

Natural Language Processing for Production-Level Conversational Interfaces

Arushi Raghuvanshi
Machine Learning Engineer

Karthik Raghunathan
Senior Technical Lead

Agenda

- Webex Intelligence
- Natural Language Processing for Conversational Interfaces
  - Data Collection
  - Conversational Application Development
  - Ongoing Model Improvement

© 2019 Cisco and/or its affiliates. All rights reserved.

Webex Intelligence

aka *“Really? Cisco works on AI?!”*

© 2019 Cisco and/or its affiliates. All rights reserved.

Cisco Webex:
Meetings, Messaging, Calling, Devices



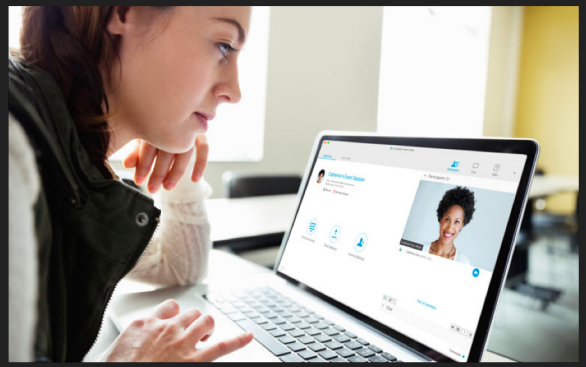








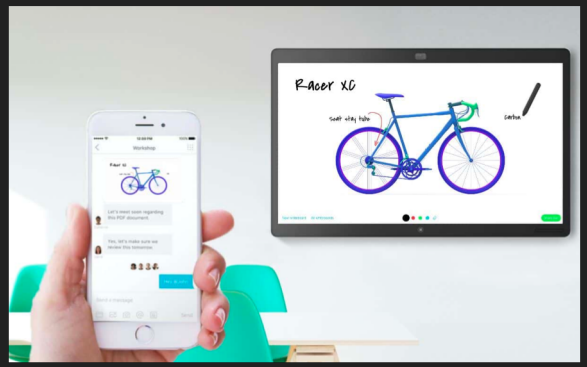
Leveraging AI to radically transform teamwork



Before



During



After

© 2019 Cisco and/or its affiliates. All rights reserved.

Face Detection



In-room analytics



Intelligent framing for clearer communication



Best Overview

Optimizes framing based on participants in the room





Speaker Tracking

Detects person speaking Selects best framing for speaker

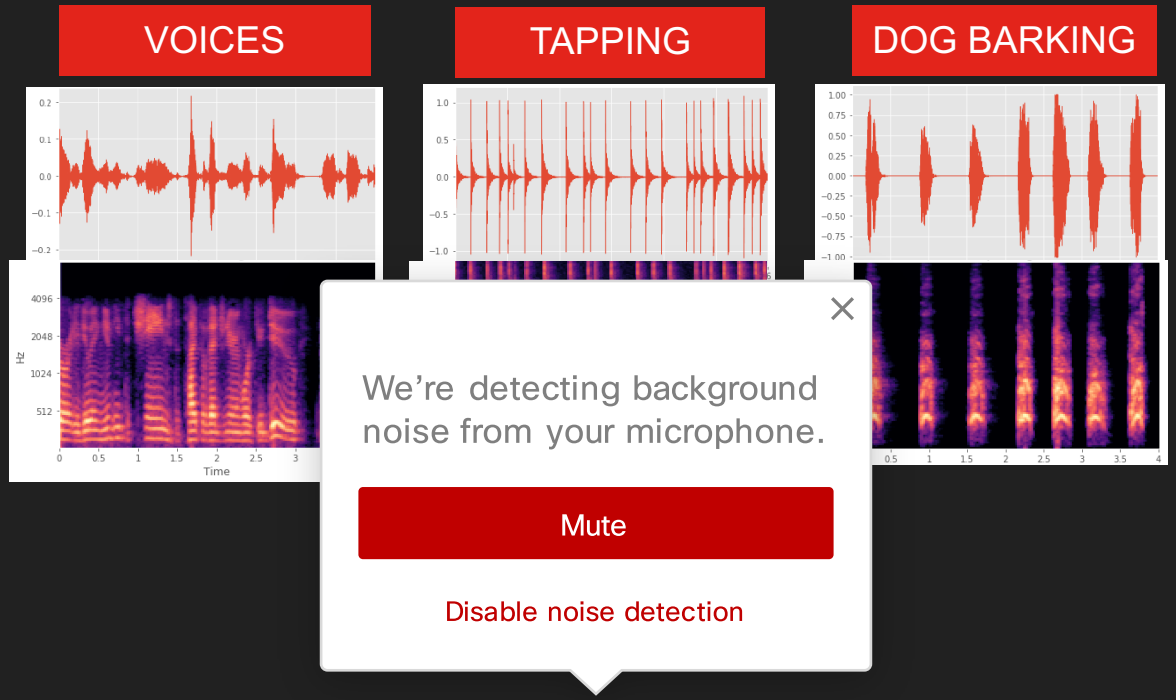


© 2019 Cisco and/or its affiliates. All rights reserved.

Face Recognition



Noise Suppression



© 2019 Cisco and/or its affiliates. All rights reserved.

Applications of Machine Learning in Collaboration

Conversational NLU	Information Retrieval	Information Extraction	Meeting Transcripts & Insights	Conversation Analytics
Gesture Recognition	Person Tracking	Object Detection	Gaze Detection	Face Recognition
Speaker Recognition	Wake word Detection	Keyword Spotting	Closed Vocabulary ASR	Noise Suppression

© 2019 Cisco and/or its affiliates. All rights reserved.

Applications of Machine Learning in Collaboration

Conversational NLU	Information Retrieval	Information Extraction	Meeting Transcripts & Insights	Conversation Analytics
Gesture Recognition	Person Tracking	Object Detection	Gaze Detection	Face Recognition
Speaker Recognition	Wake word Detection	Keyword Spotting	Closed Vocabulary ASR	Noise Suppression

© 2019 Cisco and/or its affiliates. All rights reserved.

Conversational interfaces are the next frontier

- A massive shift in user behavior is underway
  - “20% of all searches are voice searches” – Google (2016)
  - “50% of all searches will be voice searches by 2020” – Comscore

- Connected devices are changing how we interact with technology

“Alexa, lock my front door”

”Cortana, how much is a flight from Portland to Orlando?”

“Ok Google, play Handmaid’s Tale on TV”

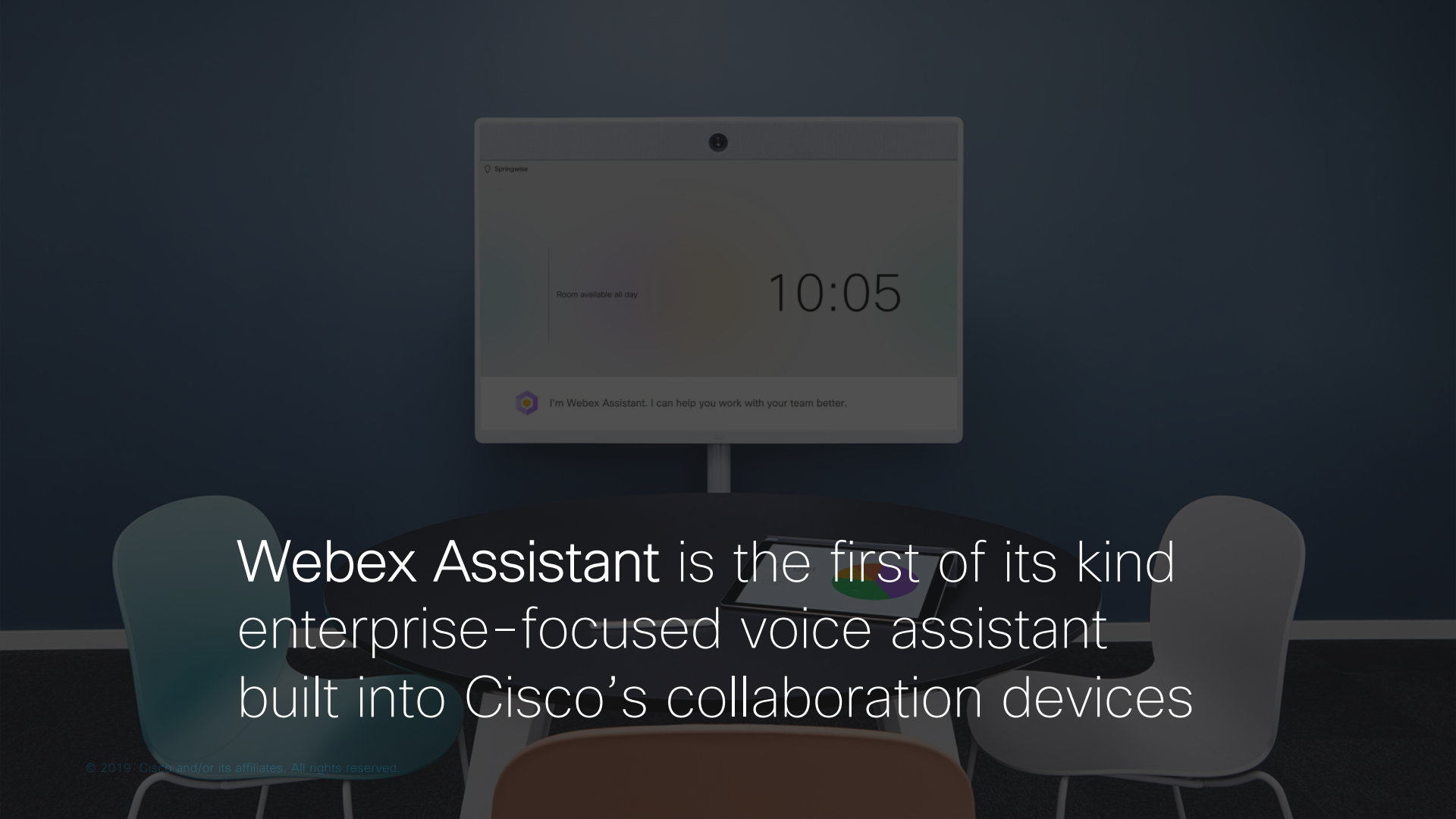
“Hey Siri, order me some Pad Thai”

© 2019 Cisco and/or its affiliates. All rights reserved.

Workplaces have largely been left out of the Conversational AI revolution



© 2019 Cisco and/or its affiliates. All rights reserved.



Webex Assistant is the first of its kind enterprise-focused voice assistant built into Cisco's collaboration devices

© 2019 Cisco and/or its affiliates. All rights reserved.

Springwise

Room available all day

10:05

Springwise

Room available all day

10:05

Listening... *Try saying* *“call John Smith”*

Springwise

Room available all day

10:05



Call Barbara

Springwise

Room available all day

10:05

Call Barbara

Which one would you like to call?



Barbara Adams

Market Manager

Option 1



Barbara Adams

Director Product M...

Option 2



Barbara Benito

N/A

Option 3

BB

Barbara Benz

Administration Coo...

Option 4

Listening... *Try saying* *"option 1"* *"show more"*

Which one would you like to call?



Barbara Adams

Market Manager

Option 1



Barbara Adams

Director Product M...

Option 2



Barbara Benito

N/A

Option 3

BB

Barbara Benz

Administration Coo...

Option 4

Show more

Which one would you like to call?



Barbara Adams

Market Manager

Option 1



Barbara Adams

Director Product M...

Option 2



Barbara Benito

N/A

Option 3

BB

Barbara Benz

Administration Coo...

Option 4

Show more

Which one would you like to call?



Barbara Bossom

Market Manager

Option 5



Barbara Elize

Engineer User Exp...

Option 6



Barbara German

Director Managem...

Option 7



Barbara Hung

Designer User Exp...

Option 8

Listening... *Try saying* *" option 5"* *"go back"*

Which one would you like to call?



Barbara Nichols

Global Staffing Man...

Option 21



Barbara Olga

Director Product M...

Option 22



Barbara Schwarzen...

Program Manager

Option 23



Barbara Wilford

Administration Coo...

Option 24

Option 22

Which one would you like to call?



Barbara Nichols

Global Staffing Man...

Option 21



Barbara Olga

Director Product M...

Option 22



Barbara Schwarzen...

Program Manager

Option 23



Barbara Wilford

Administration Coo...

Option 24

Option 22



Barbara Nichols
Global Staffing Man...
Option 21



Barbara Olga
Director Product Manager
San Francisco, CA
Building 11
barbara.olga@company.com
Option 22



Barbara Schwarzen...
Program Manager
Option 23

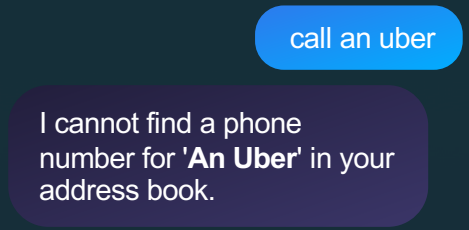
OK. Calling Barbara Olga. *To cancel say* *“OK Webex, stop”*


Springwise
10:00
User Experience Review
10:00 - 11:00
Organized by Brenda Song
5 more meetings today
Say "OK Webex, join meeting"
or join from the touch panel
Show me the calendar of this room

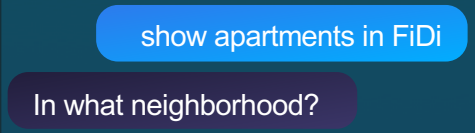
Natural Language Processing for Conversational Interfaces

© 2019 Cisco and/or its affiliates. All rights reserved.

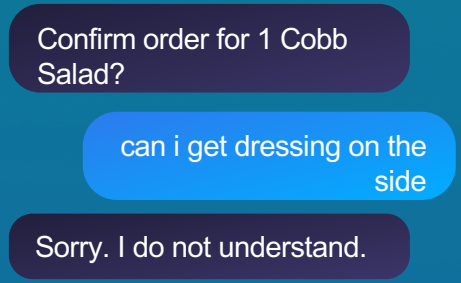
Building production-quality conversational assistants is one of today's hardest AI challenges



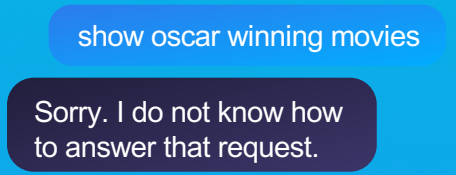
hit-or-miss accuracy quickly frustrates users



common slang can often go misunderstood



pre-set rules break when users go slightly off-script



shallow knowledge fails to answer even basic questions

© 2019 Cisco and/or its affiliates. All rights reserved.

Significant, relevant progress in academia has been made

Growing Datasets

facebook bAbI

twitter

SQuAD2.0
The Stanford Question Answering Dataset

Relevant Research

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Language Models are Unsupervised Multitask Learners

Near-Human Accuracy

$\text{BERT}_{\text{BASE}}$	81.6	-
$\text{BERT}_{\text{LARGE}}$	**86.6**	**86.3**
Human (expert)†	-	85.0
Human (5 annotations)†	-	88.0

© 2019 Cisco and/or its affiliates. All rights reserved.

With recent NLP advancements, why is building a **production level** conversational app so hard?

Papers focus on only a **single component** of the larger problem

Components that work well individually may break down as part of a **larger pipeline**

Applications may have very **limited training data** compared to standard research datasets

Domain specific application data may have different characteristics than standard datasets

© 2019 Cisco and/or its affiliates. All rights reserved.

A consistent methodology has emerged for building conversational interfaces that yields good results.

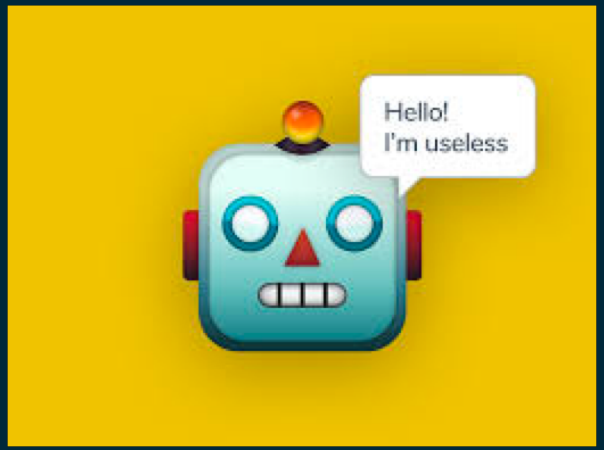
- Data Collection
- Conversational Application Development
  - Natural Language Processing Pipeline
  - Knowledge Base Creation
  - Question Answering
  - Dialogue Management
- Ongoing Model Improvement

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Data Collection

© 2019 Cisco and/or its affiliates. All rights reserved.

Selecting the right use case is essential



- Does it resemble a **real world** human **interaction**?
- Is it more **convenient** for the user?
- Is it a narrow enough domain to achieve **high accuracy**?
- But broad enough to be **useful**?
- Is it possible to collect enough **training data**?

© 2019 Cisco and/or its affiliates. All rights reserved.

Domain specific metadata is often publically or internally available

Domain Vocabulary

venti caffe latte

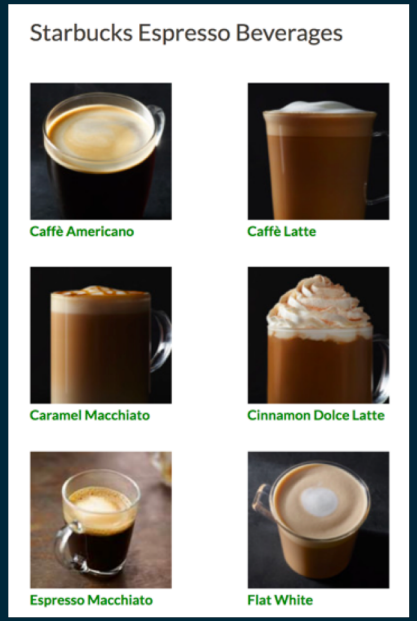
tall frappuccino

grande soy milk extra hot latte

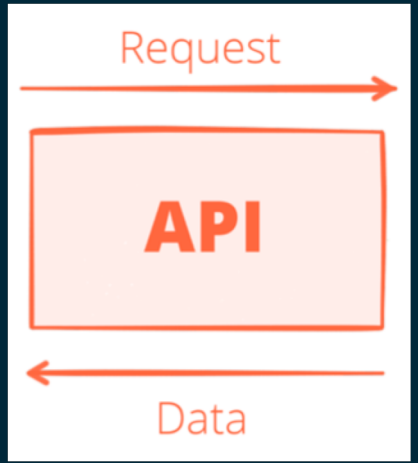
flat white

caramel macchiato

Data for a Knowledge Base



Live API



© 2019 Cisco and/or its affiliates. All rights reserved.

Conversational data is often collected via crowdsourcing tools



"Please schedule a call at 4pm in Winterfell"

"Please schedule a call at { 4pm | time } in { Winterfell | room }"

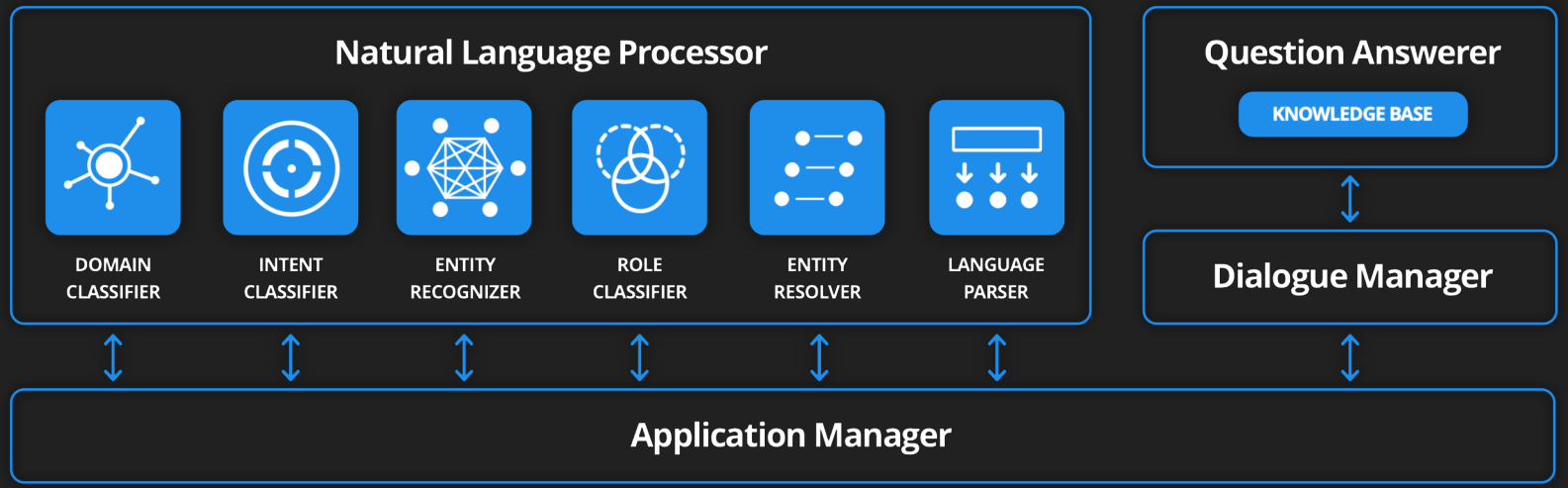
- Domain specific conversational data is virtually never available
- Generate in house or via crowdsourcing tools
- Annotate according to your app specifications
- Ensure clean data annotations

© 2019 Cisco and/or its affiliates. All rights reserved.

Conversational Application Development

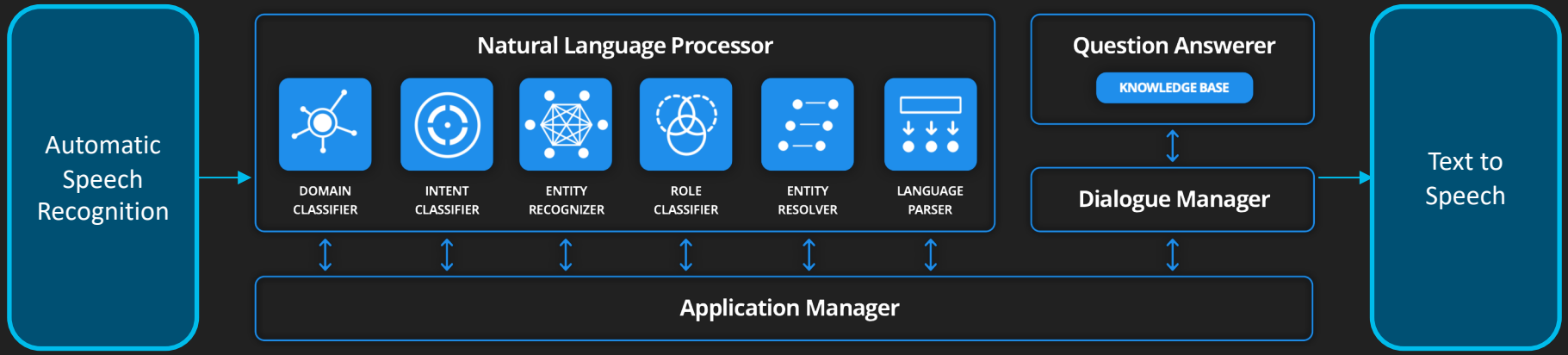
© 2019 Cisco and/or its affiliates. All rights reserved.

Key steps in a modern conversational application



© 2019 Cisco and/or its affiliates. All rights reserved.

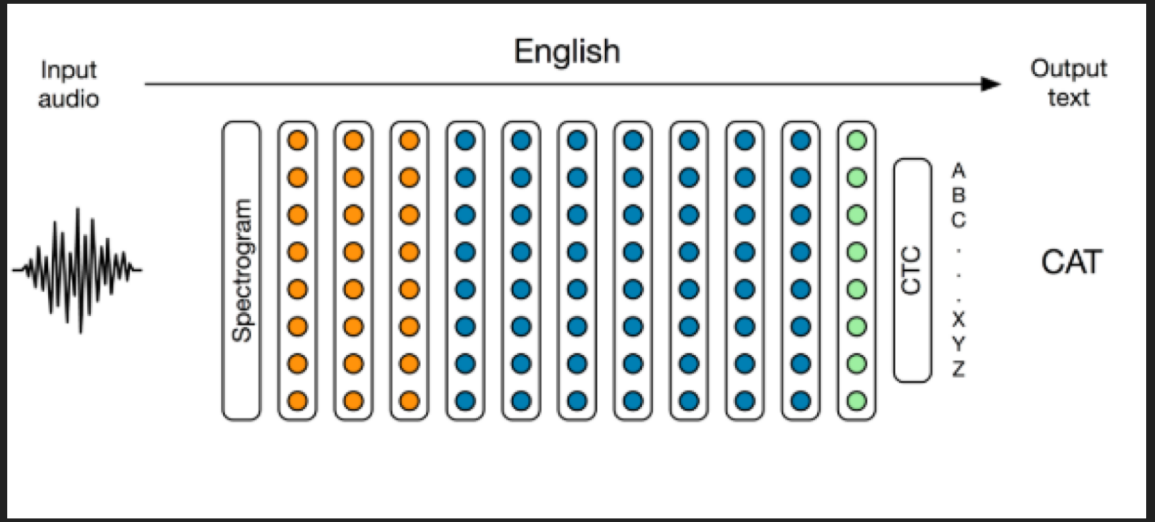
Voice based conversational applications



© 2019 Cisco and/or its affiliates. All rights reserved.

Automatic Speech Recognition

- Audio Features
- Phonetic Model
- Language Model



The cost of data collection and model development to develop an in house ASR solution is high

© 2019 Cisco and/or its affiliates. All rights reserved.

Automatic Speech Recognition

- Word Error Rate (WER) of less than 6% in English Voice Search Queries
- On transcripts with uncommon terms like person names, WER is 41%







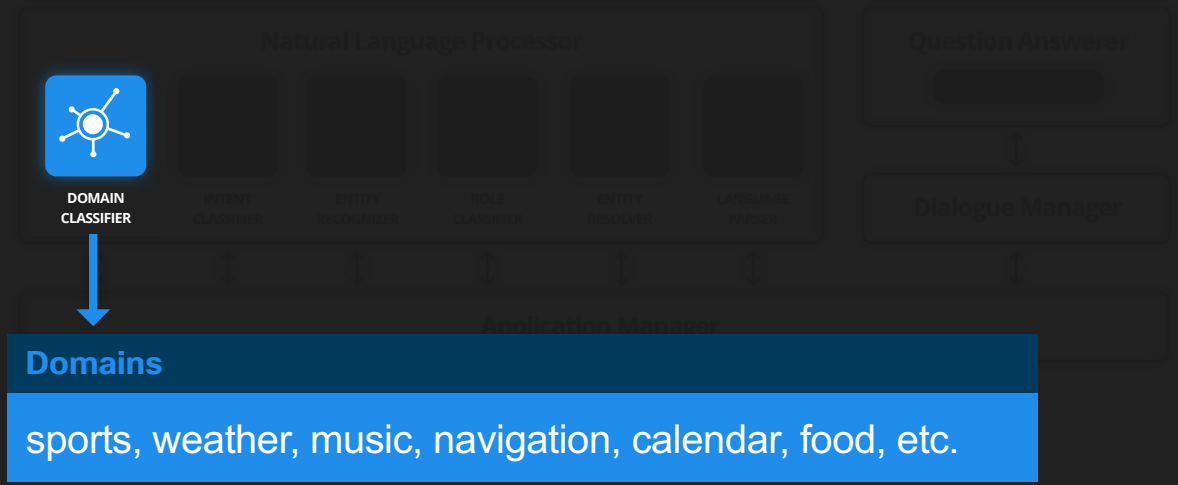






© 2019 Cisco and/or its affiliates. All rights reserved.

Domain Classifier assigns the user query to a pre-defined domain



- Performs the first level of categorization by assigning the query to one of a pre-defined set of app-supported domains
- A domain constitutes a unique area of knowledge

© 2019 Cisco and/or its affiliates. All rights reserved.

Domain Classifier

Classifier
Logistic regression
Support vector machine
Decision tree
Random forest

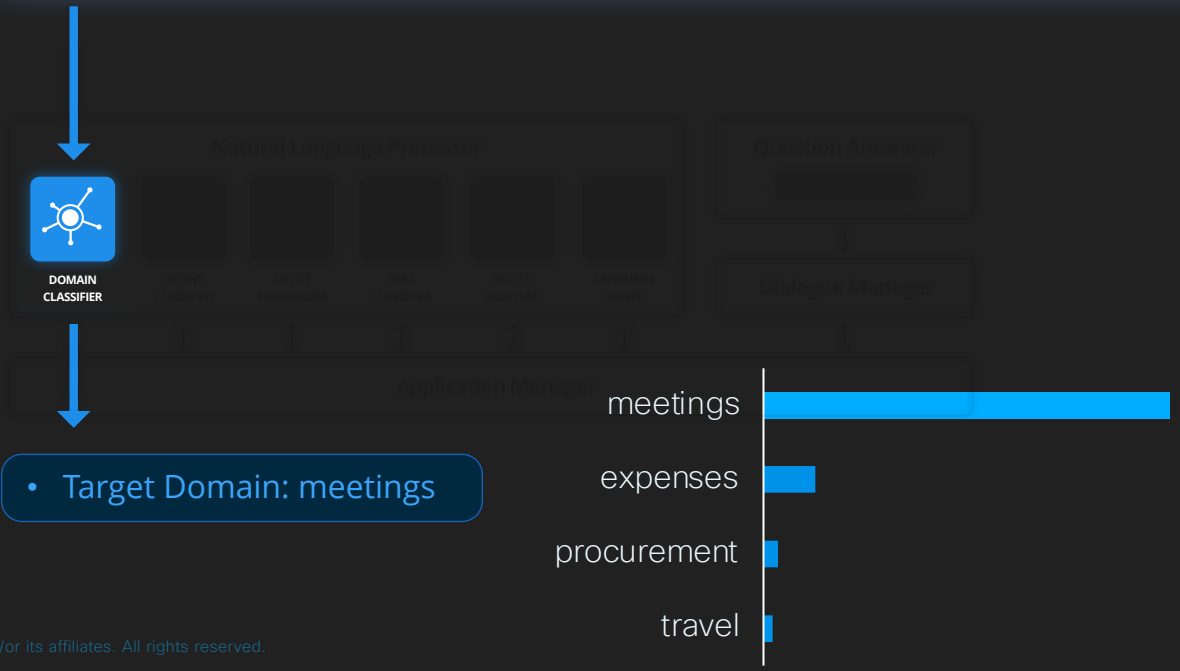
Features:

- Bag of words
- Stemming
- Word shape
- Word n-grams
- Character n-grams
- Token frequency
- Gazetteer match
- Query length

© 2019 Cisco and/or its affiliates. All rights reserved.

Domain Classification

INPUT "Schedule a meeting with Janice from accounting from 11 AM to noon in Acquerello"

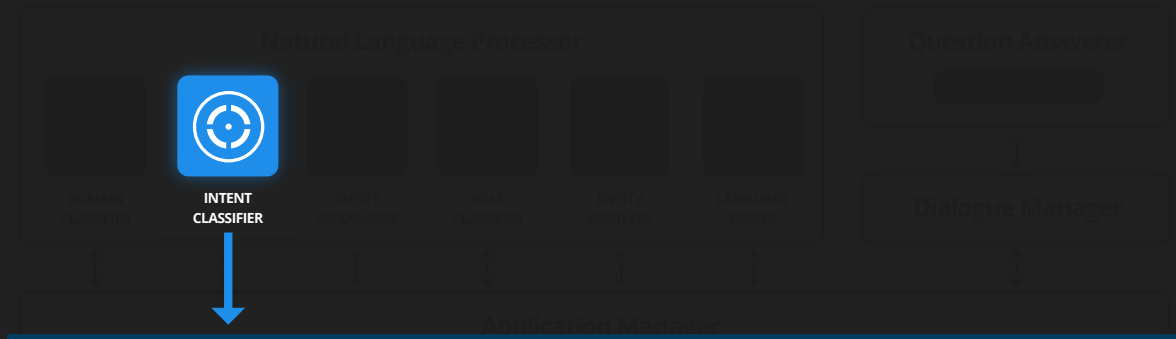


OUTPUT

- Target Domain: meetings

© 2019 Cisco and/or its affiliates. All rights reserved.

Intent Classifier determines the user intent



Domain	Intents
food	search_restaurant, get_restaurant_info, browse_dish, place_order

- Provides the next level of categorization by assigning the query to one of the intents defined for the app
- Intents reflect what a user is trying to accomplish

© 2019 Cisco and/or its affiliates. All rights reserved.

Intent Classifier

Classifier
Logistic regression
Support vector machine
Decision tree
Random forest

Features:

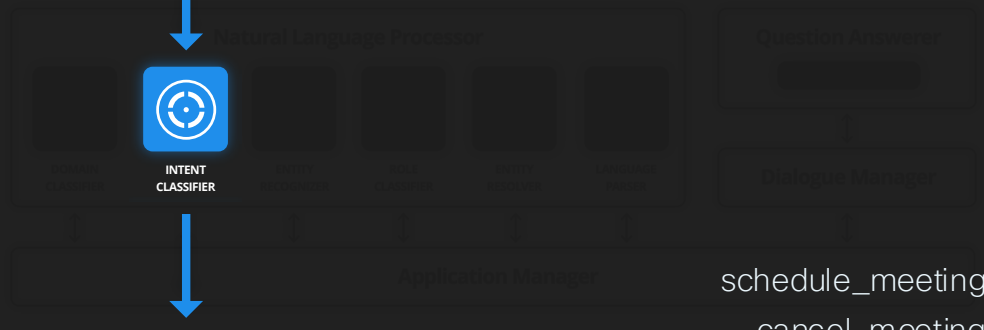
- Bag of words
- Stemming
- Word shape
- Word n-grams
- Character n-grams
- Token frequency
- Gazetteer match
- Query length

© 2019 Cisco and/or its affiliates. All rights reserved.

Intent Classification

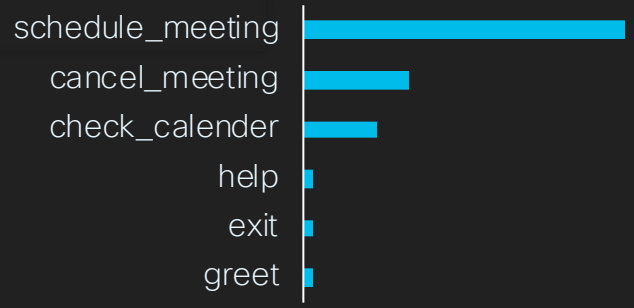
INPUT **"Schedule a meeting with Janice from accounting from 11 AM to noon in Acquerello"**

Domain: **meetings**



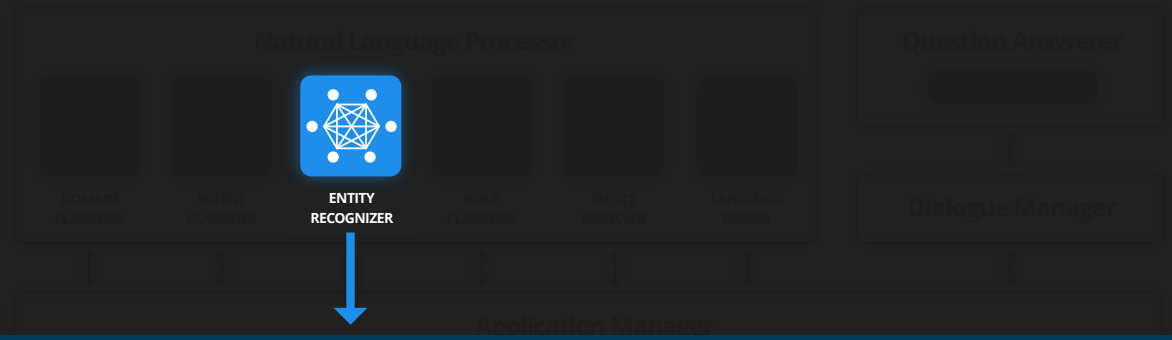
OUTPUT

- Target Intent: schedule_meeting



© 2019 Cisco and/or its affiliates. All rights reserved.

Entity Recognizer detects all relevant entities in the user query



Domain	Intent	Entity Types
food	search_restaurant	restaurant_name, cuisine, dish_name, location, price_range, rating
movies	find_movie	movie_title, genre, cast, director, release_date, rating

- Identifies every entity in the query that is relevant to the determined intent
- An entity is any word or phrase that provides information necessary to understand and fulfill the user's end goal

© 2019 Cisco and/or its affiliates. All rights reserved.

Entity Recognizer

Classifier
Maximum Entropy Markov Model
Conditional Random Field
Long Short-Term Memory

Features:

- Bag of words
- Character n-grams
- Gazetteer match
- ...

- Character embedding
- Word embedding
- ...

© 2019 Cisco and/or its affiliates. All rights reserved.

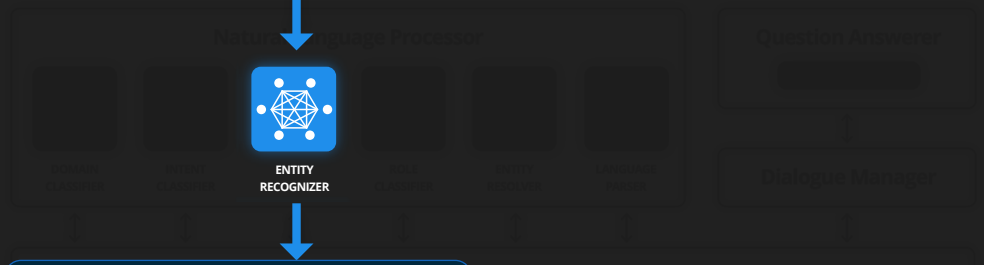
Entity Recognition

INPUT

"Schedule a meeting with Janice from accounting from 11 AM to noon in Acquerello"

Domain: **meetings**

Intent: **schedule_meeting**

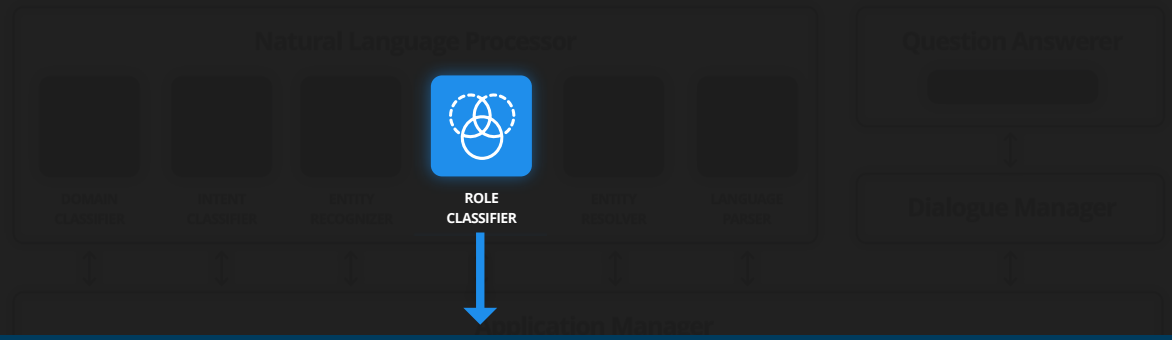


OUTPUT

- Person: **Janice**
- Department: **accounting**
- Time: **11 AM**
- Time: **noon**
- Room: **Acquerello**

© 2019 Cisco and

Role Classifier assigns role labels to extracted entities



Domain	Intent	Entity Type	Role Types
meetings	schedule_meeting	time	start_time, end_time
travel	book_flight	location	origin, destination

- Assigns a differentiating label, called a role, to the entities in the query
- Useful when a particular entity type can have multiple meanings based on the context

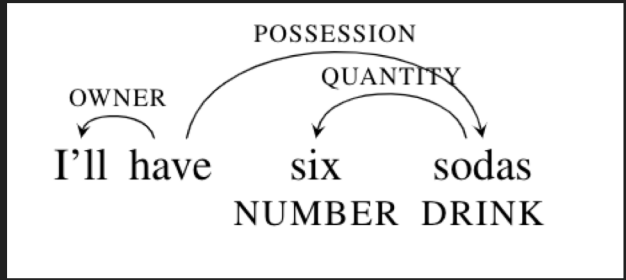
© 2019 Cisco and/or its affiliates. All rights reserved.

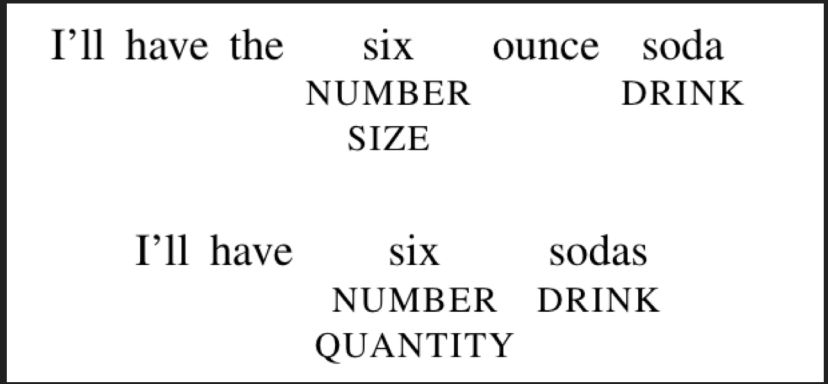
Role Classifier

We extract just the relation type rather than a full semantic parse.

This is a simple classification model:

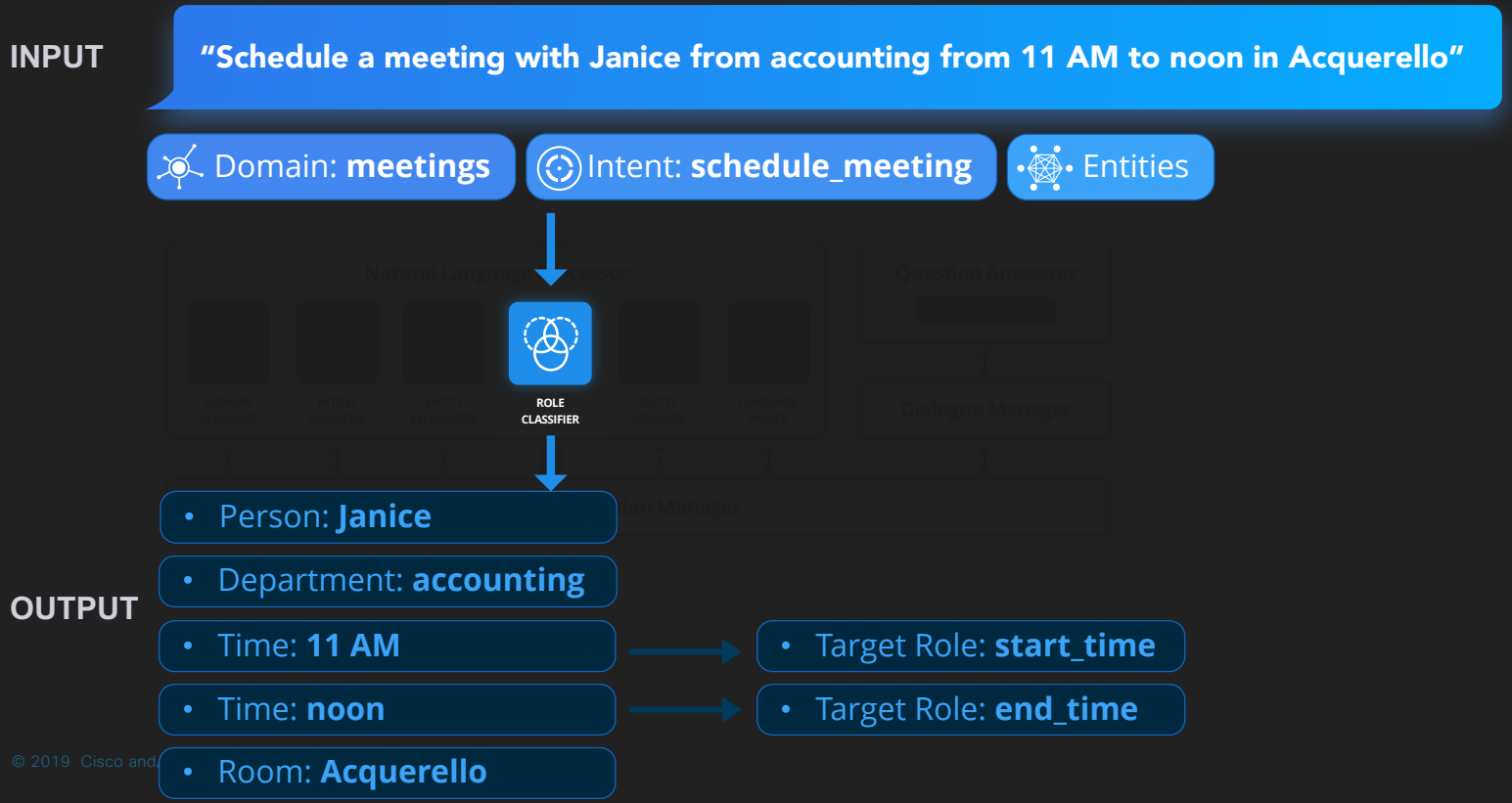
- Logistic regression
- SVM
- Decision Tree
- Random Forest





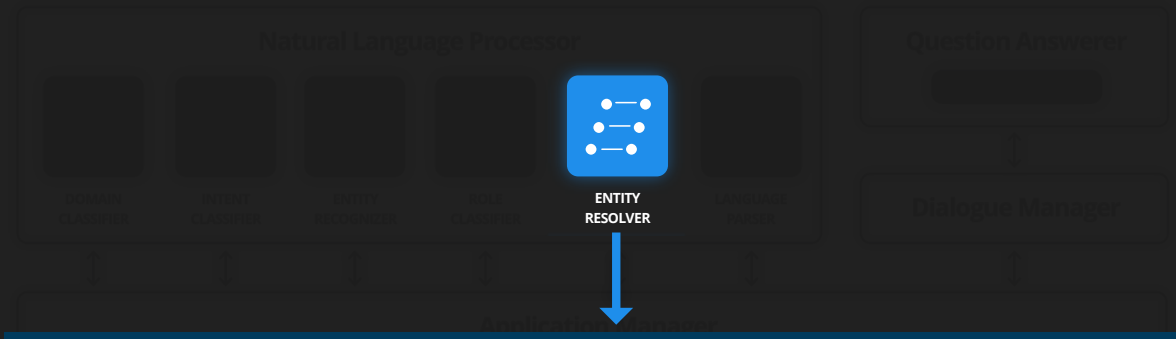
© 2019 Cisco and/or its affiliates. All rights reserved.

Role Classification



© 2019 Cisco and

Entity Resolver transforms each extracted entity into its canonical form



Extracted Entity	Resolved Canonical Value
“SF”	San Francisco, CA
“lemon bread”	Iced Lemon Pound Cake (SKU: 470)

- Maps each identified entity to a canonical entry in an official catalog or database
- Uses sophisticated text relevance algorithms to match against a comprehensive set of entity synonyms such as abbreviations, nicknames, slangs, and other aliases

© 2019 Cisco and/or its affiliates. All rights reserved.

Entity Resolution on Noisy Voice Transcripts

Entities are the uncommon, domain-specific terms that are most often mistranscribed by ASR

- “Can you connect me to Kiran Prakash’s room” → “Can you connect me to Corrine precautious room”
- “Can you get a hold of Mahojwal” → “Can you get ahold of my jaw”
- “Can you call Didi” → “can you call Stevie”
- “connect me with Heather” → ”connect Merriweather”

© 2019 Cisco and/or its affiliates. All rights reserved.

Entity Resolution – Textual Features

Fuzzy matching	Normalization, character n-grams, word n-grams
Semantic synonym matching	e.g. ‘pasta with tomato sauce’ -> ‘spaghetti marinara’

© 2019 Cisco and/or its affiliates. All rights reserved.

Entity Resolution – Phonetic Features

Double Metaphone	‘Smith’ -> ‘SMO XMT’	e.g. ‘Corrine precautious‘ -> ‘Kiran Prakash’s’
Grapheme to Phoneme	‘Michael’ -> ‘M AY K AH L’	e.g. ‘croissant ready’ -> ‘Prasanth Reddy’
N-best ASR Alternates	Gold (user’s original intention): I want to talk with Sheetal N-best Alternate Transcripts from ASR: I want to talk with sheet all I want to talk with she tall I want to talk with sheetal I want to talk with XI tall I want to talk with Shi toll I want to talk with shuttle ...	e.g. ‘Stevie’ -> ‘Didi’

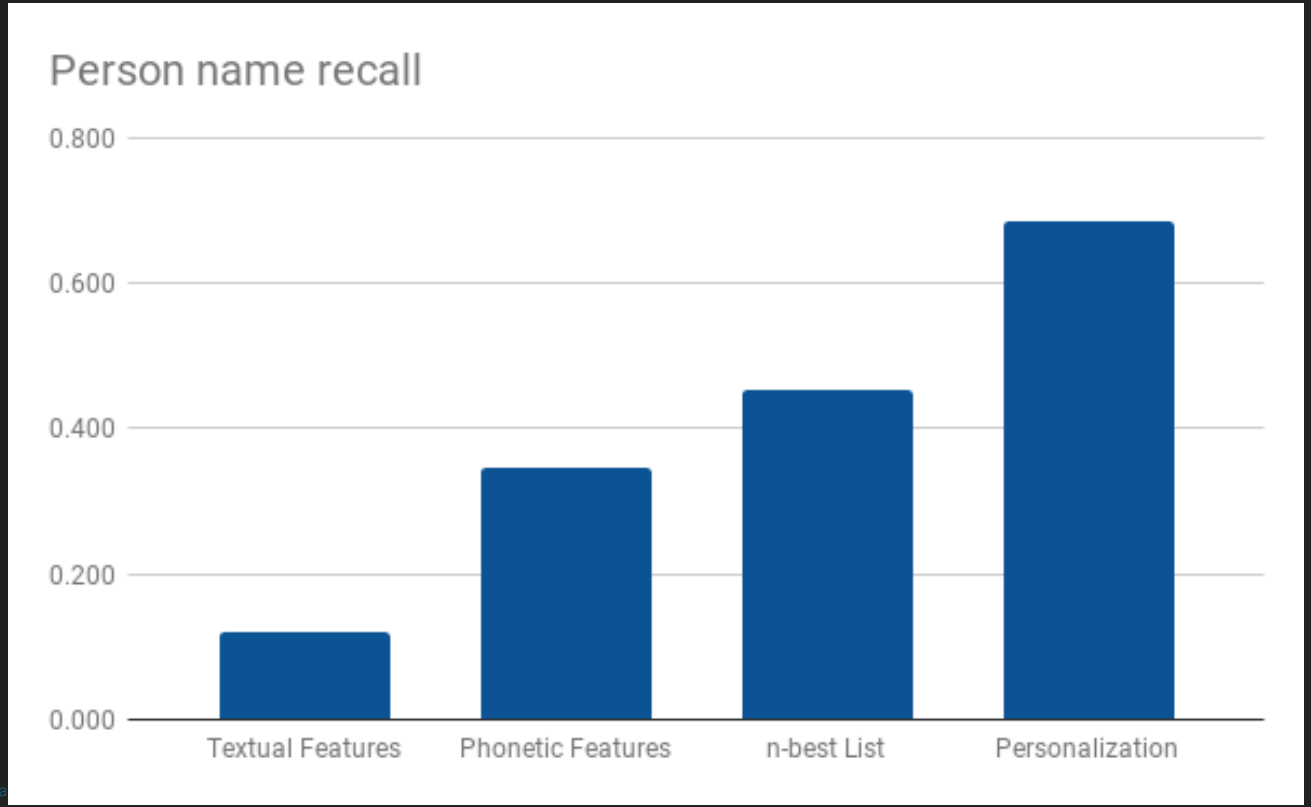
© 2019 Cisco and/or its affiliates. All rights reserved.

Entity Resolution – Personalization Features

Use Case	App specific context
Call person	Relationship between the caller and callee - previous interactions - distance in organizational hierarchy
Food ordering	Popularity of dish Order history Location of restaurant
Music discovery	Song popularity Number of previous plays

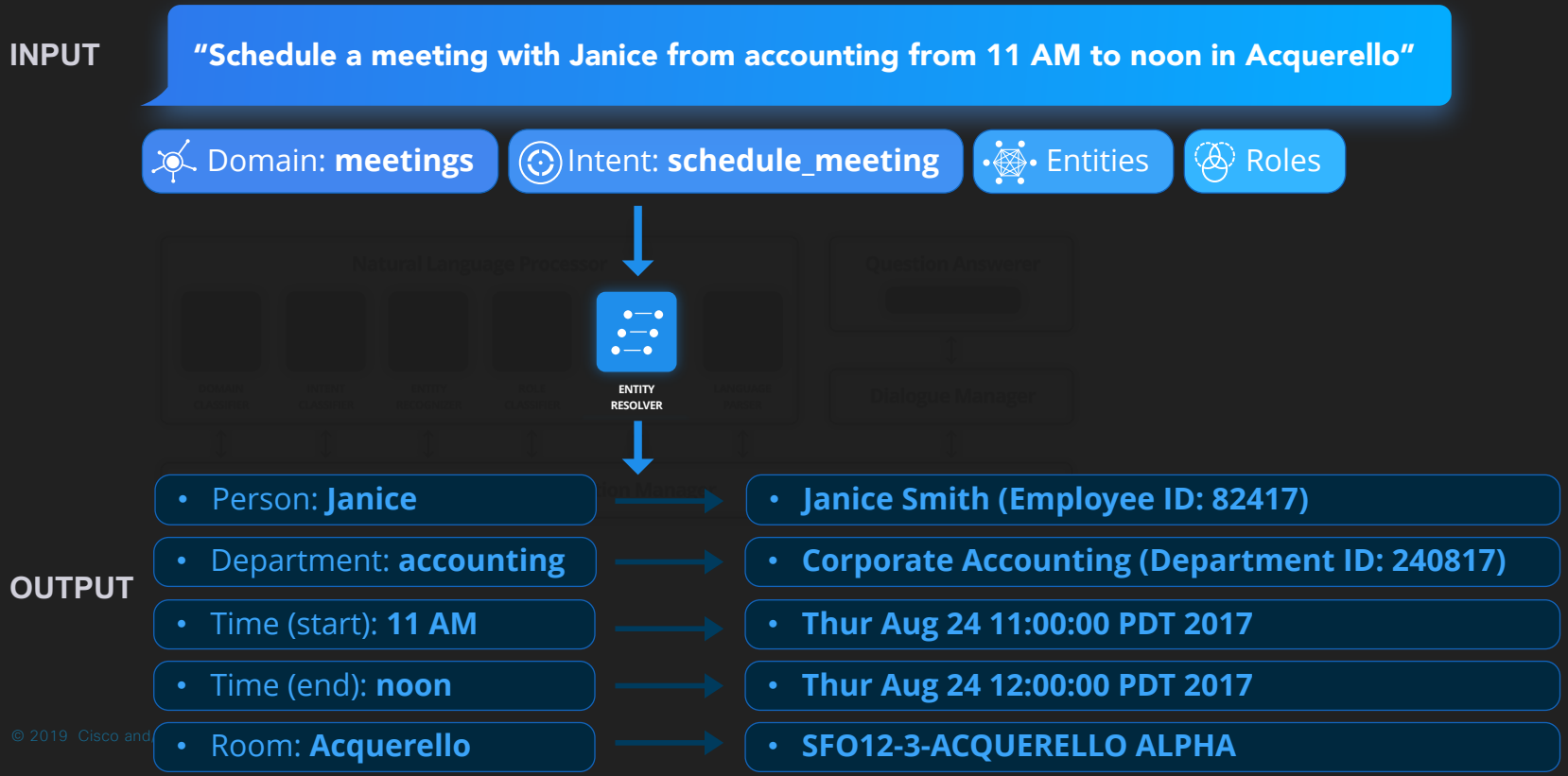
© 2019 Cisco and/or its affiliates. All rights reserved.

Entity Resolution on Noisy Voice Transcripts



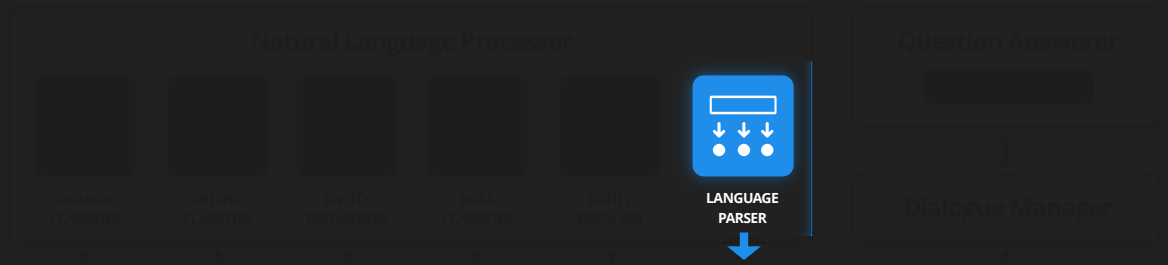
© 2019 Cisco and/or its a

Entity Resolution



© 2019 Cisco and

Language Parser clusters extracted entities into meaningful entity groups



Query
“I want a pepperoni pizza with extra cheese, a calzone and two diet coke from Domino’s on Geary”

Entities	Entity Group #1	Entity Group #2	Entity Group #3	Entity Group #4
“pepperoni pizza”, “extra cheese”, “calzone”, “two”, “diet coke”, “Domino’s”, “Geary”	pepperoni pizza • extra cheese	calzone	diet coke • two	Domino’s • Geary

- Finds relationships between the entities in a query and clusters them into meaningful entity groups
- Each entity group has an inherent hierarchy, representing a real-world organizational structure

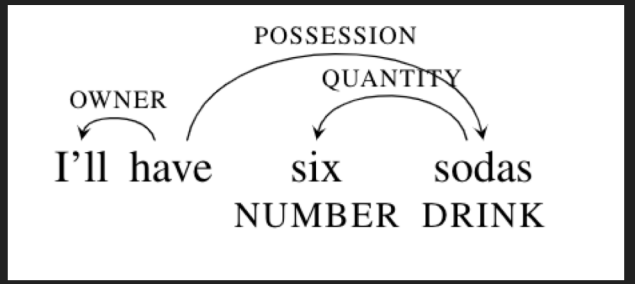
© 2019 [illegible]

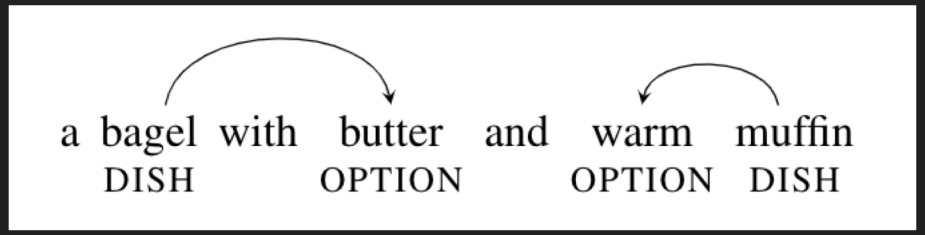
Language Parser

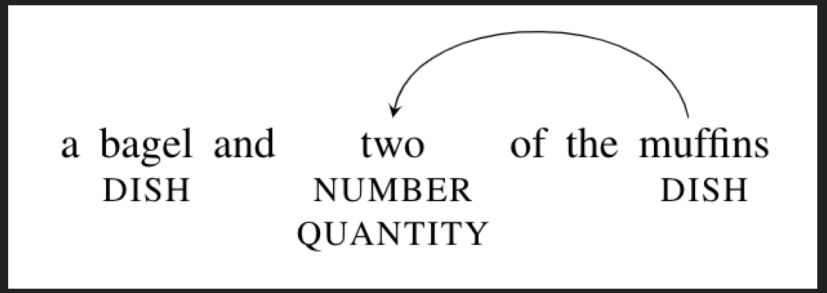
We extract just the grouping rather than a full semantic parse.

Both relation type and grouping can be found via role classification and language parsing

Higher accuracy by decomposing the problem







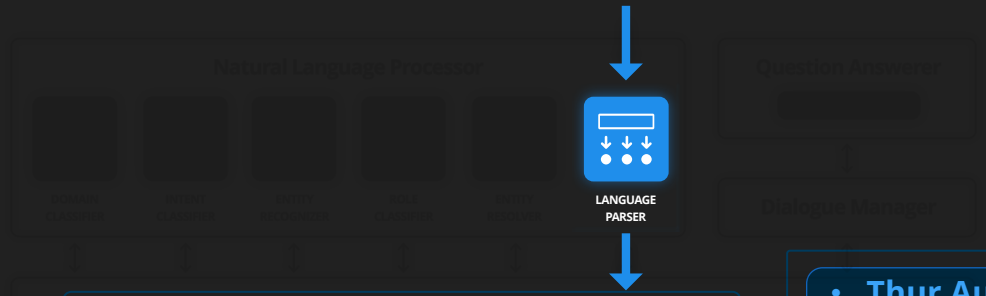
© 2019 Cisco and/or its affiliates. All rights reserved.

Language Parsing

INPUT

"Schedule a meeting with Janice from accounting from 11 AM to noon in Acquerello"

Domain: **meetings** | Intent: **schedule_meeting** | Entities | Roles

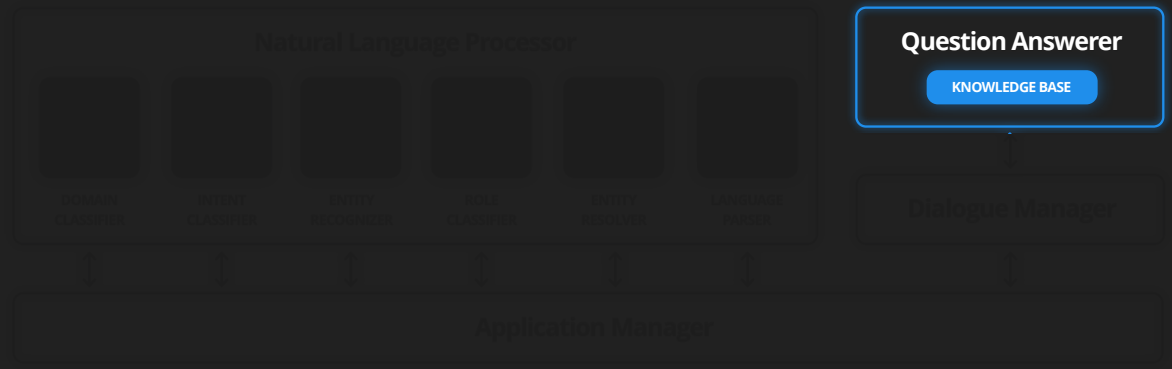


OUTPUT

- **Janice Smith (Employee ID: 82417)**
  - **Corporate Accounting (Department ID: 240817)**
- **Thur Aug 24 11:00:00 PDT 2017**
- **Thur Aug 24 12:00:00 PDT 2017**
- **SFO12-3-ACQUERELLO ALPHA**

© 2019 Cisco and/or its affiliates. All rights reserved.

Question Answerer retrieves answer candidates from the Knowledge Base



- Retrieves information from the knowledge base to identify the best answer candidates that satisfy the user-specified constraints
- Knowledge Base is a comprehensive repository of all the world knowledge that is important for a given application use case

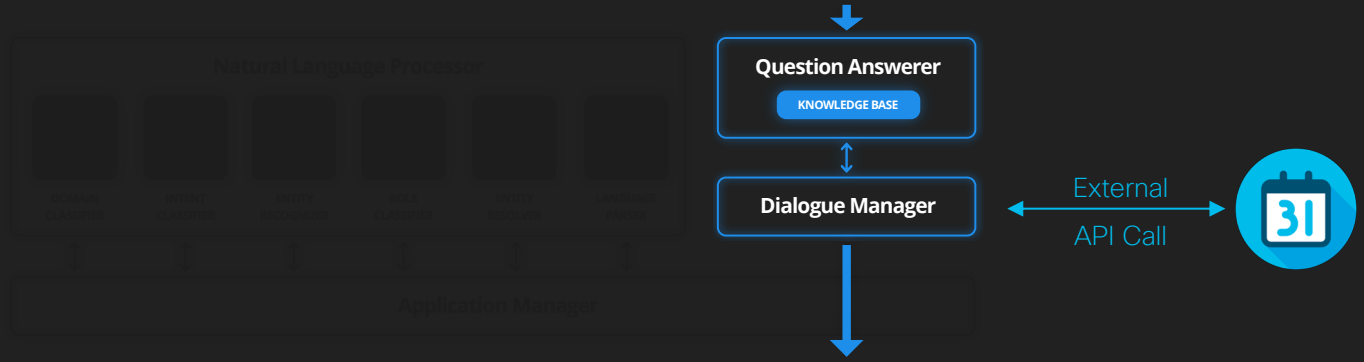
© 2019 Cisco and/or its affiliates. All rights reserved.

Question Answering

INPUT

“Schedule a meeting with Janice from accounting from 11 AM to noon in Acquerello”

Domain: **meetings** | Intent: **schedule_meeting** | Resolved Entity Groups | Roles



ACTIONS

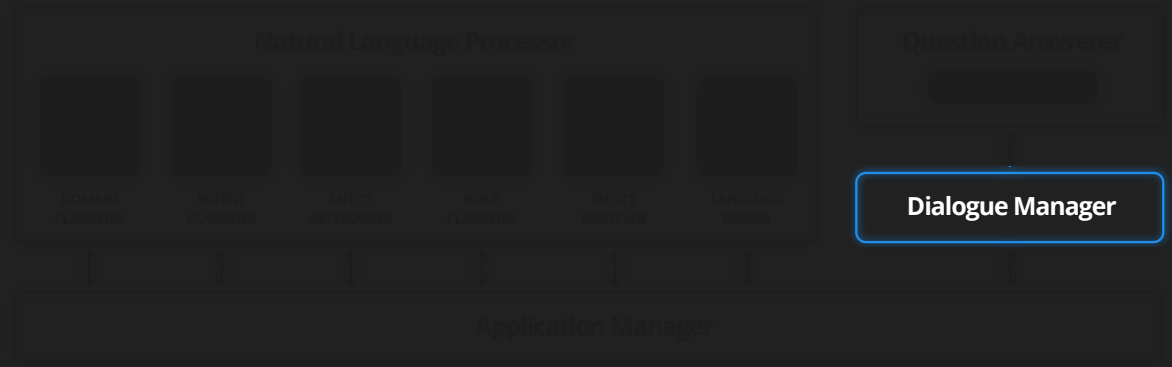
```
invitee = question_answerer.get(index='employees', emp_id='82417', dept_id='240817')

location = question_answerer.get(index='rooms', id='SFO12-3-ACQUERELLO ALPHA')

result = meeting_api.create_appointment(invitee, location, start_time, end_time)
```

© 2019 Cisco and/or its affiliates. All rights reserved.

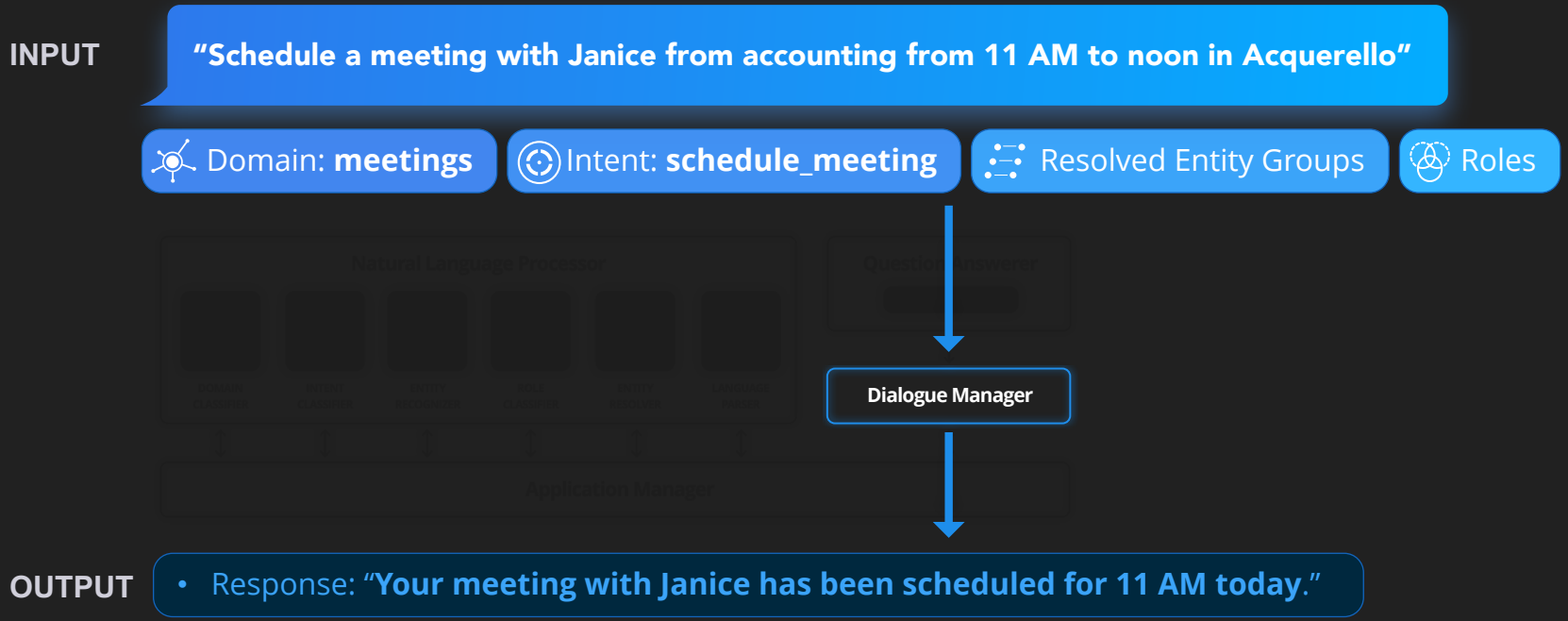
Dialogue Manager directs the flow of the conversation



- Maps NLP-analyzed user queries to appropriate dialogue states
- Moves the conversation forward by invoking the relevant dialogue state handler to execute the appropriate logic and return a response to the user

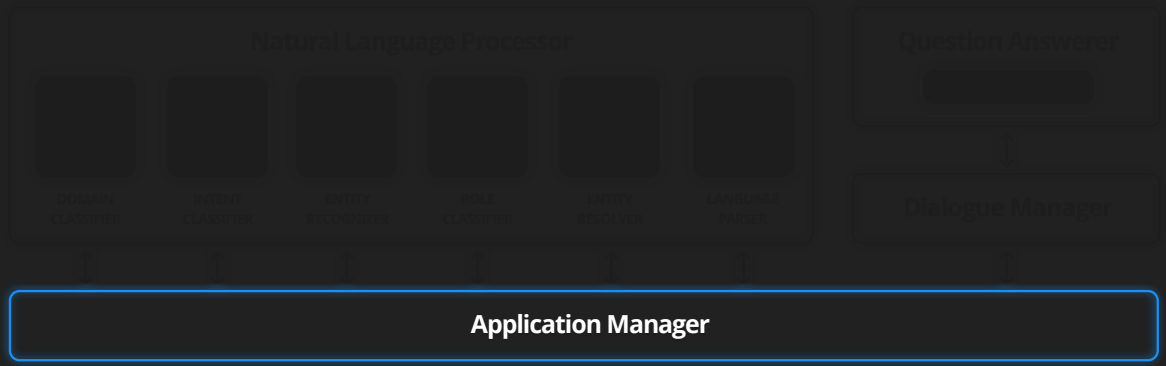
© 2019 Cisco and/or its affiliates. All rights reserved.

Dialogue Management



© 2019 Cisco and/or its affiliates. All rights reserved.

Application Manager acts as the core orchestrator



- Receives the client request from a supported endpoint
- Processes the request by passing it through all the components in the pipeline
- Returns the final response to the endpoint once processing is complete

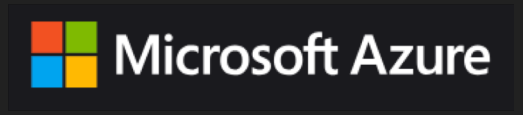
© 2019 Cisco and/or its affiliates. All rights reserved.

Text to Speech

- Third party APIs work reasonably well













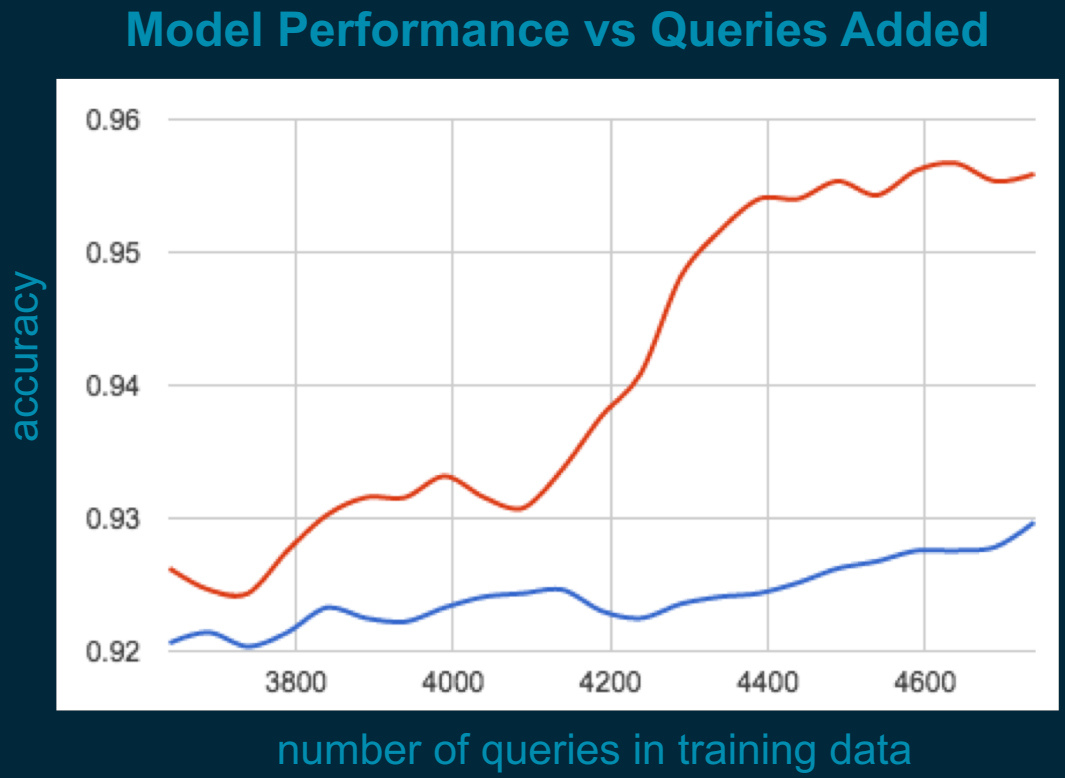
© 2019 Cisco and/or its affiliates. All rights reserved.

Ongoing Model Improvement

© 2019 Cisco and/or its affiliates. All rights reserved.

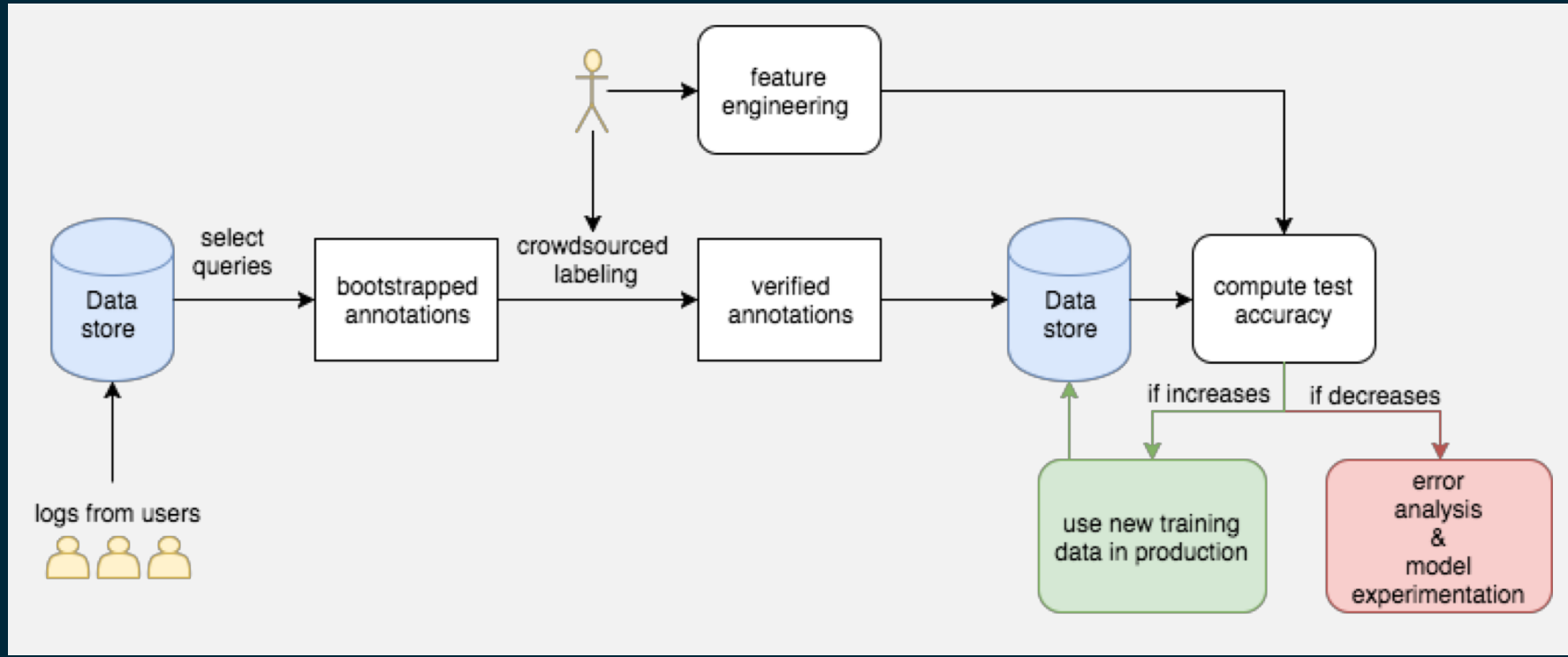
Active learning helps efficiently improve models in production

- In production, raw data is cheap but annotations are not
- We can use model uncertainty to select which queries to prioritize



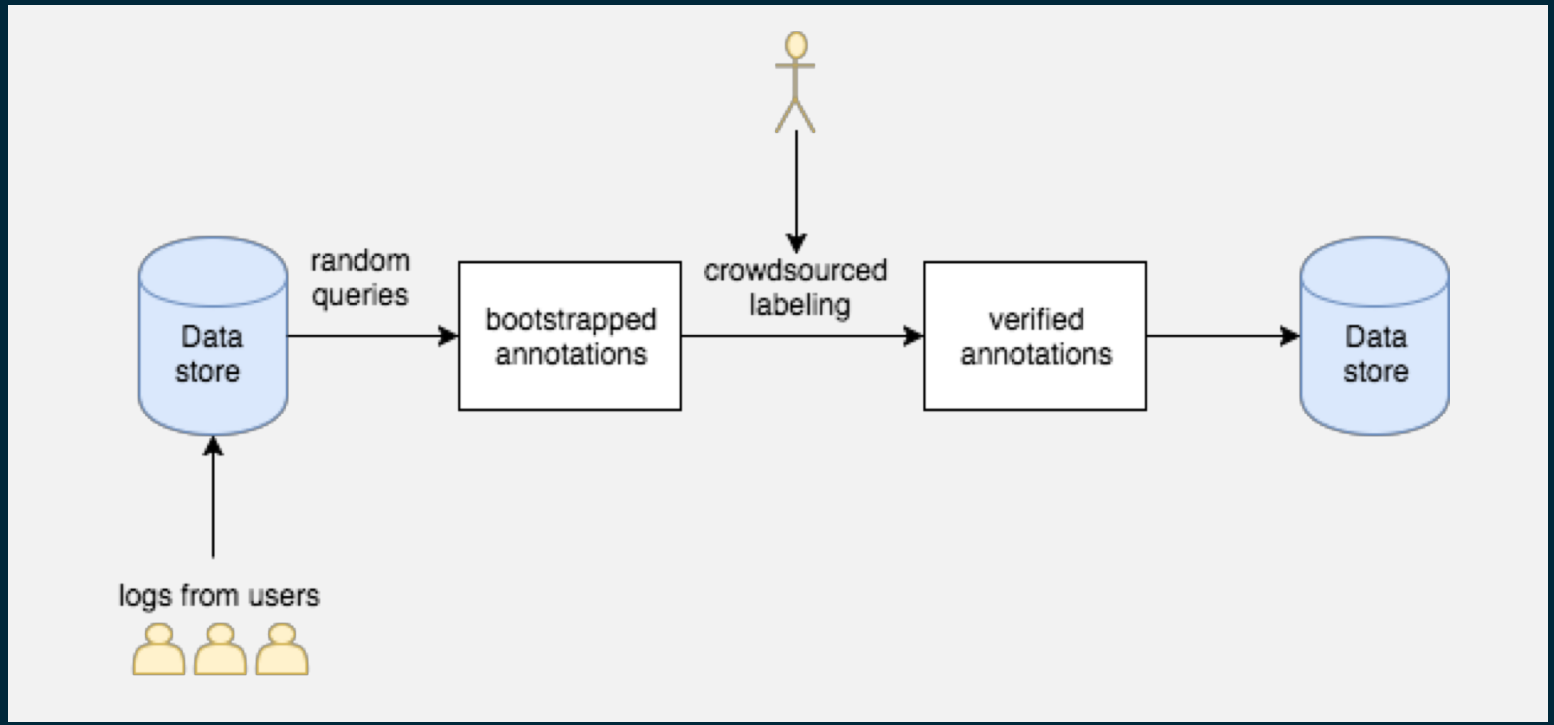
© 2019 Cisco and/or its affiliates. All rights reserved.

A semi-automated pipeline for adding training data ensures that your model continues to learn



© 2019 Cisco and/or its affiliates. All rights reserved.

A semi-automated pipeline for adding test data ensures you are evaluating your model correctly



© 2019 Cisco and/or its affiliates. All rights reserved.

Conclusion

- A massive shift in user behavior towards voice interfaces is underway, and voice assistants are starting to enter the enterprise space
- Developing a production-level conversational interface is a complex machine learning problem that involves:
    - Conversational data collection
    - Natural language processing
    - Question answering and dialog management
- Current best practices produce reliable results on limited training data, but these models will evolve as more conversational data becomes available

© 2019 Cisco and/or its affiliates. All rights reserved.

Questions?

© 2019 Cisco and/or its affiliates. All rights reserved.


CISCO