017 Northwoods Software  
Not for distribution or production use  
nwoods.com



## Intent Name: SearchResults\_2 ...

### Intent ID:

SearchCriteria-SearchResults

### Slot Name: search query ...

### Slot Name: city ...

#### Slot ID:

com\_opentable\_activities\_search\_SearchCriteriaActivity-to-con

#### Slot Possible Values:

Miami, FL



New York, NY



#### \* Prompt in use:

Which city do you want to choose?



#### Alternative prompts:

where are you located ?



which location ?



...

where are you ?



where do you usually go for thai food ?



which location did you visit ?



what is your location ?



Intent Name: SearchResults\_2  ...

Intent ID:

**Slot Name:** search query  ...

**Slot Name:** city  ...

**Slot ID:**

com\_opentable\_activities\_search\_SearchCriteriaActivity-to-con

**Slot Possible Values:**

Miami, FL



New York, NY



where are you ?



where do you usually go for thai food ?



which location did you visit ?



what is your location ?



Intent Name: SearchResults\_2  ...

Intent ID:

SearchCriteria-SearchResults

Slot Name: search query  ...

**\* Prompt in use:**

Which city do you want to choose?



*Alternative prompts:*

where are you located ?



which location ?



...

where are you ?



where do you usually go for thai food ?



which location did you visit ?



what is your location ?



Intent Name: SearchResults\_2 

Intent ID:

SearchCriteria-SearchResults

Slot Name: search query  ...

Slot Name: city  ...

Slot ID:

com\_opentable\_activities\_search\_SearchCriteriaActivity-to-con

Slot Possible Values:

Miami, FL



New York, NY



...

localhost:8000 says

Please enter the new intent name:

Search for restaurant by criteria

Cancel

OK



## Bot Preview

Welcome to the bot!

Do you want to "search for restaurants by criteria" or "choose from recommended restaurants"?

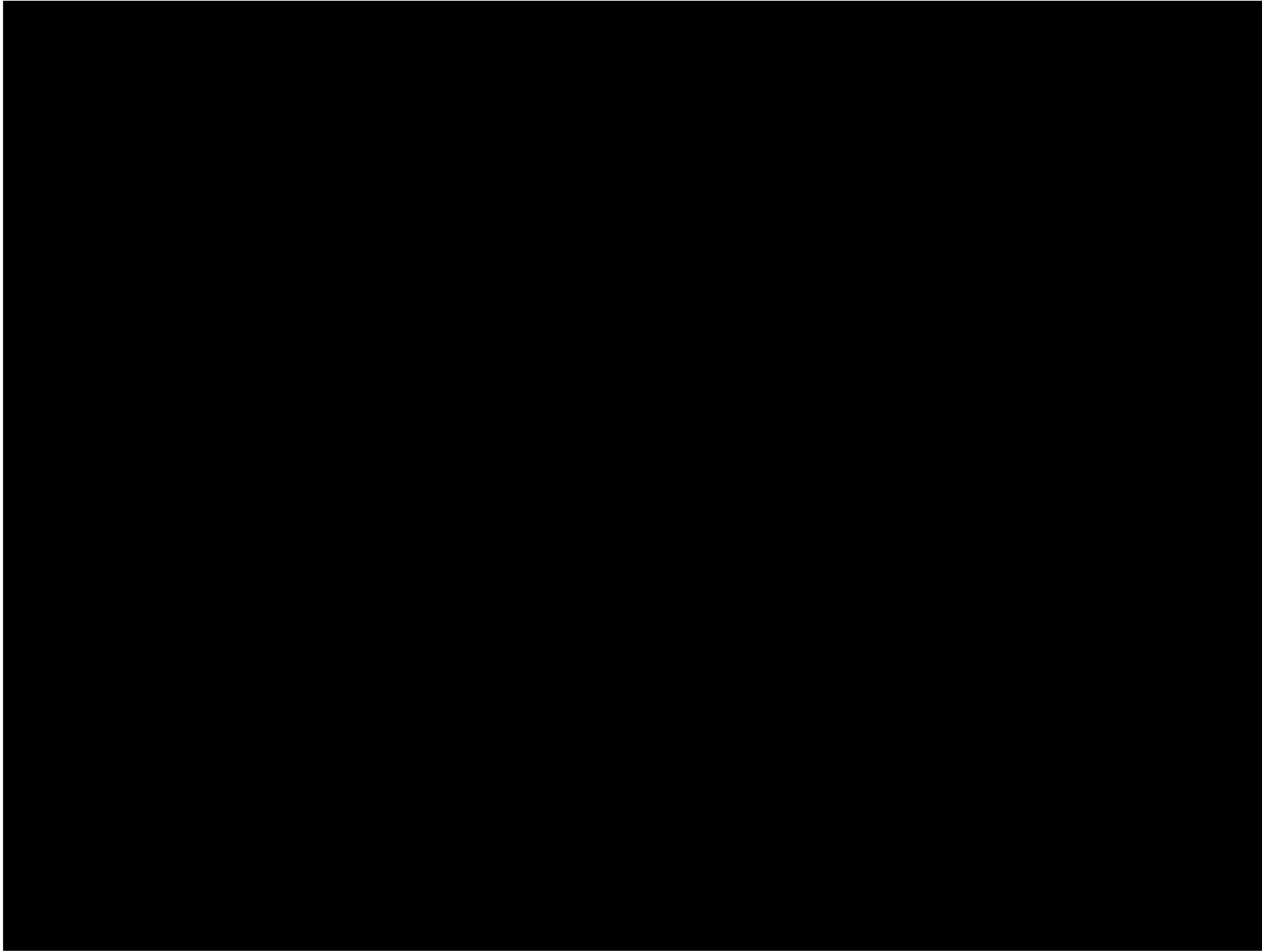
|

Send

# Multi-modal interfaces

- Interfaces that offer multiple modalities (e.g., speech, gaze, gesture, hand writing, touch, mouse pointing...) for interacting with the system
- Multi-modal interfaces provide users with greater expressive power, naturalness, flexibility, and portability
- Mutual disambiguation: use inputs from one modality to disambiguate inputs from another modality, and vice versa

Sharon Oviatt. Mutual disambiguation of recognition errors in a multimodel architecture. In *CHI '99*.  
Sharon Oviatt. 1999. Ten myths of multimodal interaction. *Commun. ACM* 42, 11 (November 1999), 74–81



## Put That There (1980) <https://youtu.be/sC5Zg0fU2e8>

Bolt, Richard A. "Put-that-there": Voice and gesture at the graphics interface. SIGGRAPH Comput. Graph. Vol. 14. No. 3. ACM, 1980.

**DreamSpace, 1998**  
**Natural Interaction**  
**Mark Lucente**  
**IBM Research**

Yorktown Heights, New York USA

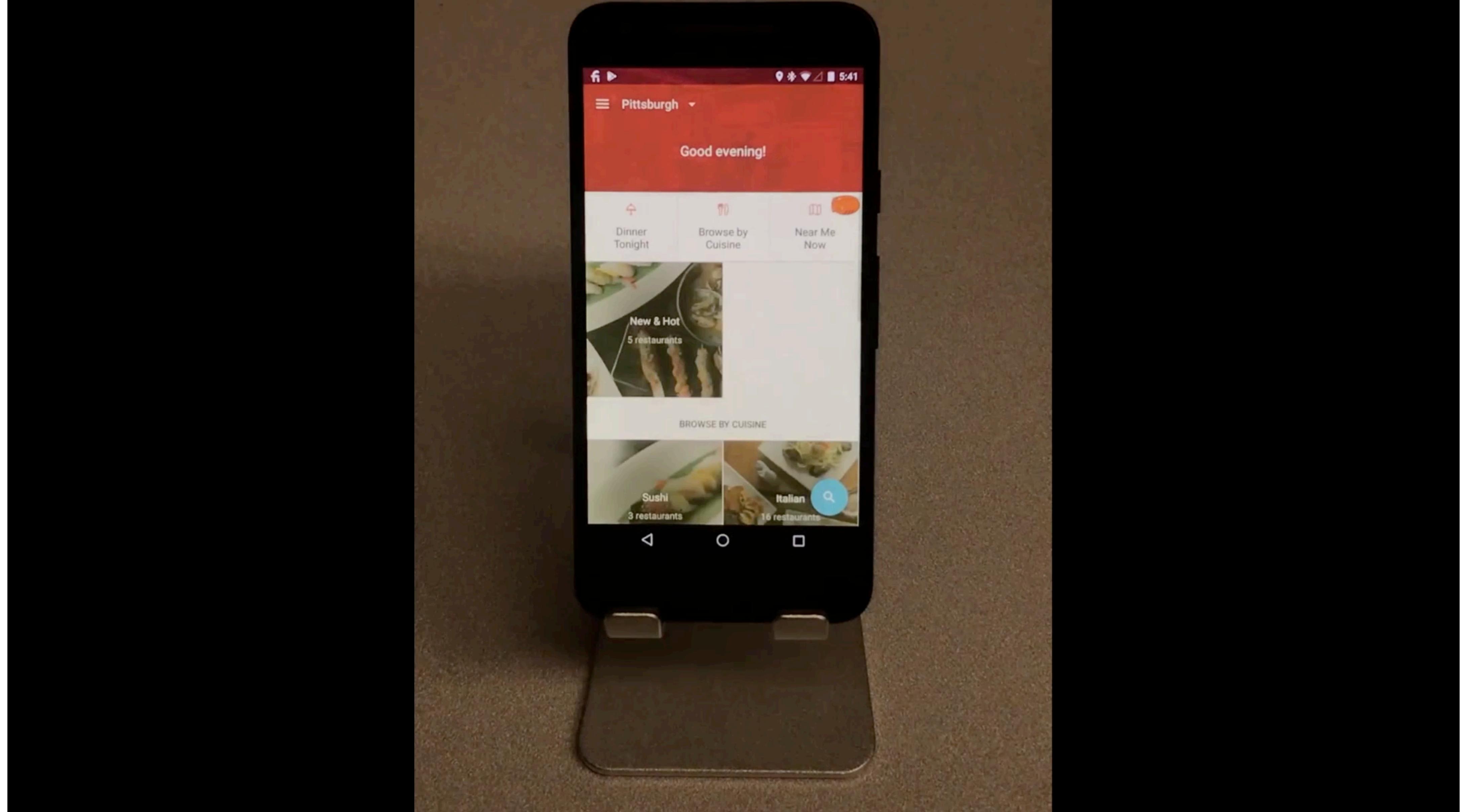
These video clips document the DreamSpace system,  
created by (and demonstrated by) Mark Lucente.

[www.lucente.us](http://www.lucente.us)

DreamSpace (1998)

<https://youtu.be/AO5zl0i6vzc>

Lucente, Mark, Gert-Jan Zwart, and Andrew D. George. "Visualization space: A testbed for deviceless multimodal user interface." Intelligent Environments Symposium. Vol. 98. 1998.

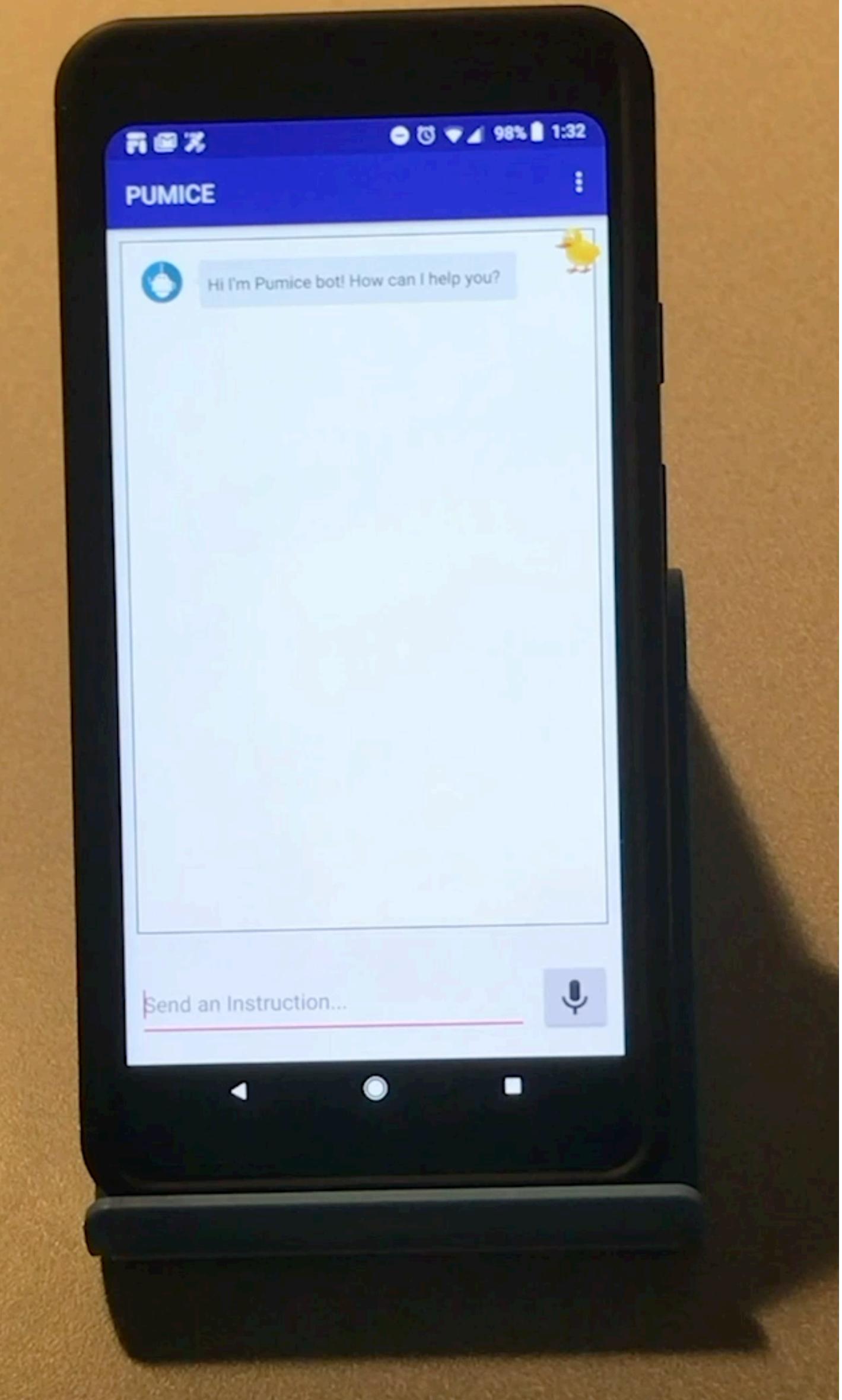


## APPINITE (2018) <https://youtu.be/2GqMUiPvidU>

Toby Jia-Jun Li, Igor Labutov, Xiaohan Nancy Li, Xiaoyi Zhang, Wenze Shi, Wanling Ding, Tom M. Mitchell, and Brad A. Myers. APPINITE: A Multi-Modal Interface for Specifying Data Descriptions in Programming by Demonstration Using Natural Language Instructions. In VL/HCC 2018

**Task: Order hot or iced coffee depending on the current temperature**

**User:** *If the weather is hot, order a cup of iced coffee.*



**PUMICE (2019)** <https://youtu.be/BAC2ZuJGY4M>

Toby Jia-Jun Li, Marissa Radensky, Justin Jia, Kirielle Singarajah, Tom M. Mitchell, and Brad A. Myers. PUMICE: A Multi-Modal Agent that Learns Concepts and Conditionals from Natural Language and Demonstrations. In *UIST 2019*.



WorldGaze (2020) <https://youtu.be/kjACtQK3D-k>

Mayer, Sven; Laput, Gierad; Harrison, Chris. 2020. Enhancing Mobile Voice Assistants with WorldGaze. CHI 2020.

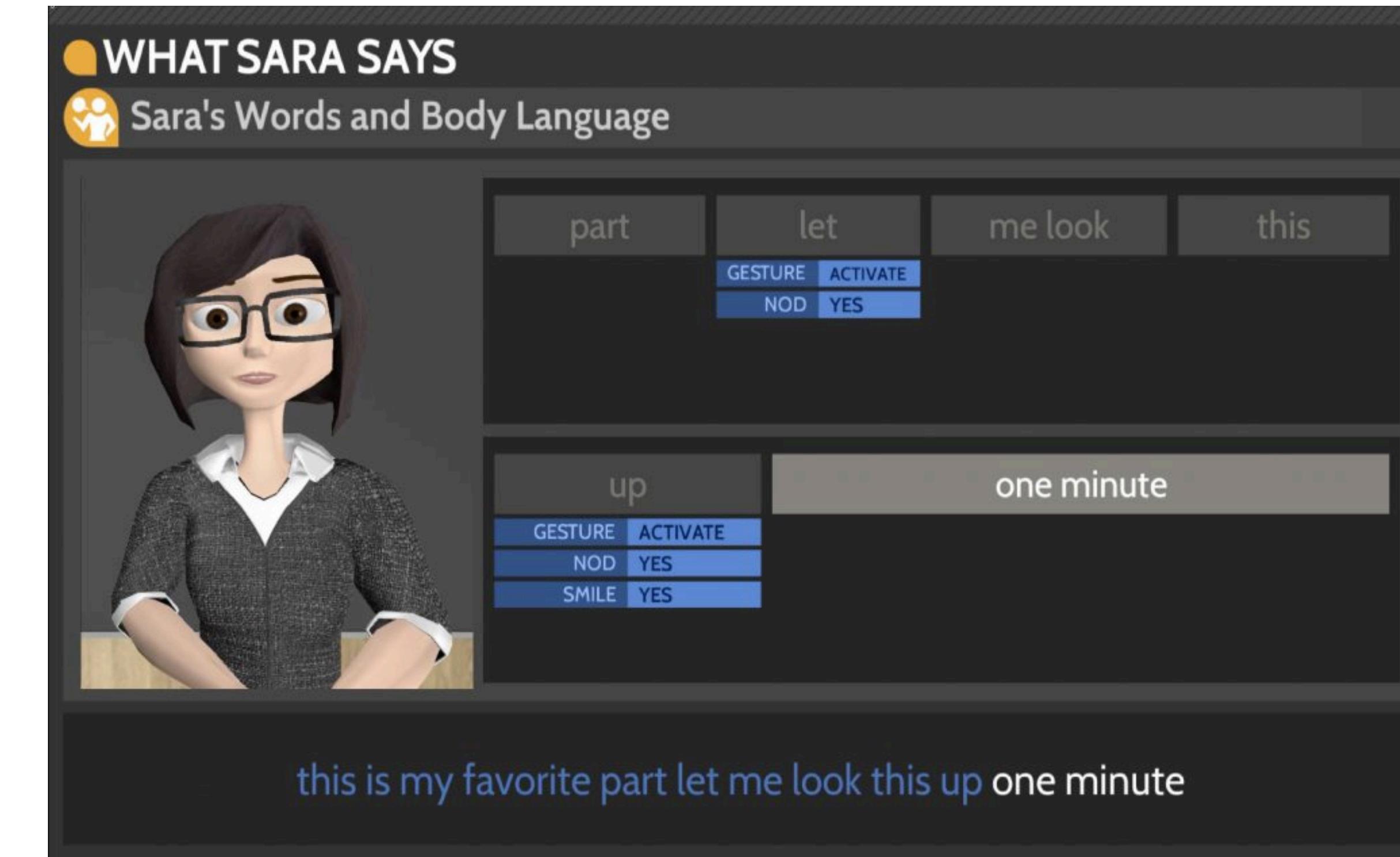
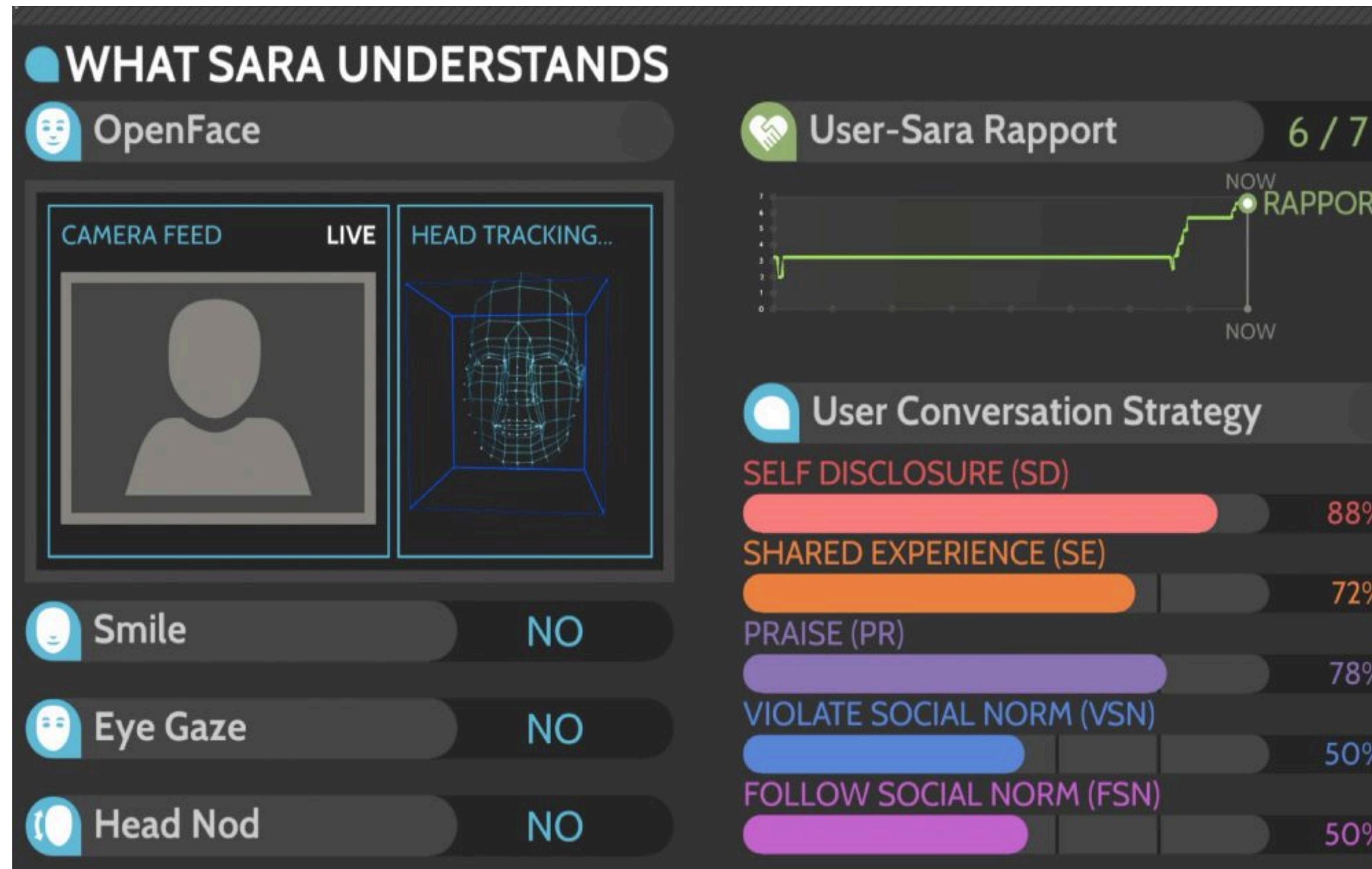
# Embodied conversational agents



jibo



# Embodied conversational agents



SARA (2016)

Zhao, R., Sinha, T., Black, A., & Cassell, J. Socially-Aware Virtual Agents: Automatically Assessing Dyadic Rapport from Temporal Patterns of Behavior, 16th International Conference on Intelligent Virtual Agents (IVA 2016)