About the Authors

Srikanth Machiraju has over nine years of experience as a developer, architect, technical trainer and community speaker. He is currently working with Microsoft Hyderabad designing and preaching modern application development using cutting-edge platforms and technologies. Prior to Microsoft he was worked with BrainScale as Technical training/SME on application building and migrations with Azure. He is a tech savvy developer who is passionate about embracing new technologies and sharing his learning as blogs or wider communities. He has also authored "Learned Windows Server Containers", blogs at https://vishwanathsrikanth.wordpress.com and active on LinkedIn at https://www.linkedin.com/in/vishsrik/

Ritesh Modi is an ex-Microsoft Customer software engineering senior technical evangelist. He is an Architect, a Senior Evangelist, Cloud Architect, published author, speaker, and a known leader for his contributions towards Datacenter, Azure, Bots, Blockchain, Cognitive services, DevOps, Artificial Intelligence and automation. Currently, he is working as Principal Consultant with Infront Consulting. He has more than a decade of experience in building and deploying enterprise solutions for customers. He has authored multiple books on Azure, DevOps and Automation. Some of his books includes "Azure for Architects", "DevOps with Windows Server 2016".

He has spoken at multiple conferences including TechEd and PowerShell Asia, very active in meetups and communities and s a writer for MSDN magazine. His blogs are available at https://automationnext.wordpress.com. His twitter handle is #automationnext.

Excerpt from Book.

Bots can be broadly divided into two categories chatty bots and smart Bots. Chat Bots respond using predefined set of rules and hence the responses are limited. For example, if you build a bot for reserving a table in a restaurant, the Bot would always ask the basic questions of data and time, number of people and seating preference (indoor/outdoor). If you throw a random request at the chat bot it might not respond with a meaningful message or just respond with a generic message. Smart Bots are more intelligent, they work upon a large and wide variety of information and generate more human like responses. Smart Bots are designed to learn from the conversation and provide more useful answers as the conversation progresses, leaving an impression of having more like a human-to-human conversation on the end-user. Bots can be designed to create smarter responses by using cutting-edge Artificial Intelligence algorithms. Authoring AI algorithms is a complex task, it requires a varied skill set and lots of analysis to build an AI algorithm which can perform tasks like Natural Language processing, sentiment analysis or to generate recommendations. Microsoft Cognitive Services provides a basket of AI algorithms which can be integrated into any application. These algorithms are developed by an expert team and caters across the fields of computer vision, speech, text analysis, natural language processing, knowledge extraction and web search. In this chapter, we will learn to build Smarter Bots using Microsoft Cognitive Services and understand the capabilities of evolving list of powerful AI algorithms.

The following topics will be discussed in the following chapter.

- Introduction to Microsoft Cognitive Services
- Getting Started, APIs, Language Support
- Building a Bing Web Search Bot
- Building an OCR Bot.

Introduction to Microsoft Cognitive Service

Microsoft Cognitive Services is an exhaustive list of intelligent APIs which can be easily integrated into any type of application. Formerly known as Project Oxford, Microsoft Cognitive Services is built on top of Azure Machine Learning. Microsoft Cognitive Services contains highly complicated, State of the art intelligent ML algorithms exposed as uniform and simple to use REST APIs and available as SDKs for few languages. REST APIs can be used in any type of application written in any language just by adding few lines of code. With Cognitive Services Microsoft is aiming to increasing the productivity of every individual and organization by allowing developers to build applications with rich intelligence by which applications can hear, see, speak and think like humans, the goal is to simplify and merge the much-complicated ML based algorithms into the world of mainstream computing thus bringing the benefits of AI to every individual.

Microsoft Cognitive Services lets you build all the smartness into the Bot application which can differentiate your Bots from the rest while all the complexity is handled within the APIs. Bots can integrate with Microsoft Cognitive Services to provide a personalized and rich conversation experience to the user. They allow you to perform complex operations like image processing, pattern matching, recommendations, speech to text, tagging images etc. using simple API calls. Microsoft Cognitive Services is broadly divided into Vision, Speech, Language, Knowledge and Search APIs. Each of these categories contain discrete APIs catering to various needs. While you are designing your Bots for intelligence or personalization it is recommended to go through the below exhaustive list of APIs to see which one fits your needs or which among these can bring a more personalized and rich conversation experience for your Bot. For example, if I'm building a Bot for which automatically places an order for my monthly list of prescribed medicines, I would try to use OCR to read text from digital prescriptions, handwriting recognition to read text from hand written prescriptions, speech to text to accept voice inputs from the user, text analytics and Bing search APIs to provide a better information on the entities extracted from user's conversation.

Before you know which APIs to use you should know the breadth of services provided by MCS. Table 4-1 lists what each API provides.

Table 4-1. Microsoft Cognitive Services API Classification

Vision	Computer Vision API	This API helps you extract actionable information from the images. Based on the uploaded image or the image URL. The Vision API can analyze visual data and extract textual/objects/analytics in the form JSON response. For example: • Identifying objects, living beings from an image. • Categorizing images based on visual aspects into groups like Indoor, Dark, Sky. Vision API supports different 86 categories. • Identifying images as clip art, non-clipart or ambiguous. • Recognizing faces, gender and age from images with human faces. • Generating descriptions of an image • Recognizing Handwritten text and converting into digital format • Generate thumbnails from an image for different form factors • Perceiving color schemes of an image. • Flagging Adult and racy content Extracting text from image using Optical Character Recognition (OCR)
	Emotion API	This API can be used to extract emotions like happiness, sadness, surprise, anger, fear, contempt, disgust or neutral from an uploaded image or URL with human faces, each human face will also have a bounding box to help relate the response. The emotions for each human face are normalized in such a way they sum up to 1, so you should always consider the emotion with higher confidence. The Emotion API works across all cultures.
	Face API	This API can be used for face detection and face recognition from a image. The response contains the coordinates of the bounded rectangle (left, right top, bottom) which indicates the face within an image and few face related attributes like age, gender, pose, head pose, facial hair and glasses. It can also perform one-to-one face verification, check for similar looking faces, face grouping, face identification from database of faces and face storage for storing faces.
	Video API	This API allows you to detect faces, emotions from a video or URL of a video. You can also stabilize a video or generate thumbnail.
Speech	Bing Speech API	This API helps you to do enable voice input for your applications. The API allows you to convert speech to text and vice-versa. It supports around 29 different languages (at the time of writing) for speech to text and around 40 languages for speech to text.

	Speaker Recognition API	This API provides speaker identification and verification services. Using Speaker Identification, you can provide sophisticated voice based authentication to your users. Each user should enroll by recording a specific phrase, during the authentication the voice for the same phrase is compared with the voice and phrase (and few other features) recorded during enrollment. For speaker identification each speaker should be enrolled by speaking any phrase, for any given voice input the API compares the voice using the enrolled list and returns an identity if found.
Language	Bing Spell Check API	This API can be used to provide contextual grammar and spelling corrections using web based spell checker built using Azure ML. Traditionally spell check is performed using dictionary based rule-sets, as in Microsoft Word but Bing Spell check uses web documents to provide a real-time spelling and contextual grammar checks. It supports many sophisticated features like slang, word breaking etc. which is otherwise not possible with regular spell checkers.
	Text Analytics API	This API can be used to perform sentiment analysis from a text, extract key phrases from a sentence using Natural Language Processing Kit, extract topics from a discussion or user review or for language detection from a hand-written text (around 120 languages are supported)
	Web Language Model API	This API can be used to build a Language Model using the web scale corpus collected by Bing in the en-us market. The data set is available as XML Web Service, it is divided into 4 categories namely Body Text, Title Text, Anchor Text and Query Text.
	Linguistics Analysis API	This API can be used to identify structure of text, the following services are provided by the API. 1. Sentence Separation and Tokenization 2. Part-of-speech tagging 3. Constituency Parsing
Knowledge	Recommendations API	This API can be used to recommend items to your customer based on the customer activity and trained catalog of items in your store. The API supports scenarios like frequently brought together, recommend related items and recommend items based on customer's past activity
	Entity Linking Intelligence Service	This API can be used to identify entities within a specific paragraph and context. For example, in the sentence "The Bermuda Triangle, also known as the Devil's Triangle, is a loosely-defined region in the western part of the North Atlantic Ocean" the API recognizes entities like "Bermuda Triangle", "Devil's Triangle", "North Atlantic Ocean". This can be combined with Bing search API to link discovered entities with contextual information from web.

	Academic Knowledge API	This API can be used to interpret search queries related with academic intent using Microsoft Academic Graph Knowledge base. The knowledge base is divided into field of study, author, institution, paper, venue and event. The MAG knowledge base is continuously indexed and mined by using Bing.
Search	Bing Search API	This API provides search capabilities like Bing.com which can be integrated into any application. Using this API, you can perform searches related to web pages, images, videos and news. Bing Search API also provides content-specific APIs each for images, web, vides and news for more specific search.
	Bing Auto Suggest API	This API can be used to provide Auto Suggest capabilities for search queries to your users using the data from the web. The suggested terms is picked from the query searches performed by other users and user intent.

Apart from the above Microsoft also allows you to have a take an early look at their biding Cognitive Services like Project Prague, Project Johannesburg, Project Wollongong for more details on each of these projects visit https://labs.cognitive.microsoft.com/.