

如何建立面向任务的自然语言理解模型

曹涌

微软亚洲研究院

大数据挖掘组

什么是面向任务？



Hi. I'm Cortana.



Siri



EDI

通过自然语言交互
Natural Language Dialogue

完成特定的工作
Task Completion

具备一定的知识和推理能力
Knowledge & Reasoning



组织会议



设定提醒



园区班车



知识问答



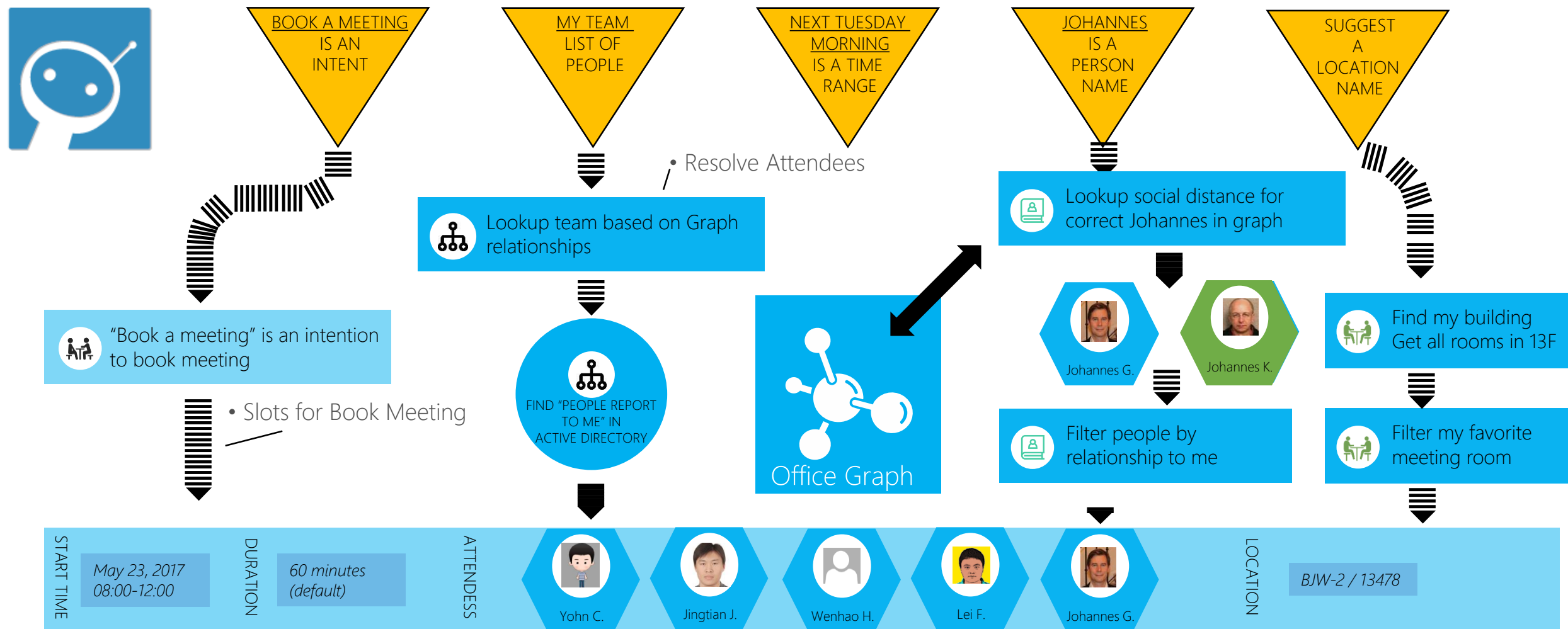
EDI – Enterprise Deep Intelligence

面向企业用户的人工智能助理



举个例子

Schedule a meeting for my team next Tuesday morning with Johannes at 13 F



什么是自然语言理解模型？

Intent

- Understanding User Intent
- 用户需要我做什么？

Entity

- Detecting Task-related Entities
- 用户给了我哪些参数？

解决方案

基于规则

- 撰写规则
- 适合冷启动
- 成长困难

基于数据

- 应用机器学习
- 冷启动困难（缺乏数据）
- 自学成材

机器学习算法

- Intent
 - 文本分类 (Text Classification) 问题
 - Naïve Bayes, Logistic Regression ...
 - Convolutional Neural Networks, LSTM ...
- Entity
 - 序列标注 (Sequence Labeling) 问题
 - Conditional Random Fields (CRFs), Semi-CRFs ...
 - grSemi-CRFs ...

基于数据的解决方案

Language Understanding & Intelligent Service

<http://www.luis.ai>

LUIS - 平台化的机器学习解决方案

- 简单易上手
 - 不需要了解算法细节
- 最少的特征工程
 - 内建基本特征
 - 深度学习（ Deep Learning ）：自动学习高阶语言表示
 - 可提供Entity列表，正则表达式
- Prebuilt domain/entity
 - 基本工具：时间，词性等标注器
 - 基本领域：Calendar, Events, Music etc.
- 最大化数据标注收益
 - 主动学习（ Active Learning ）

LUIS Demo

实用招式

兵无常势：基于规则还是基于数据？

- 并非2选1的问题
 - 各有千秋
- 针对具体的任务
 - 例如：时间，数字，比分，年龄等都比较适合用规则来抽取
- $1 + 1 > 2$
 - 规则启动，机器学习跟进
 - 强规则+机器学习
 - 机器学习+后处理规则

循序渐进：建立合适的Intent/Entity模型

尽早测试，及时得到测试用户的反馈，用以调整（增修删）Intent/Entity模型

应用反馈

需要支持哪些Intent?

Bot能够完成哪些功能？
通常一个单独的功能对应一个Intent

每个Intent需要定义哪些Entity？

对应的功能需要用户提供参数（也可以没有参数）

哪些Entity可以在Intent之间共用？

对于同类参数，在Intent之间共享可以帮助机器更好地检测它

制定/更新统一的标准

对于Intent/Entity的需要明确一个统一的标准，特别是多人合作时。

三思而行：如何合理的制定Intent/Entity列表

- Intent
 - 普通人是否可以轻松区分出所制定的不同Intent
 - “搜索Intent” vs “问答Intent”
 - 制定的Intent是否太细
 - “北京天气查询”, “上海天气查询”, ...
- Entity
 - 属于同一个Entity的文字通常在语法上是可以相互替换的
 - 被标记为“学校名称”的文字可以被其他同类型的文字替换而保持语法正确
 - 尽量在不同Intent中共享Entity
 - Reminder和Meeting都有起始时间

学海无涯：Machine Learns Forever

- 养不教，父之过
- 数据标注是教授机器的最主要途径
 - 保证数据标注的准确性
 - 增加标注数据（机器的阅读量）
- 谁是组里的“首席数据官”（Chief Data Officer）？

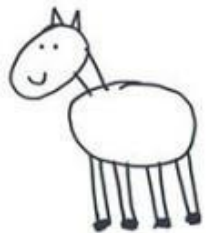
怎样画马



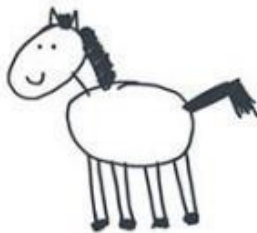
① 画两个圆圈



② 画上脚



③ 画上脸



④ 画上毛发



⑤ 再添加其他细节
就大功告成了!

细节决定成败

Reading More



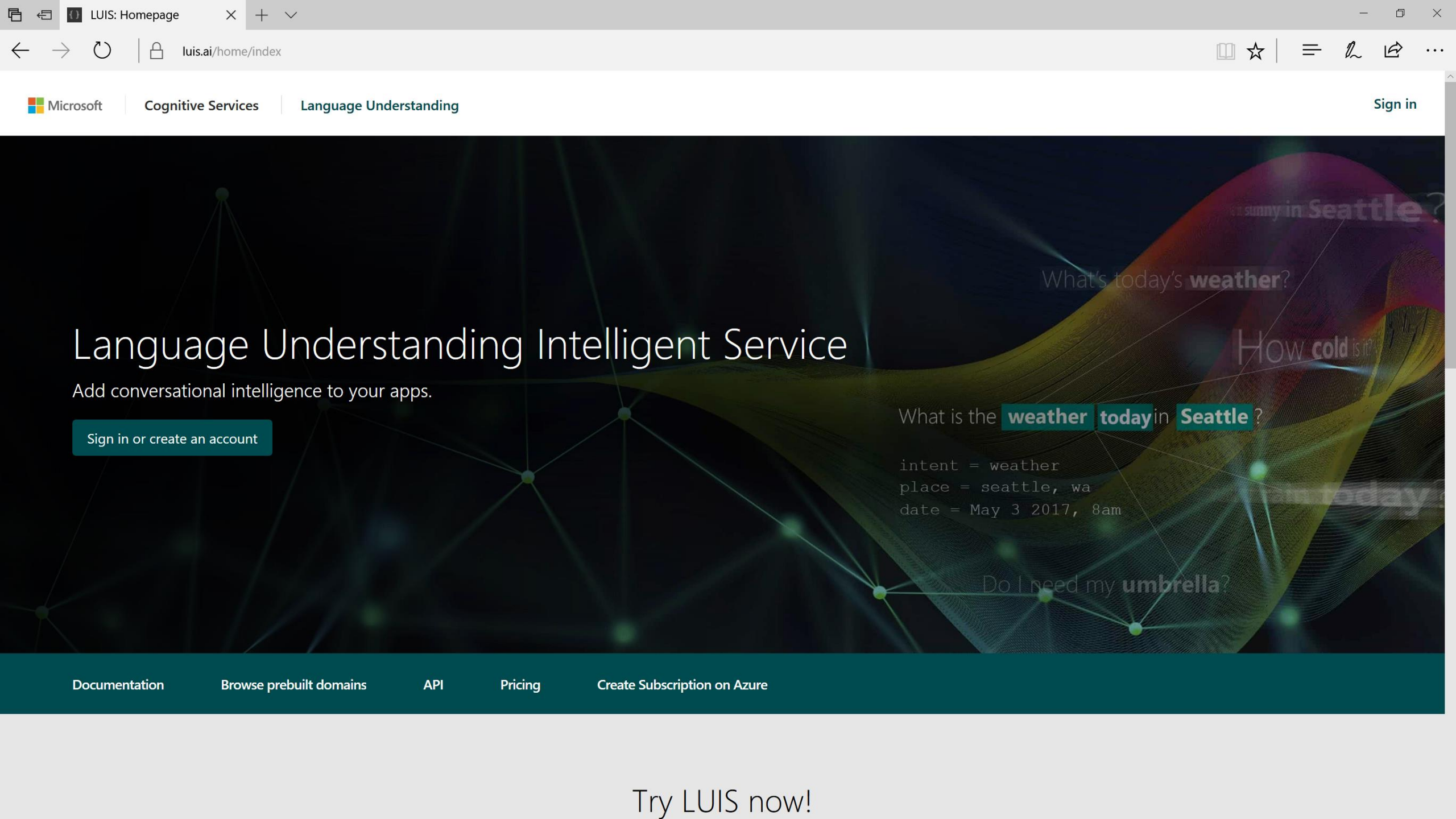
企业大数据挖掘：为员工
构建职场知识图谱



干货|如何让人工智能助理
杜绝“智障”

谢谢大家

LUIS Demo (Static)



Language Understanding Intelligent Service

Add conversational intelligence to your apps.

Sign in or create an account

What is the **weather** **today** in **Seattle** ?

```
intent = weather
place = seattle, wa
date = May 3 2017, 8am
```

[Documentation](#)

[Browse prebuilt domains](#)

[API](#)

[Pricing](#)

[Create Subscription on Azure](#)

Try LUIS now!

Language Understanding

My apps

My keys

Docs

Pricing

Support


About

My Apps

Create and manage your LUIS applications ... [Learn more](#)

New App

Import App

Cortana prebuilt apps 

Name 

Culture

Created date

Create a new app



Name (REQUIRED)

Meeting Bot

Culture (REQUIRED)

English



* App culture is the language that your app understands and speaks, not the interface language.

Description (OPTIONAL)

A meeting bot for demo



Key to use (OPTIONAL)

default



Create



Cognitive Services

[Language Understanding](#)[My apps](#)[My keys](#)[Docs](#)[Pricing](#)[Support](#)[About](#)

Meeting Bot

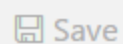
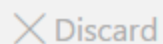
[Dashboard](#)[Intents](#)[Entities](#)[Features](#)[Train & Test](#)[Publish App](#)

Setup meeting

Here you are in full control of this intent; you can manage its utterances, used entities a

[Utterances](#) [Entities in use](#) [Suggested utterances](#)

please book a meeting for me and Jingtian early next week

[Save](#)[Discard](#)[Delete](#)[Reassign Intent](#)

Utterance text

Language Understanding

My apps

My keys

Docs

Pricing

Support

About

Meeting Bot

Dashboard

Intents

Entities

Features

Train & Test

Publish App

← Back to App list

Entities

Manage a list of entities in your application and track and control their instances within utterances ... [Learn more](#)

Entities list

Labeled utterances

Suggested utterances

Add custom entity

Add prebuilt entity

Entity Name ↓	Entity Type	
Building	Simple	
Capacity	Simple	
Duration	Simple	
Floor	Simple	
Group-of-Person	Simple	
Person	Simple	
Range	Simple	
Range:end	Simple	
Range:start	Simple	
Room	Simple	




Meeting Bot

[Dashboard](#)
[Intents](#)
[Entities](#)
[Features](#)
[Train & Test](#)
[Publish App](#)
[← Back to App list](#)

Setup meeting

Here you are in full control of this intent; you can manage its utterances, used entities and sug

Utterances (1) Entities in use Suggested utterances

 Save
  Discard
  Delete
 Reassign Intent ▾

☐ Utterance text

☐ please book a meeting for [me] and jingtian early next week

[Wrap in composite entity](#)

Group-of-Person

Person

Range

Range:end

Range:start

Room

1

Setup meeting

Here you are in full control of this intent; you can manage its utterances, used entities and suggested utterances ... [Learn more](#)

Utterances (60) Entities in use (9) Suggested utterances

Type a new utterance & press Enter ... ✕

<div> <div>Save</div> <div>Discard</div> <div>Delete</div> <div>Reassign Intent</div> </div> <div> <div>Labels view (Ctrl+E):</div> <div>Entities</div> <div>Search in utterances ...</div> </div>	
Utterance text	Predicted Intent
<input type="checkbox"/> schedule a meeting with [Person] at [Range:start] [Range]	Not trained
<input type="checkbox"/> i need a conference room on [Range] at [Range:start] .	Not trained
<input type="checkbox"/> book a meeting with [Person] on [Range]	Not trained
<input type="checkbox"/> book a meeting room [Range] [Range:start]	Not trained
<input type="checkbox"/> please schedule a meeting with [Person] from [Range:start] to [Range:end] [Range]	Not trained
<input type="checkbox"/> please schedule a meeting with [Person] on [Range] [Range]	Not trained
<input type="checkbox"/> book [Person] a meeting every work day from [Range:start] to [Range:end] with [Person] and [Person]	Not trained
<input type="checkbox"/> could you please book room [Building] [Room] at [Range:start] on [Range] for [Duration] ?	Not trained
<input type="checkbox"/> book a room from [Range:start] to [Range:end] on the [Floor]	Not trained
<input type="checkbox"/> I want to book a room [Room] on [Date] from [Time] to [Time] with [Person]	Not trained



Meeting Bot

Version: 0.1

Settings

Dashboard

Intents

Entities

Prebuilt domains PREVIEW

Features

Train & Test

Publish App

← Back to App list

Edit Phrase list

Phrase list name (REQUIRED)

person name

Value (REQUIRED)

Type and press enter ...

☒ Hide related values

Phrase list values (REQUIRED)

john × tom × rose × yohn × james × paul × david ×
robert × jones × thomas × chris × michael × jim ×
andrew × peter ×

Related Values

[Add all](#) [Recommend](#)

richard + george + brian + mark + scott + kevin +
matthew + daniel + keith + joseph +

☒ Is exchangeable?

☒ Is active?

Save

Cancel

Meeting Bot

[Dashboard](#)
[Intents](#)
[Entities](#)
[Features](#)
[Train & Test](#)
[Publish App](#)
[← Back to App list](#)

Features

Use these advanced features to improve performance and avoid mistakes in identifying and interpreting utterances ... [Learn more](#)

Phrase list features (1/10)

Pattern features (0/10)

Add pattern feature

Active

Patte

Add Pattern



Pattern name (REQUIRED)

24-hour time

Pattern value (REQUIRED)

\d{2}:\d{2}



Is active?

[Learn about supported regex syntax](#) ▾

Save

Cancel

ern feature

① Training app ... 8 / 12 completed

Cognitive Services

Language Understanding

My apps

My keys

Docs

Pricing

Support

About

Meeting Bot

Dashboard

Intents

Entities

Features

Train & Test

Publish App

← Back to App list

Test your application

Use this tool to test the current and published versions of your application, to check if you are progressing on the right track ... [Learn more](#)

Train Application

Please train your application before testing.

Interactive Testing Batch Testing

☐ Enable published model

Labels view (Ctrl

Please train your application before testing.

Train

Meeting Bot

[Dashboard](#)[Intents](#)[Entities](#)[Features](#)[Train & Test](#)[Publish App](#)[← Back to App list](#)

Test your application

Use this tool to test the current and published versions of your application, to check if you are progressing on the right track ... [L](#)

[Train Application](#)

Last train: Apr 25, 2017, 9:50:46 PM | Last publish: **Not published yet.**

Interactive Testing Batch Testing

☐ [Enable published model](#)

Type a test utterance & press Enter



please set up a meeting [tomorrow]

Current version results

Top scoring intent
Setup meeting (0.95)

Other intents
None (0.04)

Language Understanding My apps My keys Docs Pricing Support About

Meeting Bot

Dashboard

Intents

Entities

Features

Train & Test

Publish App

← Back to App list

Test your application

Use this tool to test the current and published versions of your application, to check if you are progressing on the right track ... [Learn more](#)

Train Application

Last train: Apr 25, 2017, 9:50:46 PM | Last publish: **Not published yet.**

Interactive Testing Batch Testing

☐ Enable published model

Labels view (Ctrl+E)

Entities

Type a test utterance & press Enter



please set up a meeting [\$Range]

Current version results

Top scoring intent

Setup meeting (0.95)

Other intents

None (0.04)

Meeting Bot

Dashboard

Intents

Entities

Features

Train & Test

Publish App

← Back to App list

Publish App

Publish your app as a web service or as a chat bot. You can publish a new app or an updated version of a published app ... [Learn more](#)

Essentials

Latest publish: You haven't published your application yet

Endpoint Key (REQUIRED)

Choose a key to assign to application ... ▼

[Add a new key to your account](#)

Publish settings

Endpoint slot

Production ▼

This slot has no published application.

Train

Publish

Please assign a key to your app to publish.

Dashboard

Intents

Entities

Features

Train & Test

Publish App

← Back to App list

Essentials

Latest publish: Just now

Endpoint Key (REQUIRED)

default ▼

This key is for experimental use and is valid for 30 days and 1000 endpoint hits only.

[Add a new key to your account](#)

Publish settings

Endpoint slot

Production ▼

Slot info

Published version Id: 0.1

Published date: Apr 25, 2017, 9:57:43 PM

Endpoint url

<https://westus.api.cognitive.microsoft.com/luis/v2.0/apps/0120a8a7-b8d5-4149-be8b-80e3d288fa8f?subscription-key=d77311f8d77947ac919120ce69d0d0b9&verbose=true&timezonedOffset=0&q=>

☒ Add verbose flag

☐ Enable Bing spell checker

Timezone

(GMT) Western Europe Time, London, Lisbon, Casal ▼

Train

Publish

RunnerImport

BuilderTeam Library

SYNC OFF

Sign In

Filter

HistoryCollections

Today

GET https://westus.api.cognitive.microsoft.com/luis/v2.0/apps/0120a8a7-b8d5-4149-be8b-80e3d288fa8f?subscription-key=...

https://westus.api.cog x +

No Environment

GET https://westus.api.cognitive.microsoft.com/luis/v2.0/apps/0120a8a7-b8d5-4149-be8b-80e3d288fa8f?subscription-key=... Params Send Save

Key	Value	Bulk Edit
subscription-key	d77311f8d77947ac919120ce69d0d0b9	
verbose	true	
timezoneOffset	0	
q	please set up a meeting tomorrow	
New key	value	

AuthorizationHeadersBodyPre-request ScriptTestsCode

TypeNo Auth

BodyCookiesHeaders (9) TestsStatus: 200 OKTime: 1439 ms

PrettyRawPreviewJSON

```
1 {
2   "query": "please set up a meeting tomorrow",
3   "topScoringIntent": {
4     "intent": "Setup meeting",
5     "score": 0.951010644
6   },
7   "intents": [
8     {
9       "intent": "Setup meeting",
10      "score": 0.951010644
11     },
12     {
13       "intent": "None",
14       "score": 0.04201392
15     }
16   ],
17   "entities": [
18     {
19       "entity": "tomorrow",
20       "type": "Range",
21       "startIndex": 24,
22       "endIndex": 31,
23       "score": 0.5537152
24     }
25   ]
26 }
```

Language Understanding

My apps

My keys

Docs

Pricing

Support

About

Meeting Bot

Dashboard

Intents

Entities

Features

Train & Test

Publish App

← Back to App list

Setup meeting

Here you are in full control of this intent; you can manage its utterances, used entities and suggested utterances ... [Learn more](#)

Utterances (60) Entities in use (9) Suggested utterances

Save Discard Delete Reassign Intent ▾			Labels view (Ctrl+E):	Entities ▾	Search in utterances ... 🔍	⌵
<input type="checkbox"/>	Utterance text	Suggested Intent				
<input type="checkbox"/>	please set up a meeting [\$Range]	0.95 Setup meeting				