

Chapter 26: Conversational Agents

Conversational Agents AKA Dialog Agents

Phone and voice based personal assistants
SIRI, Alexa, Cortana, Google Assistant
Talking to your car
Communicating with robots
Clinical uses for mental health
Chatting for fun

Two classes of systems

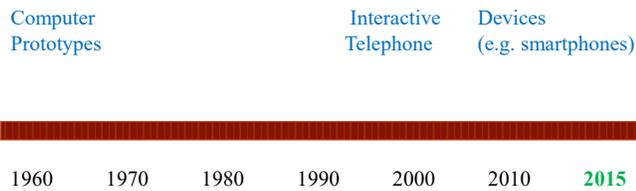
1. Chatbots
2. (Goal-based) Dialog agents
 - *SIRI, interfaces to cars, robots,*
 - *booking flights or restaurants*

The word "chatbots" is sometimes used in the popular press for both. We'll use it only for #1.

Spoken Dialogue Systems

- Computer systems that can engage in extended human-machine conversations
- Benefits of speech as an interface
 - Highly intuitive
 - Eyes and hands free
 - Small devices
 - Rich communication channel

Dialogue Systems: A Brief History



Dialogue Systems: A Brief History

ELIZA
(Chatbots)

*Men are all alike.
IN WHAT WAY
They're always bugging us about something or other.
CAN YOU THINK OF A SPECIFIC EXAMPLE*

[Weizenbaum, 1966]

Dialogue Systems: A Brief History

ELIZA
(Chatbots)

SHRDLU
(Artificial Intelligence)

Pick up a big red block.

OK

Grasp the pyramid.

I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN

[Winograd, 1971]



Dialogue Systems: A Brief History

ELIZA
(Chatbots)

SHRDLU
(Artificial Intelligence)

VODIS, VOYAGER
(Speech)

How many hotels are there in Cambridge.

I KNOW OF SIX HOTELS IN CAMBRIDGE

[Glass et al., 1995]



Dialogue Systems: A Brief History

ELIZA
(Chatbots)

SHRDLU
(Artificial Intelligence)

VODIS, VOYAGER
(Speech)

Startups



Dialogue Systems: A Brief History

ELIZA
(Chatbots)

SHRDLU
(Artificial Intelligence)

SIRI
(hybrid approach)

VODIS, VOYAGER
(Speech)

Startups



Dialogue Systems: A Brief History

ELIZA
(Chatbots)

SHRDLU
(Artificial Intelligence)

SIRI
(hybrid approach)

Alexa Challenge
(social bots)

VODIS, VOYAGER
(Speech)

Startups



Spoken Dialogue Systems: Examples



[a] Cortana: Microsoft's new personal assistant for Windows Phone 8.1.



[b] MMI: In-car multimedia and navigation system from Audi.



[c] Ada and Grace, the virtual guides at the Museum of Science in Boston, developed by the USC Institute for Creative Technologies [2].



[d] Billie, an interactive scheduling assistant [3].



[e] Nao, the interactive robot created by Aldebaran Robotics.

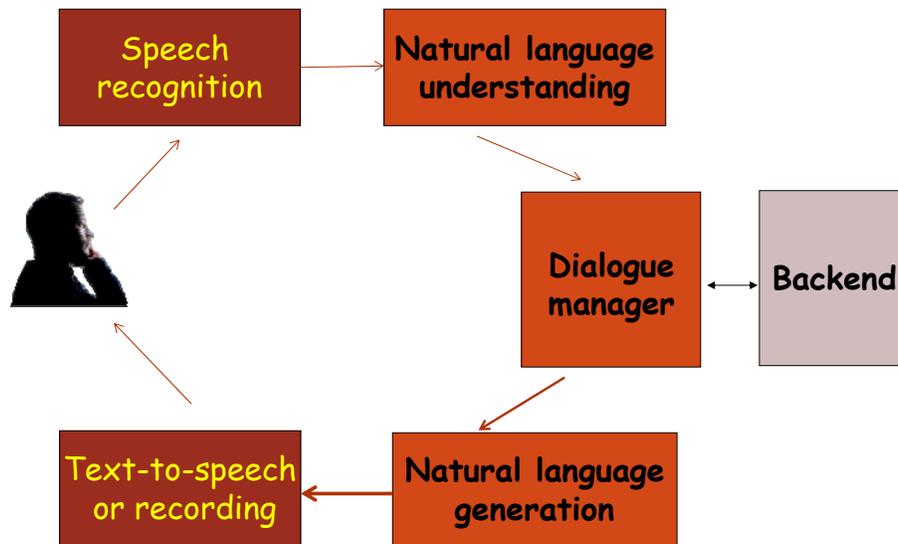
[Lison and Meena, 2014]

Are we done yet?

The Future Directions of Dialogue-Based Intelligent Personal Assistants (SIGdial 2016 panel)

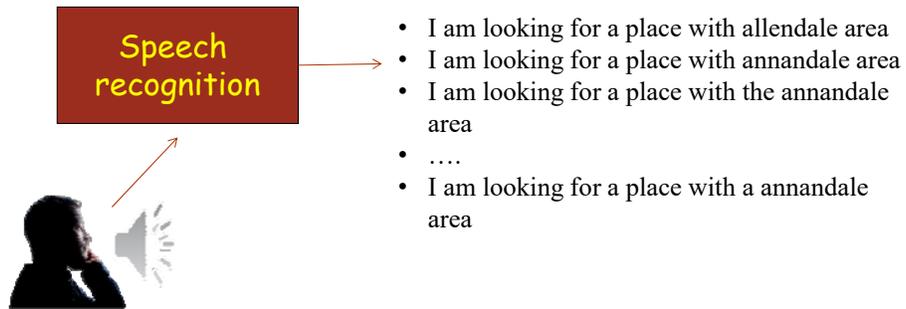
- Yes (one of the panelists)
 - no bottleneck but privacy!
- No (audience members - especially senior NLPers)
 - pragmatics, social dimensions, ...

Typical Architecture



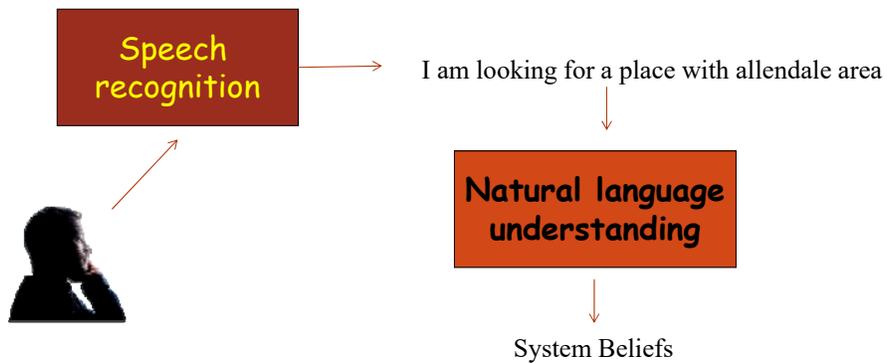
14

Typical Architecture



15

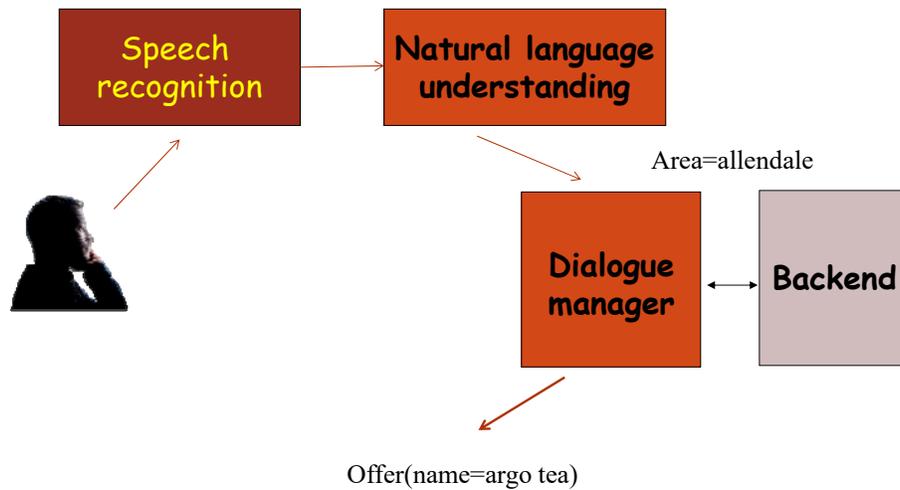
Typical Architecture



Name	-	.999
Area	allendale	.997
Food	-	.999
Area Code	-	.999
Requestable	-	.053

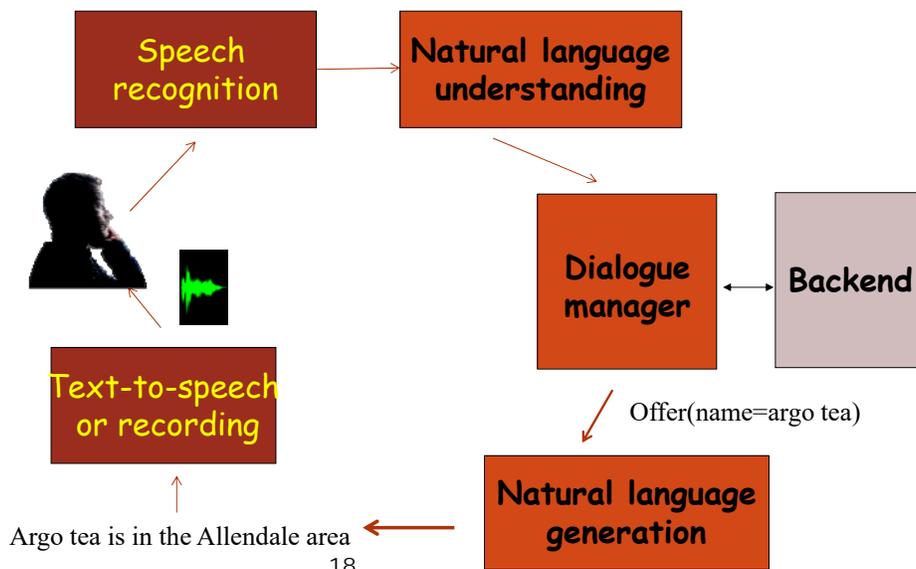
16

Typical Architecture



17

Typical Architecture



18

Challenges

- Input errors

Hello, what kind of laptop are you after?

SPEECH RECOGNITION: I WANT IT FOR OF IS THAT



What product family do you have in mind ...

Statistical Speech Recognition

- Let's assume
 - X is the entire sequence of speech data (features) you observe
 - W is one possible word sequence that it corresponds to
 - ... then the most likely word sequence would maximize the probability of occurrence of that sequence (among all possible word sequences) given that you observed the speech data X
- $P(W | X) = (P(X|W)P(W)) / P(X)$ (Bayes Rule)
- $\text{Argmax}_W P(X|W) P(W)$ (acoustic model and language model)

Speech Recognition, continued

- Acoustic models predict the probability of observing a given speech feature stream assuming it was generated by a particular word sequence
 - $P(X|W)$
- Language models help disambiguate by capture the relative probability of different word sequences, e.g. *recognize speech vs. wreck a beach*
 - $P(W)$

Challenges

- Input errors
 - Speech recognition (and turn-taking – time-out, barge-in)
 - Adverse acoustic conditions (speaker phone, background noise)
 - Natural language understanding (out of vocabulary)
- Other limitations
 - Restricted domains and tasks
 - Behave naturally (back-channel)
 - System components are typically ‘hand-crafted’
 - costly, don’t easily transfer
- A ‘big data’ alternative: statistical systems
 - System components are trained from data
 - “Deploy, Collect Data and Improve” [Young, 2014]

Part I: Chatbots!

- ELIZA (1966)
- PARRY (1968)
 - The first system to pass the Turing test?
- ALICE
- CLEVER
- Microsoft Little Bing 小冰

Chatbot Architectures

Rule-based

1. Pattern-action rules (Eliza)
 - + a mental model (Parry)

Corpus-based (from large chat corpus)

2. Information Retrieval
3. Neural network encoder-decoder

Chatbots

Sometimes for fun

But originally designed as a testbed
for clinical therapy
(And still used that way!)

Eliza: Weizenbaum (1966)

Men are all alike.

IN WHAT WAY

They're always bugging us about something or other.

CAN YOU THINK OF A SPECIFIC EXAMPLE

Well, my boyfriend made me come here.

YOUR BOYFRIEND MADE YOU COME HERE

He says I'm depressed much of the time.

I AM SORRY TO HEAR YOU ARE DEPRESSED

...

WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU?

My father

YOUR FATHER

You are like my father in some ways.

WHAT RESEMBLANCE DO YOU SEE

You are not very aggressive but I think you don't want me to notice that.

WHAT MAKES YOU THINK I AM NOT AGGRESSIVE

You don't argue with me.

WHY DO YOU THINK I DON'T ARGUE WITH YOU

You are afraid of me.

DOES IT PLEASE YOU TO BELIEVE I'M AFRAID OF YOU

My father is afraid of everybody.

WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER

Bullies.

DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR BOYFRIEND MADE YOU COME HERE

How does it work?

Eliza uses this one weird trick

That trick: be a Rogerian psychologist

Draw the patient out by reflecting patient's statements back at them

Rare type of conversation in which one can "assume the pose of knowing almost nothing of the real world"

Rogerian psychologist

Patient: "I went for a long boat ride"

Psychiatrist: "Tell me about boats"

- You don't assume she didn't know what a boat is
- You assume she had some conversational goal
- Chatbots trying to pass the Turing test usually choose some such domain

Eliza pattern/transform rules

(0 YOU 0 ME) [pattern]

→

(WHAT MAKES YOU THINK I 3 YOU) [transform]

0 means Kleene *

The 3 is the constituent # in pattern

You hate me

WHAT MAKES YOU THINK I HATE YOU

function ELIZA GENERATOR(*user sentence*) **returns** *response*

Find the word *w* in *sentence* that has the highest keyword rank

if *w* exists

 Choose the highest ranked rule *r* for *w* that matches *sentence*

response ← Apply the transform in *r* to *sentence*

if *w* = 'my'

future ← Apply a transformation from the 'memory' rule list to *sentence*

 Push *future* onto memory stack

else (no keyword applies)

either

response ← Apply the transform for the NONE keyword to *sentence*

or

response ← Pop the top response from the memory stack

return(*response*)

Keywords are ranked from specific to general

I know everybody laughed at me

- “I” is a very general keyword:

I: (I *) -> (You say you 2)

YOU SAY YOU KNOW EVERYBODY LAUGHED AT YOU

- “Everybody” is much more interesting (someone using universals like everybody/always is probably “referring to some quite specific event or person”)

WHO IN PARTICULAR ARE YOU THINKING OF?

- Implementation: keywords stored with their rank

Everybody 5 (*transformation rules*)

I 0 (*transformation rules*)

NONE

PLEASE GO ON

THAT'S VERY INTERESTING

I SEE

Memory

```
(MEMORY MY  
(0 YOUR 0 = LETS DISCUSS FURTHER WHY YOUR 3)  
(0 YOUR 0 = EARLIER YOU SAID YOUR 3)
```

- Whenever “MY” is highest keyword
 - Randomly select a transform on the MEMORY list
 - Apply to sentence
 - Store on a stack
- Later, if no keyword matches a sentence
 - Return the top of the MEMORY queue instead
- A hierarchical model of discourse

Other Eliza stuff

- Rules can refer to classes of words
Family = mother, father, brother, sister
NOUN = ...
- Don't reuse transforms in the same conversation
 - Whenever we use a transform associated with a pattern
 - We increment a counter for that rule
 - So the next time we use the next ranked transform
- Some basic transforms happen during input processing
 - I -> YOU
 - YOU -> I

Some implications

- People became deeply emotionally involved with the program
- Weizenbaum tells the story of his secretary who would ask Weizenbaum to leave the room when she talked with ELIZA
- When he suggested that he might want to store all the ELIZA conversations for later analysis, people immediately pointed out the privacy implications
 - Suggesting that they were having quite private conversations with ELIZA

Woebot

Fitzpatrick, Darcy, Vierhile, 2017. Delivering Cognitive Behavior Therapy to Young Adults with Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Ment Health* 4:2.

- Chatbot for delivering Cognitive Behavior Therapy via brief daily conversations
 - Little decision tree of language and menu responses
- "What's going on in your world right now?"
- "How are you feeling?"
- In response to endorsed loneliness:
- "I'm so sorry you're feeling lonely. I guess we all feel a little lonely sometimes"
- In a 2-week experiment seemed to reduce depression, but not anxiety

- Loebner Prize



- Sofia

IR-based chatbots

Idea: Mine conversations of human chats or human-machine chats

Microblogs: Twitter or Weibo (微博)

Movie dialogs

- Cleverbot (Carpenter 2017 <http://www.cleverbot.com>)
- Microsoft Xiaolce
- Microsoft Tay

A Cleverbot conversation

User:	Do you like Doctor Who?
Cleverbot:	Yes, so funny
User:	What's your favorite show?
Cleverbot:	I don't have one
User:	What's your name?
Cleverbot:	Jason I already told you.
User:	No your real name.
Cleverbot:	Stephanie IS my real name.
User:	I thought it was Jason?
Cleverbot:	How did you come to that conclusion.
User:	Because you said so.
Cleverbot:	I have to do my homework.

Two IR-based chatbot architectures

- Return the response to the most similar turn
 - Take user's turn (q) and find a (tf-idf) similar turn t in the corpus C

$q = \text{"do you like Doctor Who"}$

$t = \text{"do you like Doctor Strangelove"}$

- Grab whatever the response was to t .

$$r = \text{response} \left(\operatorname{argmax}_{t \in C} \frac{q^T t}{\|q\| \|t\|} \right) \quad \text{Yes, so funny}$$

- Return the most similar turn

$$r = \operatorname{argmax}_{t \in C} \frac{q^T t}{\|q\| \|t\|} \quad \text{Do you like Doctor Strangelove}$$

IR-based models of chatbots

- Also fine to use other features like user features, or prior turns
- Or non-dialogue text
 - COBOT chatbot (Isbell et al., 2000)
 - sentences from the Unabomber Manifesto by Theodore Kaczynski, articles on alien abduction, the scripts of “The Big Lebowski” and “Planet of the Apes”.
 - Wikipedia text

Neural chatbots

- Think of response generation as a task of *transducing* from the user’s prior turn to the system’s turn (like machine translation, ML version of Eliza)
- Train on:
 - movie dialogue databases
 - Twitter conversations
- Train a deep neural network
 - map from user1 turn to user2 response

Sample neural chatbot output

Input	tell me ... how long have you had this falling sickness ?
Vanilla-SEQ2SEQ	i'm not a doctor.
Adversarial	a few months, i guess .
Input	so i had the doctors test sammy 's response to conditioning .
Vanilla-SEQ2SEQ	sammy wrote the test sammy wrote the test .
Adversarial	so he took the pills .
Input	they didn 't have much success with this problem commander
Vanilla-SEQ2SEQ	they 're not the only ones who have been in the system .
Adversarial	can we find someone else ?

Chatbots: pro and con

- Pro:
 - Fun
 - Applications to counseling
 - Good for narrow, scriptable applications
- Cons:
 - They don't really understand
 - Rule-based chatbots are expensive and brittle
 - IR-based chatbots can only mirror training data
 - The case of Microsoft Tay
 - (or, Garbage-in, Garbage-out)
- The future: combining chatbots with frame-based agents

II: Frame-based dialog agents

A travel dialog: Communicator

S1: Hello. You've reached the [Communicator]. Tell me your full name
U2: *hi i'd like to fly to seattle tuesday morning*
S3: Traveling to Seattle on Tuesday, August eleventh in the morning. Your full name?
U4: *[John Doe]*
S5: Hello, [John], I'm glad to hear from you again. You can now ask me for help.
S6: OK. Let's see I have a United flight departing Pittsburgh at ten forty five a.m. arriving into Seattle at one fifty seven p.m. Is that OK?
U7: *Yes*
S8: Will you return to Pittsburgh from Seattle?
U9: *no i will go to Los Angeles*
S10: Traveling to Los Angeles. on what date
U11: *Sorry can we go back to talk about the previous leg again*
S12: For the first leg, you are on a United flight 1446 departing Pittsburgh at ten forty five A.M. tomorrow, arriving in Seattle at one fifty seven P.M. Do you want to take that?
U13: *actually i want to leave on wednesday*
S14: a flight on wednesday.
S15: I have a flight departing Pittsburgh at ten forty five a.m., arrives Seattle at one fifty seven p.m. Is that OK?
U16: *Yes*

Figure 19.1 The travel domain: a fragment from a successful conversation between a user (U) and the Communicator system (S) of Xu and Rudnicky (2000).

Call routing: ATT HMIHY

S: How may I help you?
U: *can you tell me how much it is to tokyo?*
S: You want to know the cost of a call?
U: *yes that's right*
S: Please hold on for rate information.
S: How may I help you?
U: *what's the area code for chicago?*
S: You want an area code.
U: *(silence)*
S: The area code for Chicago is 312.

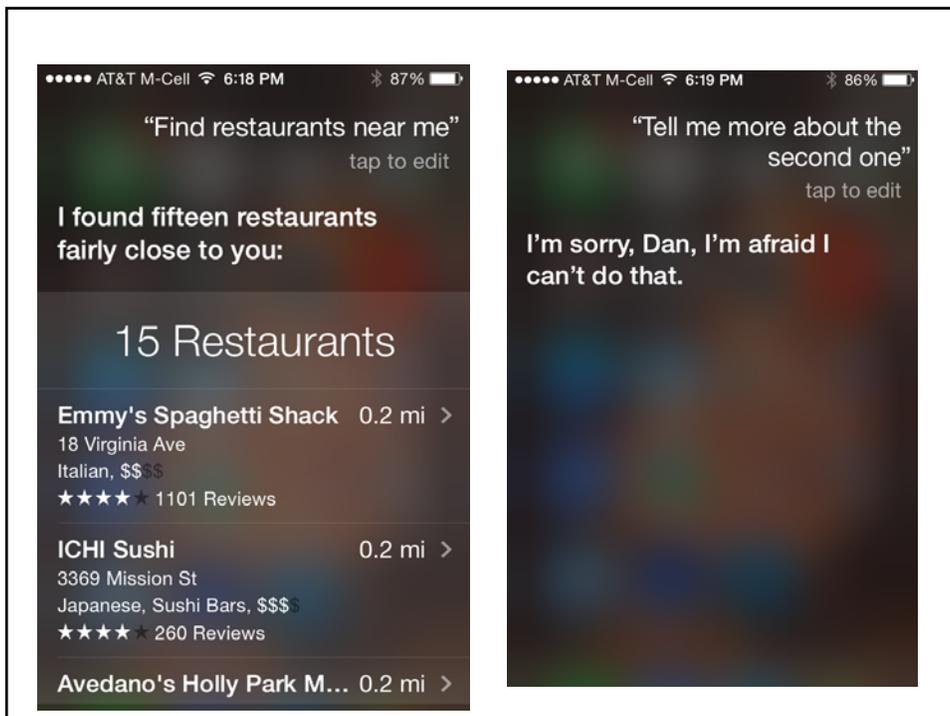
Figure 19.2 The call routing/telephone information domain: two actual conversations between a user (U) and the AT&T How May I Help You (HMIHY) system (S) (Gorin et al., 1997).

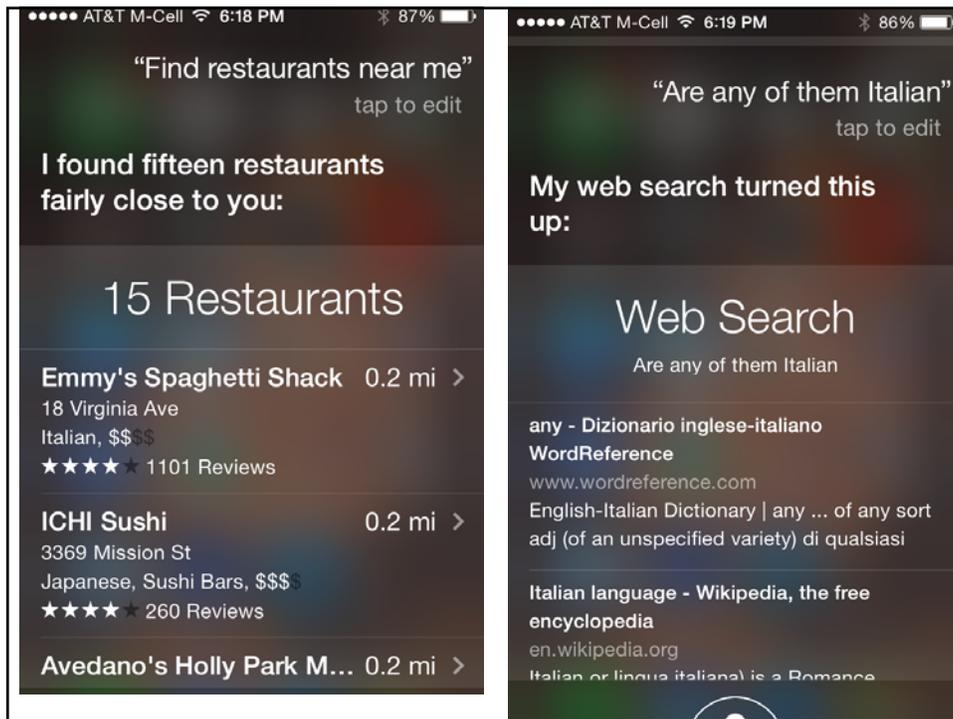
A tutorial dialogue: ITSPOKE

S: I didn't understand. What force acts on all objects within close proximity to earth?
U: *uh the force of gravity*
S: Fine. Besides the Earth's gravitational force, are there any other forces acting on the person, keys, and elevator after the release?
U: *no*
S: Yep. Thus, keep in mind that after the release, the only force on the person, keys, and elevator is the force of gravity. If gravity is the only force acting on an object, will it be moving or staying still?
U: *it will be moving*
S: Fine. Not only are the person, keys, and elevator moving, they have only gravitational forces on them. When an object is falling and has only gravitational force on it, it is said to be in what?
U: *freefall*

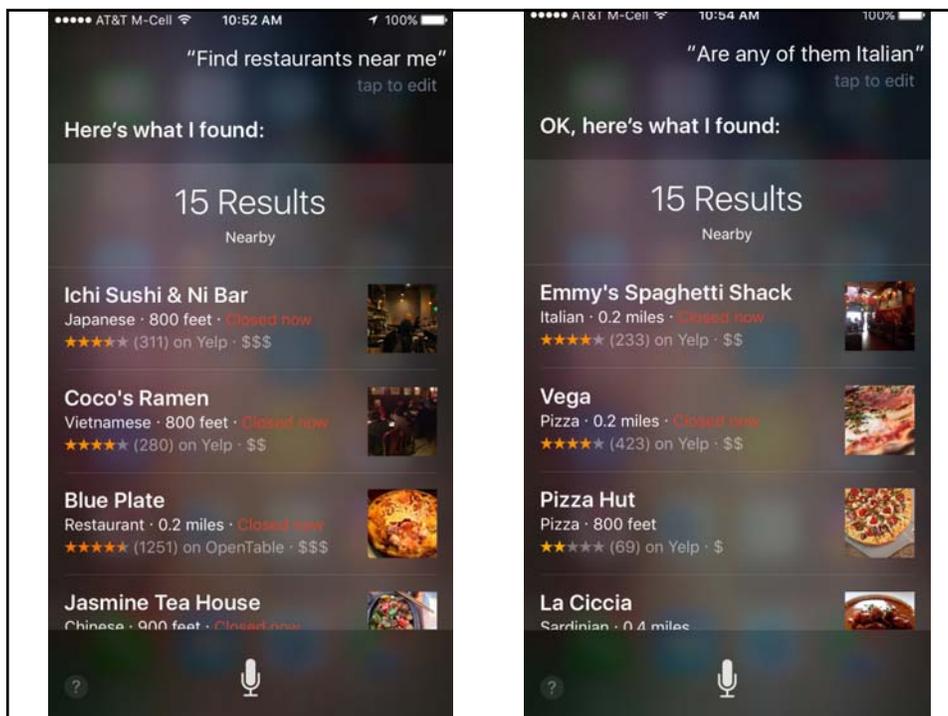
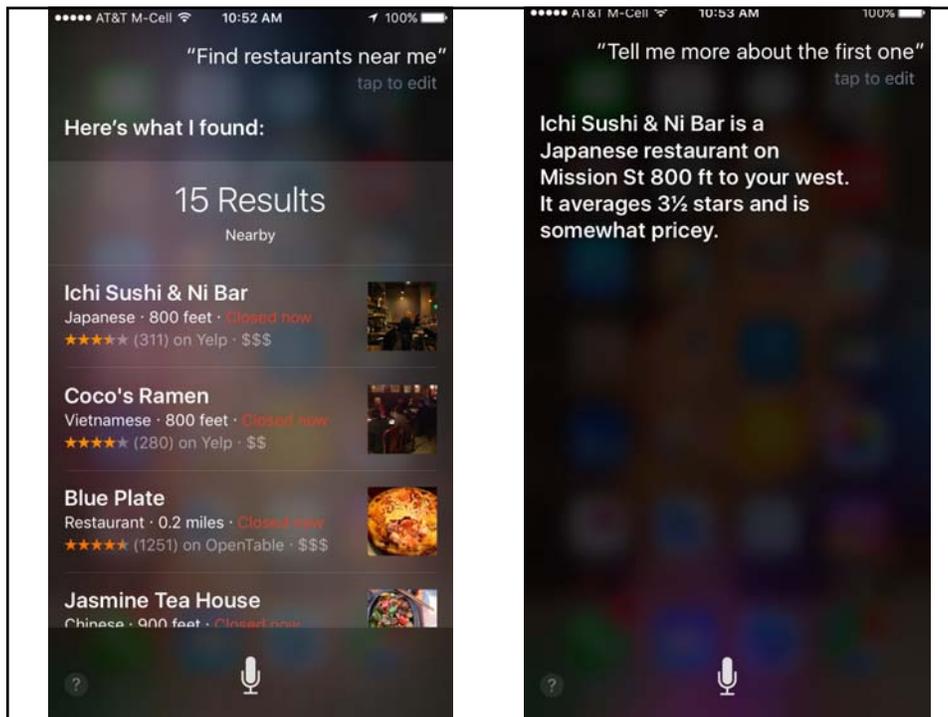
Figure 19.3 The tutoring domain: part of an actual conversation between a student user (U) and the ITSPOKE system (S) of (Litman and Silliman, 2004), based on the Why2-Atlas text-based tutoring system (?).

SIRI around 2014





SIRI in February 2017



Frame-based dialog agents

- Sometimes called "task-based dialog agents"
- Based on a "domain ontology"
 - A knowledge structure representing user intentions
- One or more **frames**
 - Each a collection of **slots**
 - Each slot having a **value**

The Frame

- A set of **slots**, to be filled with information of a given **type**
- Each associated with a **question** to the user

Slot	Type	Question
ORIGIN	city	What city are you leaving from?
DEST	city	Where are you going?
DEP DATE	date	What day would you like to leave?
DEP TIME	time	What time would you like to leave?
AIRLINE	line	What is your preferred airline?

Frame-based dialogue agents

- 1977:

*versation with a client who wants to make a simple return trip to a single city in California.
There is good reason for restricting the domain of discourse for a computer system which is to
engage in an English dialog. Specializing the subject matter that the system can talk about permits
it to achieve some measure of realism without encompassing all the possibilities of human knowledge
or of the English language. It also provides the user with specific motivation for participating in the
conversation, thus narrowing the range of expectations that GUS must have about the user's pur
poses. A system restricted in this way will be more able to guide the conversation within the boundarie
of its competence.*

Artificial Intelligence Journal, 1977

- Still the industrial state of the art
 - SIRI based on GUS architecture

GUS: Hello. My name is GUS. I can help you plan a simple trip by air. Where do you want to go ?

Client: I want to go to San Diego on May 28

GUS: What time do you want to leave?

Client: I must be in San Diego before 10 am

GUS: Would you like P.S.A. flight 102 that arrives at 9.15 am ?

Client: What is the next flight ?

GUS: Air California flight 310 that leaves at 8.30am and arrives at 10.00am

Client: I'll take the first one

GUS: In what name should I make a reservation, please?

Client: Dan Bobrow

GUS: I have confirmed the following flight: P.S.A. flight 102 on Wednesday May 28 from San Jose to San Diego that leaves at 7.30 am and arrives at 9.15 am.

GUS: What date do you want to return on ?

Client: On Friday in the evening.

GUS: Would you like the flight that leaves at 7.45 pm ?

Client: That's fine.

GUS: I have confirmed the following flight: P.S.A. flight 307 on Friday May 30 from San Diego to San Jose that leaves at 7.45 pm and arrives at 9.30 pm Thank you for calling. Goodbye

The state of the art in 1977 !!!!

Slot types can be complex

- The type *DATE*

```
DATE
  MONTH NAME
  DAY (BOUNDED-INTEGER 1 31)
  YEAR INTEGER
  WEEKDAY (MEMBER (SUNDAY MONDAY TUESDAY WEDNESDAY THURSDAY FRIDAY SATURDAY)]
```

Control structure for frame-based dialog

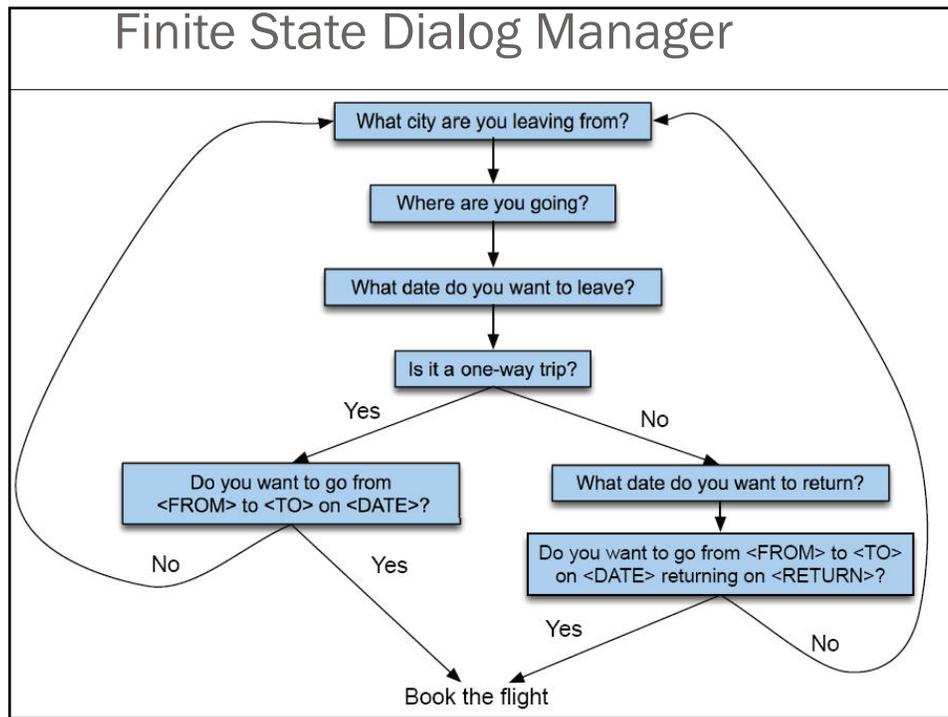
Consider a trivial airline travel system:

Ask the user for a departure city

Ask for a destination city

Ask for a time

Ask whether the trip is round-trip or not



Finite-state dialog managers

- System completely controls the conversation with the user.
- It asks the user a series of questions
- Ignoring (or misinterpreting) anything the user says that is not a direct answer to the system's questions

Summary

- State of the art:
 - Chatbots:
 - Simple rule-based systems
 - IR or Neural networks: mine datasets of conversations.
 - Frame-based systems:
 - hand-written rules for slot fillers
 - ML classifiers to fill slots
- What's the future?
 - Key direction: Integrating goal-based and chatbot-based systems