# Azure Cognitive Services

Elérhető Cognitive Services modulok:

* **Computer Vision (számítógépes látás)**: rengeteg, szinte minden alkalmazást lefedő szolgáltatást biztosít, köztük: felnőtt tartalom detekció, márka detekció, színdetekció, kép körülírása, arcdetekció (arc kontúr koordináták, kor, nem), objektumdetekció, híresség/híres helyszín beazonosítás, optikai karakterfelismerés, képhez tartozó címkék listázása, ...
* **Face recognition (arcfelismerés)**: naprakész technikák és algoritmusok segítségével egy bizonyos konfidencia szinten arcokat ismer fel. Képes meglévő arckép adatbázisokban hasonló vagy azonos személyhez tartozó arcokat keresni. Képes bizonyos vonásokat, jellemzőket kinyerni, mint például: nem, kor, szemszín, stb…
* **Speaker recognition (beszélő felismerés):** egy meglévő adatbázisban keres egy olyan személyt, akinek a beszéde legjobban közelíti a kérdéses beszélőét. A modul célja részben autentikációs rendszerekben való alkalmazás.
* **Speech recognition (beszédfelismerés)**: beszédfelvételből szöveget állít elő, figyelembe véve a nyelvi szabályokat.
* QnA maker (kérdések & válaszok generátor)
* Text Analytics (szövegelemzés, érzelem analízis)
* Text Translator (fordító)
* Anomaly Detector (anomáliadetekció, anomáliapredikció)
* Bing Autosuggest (keresési javaslatok generálása)
* Bing Custom Search (testreszabható intelligens keresőmotor)
* Bing Spellcheck (Nyelvi hibák automatikus javítása)
* Content Moderator (intelligens tartalom moderálás)
* Language Understanding (szóbeli parancsok megtanulása)
* Immersive Reader (szövegfelolvasó)

A modulok kipróbálhatók programkód írása nélkül is, az API Console menüből, ahol az API kulcs megadásával lehet requesteket végrehajtani, és a választ az oldalon megtekinteni.

### Computer Vision

A Computer Visiont az API Consoleból csak online tárhelyen levő képekkel lehet kipróbálni, a kép címének a requestbe való bemásolásával. Lokális fájlokat (képeket) a számítógépről futtatott programmal lehet felhasználni, a REST API segítségével, a POST request adat paraméterében megadva a kép byte kódját.



A Computer Vision API-nak az alábbi endpointjai érhetők el:

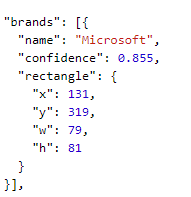
* **Analyze Image (POST)**:

Ez a szolgáltatás képes képeket analizálni, amelyen keresztül képes információt kinyerni a kép tartalmát illetően, felismerni a rajta szereplő objektumokat.

Példák:

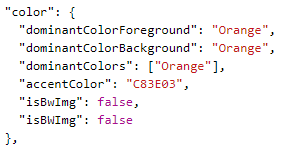
* **Brand felismerés**:

Bemenet: 

Kimenet: 

* **Színfelismerés** (domináns szín, háttérszín, fekete/fehér kép-e?)

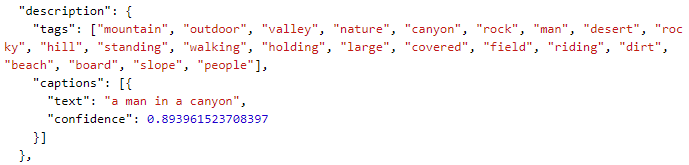
Bemenet: 

Kimenet: 

* **Kép leírása, körülírása:**

Bemenet: 

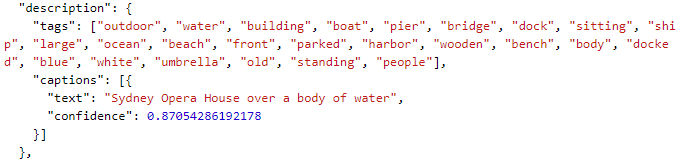
Kimenet:



Bemenet:



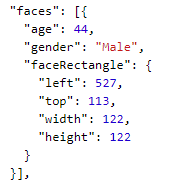
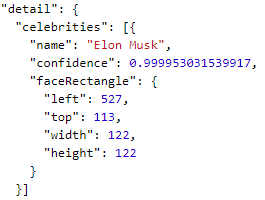
Kimenet:



* **Arcfelismerés** (híresség azonosítása, arckarakterisztikák, kor, nem)

Bemenet: 

Kimenet:



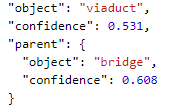
A kormeghatározás 4 évet tévedett. (48)

* **Objektumdetekció**:

Bemenet:



Kimenet:

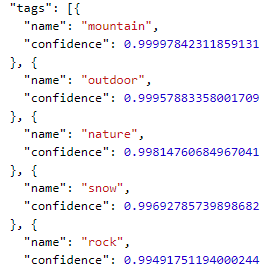


* **Címkegenerálás**

Bemenet:



Kimenet:



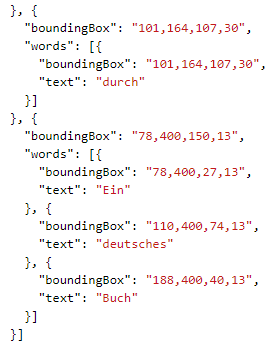
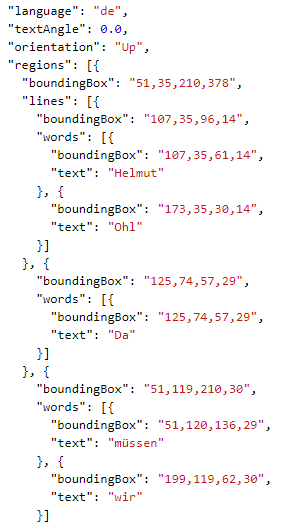
* OCR (POST):

Az optikai karakterfelimerés modul szövegfelismerést végez képeken. Bemenete egy kép, kimenete a képen felismert szöveg. A szolgáltatás képes automatikusan felismerni a szöveg nyelvét (manuálisan is meg lehet adni requestben - a magyar nyelv szerepel a támogatott nyelvek között), illetve a szöveg orientációját is felismeri.

Bemenet:



Kimenet:



Amint látható, az OCR helyesen felismeri a szöveg nyelvét (német), orientációját (0 fok), és magát a szöveget is. (Helmut Ohl / Da müssen wir durch / Ein deutsches Buch)

Bemenet:



Kimenet:

Budaconsum Kft

1025 Budapest

Csatárka út 58

Príma Kányakapu Csemege

1116 Budapest

Budaörsi út 115

ADóSZÄM: 10274019-2-41

NYUGT A

COO CBA DOPPELKEKS 500G

I DB \* 665 Ft/DB

COO THEODORA 1,51 MENTES P

I DB \* Ft/DB

BOO MILLI KAUKAZUSI KEFIR

I DB \* 219 Ft/DB

COO PRIMA INGVALLAS TASKA

I DB \* 39 Ft/DB

COO DELIJXE WC 4T BARA

I DB \* Ft/DB

öSSlESEN:

Bankkártya :

665

99

219

39

499

1 521 Ft

1 521 Ft

A hét minden napján 5.000 Ft értékü

10% azonnali enged-

vásárlás felett,

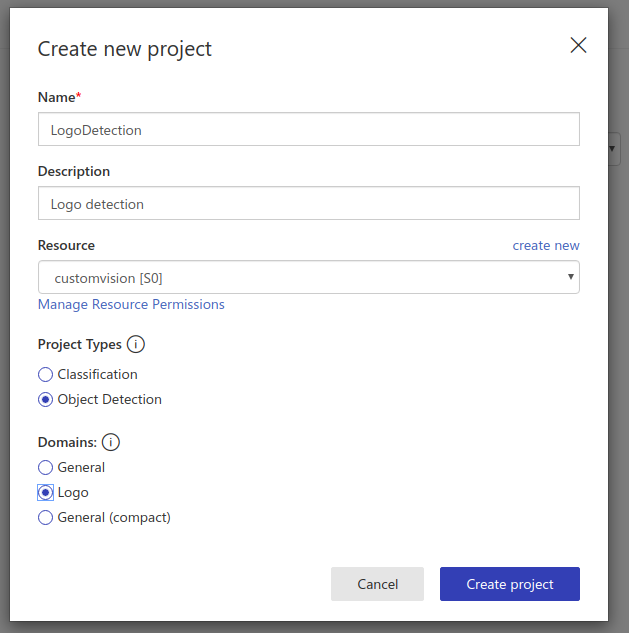
ményt adunk a vásárlás végösszegéből I

Az eredmény pontossága, minősége kiváló, az algoritmus néhány betű kivételével minden szót helyesen felismert, valamint a struktúrát is sikerült meglepően jól prezerválni.

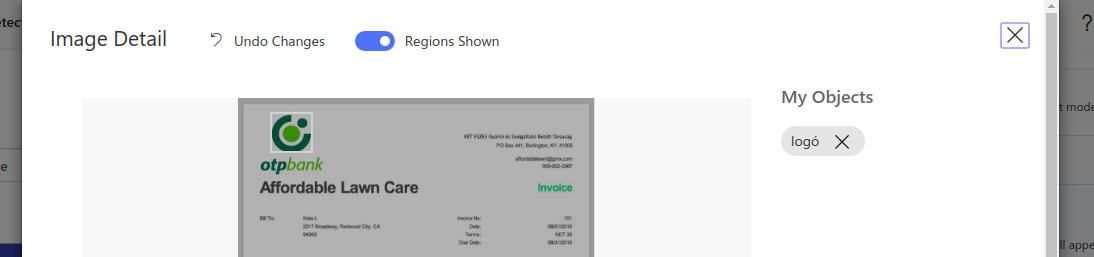
### Custom Vision

A Custom Vision szolgáltatás segítségével be lehet tanítani egy modellt saját objektumok felismerésére, klasszifikálására. A szolgáltatásnak dedikált weboldala is van (<https://www.customvision.ai/>), mely olyan segédfelületet tartalmaz, amely segítségével a feltöltött tanító képeken manuálisan be lehet jelölni a betanulandó objektumot.

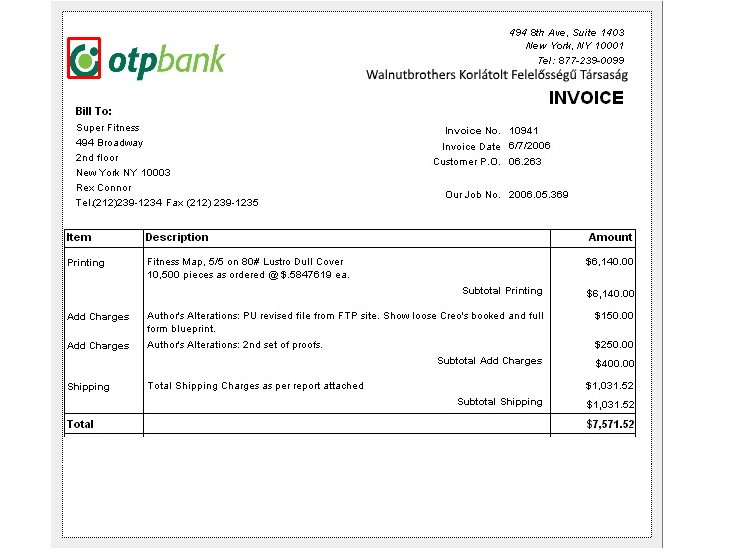
Az oldalon, új projektet lehet létrehozni, amely beállításaiban meg lehet adni a projekt típusát (klasszifikáció vagy objektumdetekció) és a felismerni kívánt objektumok jellegét (egyelőre logódetekció és általános objektumfelismerés támogatott).



Ezután lehet képeket felölteni a fájlrendszerről, majd felcímkézni a képeket egyenként vagy egy okos régiócímkézést javasló funkció segítségével, amely az összes képre javasol egy címkézést. Minden címkéhez minimum 15 képre van szükség.



Címkézés után el lehet indítani a modell betanítását, amire ismét két lehetőség van: gyors tanulás (kevés iteráció, próbál otimalizálni, alacsony költéségű), vagy általános betanítás, amelynek hossza állítható (1 óra - 24 óra, precízitás és költség az időtartammal arányos). Betanítás után a modellt a helyszínen, a felületen ki lehet próbálni, tesztképek feltöltétésével. Alternatív megoldás a modell publikálása (performace -> publish gomb), amely a modellt nyilvánosan elérhetővé teszi, így kódból el lehet érni a szolgáltatást és predikciókat végrehajtani. A kérés válaszában egy lista szerepel, amely tartalmazza a képen levő összes potenciális találatot, valószínűséggel, a körülhatároló négyzet paramétereivel és a felismert címkékkel. A négyzet paraméterei 0 és 1 közötti lebegőpontos számok, amelyek a relatív hosszokat adja meg (‘left’ = bal pixel / kép teljes szélessége).



Források, tanító példák:

* Logódetekció a felület használatával, majd a modell felhasználása Xamarin projektben, SDK segítségével, mobil környezetre: [link](https://docs.microsoft.com/hu-hu/azure/cognitive-services/custom-vision-service/logo-detector-mobile)
* Prediction API használata, publikálása a Custom Vision felületen: [link](https://docs.microsoft.com/en-us/azure/cognitive-services/custom-vision-service/use-prediction-api)
* A felület bemutatása, majd HTTP kérésen keresztüli elérése a modellnek C# nyelvben: [link](https://medium.com/@ThisisZone/image-recognition-using-the-azure-custom-vision-service-c0bfc74a3343)

### Speech Services

A Speech nevű szolgáltatás három fontos funkció gyűjtőhelye: speech-to-text (beszédből szöveg), text-to-speech (szövegből beszéd, szövegfelolvasás), speech-translation. A modul által használt technológia megegyezik a Cortana és Office termékekben használtakkal.

A Speech szolgáltatásokhoz tartozó gyorstalpalók és bevezető példák az alábbi linken elérhetők: [link](https://docs.microsoft.com/en-us/azure/cognitive-services/speech-service/#5-minute-quickstarts).



A szolgáltatás elérhető SDK fomájában, vagy REST API kérések útján. Egyes funkciók csak az egyik, vagy a másik formában elérhetők.

Az SDK könnyen telepíthető, Python környezetben például a pip install --upgrade azure-cognitiveservices-speech parancsot kell hívni.

#### Speech-to-text

A beszédfelismerő funkció egy hangfájlba rögzített beszédet szöveges formára ír át.

A jelenleg támogatott formátum a WAV fájl, és a támogatott mintavételezési frekvencia pedig 16 kHz. A funkciónak van egy rövid időtartamú (<15s) beszéded és egy hosszabb leforgású beszélgetést felismerő változata. Míg az előbbit az SDK és a REST API is támogatja, az utóbbit csak az SDK. A beszédfelismerő több nyelvet is támogat, például angol, kínai, arab, japán, finn, német, stb… A magyar nyelv nem támogatott.

A beszédfelismerő algoritmus egy előre betanított, univerzális modellt használ. Az inputban szereplő szokatlan zaj vagy specifikus dialektus megnehezíti, ellehetetleníti a felismerést. Ilyen esetben saját, személyre szabható modellt kell betanítani.

Egy saját Python programot használva, a mintakódok segítségével sikerült a Speech SDK-t kipróbálni. A felhasznált hangfelvétel a mintakódok Github oldalán megtalálható enrollment\_audio\_steve.wav fájl, melynek időtartama 19 másodperc. A felvételen egy férfi a következő szöveget mondja fel:

“*Hello it's a good day for me to teach you the sound of my voice. You have learned what I look like now you can hear what I sound like. The sound of my voice will help the transcription service to recognize my unique voice in the future. Training will provide a better experience with greater accuracy when talking or dictating. Thank you and goodbye.*”

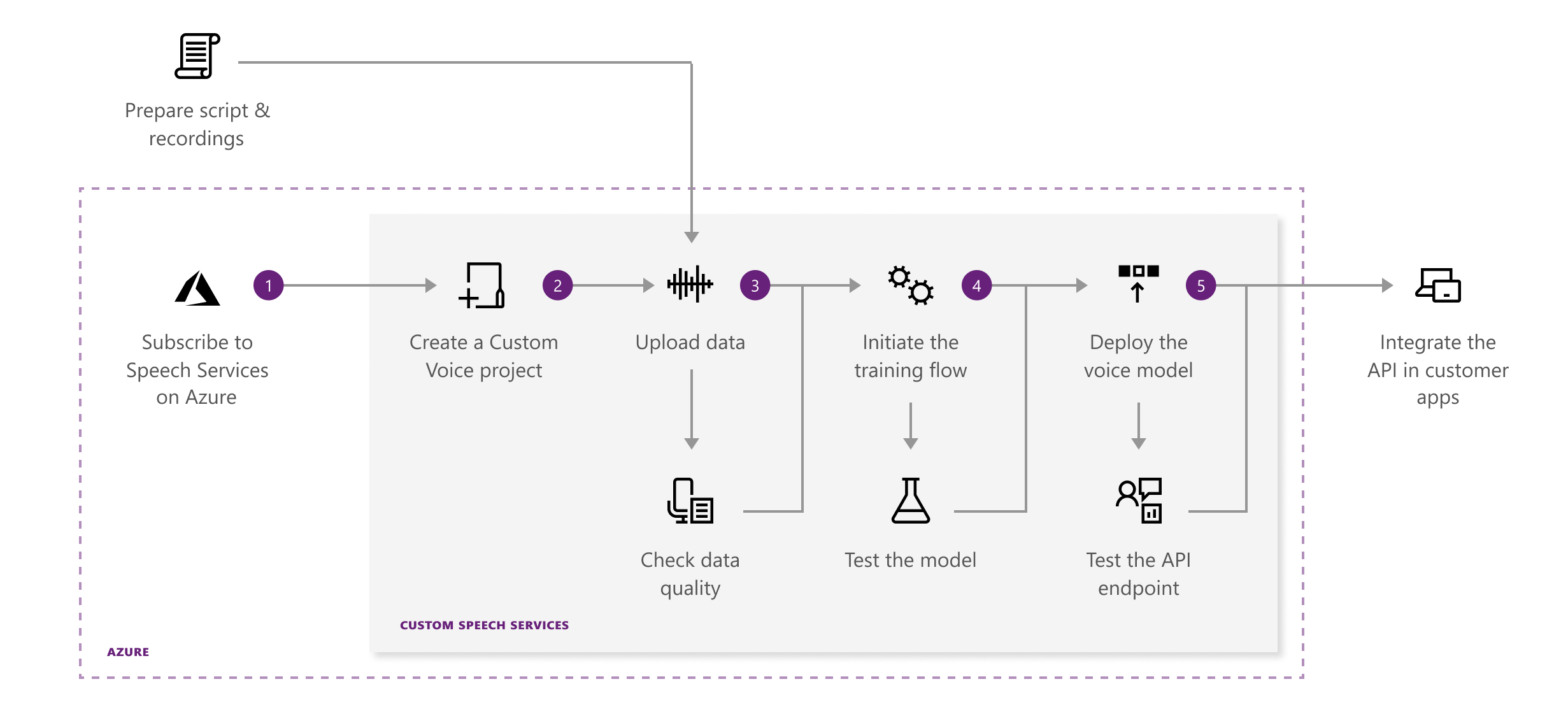
Az algoritmus szinte tökéletes helyességgel felismeri a kimondott szöveget. Az egyetlen hibás felismerés az utolsó előtti mondatban következik be, ahol a “*...talking or dictating..*.” helyett az outputban “*...talking to dictating*” szerepel.

#### Text-to-speech

Ez a szolgáltatás lényegében egy szövegfelolvasó, egy szövegblokkot szintetizált beszéddé alakít. Lehetőség van több hang közül választani, vagy akár saját hangprofilt is be lehet állítani. 75 különféle hangprofil elérhető, összesen 45 nyelven.

A szolgáltatáson belül két fajta hang létezik: standard, neurális és custom.

* Standard: normál, természetesnek hangzó beszéd, helyes, pontos kiejtés. Rövidítések, számok, dátumok és egyéb szimbólumok felolvasását is támogatja.
* Neurális: mély neuronhálókkal szintetizált hang, amely a standard beszédhez képest változatosabb, természetesebb intonációval rendelkezik. Hosszú szövegek, könyvek felolvasásához használatos Azureben.
* Custom (saját / személyre szabhatú hangprofil): ha rendelkezünk tanító adatokkal (szöveg, felolvasott szöveg felvétele), akkor egy gépi tanulás modellt betanítva, imitálni tudjuk mesterséges intelligenciával ezt a hangot.



#### Intent Recognition

Szóbeli parancs, utasítás interpretációját, és az kimondott beszédben kulcsfontosságú entitások felismerését végzi ez a szolgáltatás. Működése a Microsoft Language Undersdanding Service (LUIS) rendszeren alapszik. A LUIS a beszédfeliserő algoritmusokat használva, természetes nyelvfeldolgozást végezve hang-alapú utasításokat interpretál.

Dedikált weboldallal rendelkezik: [https://www.luis.ai](https://www.luis.ai/home).

Szemléltető példa a LUIS működéséről:

input (szóbeli utasítás): “Book me a flight to Cairo”

output:

"query": "Book me a flight to Cairo",

"topScoringIntent": {

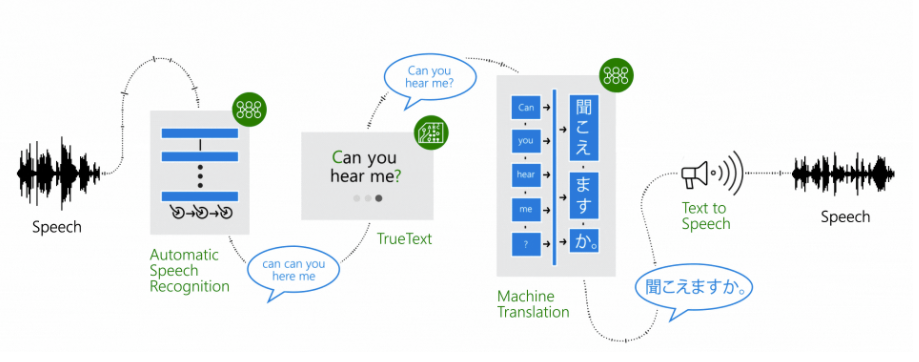
"intent": "BookFlight",

"score": 0.9887482

}

#### Speech Translation

Beszéd-beszéd fordítást végez, felhasználva a Speech API beszédfelismerő, átíró és hangszintetizáló képességeit, illetve a Text Translator API fordító képességeit. Lényegében: a beszédet szöveggé alakítja, a szöveget lefordítja a másik nyelvre, majd a lefordított szöveget beszédszintézissel felolvassa.



### Text Translator

A Translator Text API szövegek különböző nyelvek közötti fordítását végzi. A modul több, mint 60 nyelvet támogat, bármilyen platformon / operációs rendszeren használható. A szolgáltatás a Neural Machine Translation technológiát használja azaz mesterséges neurális hálók segítségével végzi a fordításokat.

Az általam kipróbált példában az alábbi angol szöveget fordítottam le német és magyar nyelvekre:

*“The history of Budapest began when an early Celtic settlement transformed into the Roman town of Aquincum,'the capital of Lower Pannonia. The Hungarians arrived in the territory in the late 9th century, but the area was pillaged by the Mongols in 1241–42. Re-established Buda became one of the centres of Renaissance humanist culture by the 15th century.The Battle of Mohács, in 1526, was followed by nearly 150 years of Ottoman rule.”*

Eredmény - német változat:

*“Die Geschichte von Budapest begann, als sich eine frühe keltische Siedlung in die römische Stadt Aquincum, die Hauptstadt von Unterpannonien, verwandelte. Die Ungarn kamen im späten 9. Jahrhundert in das Gebiet, aber das Gebiet wurde von den Mongolen in den Jahren 1241–42 geplündert. Buda wurde im 15. Jahrhundert zu einem der Zentren der humanistischen Kultur der Renaissance. Auf die Schlacht von Mohc im Jahre 1526 folgten fast 150 Jahre osmanische Herrschaft.”*

Érdemes megjegyezni, hogy a Mohács településnévből lemaradtak a magyar karakterek a német fordításban.

Eredmény - magyar változat:

*“Budapest történelme akkor kezdődött meg, amikor az alsó-Pannon főváros, Aquincum római városába átformált korai kelta település. A magyarok a 9 század végén érkeztek a területre, de a területet 1241 – 42-ben a mongolok kifoszították. Buda újra megalapitja a reneszánsz humanista kultúra egyik központává a 15. A Mohácsi csata, 1526-ben, a török uralom után közel 150 évvel.”*

A szöveg közel sem helyes mindenhol szemantikailag, de nagyon jó alap egy manuális fordításhoz.

Fordítás mellett, a szolgálatás transzliterációt (átbetűzés - pl: latin ábécéből japánba), nyelvfelismerést, valamint egyedülálló szavak fordítását is elvégez.

### Text Analytics

A Text Analytics API a Cognitive Services által nyújtott szövegelemzési szolgáltatások gyűjtőhelye. Három fontos képessége van: érzelmi analízis, entitás felismerés és kulcsfontosságú szószerkezetek kinyerése.

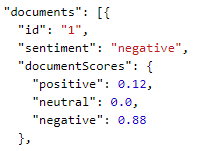
#### Sentiment Analyisis

Érzelem analízis során, a program egy nyers szövegrészletről próbálja megállapítani az érzelmi töltetét. Az Azure Text Analytics API-ja támogatja ezt a szolgáltatást olyan formában, hogy a bemeneti szövegre egy 0 és 1 közötti értékkel tér vissza, amely az érzelmi skálát reprezentálja, 0 jelenti az abszolút negatív érzelmi töltetet, 1 pedig az abszolút pozítivot.

Példa, vélemény a Game of Thrones sorozat utolsó részéről:

“*Its sad that one of the best TV shows of all time concluded the way it did. The ending wasn't bad, but the 7 season buildup only to have this mad dash to the end is awful. Nothing was explained, nothing was written well, terrible pacing! Absolutely disrespectful to fans! This last season makes it really hard to recommend the show to new viewers or repeat viewings. If the writers didn't care why should we?*”

Az elemzés eredménye:



A szoftver helyesen felismeri, hogy egy negatív véleményről van szó. Az eredményben, továbbá, az egyes mondatok külön-külön is ki vannak értékelve.

#### Entities

Szövegben szereplő különböző entitásokat (személy, cég, helyszín) ismeri fel.

Példa szöveg:

“*Dickens Hampshire-ben született, ötévesen költözött Kentbe, tízévesen Londonba a családjával. A korai évei szépen, jómódban teltek; sokat olvasott, különösen a pikareszk regényeket szerette. Azonban amikor tizenkét éves lett, apja, a tengerészeti pénztár kistisztviselője, eladósodott és börtönbe került. Dickens egy fénymázműhelyben kezdett dolgozni. Gyermekkori élményei, a látott nyomor visszatérő elemeivé váltak regényeinek. Heti hat shillingből tartotta el magát és támogatta családját.*”

Eredmény:



Az algoritmus többnyire helyes adatokat nyer ki, a magyar nyelv határozóragját nem kezeli helyesen, London helyett Londonba szót kapjuk.

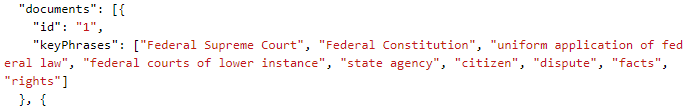
#### Key Phrases

Ez a szolgáltatás kulcsfontosságú szavakat, szócsoportokat nyer ki egy hosszabb nyers szövegből.

Példa szöveg (részlet a svájci legfelsőbb bíróságról szóló Wikipédia cikkből):

“*As a state agency, the Federal Supreme Court examines the uniform application of federal law by the cantonal and federal courts of lower instance. It protects the rights that the citizen has according to the Federal Constitution. During a dispute, the Federal Supreme Court examines the application of the law and does not examine the facts of the other courts below, unless evidently flawed*.”

Eredmény:



A kutatást az „Integrált kutatói utánpótlás-képzési program az informatika és számítástudomány diszciplináris területein” (EFOP-3.6.3-VEKOP-16-2017-00002) című projekt támogatta. A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósult meg.