TECH TALKS

AN OVERVIEW OF SPOKEN LANGUAGE UNDERSTANDING

Xuedong Huang

Chief Speech Scientist & Distinguished Engineer

Microsoft Corporation

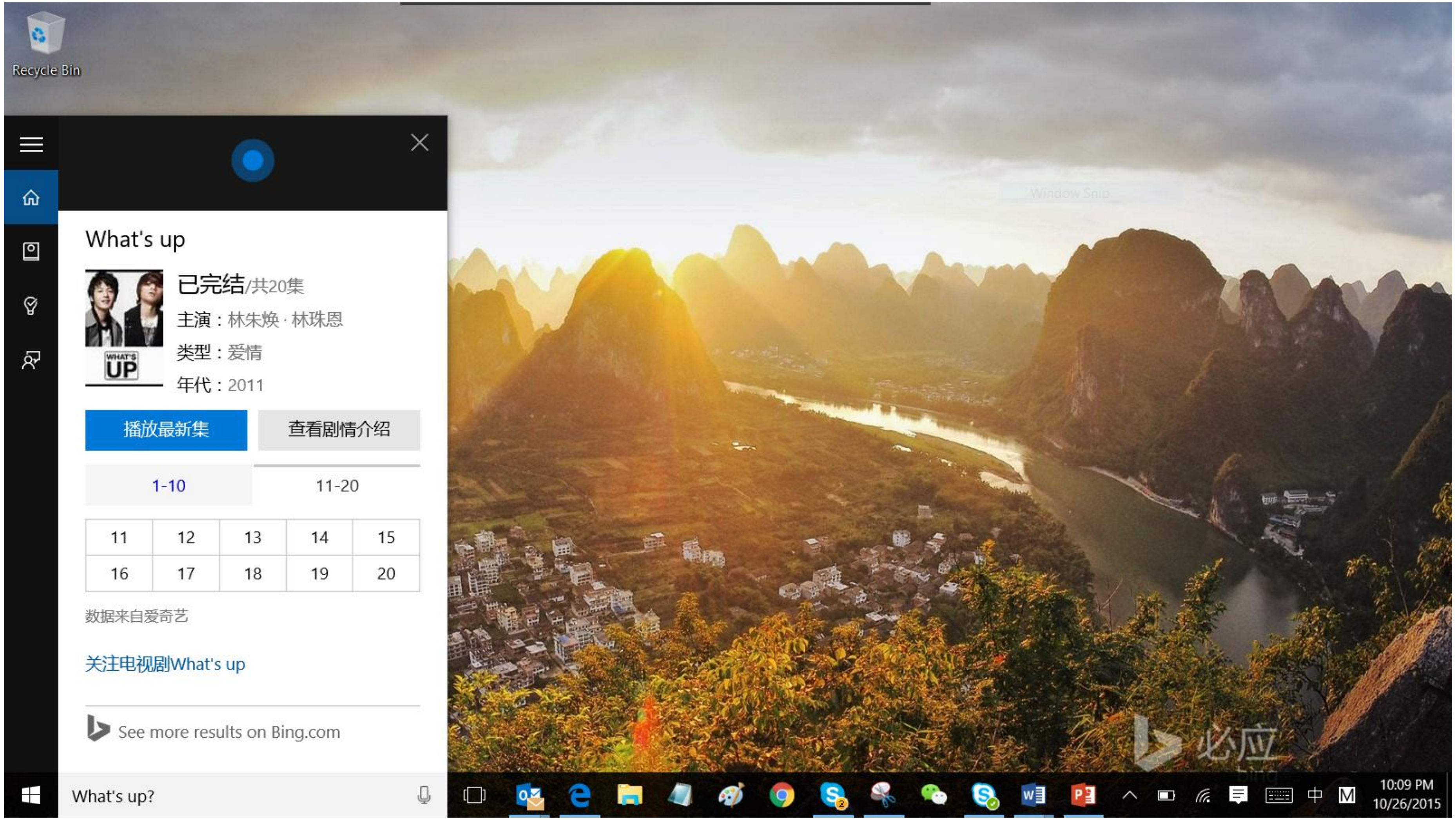
xdh@microsoft.com

An **invisible** revolution is coming

ARE WE READY?

Cloud-enabled multimodal NUI with speech, gesture, gaze…

TODAY'S STATE OF THE ART: CORTANA



What's up

已完结/共20集

主演：林朱焕 · 林珠恩

类型：爱情

年代：2011

11	12	13	14	15
16	17	18	19	20

数据来自爱奇艺

关注电视剧What's up

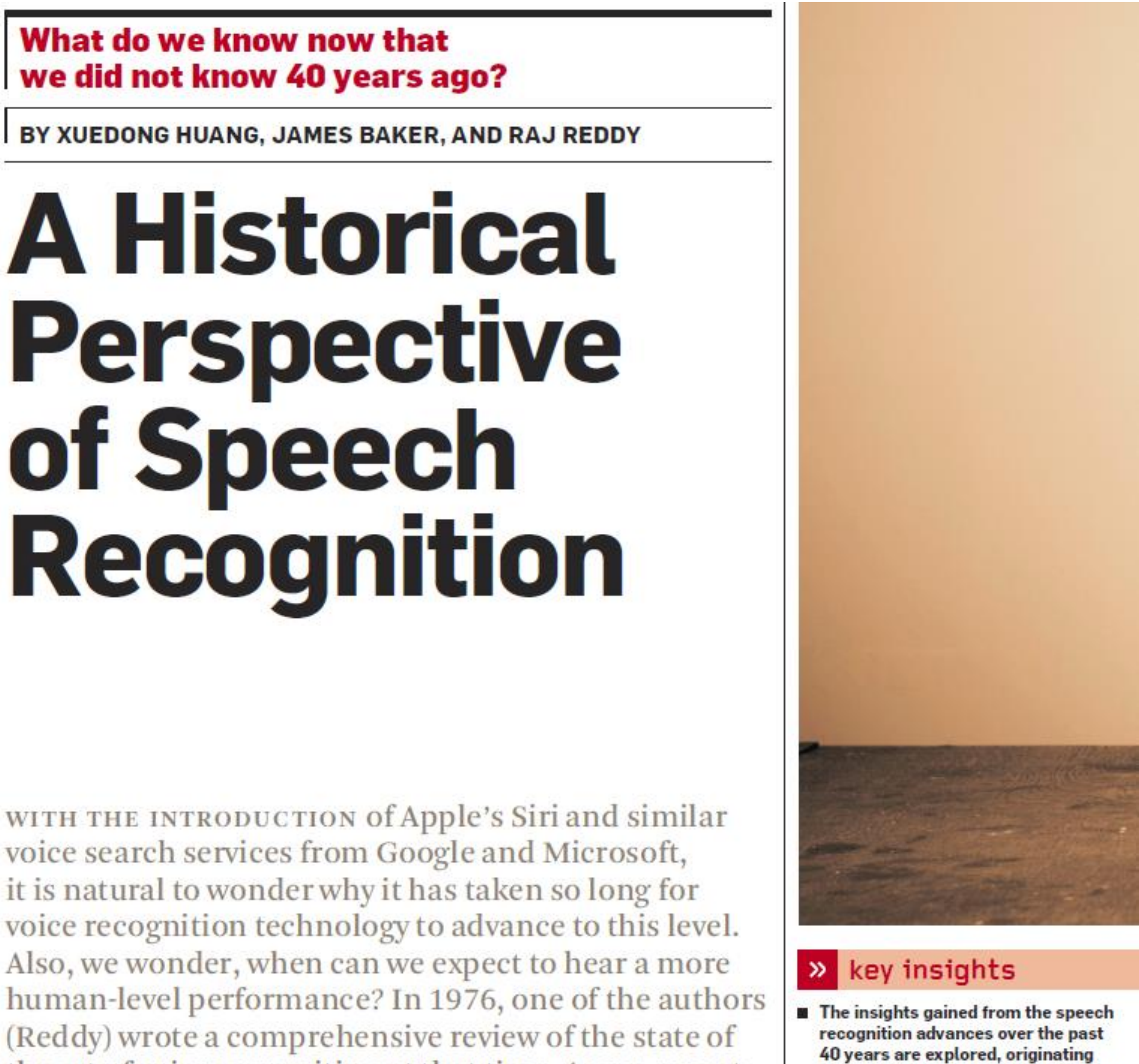
See more results on Bing.com

What's up?

TODAY'S STATE OF THE ART: SKYPE TRANSLATOR

SPEECH RECOGNITION – APPROACHING HUMAN PARITY

http://cacm.acm.org/magazines/2014/1/170863-a-historical-perspective-of-speech-recognition



What do we know now that we did not know 40 years ago?

BY XUEDONG HUANG, JAMES BAKER, AND RAJ REDDY

A Historical Perspective of Speech Recognition

WITH THE INTRODUCTION of Apple's Siri and similar voice search services from Google and Microsoft, it is natural to wonder why it has taken so long for voice recognition technology to advance to this level. Also, we wonder, when can we expect to hear a more human-level performance? In 1976, one of the authors (Reddy) wrote a comprehensive review of the state of

» key insights

- The insights gained from the speech recognition advances over the past 40 years are explored, originating

SECRET SAUCE: BIG DATA + MACHINE LEARNING + INFRASTRUCTURE

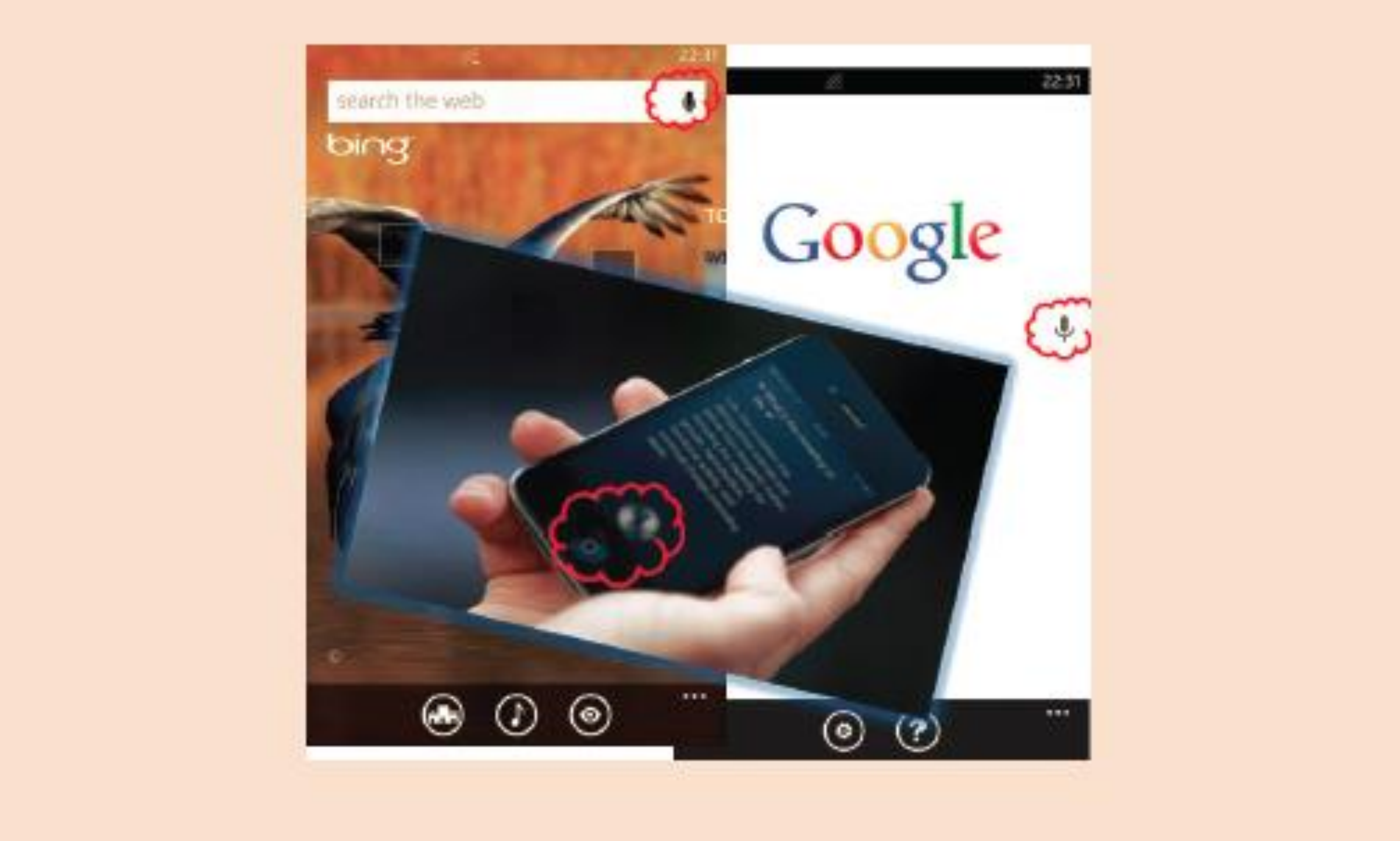
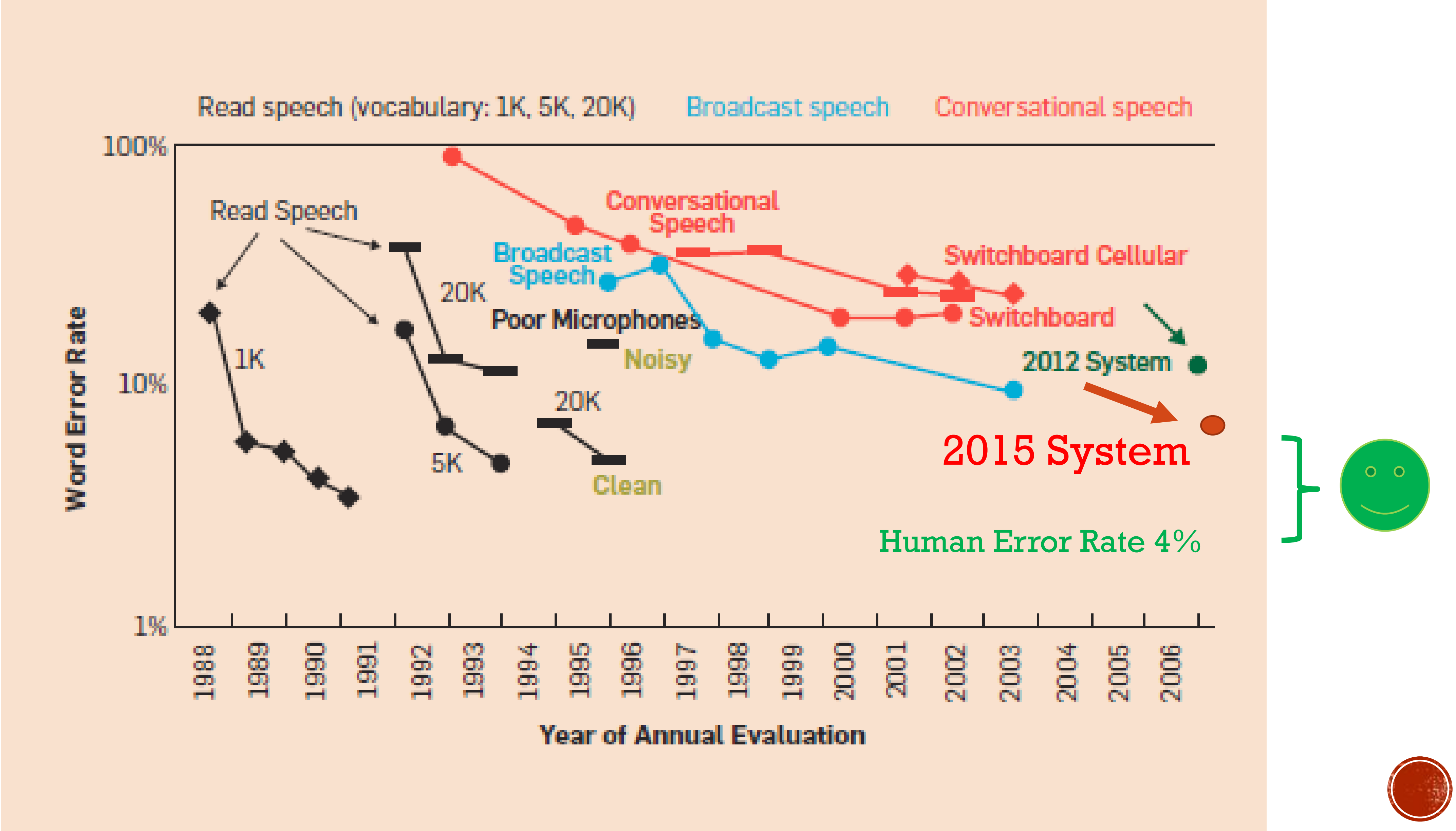


Figure 1. Historical progress of speech recognition word error rate on more and more difficult tasks.[10] The latest system for the switchboard task is marked with the green dot.

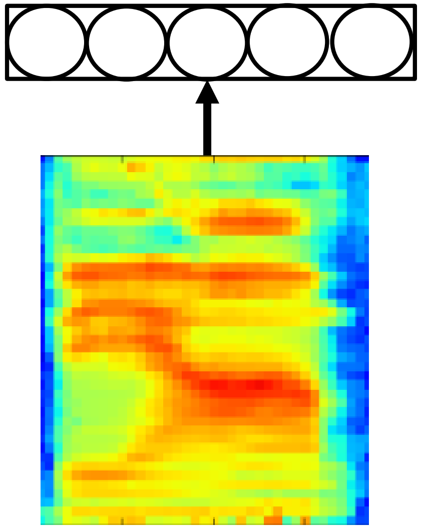


2015 System

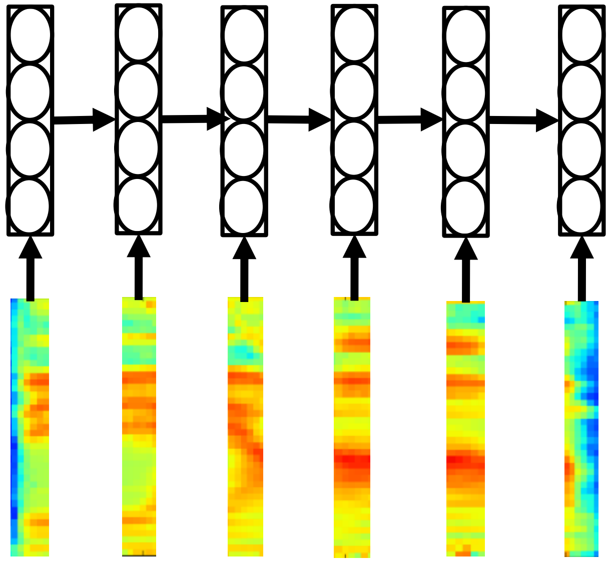
Human Error Rate 4%

BEYOND FF-DNNS

- Speech is a sequential process while FF-DNNs are not sequential in nature.
- FF-DNNs are not efficient in modeling temporal/spectral compression/stretch.
- Theoretically, recurrent models should better model speech data.
  - In RNNs, hidden representations are conditioned on all previously seen frames.



(a) FF-DNN



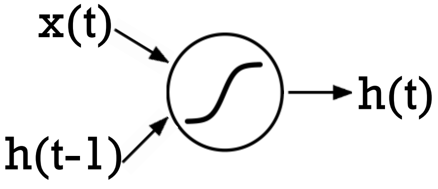
(b) RNN

FROM SIMPLE RNNS TO LSTM RNNS

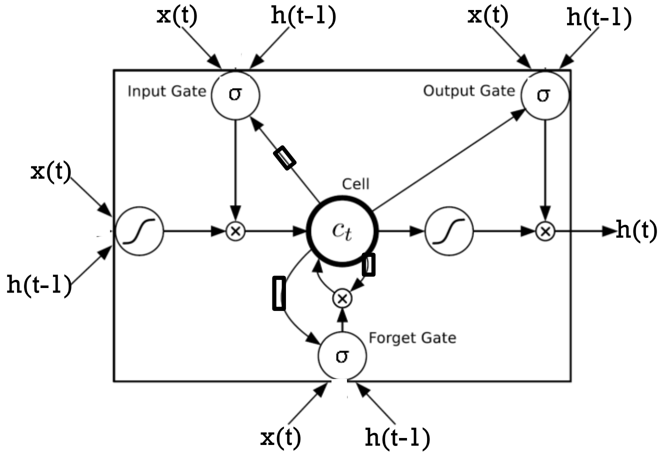
- At each time step, each hidden node in an RNN consumes the current input vector and the previous hidden vector.

- RNNs have a limited capacity to remember important distant past events. (Vanishing gradient problem)

- LSTM input, output, and forget gates combat vanishing gradient problem.



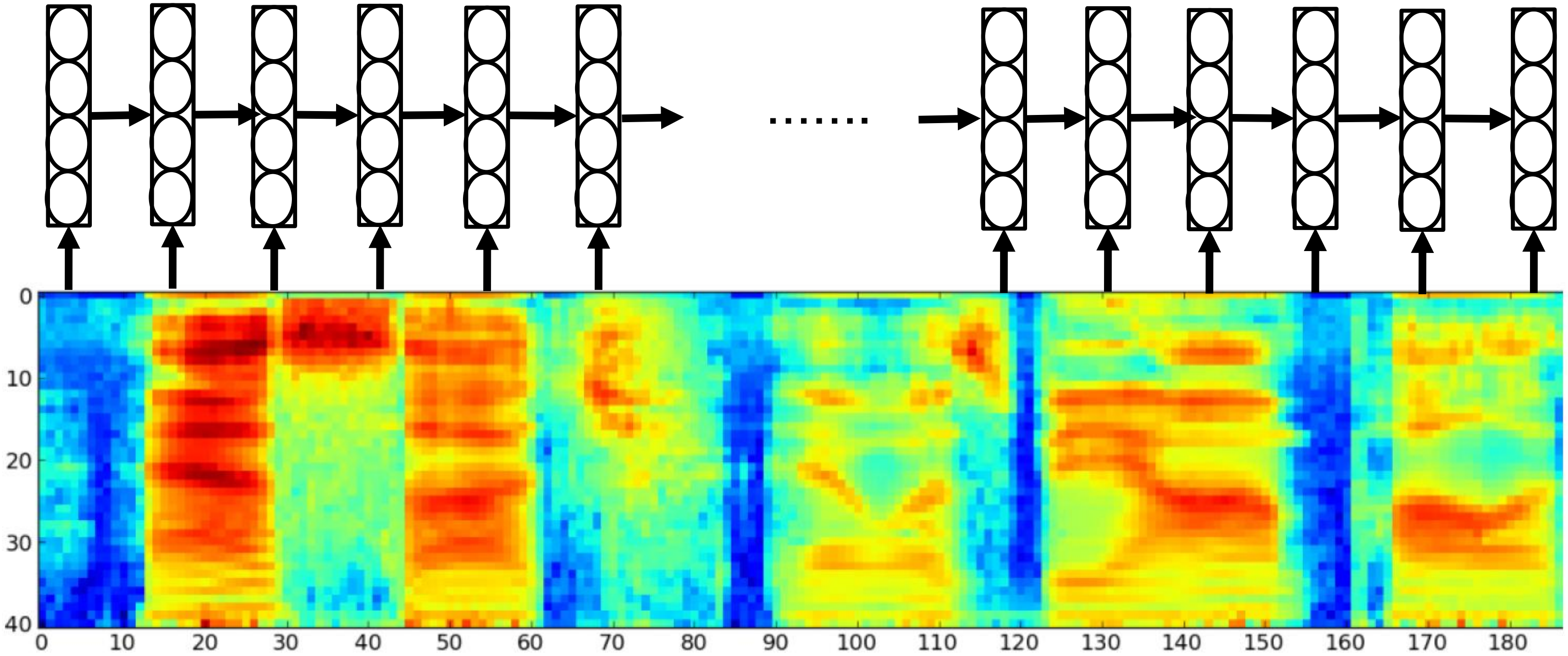
(a) RNN node



(b) LSTM node

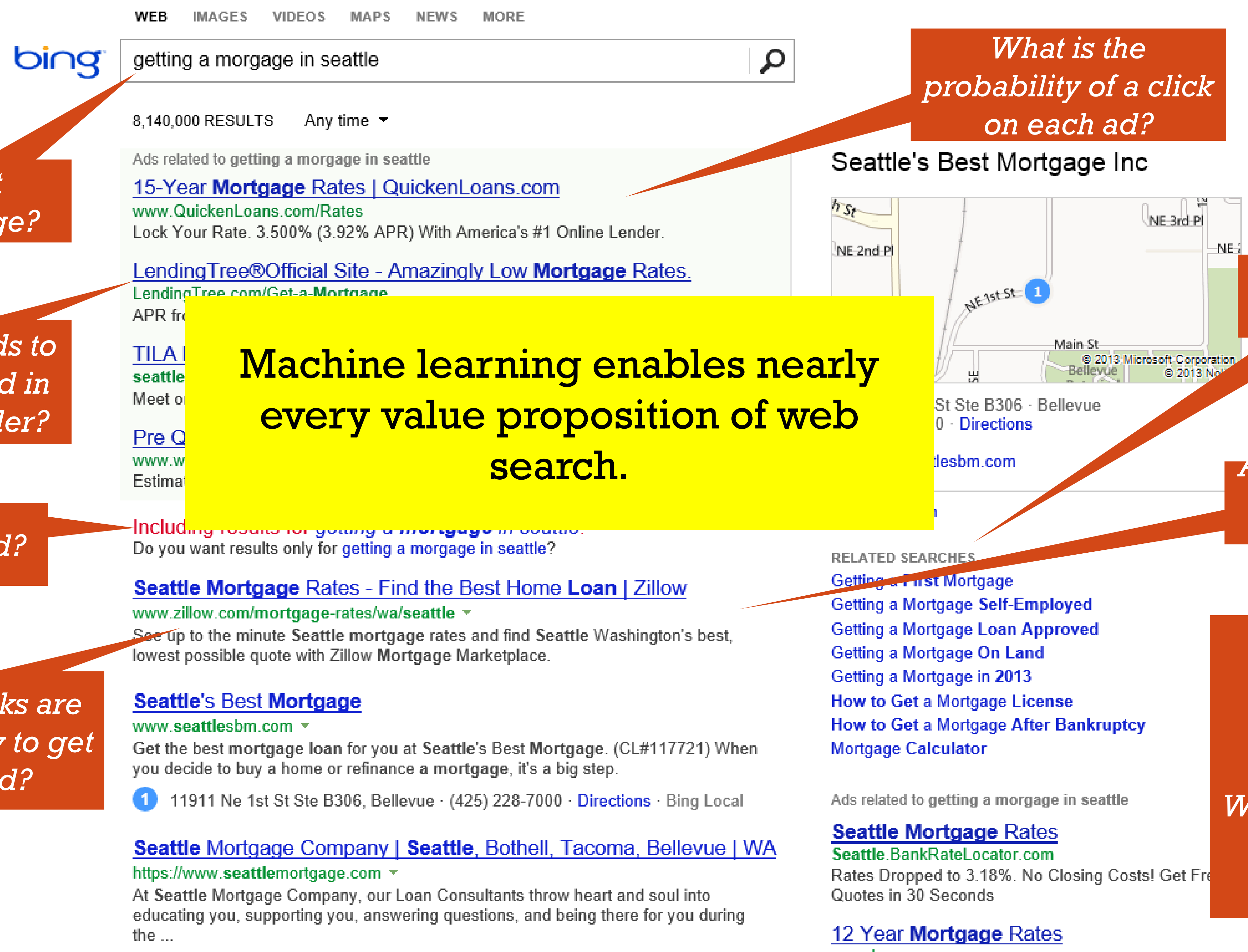
ACOUSTIC MODELING USING LSTMS

- Models long-span phenomena well
- Good performance improvements observed
- But less convenient than fixed window methods



UNDERSTANDING LANGUAGE – MUCH HARDER

Machine learning enables nearly every value proposition of web search.



WEB IMAGES VIDEOS MAPS NEWS MORE

bing

getting a morgage in seattle

8,140,000 RESULTS Any time

Ads related to **getting a morgage in seattle**

15-Year **Mortgage** Rates | QuickenLoans.com
www.QuickenLoans.com/Rates
Lock Your Rate. 3.500% (3.92% APR) With America's #1 Online Lender.

LendingTree®Official Site - Amazingly Low **Mortgage** Rates.
LendingTree.com/Get-a-Mortgage
APR fr

TILA I
seattle
Meet o

Pre Q
www.w
Estimat

Including results for *getting a mortgage in seattle*.
Do you want results only for getting a morgage in seattle?

Seattle Mortgage Rates - Find the Best Home **Loan** | Zillow
www.zillow.com/mortgage-rates/wa/**seattle**
See up to the minute **Seattle mortgage** rates and find **Seattle** Washington's best, lowest possible quote with Zillow **Mortgage** Marketplace.

Seattle's Best **Mortgage**
www.**seattle**sbm.com
Get the best **mortgage loan** for you at **Seattle**'s Best **Mortgage**. (CL#117721) When you decide to buy a home or refinance **a mortgage**, it's a big step.
1 11911 Ne 1st St Ste B306, Bellevue · (425) 228-7000 · Directions · Bing Local

Seattle Mortgage Company | **Seattle**, Bothell, Tacoma, Bellevue | WA
https://www.**seattle**mortgage.com
At **Seattle** Mortgage Company, our Loan Consultants throw heart and soul into educating you, supporting you, answering questions, and being there for you during the ...

Seattle's Best Mortgage Inc

St Ste B306 · Bellevue
0 · Directions
tlesbm.com

RELATED SEARCHES
Getting a First Mortgage
Getting a Mortgage **Self-Employed**
Getting a Mortgage **Loan Approved**
Getting a Mortgage **On Land**
Getting a Mortgage **in 2013**
How to Get a Mortgage **License**
How to Get a Mortgage **After Bankruptcy**
Mortgage **Calculator**

Ads related to **getting a morgage in seattle**

Seattle Mortgage Rates
Seattle.BankRateLocator.com
Rates Dropped to 3.18%. No Closing Costs! Get Fre
Quotes in 30 Seconds

12 Year **Mortgage** Rates

What language?

What is the probability of a click on each ad?

Which ads to show, and in what order?

What is the intent?

Misspelled?

Are any of these pages malicious?

Which links are most likely to get clicked?

What pages should we index?

What ad pricing will optimize revenue?

CORTANA

A personal assistant built around you

Cortana gets to know you over time, building a relationship with you that’s based on trust.

She tracks the stuff you care about, looks out for you throughout the day, and helps filter out the noise so you can stay on top of what matters.

Throughout, Cortana is delightful and easy to use.

CORTANA – CORE PILLARS

1 PERSONAL

Cortana…

…is your truly personal assistant

…gets to know you

…is transparent

Proof points:

- Learning
- Notebook
- Personal suggestions
- Transparency & control

2 LOOKS OUT FOR YOU

Cortana…

…looks out for you

…filters out the noise

…reminds you of what's important

Proof points:

- Useful and relevant alerts
- Planners
- Event scheduling
- Quiet hours and inner circle
- Reminders

3 DELIGHTFUL & EASY TO USE

Cortana…

…"just works"

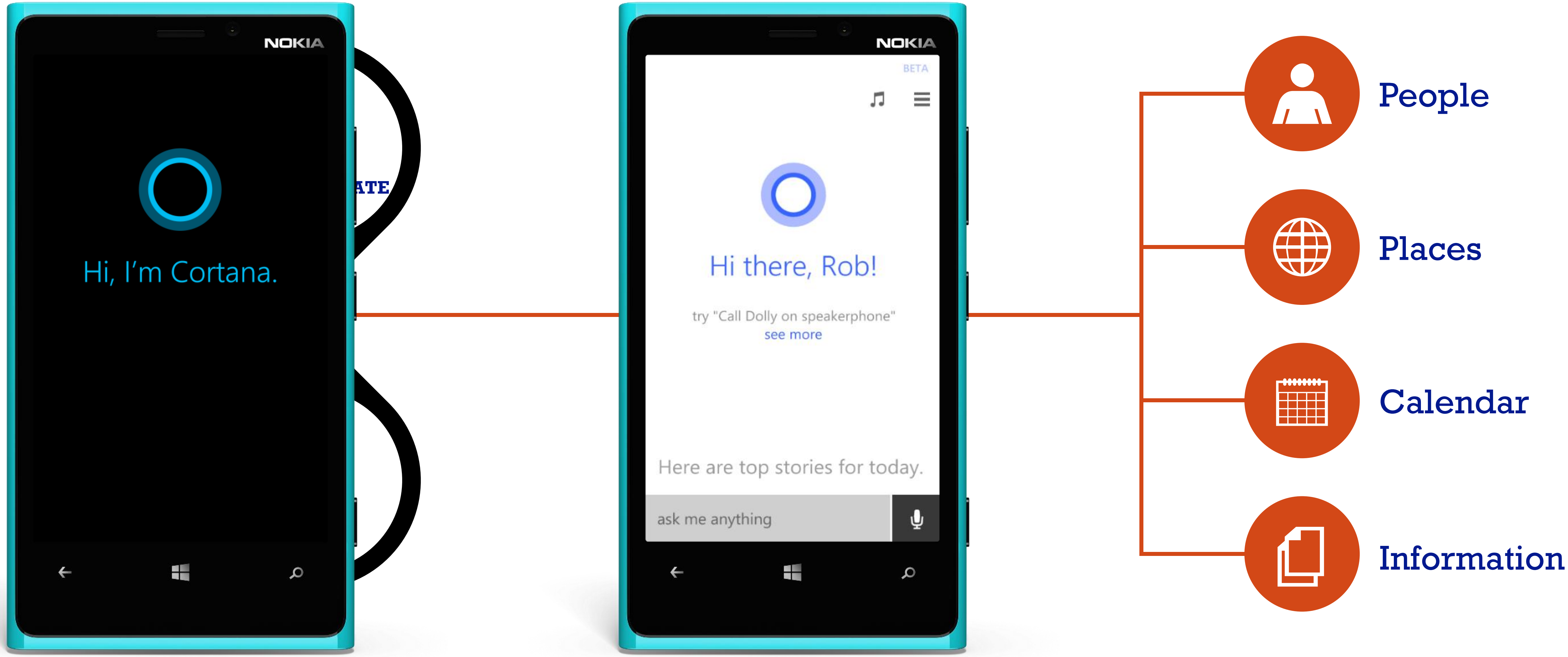
…lets you interact on your terms

…has a fun & engaging personality

Proof points:

- Voice & natural language
- Text input
- Personality (visual, spoken voice, and behavior)

Cortana: What Can You Do?



People

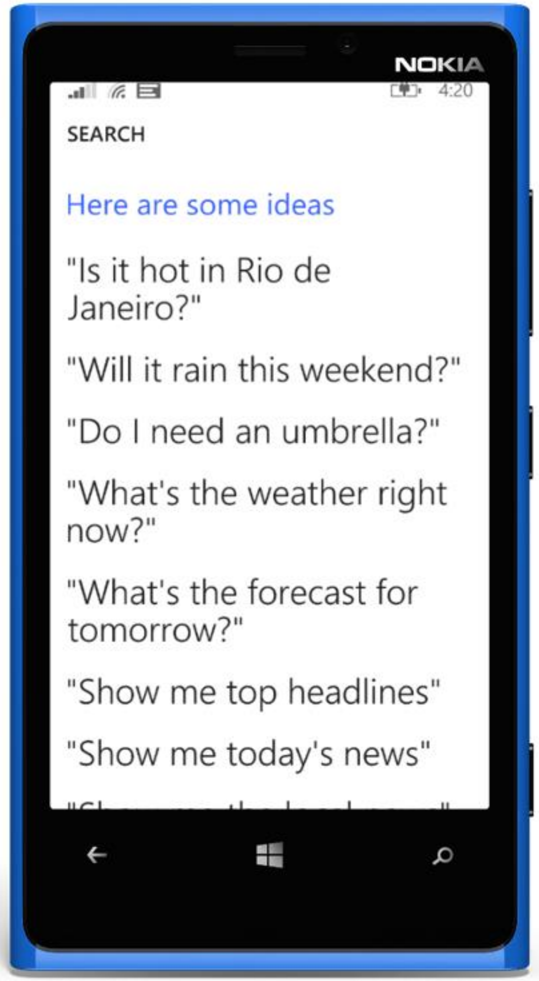
Places

Calendar

Information

Cortana: What Can You Do?

- Phone
- Messaging
- Calendar
- Reminders
- Notes
- Alarms
- Music
- Places
- Search



SEARCH

Here are some ideas

"Is it hot in Rio de Janeiro?"

"Will it rain this weekend?"

"Do I need an umbrella?"

"What's the weather right now?"

"What's the forecast for tomorrow?"

"Show me top headlines"

"Show me today's news"

PROJECT OXFORD

HTTP://WWW.PROJECTOXFORD.AI

MICROSOFT PROJECT OXFORD SERVICES

PROJECT OXFORD: http://www.projectoxford.ai

Compute APIs

Face APIs

Speech APIs

Emotion APIs

BETA

Understand your users with Emotion Recognition

Spell Check APIs

BETA

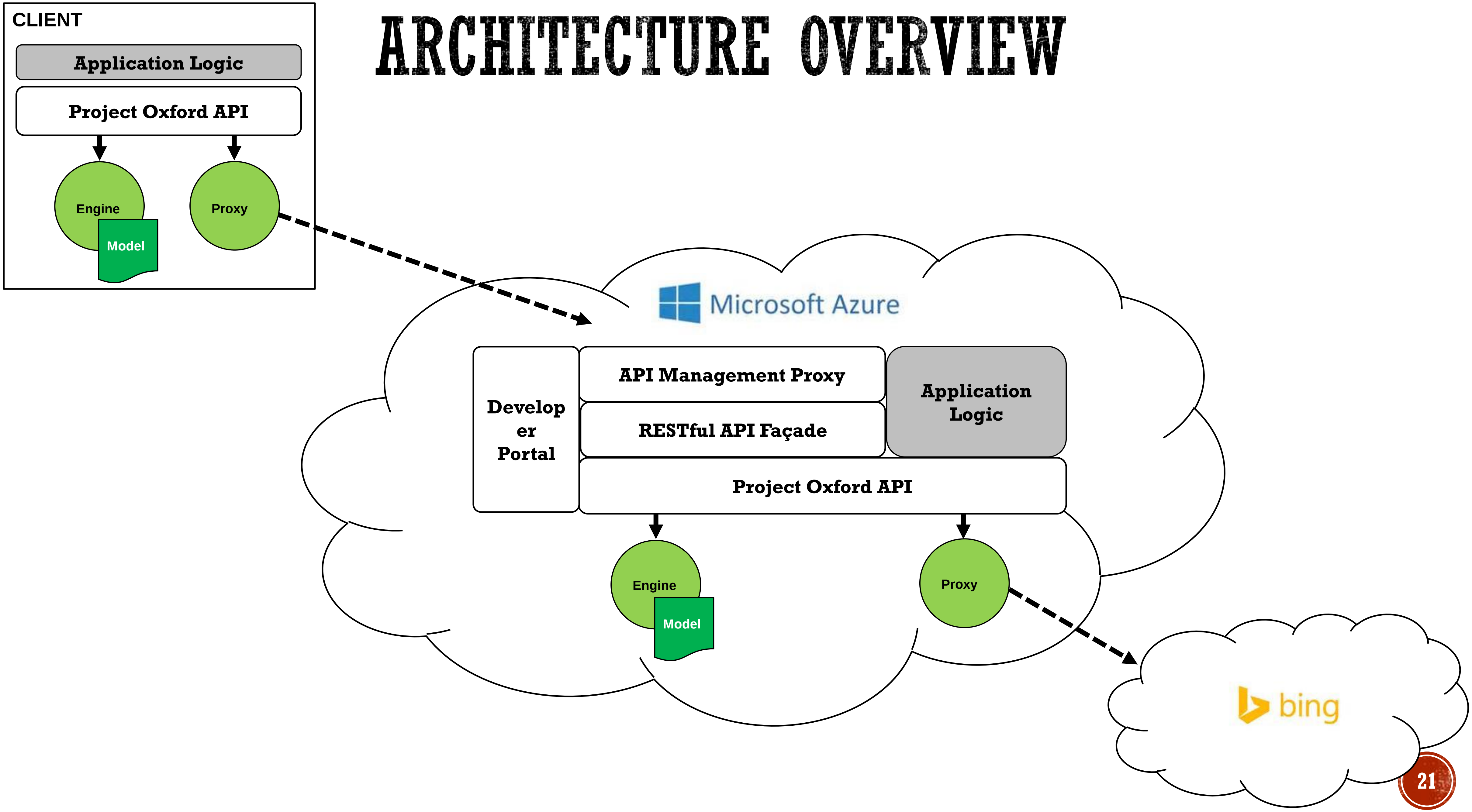
Detect and correct common and uncommon spelling errors, via the Bing document index

Language Understanding Intelligent Service (LUIS)

BETA

Understand natural language commands tailored to your application

ARCHITECTURE OVERVIEW



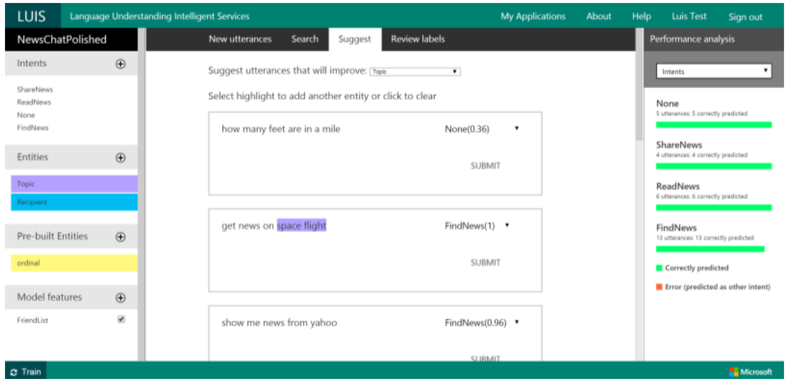
OXFORD'S LUIS - BEYOND CORTANA

http://www.luis.ai

```
{
  "entities": [
    {
      "entity": "flight delays",
      "type": "Topic"
    }
  ],
  "intents": [
    {
      "intent": "FindNews",
      "score": 0.9985384
    },
    {
      "intent": "None",
      "score": 0.07289317
    },
    {
      "intent": "ReadNews",
      "score": 0.0167122427
    },
    {
      "intent": "ShareNews",
      "score": 1.0919299E-06
    }
  ]
}
```

LUIS OVERVIEW

Train



Suggest utterances that will improve: Topic ▾

Select highlight to add another entity or click to clear

how many feet are in a mile — None(0.36) ▾ — SUBMIT

get news on space flight — FindNews(1) ▾ — SUBMIT

show me news from yahoo — FindNews(0.96) ▾ — SUBMIT

Intents ▾

None
5 utterances: 5 correctly predicted

ShareNews
4 utterances: 4 correctly predicted

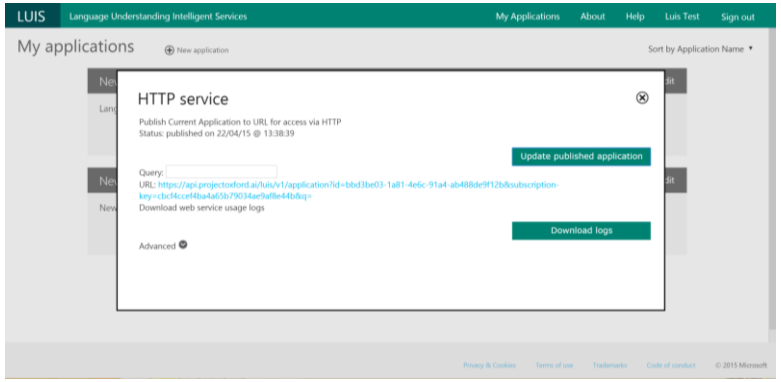
ReadNews
6 utterances: 6 correctly predicted

FindNews
13 utterances: 13 correctly predicted

🟩 Correctly predicted

🟥 Error (predicted as other intent)

Deploy



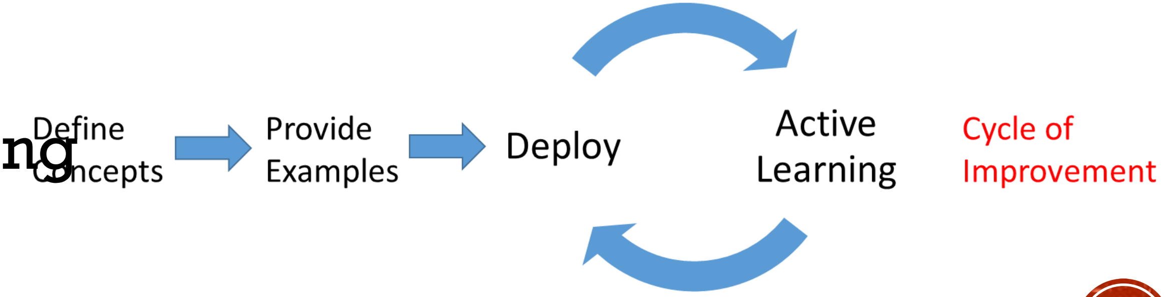
HTTP service

Publish Current Application to URL for access via HTTP
Status: published on 22/04/15 @ 13:38:39

Query:
URL: https://api.projectoxford.ai/luis/v1/application?id=bbd3be03-1a81-4e6c-91a4-ab488de9f12b&subscription-key=cbcf4ccef4ba4a65b79034ae9af8e44b&q=
Download web service usage logs

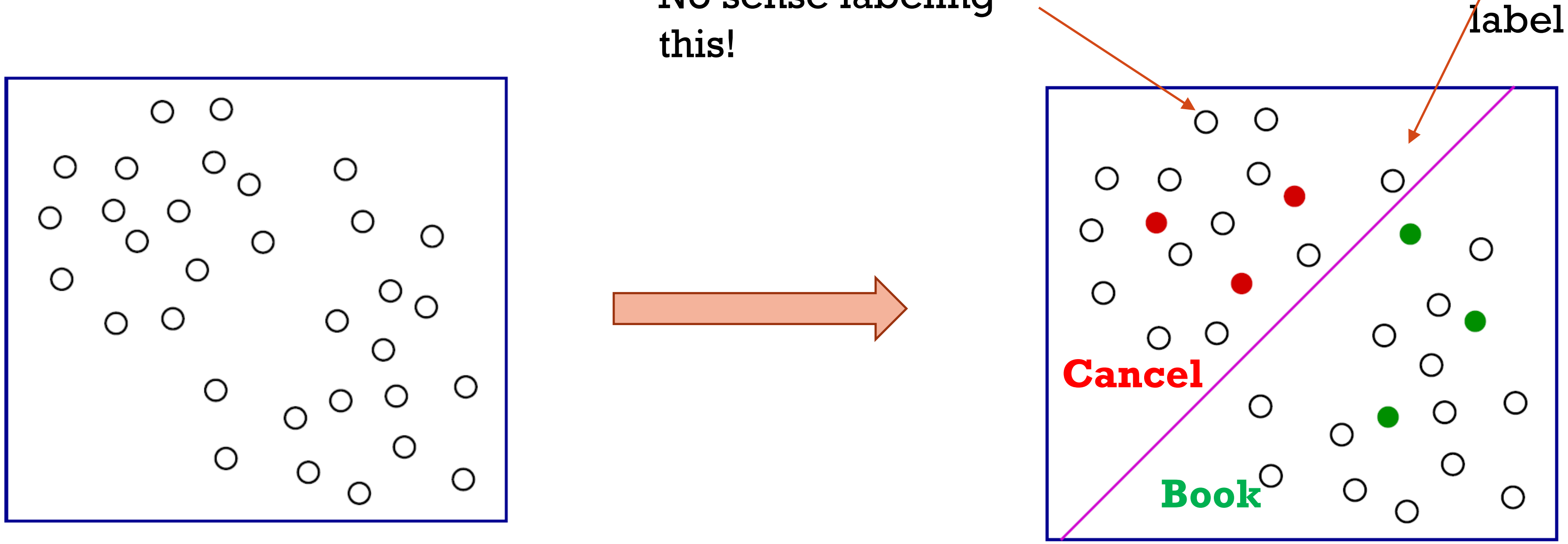
Access

- Learns models from a few examples
- Continuous refinement with active learning
- Existing Bing models for common cases



ACTIVE LEARNING

- Model parameters are estimated with labeled data
- Labeling data is expensive
- Want to infer decision boundaries quickly

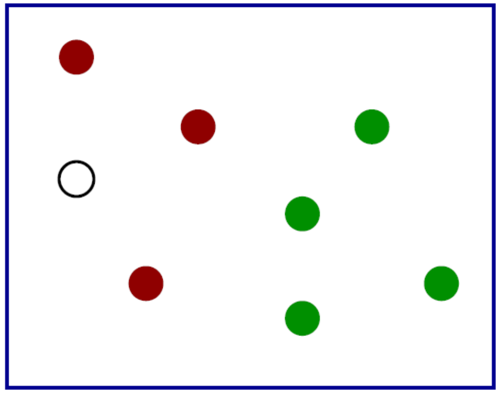


No sense labeling this!

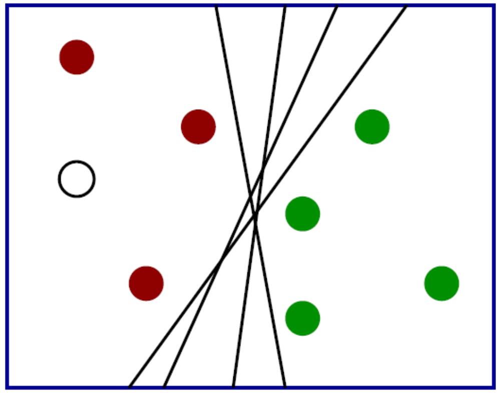
Useful to label

ACTIVE LEARNING (2)

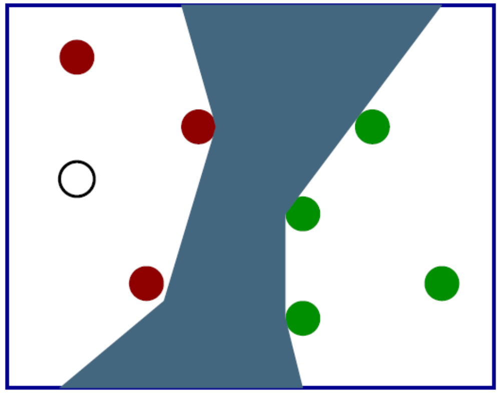
- Only label examples near the decision boundary



Is a label needed?



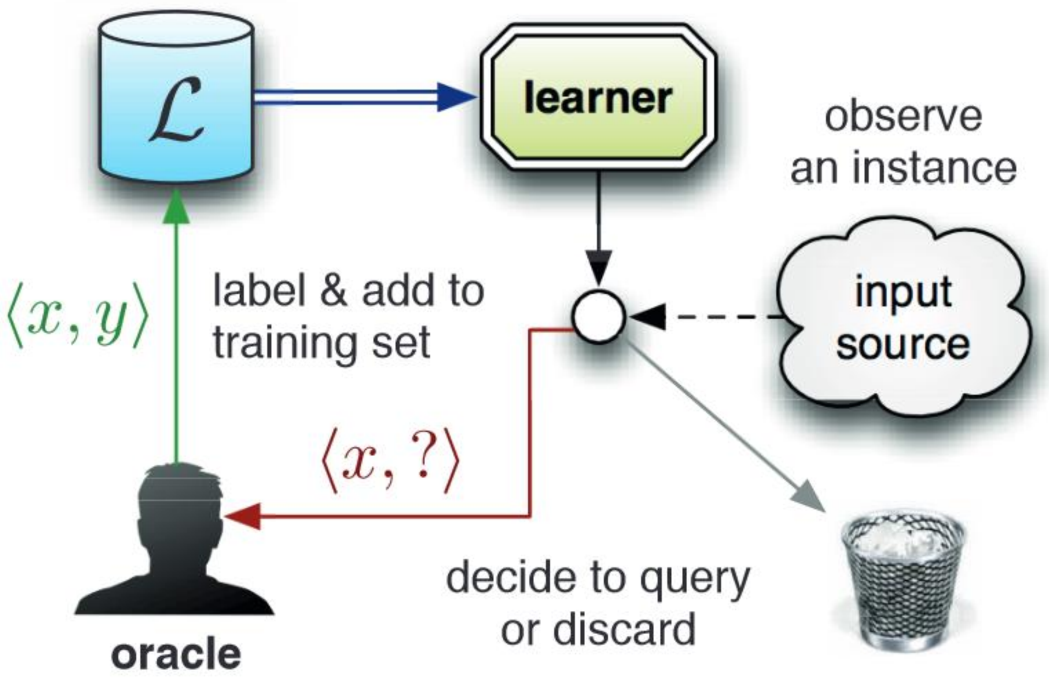
H_t = current candidate hypotheses

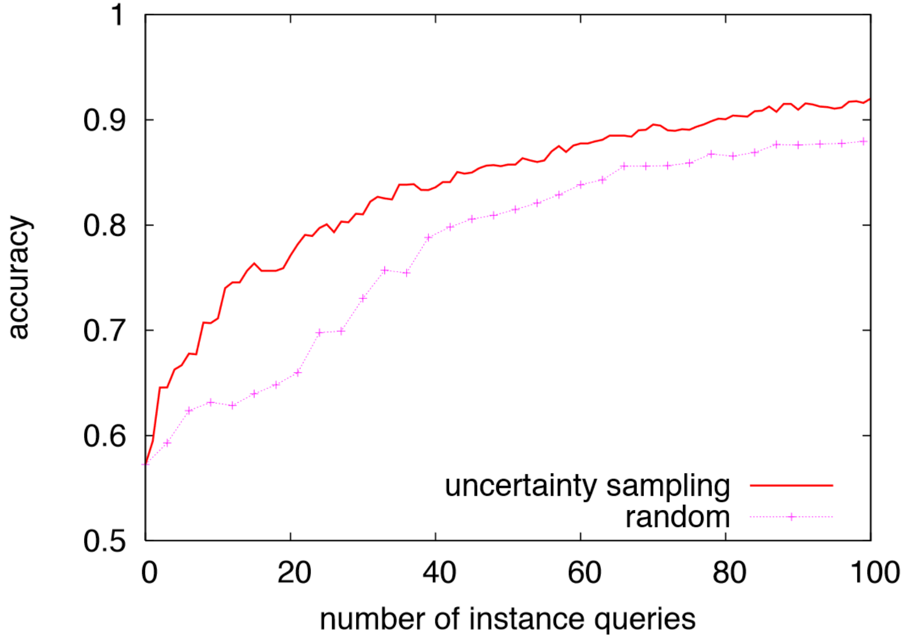


Region of uncertainty

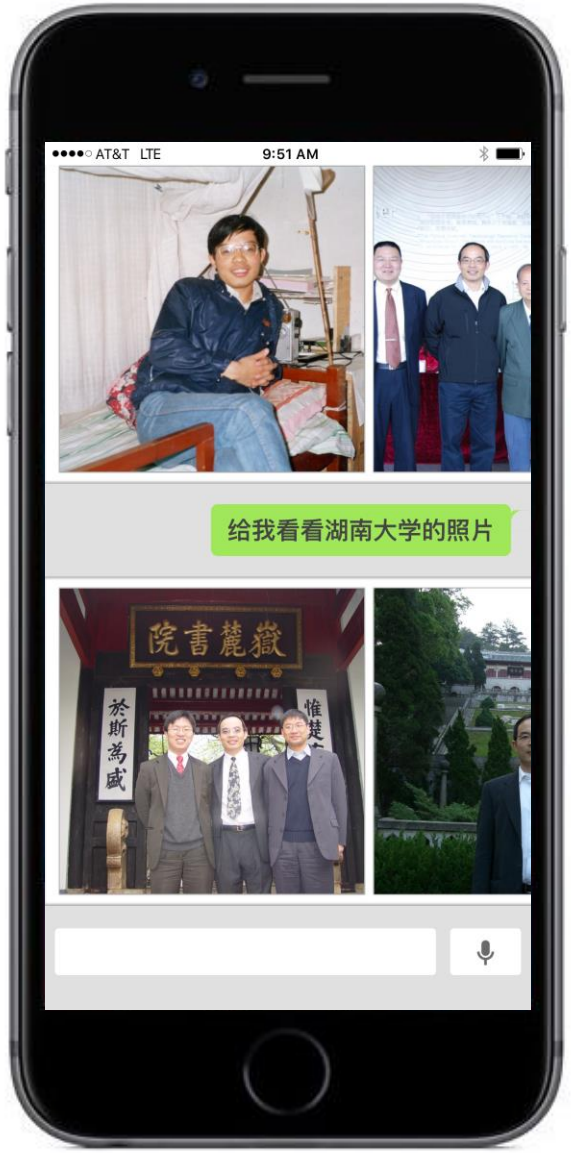
[Dasgupta & Langford 2009]

ACTIVE LEARNING == FAST LEARNING

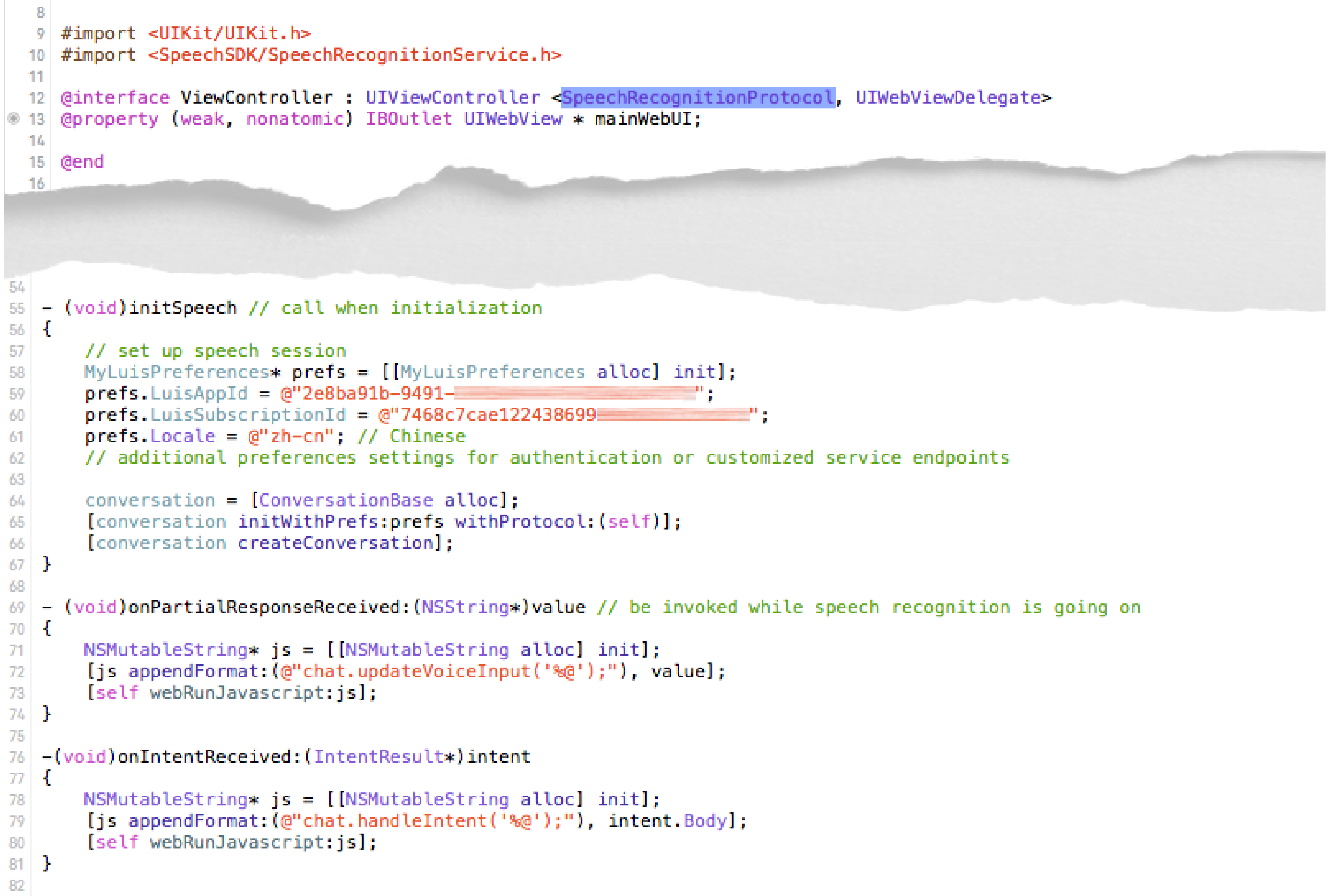




SPOKEN LANGUAGE UNDERSTANDING DEMO

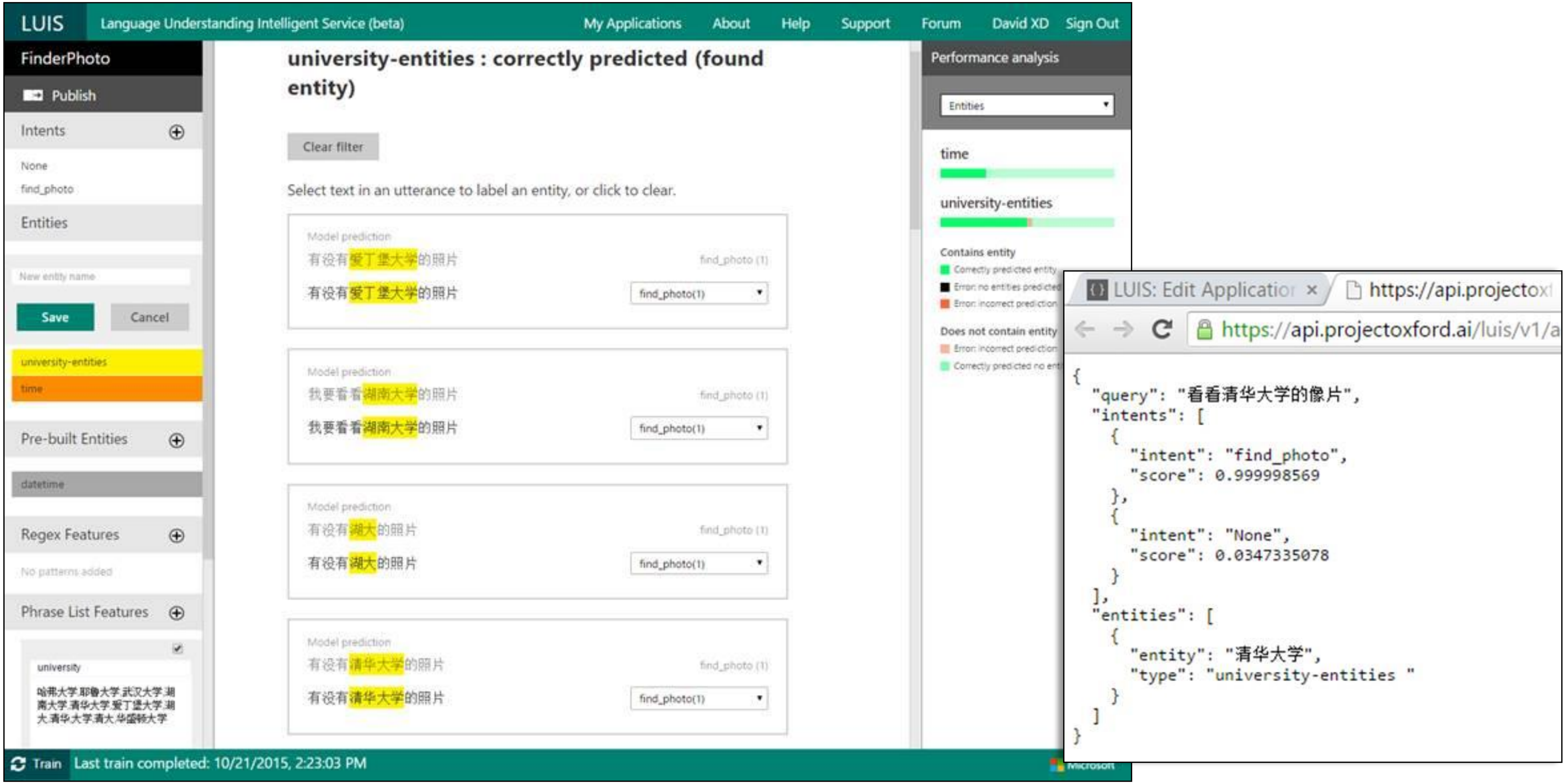


给我看看湖南大学的照片



```
#import <UIKit/UIKit.h>
#import <SpeechSDK/SpeechRecognitionService.h>

@interface ViewController : UIViewController <SpeechRecognitionProtocol, UIWebViewDelegate>
@property (weak, nonatomic) IBOutlet UIWebView * mainWebUI;

@end


- (void)initSpeech // call when initialization
{
    // set up speech session
    MyLuisPreferences* prefs = [[MyLuisPreferences alloc] init];
    prefs.LuisAppId = @"2e8ba91b-9491-[redacted]";
    prefs.LuisSubscriptionId = @"7468c7cae122438699[redacted]";
    prefs.Locale = @"zh-cn"; // Chinese
    // additional preferences settings for authentication or customized service endpoints

    conversation = [ConversationBase alloc];
    [conversation initWithPrefs:prefs withProtocol:(self)];
    [conversation createConversation];
}

- (void)onPartialResponseReceived:(NSString*)value // be invoked while speech recognition is going on
{
    NSMutableString* js = [[NSMutableString alloc] init];
    [js appendFormat:(@"chat.updateVoiceInput('%@');"), value];
    [self webRunJavascript:js];
}

-(void)onIntentReceived:(IntentResult*)intent
{
    NSMutableString* js = [[NSMutableString alloc] init];
    [js appendFormat:(@"chat.handleIntent('%@');"), intent.Body];
    [self webRunJavascript:js];
}
```

CREATING LUIS MODELS



哈佛大学,耶鲁大学,武汉大学,湖南大学,清华大学,爱丁堡大学,湖大,清华大学,清大,华盛顿大学

university-entities : correctly predicted (found entity)

Select text in an utterance to label an entity, or click to clear.

Model prediction
有没有爱丁堡大学的照片 find_photo (1)
有没有爱丁堡大学的照片 find_photo(1)

Model prediction
我要看看湖南大学的照片 find_photo (1)
我要看看湖南大学的照片 find_photo(1)

Model prediction
有没有湖大的照片 find_photo (1)
有没有湖大的照片 find_photo(1)

Model prediction
有没有清华大学的照片 find_photo (1)
有没有清华大学的照片 find_photo(1)

🟩 Correctly predicted entity
■ Error: no entities predicted
🟥 Error: incorrect prediction

Error: incorrect prediction
Correctly predicted no ent

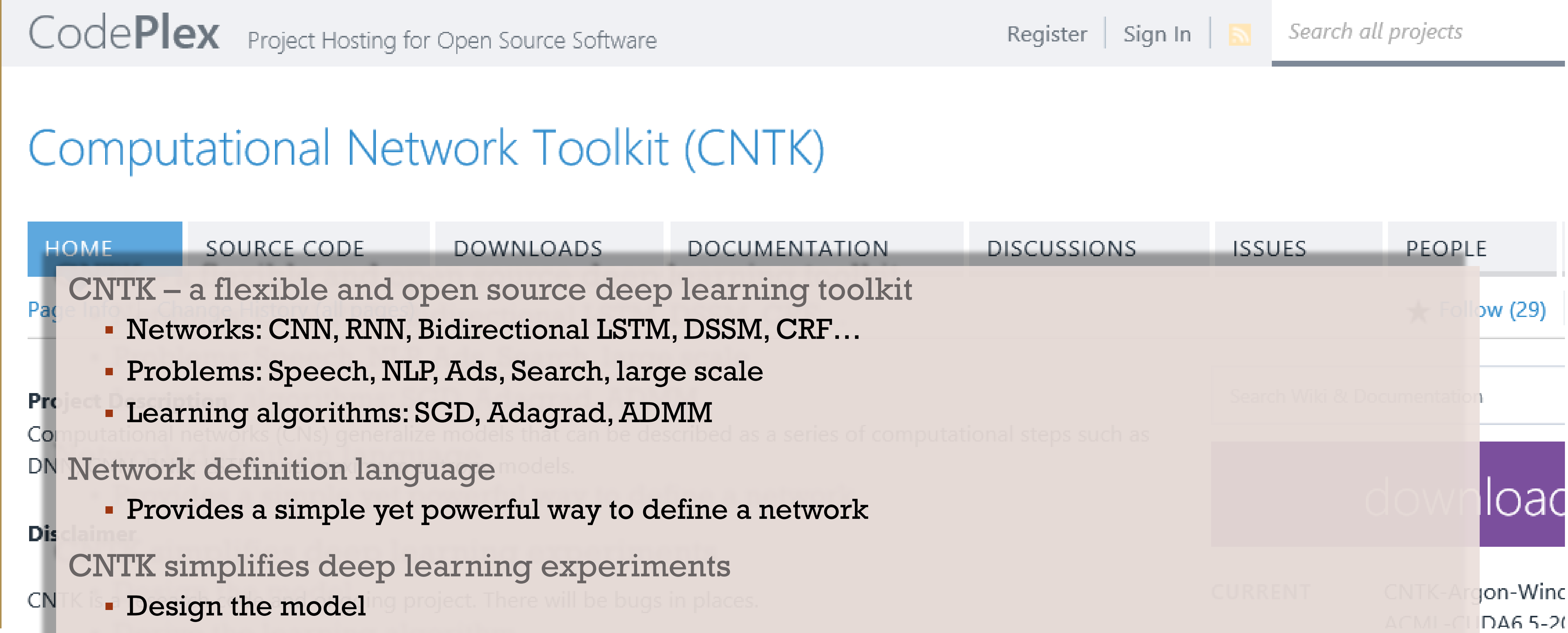
Last train completed: 10/21/2015, 2:23:03 PM

https://api.projectoxford.ai/luis/v1/a

```
{
  "query": "看看清华大学的像片",
  "intents": [
    {
      "intent": "find_photo",
      "score": 0.999998569
    },
    {
      "intent": "None",
      "score": 0.0347335078
    }
  ],
  "entities": [
    {
      "entity": "清华大学",
      "type": "university-entities "
    }
  ]
}
```

PROJECT PHILLY

Coming soon



CodePlex Project Hosting for Open Source Software

Register | Sign In

Search all projects

Computational Network Toolkit (CNTK)

HOME SOURCE CODE DOWNLOADS DOCUMENTATION DISCUSSIONS ISSUES PEOPLE

Page Info Change History (all pages)

Follow (29)

Project Description

Computational networks (CNs) generalize models that can be described as a series of computational steps such as DN … models.

Search Wiki & Documentation

download

Disclaimer

CNTK is a research code and ongoing project. There will be bugs in places.

CURRENT CNTK-Argon-Winc ACML-CUDA6.5-2

CNTK – a flexible and open source deep learning toolkit

- Networks: CNN, RNN, Bidirectional LSTM, DSSM, CRF…
- Problems: Speech, NLP, Ads, Search, large scale
- Learning algorithms: SGD, Adagrad, ADMM

Network definition language

- Provides a simple yet powerful way to define a network

CNTK simplifies deep learning experiments

- Design the model
- Derive the learning algorithm
- Implement the model
- Run the experiments

AZURE GPU CLUSTER LAB

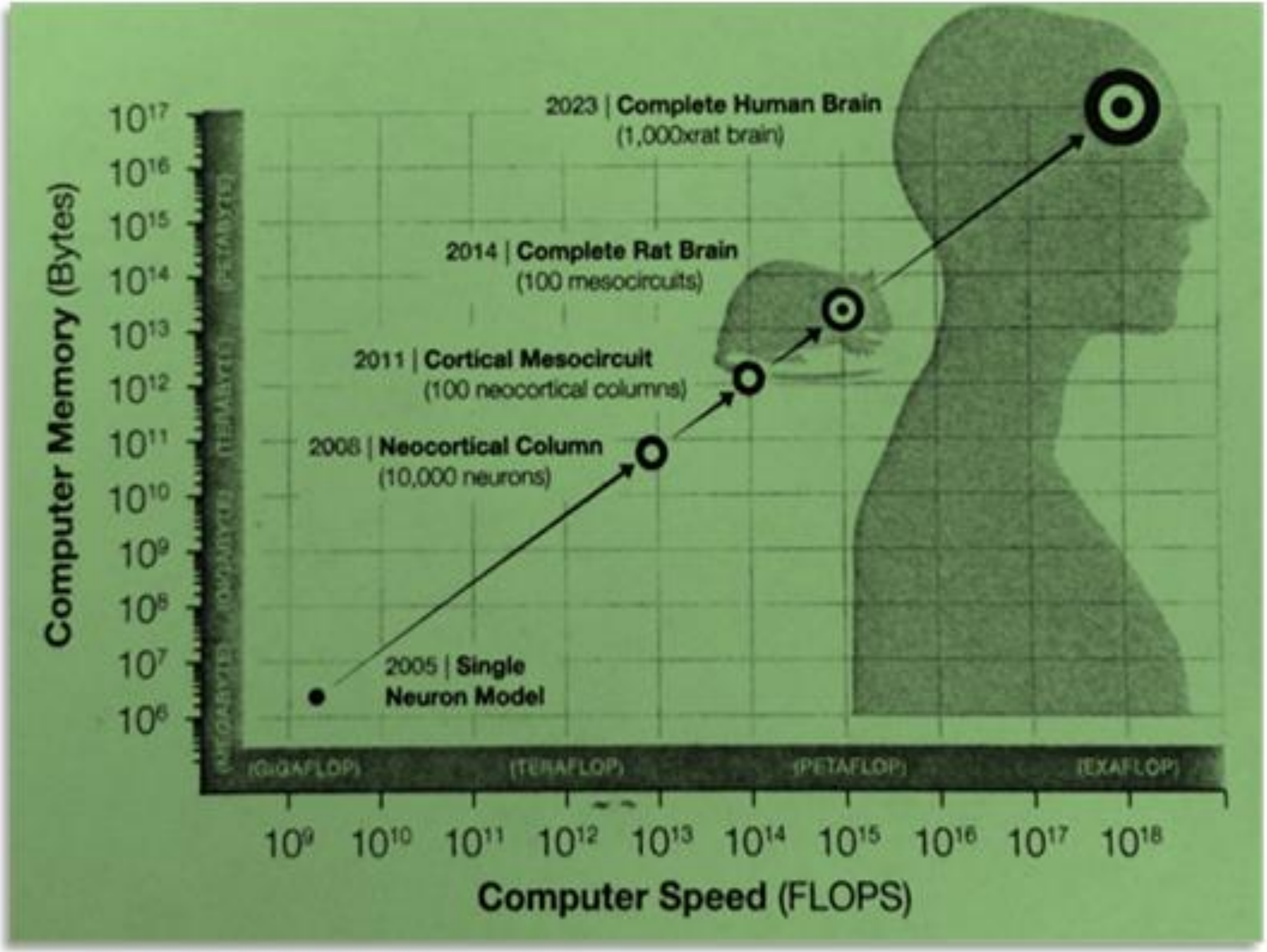
Improved speech and image recognition workloads

- Deep Learning dramatically improved both speech and image recognition
- Progress hampered by distributed computing infrastructure

AZURE GPU Labs

- Team up with researchers to scale out on deep learning
- Optimized for CNTK

WHAT IS NEXT?



2023 | Complete Human Brain
(1,000xrat brain)
2014 | Complete Rat Brain
(100 mesocircuits)
2011 | Cortical Mesocircuit
(100 neocortical columns)
2008 | Neocortical Column
(10,000 neurons)
2005 | Single
Neuron Model
Computer Memory (Bytes)
10^{17}
10^{16}
10^{15}
10^{14}
10^{13}
10^{12}
10^{11}
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
(PETABYTE)
(TERABYTE)
(GIGABYTE)
(MEGABYTE)
(GIGAFLOP)
(TERAFLOP)
(PETAFLOP)
(EXAFLOP)
10^{9}
10^{10}
10^{11}
10^{12}
10^{13}
10^{14}
10^{15}
10^{16}
10^{17}
10^{18}
Computer Speed (FLOPS)

An **invisible** revolution is coming

Q&A?

Email me any follow-up questions: xdh@microsoft.com