



Understanding **Microsoft Cognitive Services**

Use cloud-based intelligence APIs to build smart apps that see, hear, speak, and understand – in any language and on any platform

October 2016

Microsoft Cognitive Services is a set of cloud-based intelligence APIs for building richer, smarter, and more sophisticated applications. The product of years of ongoing research from data scientists at Microsoft and research institutions worldwide, these APIs offer a range of capabilities to app developers, from generating captions and search metadata for photos and recognizing faces and emotions in videos to performing sentiment analysis on social media posts and authenticating speakers via voice verification.

“Once we started using the APIs, we realized the quality was exceptional. From a developer standpoint, we got the APIs working very fast, they did what they were supposed to do, and the results were outstanding.” – *Jaan Apajalahti, CEO, Blucup*

Cognitive Services APIs can be invoked using REST calls in virtually any programming language and on any device. In addition, free SDKs targeting popular languages such as C# and Java and platforms such as .NET and Android are available for many of the APIs, simplifying the code required to invoke them and insulating the developer from the details of HTTP(S), REST, and JSON.

Developer documentation for Microsoft Cognitive Services can be found at <https://www.microsoft.com/cognitive-services/en-us/documentation>. Each method of each API is documented in detail and comes with sample code in several languages, including C#, Java, JavaScript, Objective-C, PHP, Ruby, and Python. For an example, see <http://bit.ly/2aQSf94>. Each method, of “feature,” is also accompanied by an Open API Testing Console enabling developers to call the method for discovery and experimentation purposes without writing a test harness themselves. For an example of one such console, see <http://bit.ly/2b60Ohf>.

Cognitive Services APIs

Microsoft Cognitive Services presently includes 21 API sets grouped into five categories:

- **Vision APIs** for extracting information from images, recognizing emotions in images and videos, detecting and identifying faces in images, detecting motion in videos, and more
- **Speech APIs** for converting text to speech and text to speech as well as performing voice verification and identification
- **Language APIs** for spell checking, natural-language understanding, linguistic analysis, sentiment analysis, topic detection, and more
- **Knowledge APIs** for natural-language query processing, entity linking, knowledge exploration, and personalized shopping recommendations
- **Search APIs** for offering search suggestions and performing sophisticated searches

These APIs are constantly being refined to offer the most comprehensive cognitive-services platform on the planet. The sections that follow provide an overview of each of the five API categories, complete with sample code demonstrating how selected APIs are called.

Vision APIs

The Microsoft Cognitive Services Vision APIs include the following APIs:

API	Description
Computer Vision API	Generate descriptions and search metadata for images, detect faces and recognize celebrities in images, read text in images, check images for adult content, generate smart image thumbnails, and more
Emotion API	Recognize anger, contempt, disgust, fear, happiness, sadness, and surprise in faces in images and video
Face API	Detect faces in images, determine age, gender, and other attributes, determine the likelihood that two photos depict the same person, search and identify faces, detect facial similarity, and group faces by similarity
Video API	Stabilize shaky video, detect and track faces in video, detect motion in video, and generate motion thumbnails

The Computer Vision API consists of five methods (one of which acts as two methods since it can be called with a GET or a POST) designed to extract information from images. Images can be provided to the API by uploading them in the request body or passing an image URL. For example, the **ocr** method takes an image as input and returns text extracted from the image by performing optical character recognition (OCR). Figure 1 shows an image input to the method and the text returned. Additional information about that text, such as the position in the image of each word, is available in the JSON return data.

“I have not seen anything comparable. We can use the Computer Vision API to prove to our clients the reliability of the data, so they can be confident making important business decisions based on that information.” – *Leendert de Voogd, CEO, Vigiglobe*

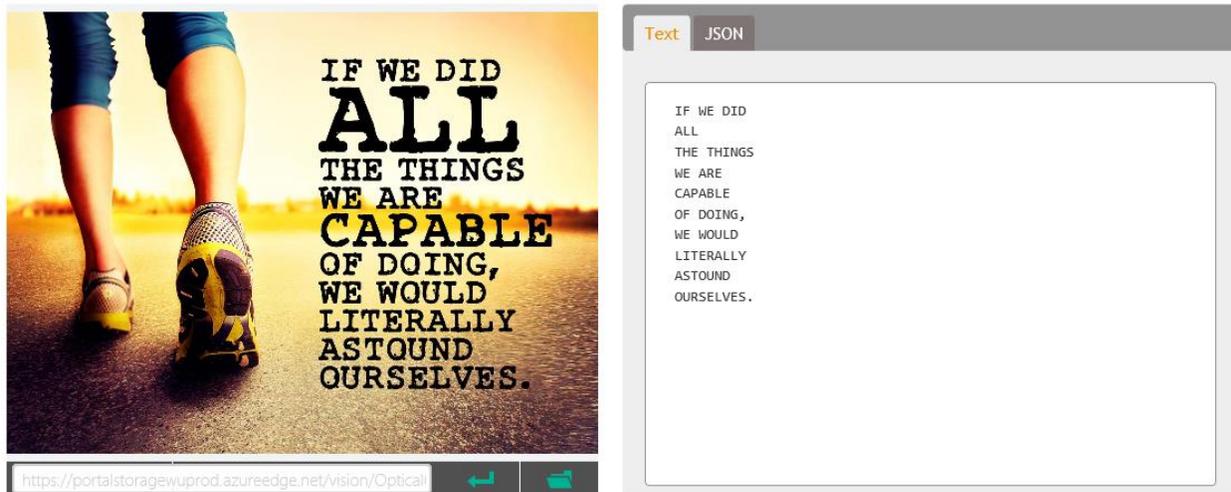


Figure 1: Text returned by passing an image to the Computer Vision API's `ocr` method

The `ocr` method recognizes more than 20 different languages, including English, French, German, Russian, Spanish, and Chinese (traditional and simplified). The following example demonstrates how the method is called from Python, with the image input via URL:

```
import http.client, urllib.request, urllib.parse, urllib.error, base64

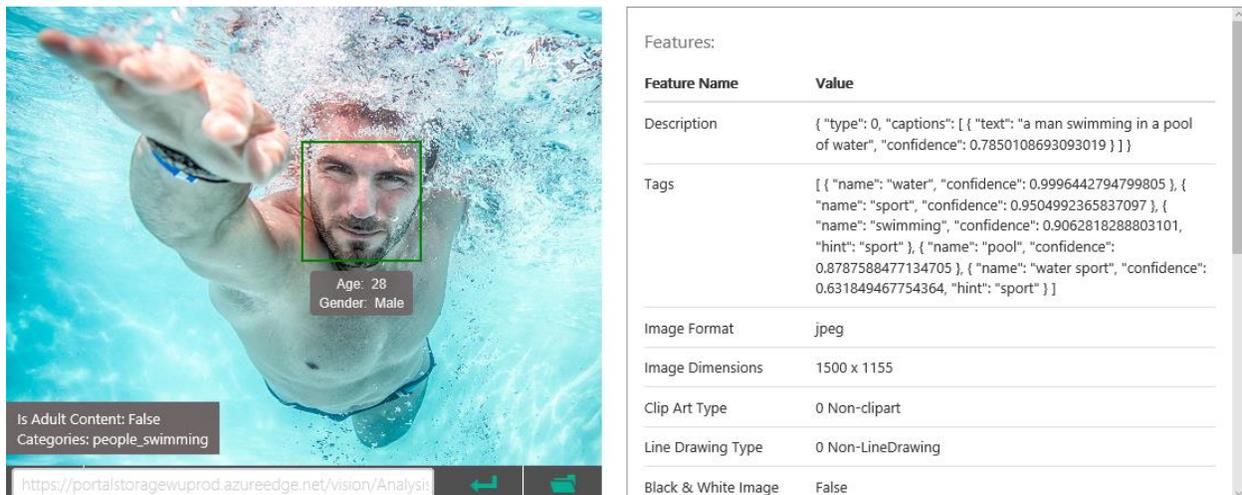
headers = {
    'Content-Type': 'application/json',
    'Ocp-Apim-Subscription-Key': '{key}',
}
body = '{"url': 'http://contoso.com/images/banner.png'}"

try:
    conn = http.client.HTTPSConnection('api.projectoxford.ai')
    conn.request("POST", "/vision/v1.0/ocr", body, headers)
    response = conn.getresponse()
    data = response.read()
    print(data)
    conn.close()
except Exception as e:
    print("[Errno {0}] {1}".format(e.errno, e.strerror))
```

The string “{key}” must be replaced with a subscription key obtained for the Computer Vision API from the Cognitive Services Web site. Subscription keys are free, but generally cap the number of uses. At the time of this writing, a free subscription key for the Computer Vision API allows up to 5,000 transactions per month and up to 20 calls per minute. Additional usage can be purchased via an Azure subscription.

Although not shown in the example above, the **ocr** method also accepts query-string parameters specifying what language the text is in (in the absence of this parameter, the API auto-detects the language) and whether the return data should include information about the angle of orientation of the words it finds in the image.

The Computer Vision API is capable of much more than extracting text from images. For example, the **analyze** method accepts an image as input and returns a variety of information about the image, from dominant colors to sentences describing the image and keywords that can be used as search metadata. It even returns scores from 0.0 to 1.0 quantifying the amount of adult content in the image and the “raciness” of the content. Figure 2 shows a sample image and a portion of the information describing it returned by the **analyze** method.



Feature Name	Value
Description	{ "type": 0, "captions": [{ "text": "a man swimming in a pool of water", "confidence": 0.7850108693093019 }] }
Tags	[{ "name": "water", "confidence": 0.9996442794799805 }, { "name": "sport", "confidence": 0.9504992365837097 }, { "name": "swimming", "confidence": 0.9062818288803101, "hint": "sport" }, { "name": "pool", "confidence": 0.8787588477134705 }, { "name": "water sport", "confidence": 0.631849467754364, "hint": "sport" }]
Image Format	jpeg
Image Dimensions	1500 x 1155
Clip Art Type	0 Non-clipart
Line Drawing Type	0 Non-LineDrawing
Black & White Image	False

Figure 2: Metadata returned by passing an image to the **analyze** method

Like other Cognitive Services methods, **analyze** can be called using REST protocols. But free libraries are available for iOS, Android, and .NET, enabling Swift, Java, and C# programmers to call the API more with greater ease and less code. Here is a call to the **analyze** method written in C# that utilizes the Microsoft.ProjectOxford.Vision package available for free from NuGet to retrieve a computer-generated image caption:

```

VisionServiceClient vision = new VisionServiceClient("{key}");
var features = new VisualFeature[] { VisualFeature.Description };
var result = await vision.AnalyzeImageAsync
    (http://contoso.com/images/banner.png, features);
string caption = result.Description.Captions[0].Text;

```

The parameter passed to **AnalyzeImageAsync** is the image URL. The second parameter, which is optional, narrows the scope of the call to limit the amount of information returned.

The **analyze** method can detect faces in images and even estimate those faces' age and gender. But the Face API, which is also part of the Vision APIs, goes further by allowing faces to be identified by comparing them to previously seen facial images. Imagine that you are building a photo-sharing site and you want to 1) automatically generate search keywords for the images that users upload, 2) identify faces in the images for possible tagging and identification, and 3) filter out potentially offensive images. The Computer Vision API and the Face API facilitate all of this and more. Now imagine that you had to do all of this without Cognitive Services – by coding the algorithms yourself. *That* is the power of the Vision APIs.

“Face API has enabled us to validate more than 70% of credit transactions automatically, build a robust fraud detection system, and free our call centers to focus on supporting our customers.” – *Juan Rafael Sierra Escobar, Subgerente de Tecnología, SISTECREDITO S.A.S.*

Speech APIs

The Microsoft Cognitive Services Speech APIs include the following APIs:

API	Description
Bing Speech API	Convert speech to text, understand intent in speech, and convert text to speech
Custom Recognition API	Create a customized acoustic environment for speech and speaker recognition
Speaker Recognition API	Identify speakers or use speech as a means of authentication

There are many libraries available for translating text to speech and speech to text, but the Bing Speech API offers capabilities that most libraries don't. It supports streaming of input audio so

textual results can be returned in real-time. (In fact, it is used by Cortana, Skype Translator, and Bing Torque for Android.) It supports a wide range of languages, including Russian, Korean, and Chinese. And it includes Language Understanding Intelligence Services (LUIS) support, which employs active machine learning to extract intent from spoken commands.

The Speaker Recognition API offers the ability to verify speakers from spoken passphrases. You can test-drive this feature for yourself on the [Speech Recognition API](#) page by recording three voice samples, and then submitting a fourth sample and gauging the API's ability to identify your voice.

Implementing a similar feature yourself is a two-step process. First you write code to create a verification profile and upload three or more voice samples containing spoken passphrases, a process known as “enrollment.” Once enrollment is complete, you call the **verify** method with a spoken passphrase to perform voice verification. The following code sample, written in Ruby, demonstrates how to call **verify** to determine whether the vocal recording in the body of the request (“{audio}”) belongs to a specific user identified by “{id}”:

```
require 'net/http'

uri = URI('https://api.projectoxford.ai/spid/v1.0/verify?verificationProfileId={Id}')
uri.query = URI.encode_www_form({})

request = Net::HTTP::Post.new(uri.request_uri)
request['Content-Type'] = 'application/octet-stream'
request['Ocp-Apim-Subscription-Key'] = '{key}'
request.body = "{audio}"

response = Net::HTTP.start(uri.host, uri.port, :use_ssl =>
  uri.scheme == 'https') do |http|
  http.request(request)
end

puts response.body
```

The response is a JSON-encoded object containing three fields: “result,” “confidence,” and “phrase.” The “result” field contains the string “Accept” if the speaker is verified, or “Reject” if they are not. If the “result” field contains “Accept,” then the “confidence” field holds the string “Low,” “Normal,” or “High” indicating the level of confidence in the result. The “phrase” field

contains the phrase used to verify the user. Here is a JSON response for the phrase “My name is Satya Nadella” indicating that the speaker was verified with a high degree of confidence:

```
{
  "result" : "Accept",
  "confidence" : "High",
  "phrase": "My name is Satya Nadella"
}
```

The Speaker Recognition API supports an alternate form of voice identification in which arbitrary voice samples are uploaded to various users’ profiles, and the **identify** method is called with a voice sample to determine which of several users the voice belongs to.

Language APIs

The Microsoft Cognitive Services Language APIs include the following APIs:

API	Description
Bing Spell Check API	Correct spelling errors, understand homophones, and more
Language Understanding API	Understand language contextually, enabling apps to communicate with people the way they speak
Linguistic Analysis API	Simplify complex language concepts (for example, separate sentences and tag parts of speech) and parse text
Text Analytics API	Extract sentiment, key phrases, and topics from text, and identify language
Web Language Model API	Automate a variety of natural-language processing tasks using state-of-the-art language modeling APIs

The Text Analytics API contains five methods, each of which poses intriguing possibilities for developers. For example, suppose you wish to write an app that examines a Twitter stream for reactions to a product that your company recently introduced. The **sentiment** method analyzes text input to it and returns a numeric score from 0.0 to 1.0 indicating the sentiment expressed in that text, with higher values indicating higher positivity. The following code demonstrates how to call the **sentiment** method from JavaScript using jQuery’s \$.ajax function:

```

var input = '{"documents':[{'language':'en','id':'1000','text':" +
    twitterText + "}]}'";

$.ajax({
  url:"https://westus.api.cognitive.microsoft.com/text/analytics/v2.0/sentiment",
  beforeSend: function(xhr) {
    xhr.setRequestHeader("Content-Type", "application/json");
    xhr.setRequestHeader("Ocp-Apim-Subscription-Key", "{key}");
  },
  type: "POST",
  data: input
}).done(function(data) {
  alert(data.documents[0].score); // Sentiment from 0.0 to 1.0
}).fail(function() {
  alert("error");
});

```

Other Text Analytics API methods, including **topics**, which extracts a list of topics from text input, and **keyPhrases**, which generates a list of key phrases, could be used to perform further analysis on the Twitter stream and generate actionable intelligence from what users are saying.

Knowledge APIs

The Microsoft Cognitive Services Knowledge APIs include the following APIs:

API	Description
Academic Knowledge API	Interpret natural-language query strings and evaluate query expressions using the Microsoft Academic Graph
Entity Linking API	Identify related words and phrases within a paragraph
Knowledge Exploration API	Enable interactive search experiences over structured data using natural-language inputs
Recommendations API	Identify items that are frequently purchased together and items that a customer is likely to be interested in

The Recommendations API employs machine learning to handle the heavy lifting in including suggestions and recommendations in e-commerce applications. You create models and train them with information regarding completed transactions. Then you call methods in the API to

“By leveraging Azure Machine Learning and the Recommendations API, we have launched a new Personalized Commerce Experience for retailers that grows shopper conversion and engagement.” – *Frank Kouretas, Chief Product Officer at Orckestra*

retrieve item-to-item recommendations (“customers who purchased this item also liked the following items”) or user-to-item recommendations (“based on your purchase history, you may also be interested in the following items”). Here is a Curl command that requests up to eight recommendations for the item whose ID is 1000 from the model named “contoso:”

```
curl -v -X GET
"https://westus.api.cognitive.microsoft.com/recommendations/v4.0/models/contoso/recommend/item?itemIds=1000&numberOfResults=8&minimalScore=0.8"
-H "Ocp-Apim-Subscription-Key: {key}"
```

The return value is a JSON string containing an array of up to 8 items, each of which contains an item ID, an item name, and an item rating (a value from 0.0 to 1.0 with the calculated relevance of the item). Since the input specified a minimal score of 0.8, no items with a rating less than 0.8 are returned.

Search APIs

The Microsoft Cognitive Services Search APIs include the following APIs:

API	Description
Bing Autosuggest API	Generate search suggestions from partial search terms
Bing Image Search API	Implement advanced image search in an app or Web site
Bing News Search API	Turn any app or Web site into a news desk with news grouped and filtered by topic and searchable metadata
Bing Video Search API	Implement advanced video search in an app or Web site
Bing Web Search API	Implement intelligent search by leveraging Bing

Figure 3 shows the Bing Autosuggest API in action. Given a partial search term as input, it returns a JSON-encoded list of suggested search terms, complete with links for executing a search with each of the suggested terms.

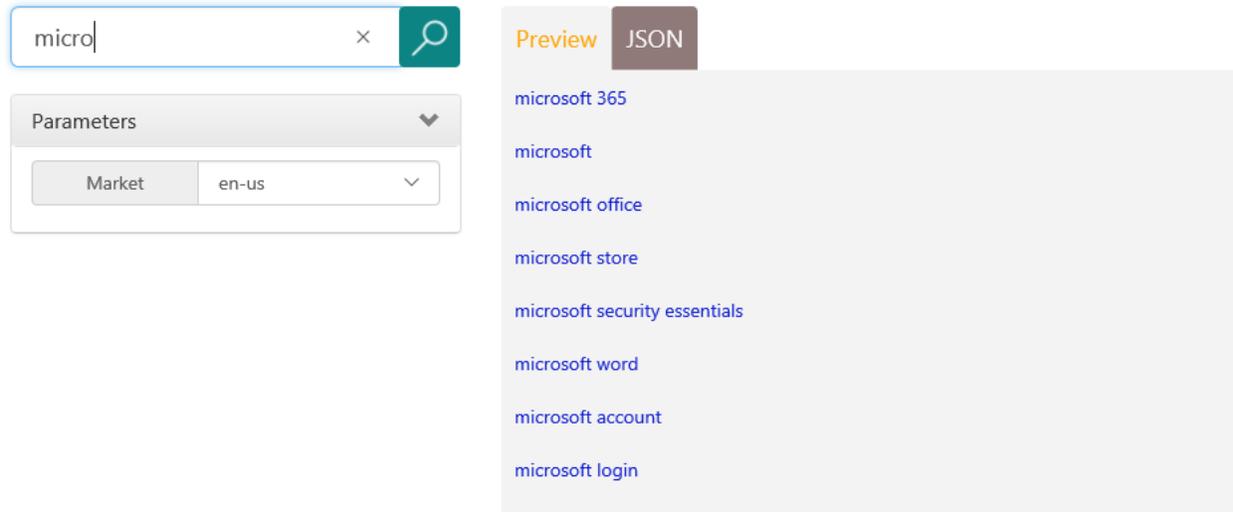


Figure 3: Search suggestions returned by the Bing Autosuggest API

Here's the C# code one would write to pass the string "micro" to the Bing Autosuggest API to generate a list of auto-completion terms, with a helping hand from the popular [Json.NET](#) library for extracting individual search suggestions from the JSON result:

```
var client = new HttpClient();
client.DefaultRequestHeaders.Add("Ocp-Apim-Subscription-Key", "{key}");
var uri =
    "https://api.cognitive.microsoft.com/bing/v5.0/suggestions?q=micro";
var response = await client.GetAsync(uri);
string json = await response.Content.ReadAsStringAsync();
dynamic result = JValue.Parse(json);
foreach (var suggestion in result.suggestionGroups[0].searchSuggestions)
{
    Console.WriteLine(suggestion.displayText); // Output each suggestion
}
```

Other Search APIs may be used to implement rich search experiences around images, videos, news, and Web content, essentially allowing you to harness the functionality of Bing in your own applications.

Summary

Embodying years of AI research and the development of sophisticated machine-learning algorithms designed to extract information from images, videos, text, speech, and even custom data sources, the Microsoft Cognitive Services APIs offer unprecedented power to build apps that communicate with users in an intelligent manner. Cognitive Services raises the bar for human-machine interaction and represents the next frontier in Software-as-a-Service (SaaS) offerings in the cloud. Put it to work for you – and your customers – today.

Need Help with Microsoft Cognitive Services?

[Wintellect](#) is an Azure Gold Partner with years of experience building cloud-enabled software and training others to do the same. We employ multiple Azure MVPs and we practice what we preach,

having migrated our own internal infrastructure to Azure while realizing a cost savings of more than 70%. We also develop extensive Azure training content for Microsoft and deliver it to customers all over the world. Want to work with the Azure experts who Microsoft trusts to know Azure inside and out? Send us an email at consulting@wintellect.com or call 1-865-966-5528 for more info.

