Consuming 3rd Party Services

To assist in your research and development work

Agenda

- On-demand "machine learning"
- Complementing our research and development
- Microsoft Cognitive Services
  - Signing Up
  - Image Analysis
  - Text Analytics with Python
  - Web Language Model with R

Introduction

- On-demand machine learning
  - Google’s Prediction API (2012)
  - Microsoft’s Cortana Intelligence Suite (2015)
  - Amazon’s Machine Learning (2015)

On-Demand Machine Learning

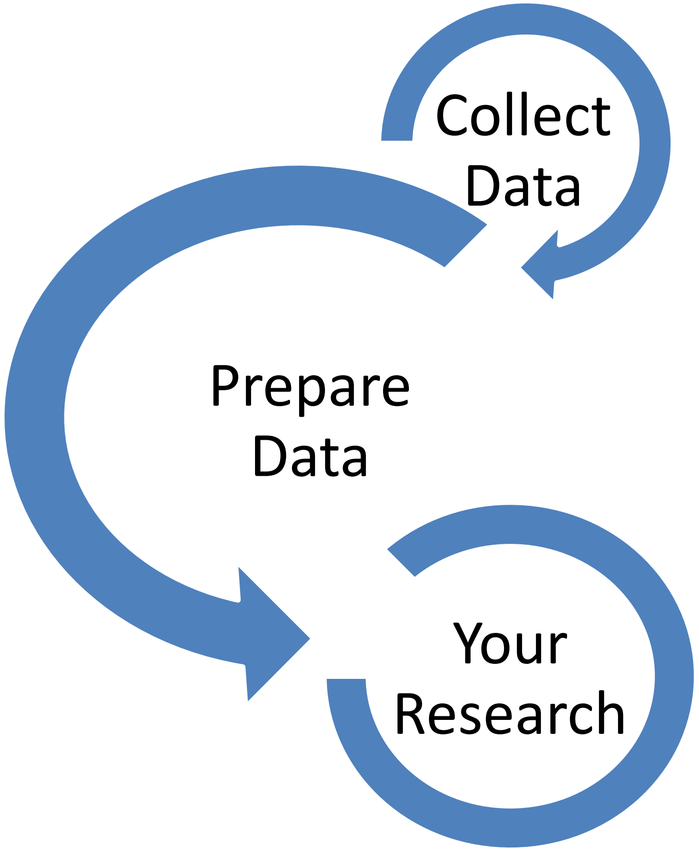
✓No need to know complex AI algorithms.

✓Some programming needed.

✓No installation of software, calls are generally via RESTful interface.

Complementing R&D

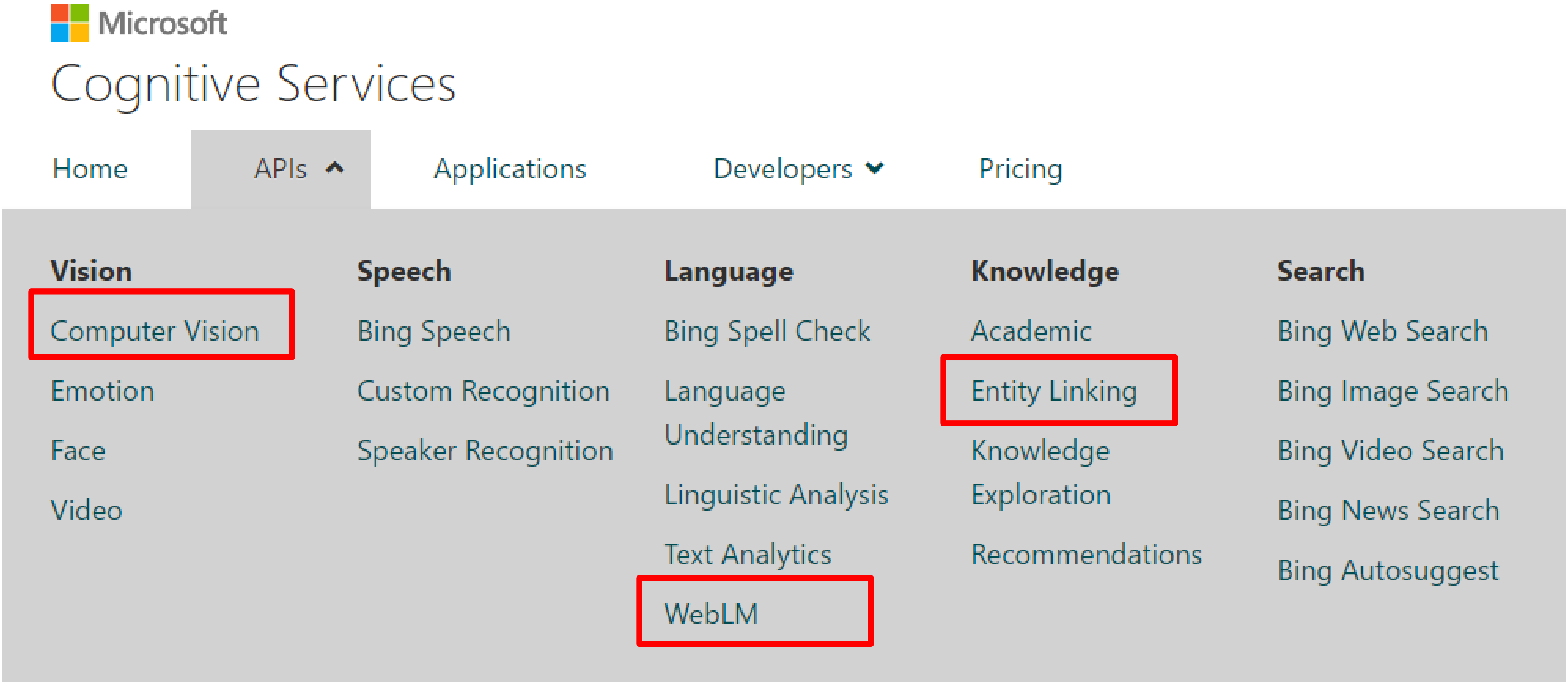
- A small part of a big picture!
- Typically, you spend a lot of time to prepare your data.



Preparing Data

- Can involve, for example
  - Filtering for language
  - Selecting specific images
  - Named entity or keyword extraction

Microsoft Cognitive Services



Microsoft

Cognitive Services

Home | APIs | Applications | Developers | Pricing

Vision

- Computer Vision
- Emotion
- Face
- Video

Speech

- Bing Speech
- Custom Recognition
- Speaker Recognition

Language

- Bing Spell Check
- Language Understanding
- Linguistic Analysis
- Text Analytics
- WebLM

Knowledge

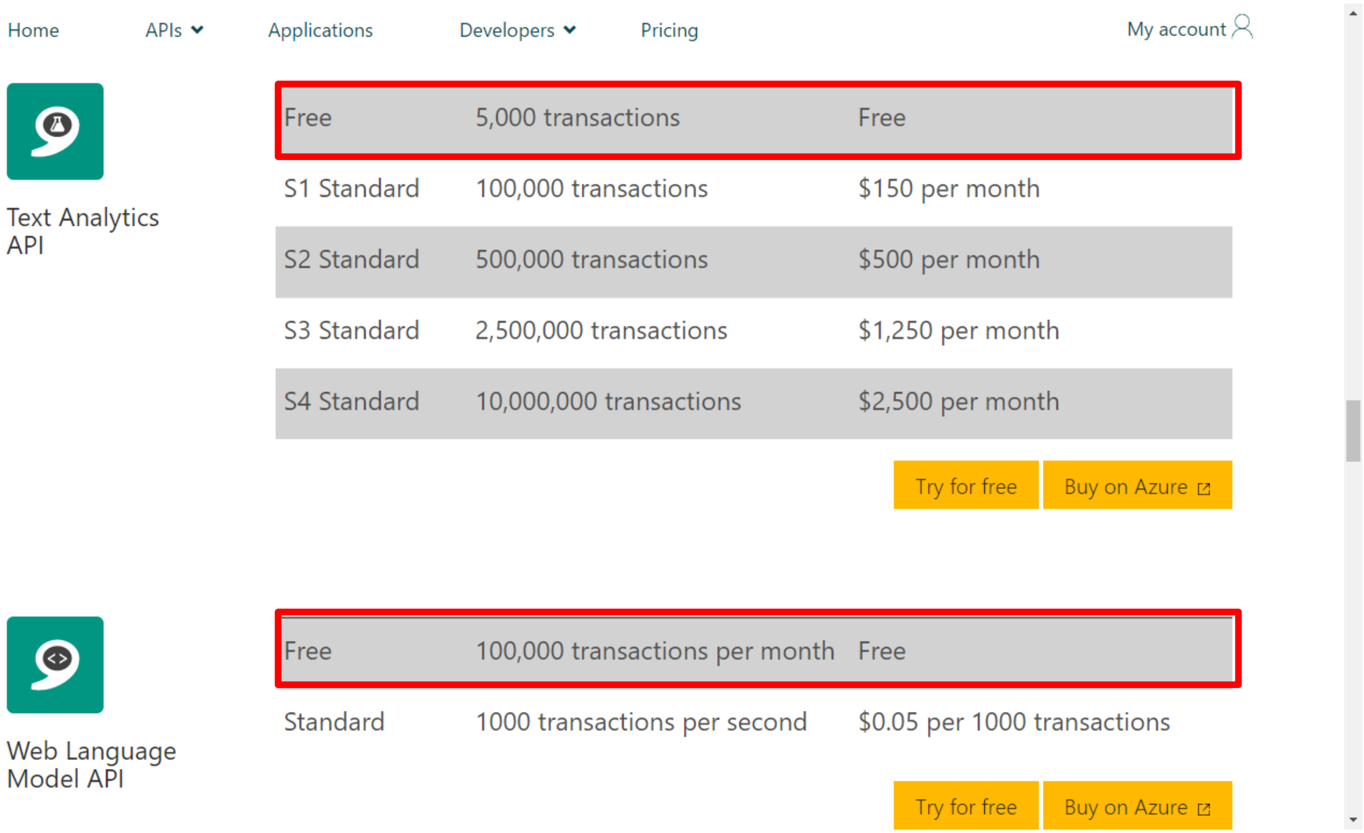
- Academic
- Entity Linking
- Knowledge Exploration
- Recommendations

Search

- Bing Web Search
- Bing Image Search
- Bing Video Search
- Bing News Search
- Bing Autosuggest

Microsoft Account

- @hotmail.com, @live.com, @outlook.com



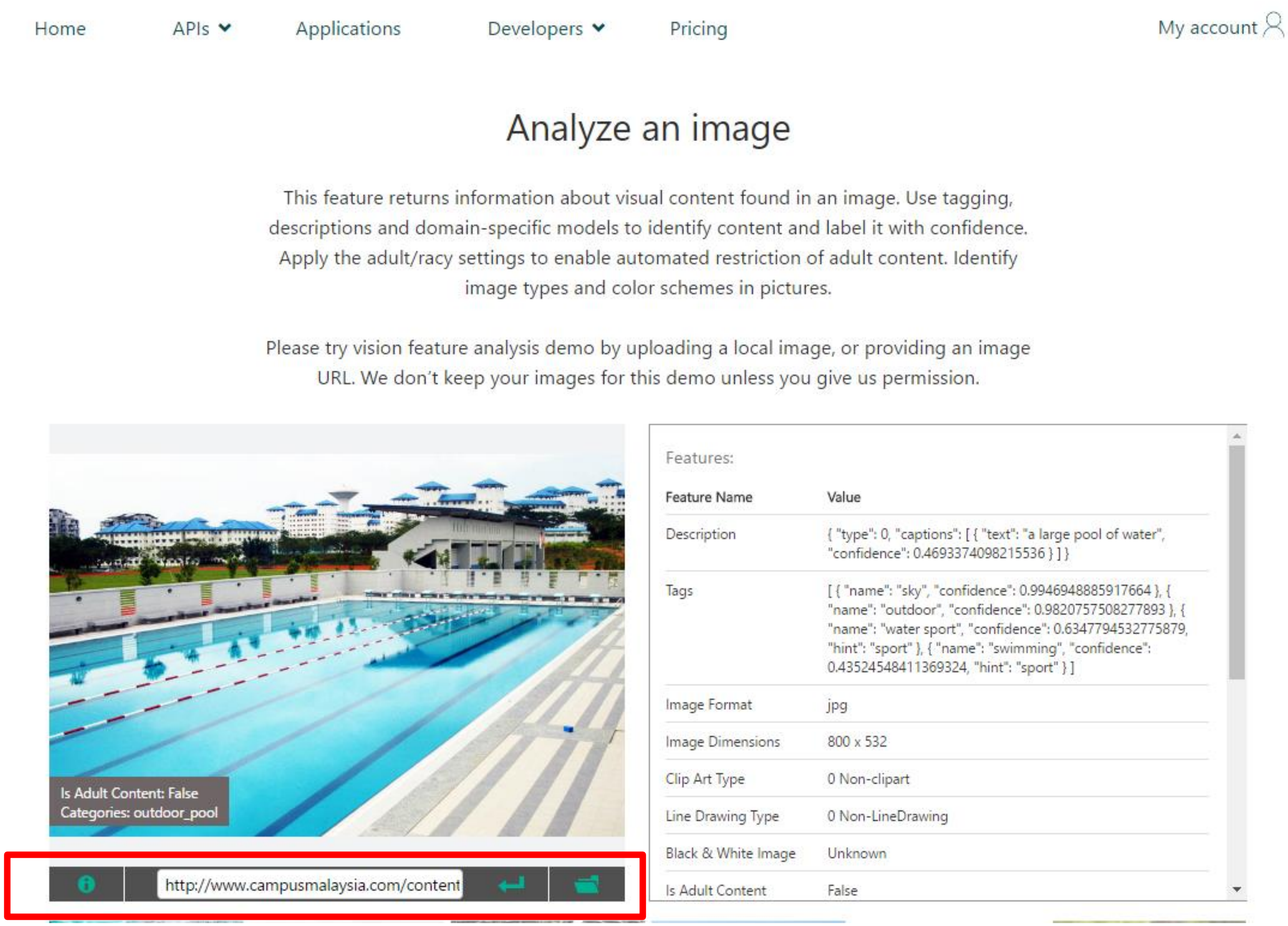
Text Analytics API

Free	5,000 transactions	Free
S1 Standard	100,000 transactions	$150 per month
S2 Standard	500,000 transactions	$500 per month
S3 Standard	2,500,000 transactions	$1,250 per month
S4 Standard	10,000,000 transactions	$2,500 per month

Web Language Model API

Free	100,000 transactions per month	Free
Standard	1000 transactions per second	$0.05 per 1000 transactions

Image Analysis



Analyze an image

This feature returns information about visual content found in an image. Use tagging, descriptions and domain-specific models to identify content and label it with confidence. Apply the adult/racy settings to enable automated restriction of adult content. Identify image types and color schemes in pictures.

Please try vision feature analysis demo by uploading a local image, or providing an image URL. We don't keep your images for this demo unless you give us permission.

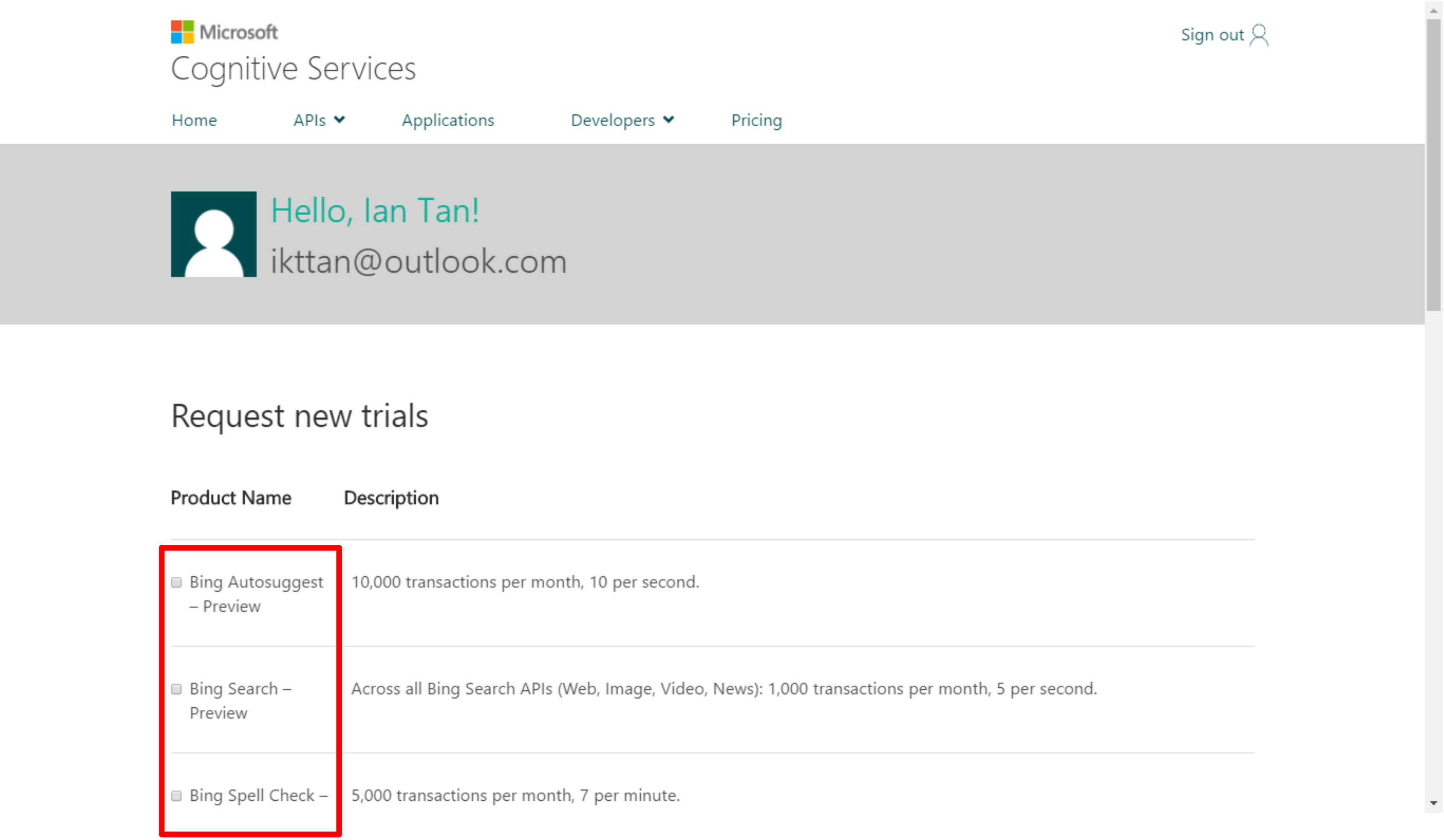
Is Adult Content: False
Categories: outdoor_pool

http://www.campusmalaysia.com/content

Features:

Feature Name	Value
Description	{ "type": 0, "captions": [{ "text": "a large pool of water", "confidence": 0.4693374098215536 }] }
Tags	[{ "name": "sky", "confidence": 0.9946948885917664 }, { "name": "outdoor", "confidence": 0.9820757508277893 }, { "name": "water sport", "confidence": 0.6347794532775879, "hint": "sport" }, { "name": "swimming", "confidence": 0.43524548411369324, "hint": "sport" }]
Image Format	jpg
Image Dimensions	800 x 532
Clip Art Type	0 Non-clipart
Line Drawing Type	0 Non-LineDrawing
Black & White Image	Unknown
Is Adult Content	False

Get Key



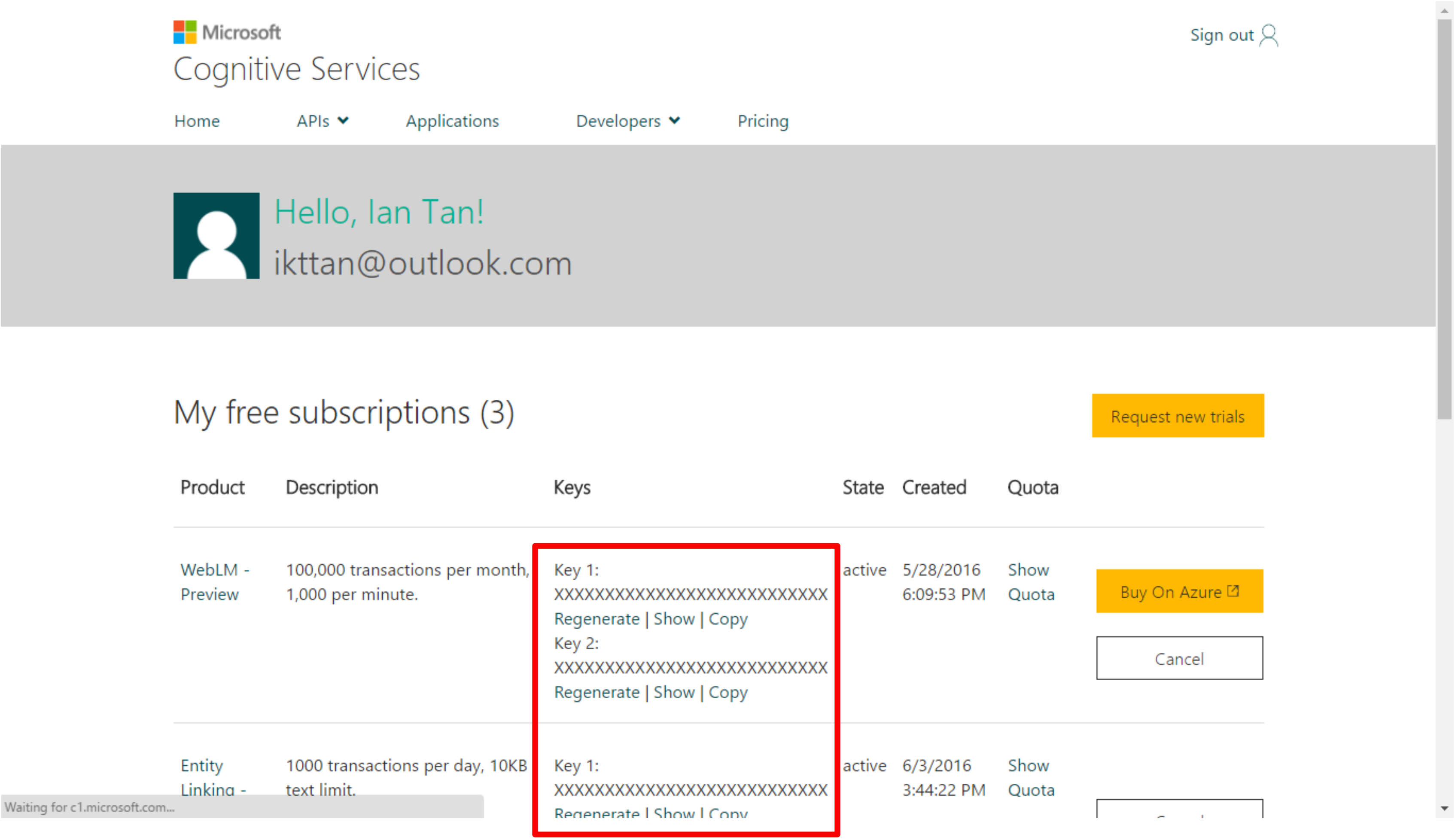
Hello, Ian Tan!

ikttan@outlook.com

Request new trials

Product Name	Description
□ Bing Autosuggest – Preview	10,000 transactions per month, 10 per second.
□ Bing Search – Preview	Across all Bing Search APIs (Web, Image, Video, News): 1,000 transactions per month, 5 per second.
□ Bing Spell Check –	5,000 transactions per month, 7 per minute.

Keys

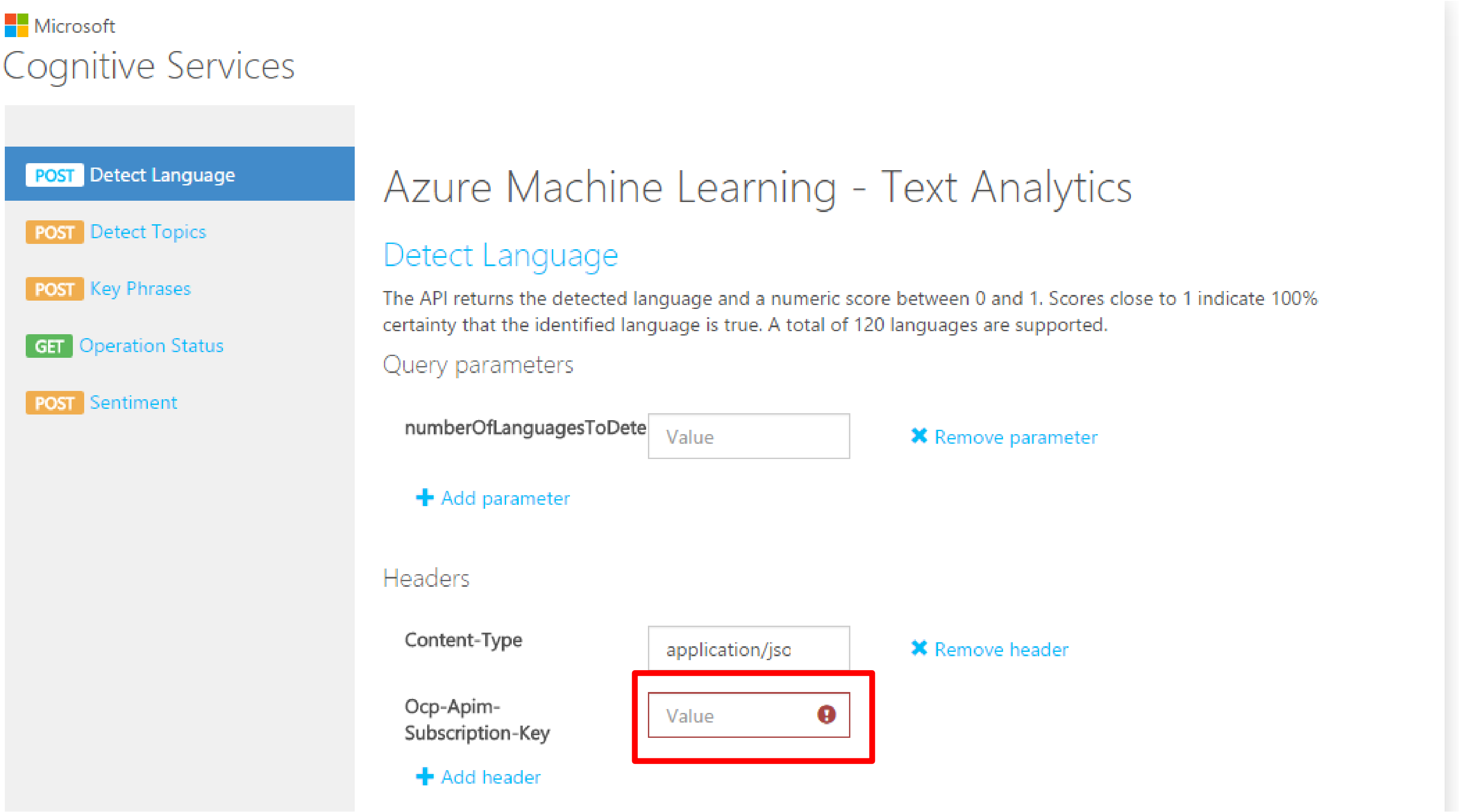


Hello, Ian Tan!
ikttan@outlook.com

My free subscriptions (3)

Product	Description	Keys	State	Created	Quota	
WebLM - Preview	100,000 transactions per month, 1,000 per minute.	Key 1: XXXXXXXXXXXXXXXXXXXXXXXXXXXX Regenerate | Show | Copy Key 2: XXXXXXXXXXXXXXXXXXXXXXXXXXXX Regenerate | Show | Copy	active	5/28/2016 6:09:53 PM	Show Quota	Buy On Azure Cancel
Entity Linking -	1000 transactions per day, 10KB text limit.	Key 1: XXXXXXXXXXXXXXXXXXXXXXXXXXXX Regenerate | Show | Copy	active	6/3/2016 3:44:22 PM	Show Quota	

Language Filtering



Microsoft
Cognitive Services

Azure Machine Learning - Text Analytics

Detect Language

The API returns the detected language and a numeric score between 0 and 1. Scores close to 1 indicate 100% certainty that the identified language is true. A total of 120 languages are supported.

Query parameters

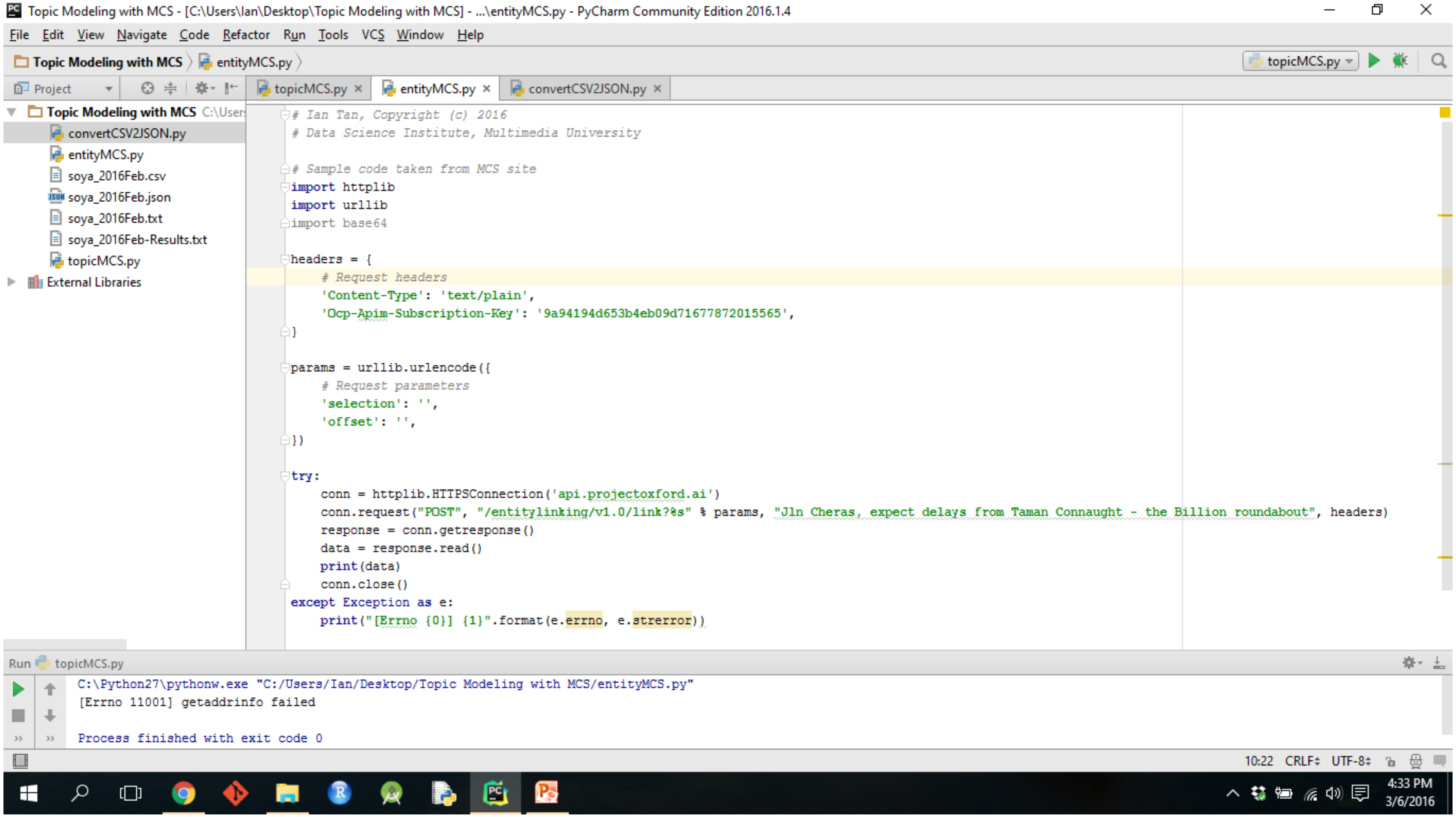
numberOfLanguagesToDete: Value

Headers

Content-Type: application/jsc

Ocp-Apim-Subscription-Key: Value

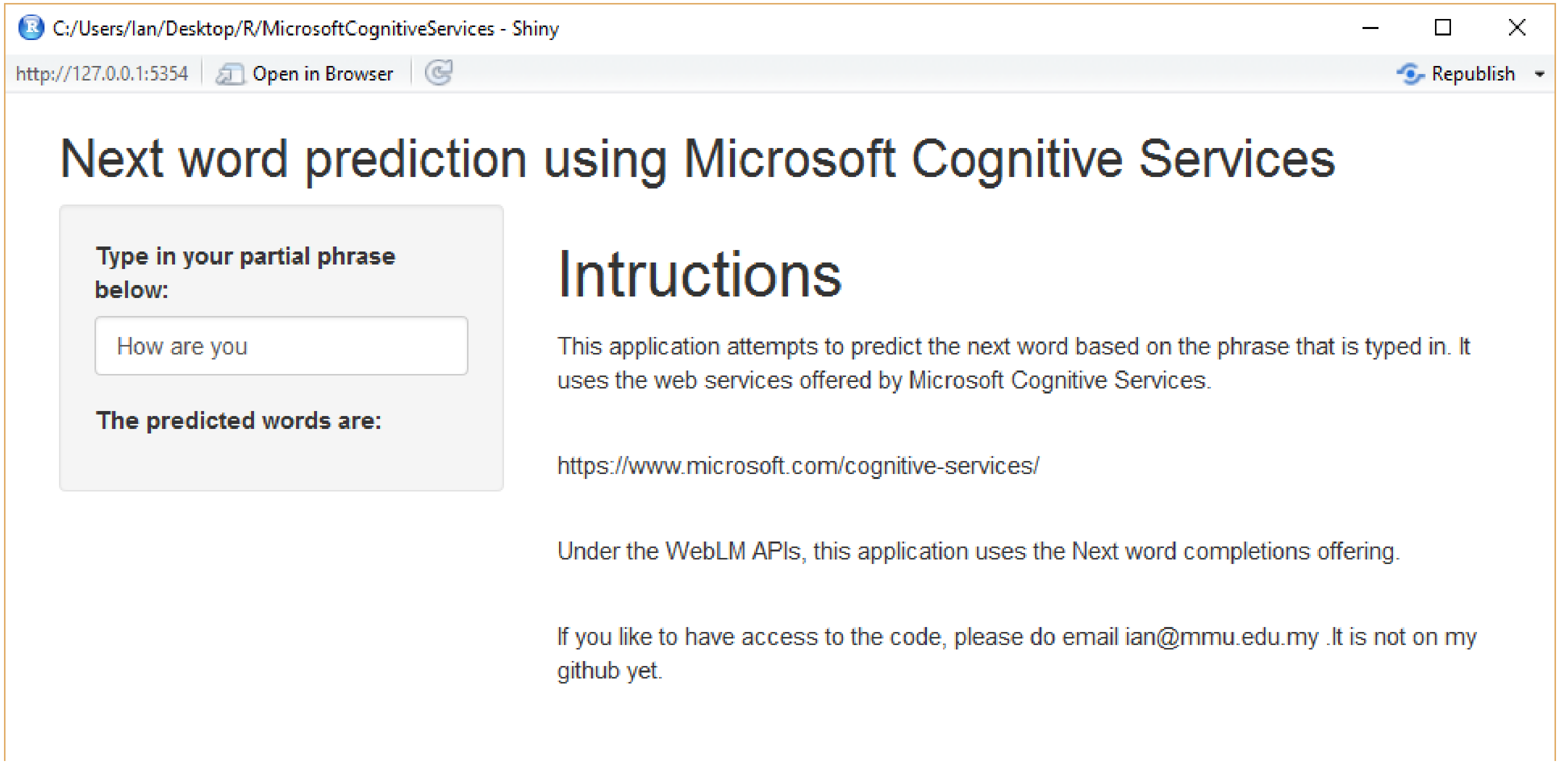
Entity Linkage (Python)



```python
# Ian Tan, Copyright (c) 2016
# Data Science Institute, Multimedia University

# Sample code taken from MCS site
import httplib
import urllib
import base64

headers = {
    # Request headers
    'Content-Type': 'text/plain',
    'Ocp-Apim-Subscription-Key': '9a94194d653b4eb09d71677872015565',
}

params = urllib.urlencode({
    # Request parameters
    'selection': '',
    'offset': '',
})

try:
    conn = httplib.HTTPSConnection('api.projectoxford.ai')
    conn.request("POST", "/entitylinking/v1.0/link?%s" % params, "Jln Cheras, expect delays from Taman Connaught - the Billion roundabout", headers)
    response = conn.getresponse()
    data = response.read()
    print(data)
    conn.close()
except Exception as e:
    print("[Errno {0}] {1}".format(e.errno, e.strerror))
```

```
C:\Python27\pythonw.exe "C:/Users/Ian/Desktop/Topic Modeling with MCS/entityMCS.py"
[Errno 11001] getaddrinfo failed

Process finished with exit code 0
```

Next Word Prediction (R)



Next word prediction using Microsoft Cognitive Services

Type in your partial phrase below:

How are you

The predicted words are:

Intructions

This application attempts to predict the next word based on the phrase that is typed in. It uses the web services offered by Microsoft Cognitive Services.

https://www.microsoft.com/cognitive-services/

Under the WebLM APIs, this application uses the Next word completions offering.

If you like to have access to the code, please do email ian@mmu.edu.my .It is not on my github yet.

Next Word Prediction (R)

```
# Ian Tan, Copyright (c) 2016
# Data Science Institute, Multimedia University

library(httr)
library(jsonlite)
shinyServer(function(input, output) {
  output$predictedText <- renderText({
     # input$entry (or ngrams) should be preprocessed first actually.
    ngrams <- input$entry
    url <-
paste0('https://api.projectoxford.ai/text/weblm/v1.0/generateNextWords?model=body&words=',ngrams,'&order=5&maxNumOfCandidatesReturned=5')
    url <- URLencode(url)
    ans <- POST(url, body = NULL, add_headers("Ocp-Apim-Subscription-Key"="f8f16618074948828176ef9063ba06a0"))

    json_ans <- toJSON(content(ans, "parsed", "application/json"))
    next_words <- unlist(fromJSON(json_ans)[[1]][[1]])
  })
})
```

Concluding Remarks

- We can make use of tools to assist us in our research work. We don’t have to reinvent the wheel.
- We can use these as a reference for comparison if we are working on the same area.
  - These tools are rather English centric and hence they have limitations in application to our context usually.