

Projet 01 : Découvrez le métier d'Ingénieur IA

Mohamed A

Présentation de la banque 'Credit' et du contexte commercial :

- Usage 100% digital des consommateurs.
- Concurrence des banques en ligne agressive

Présence dans +100 pays sur les cinq continents



Relation client de plus en plus digital (+10% an sur 3 ans)



Concurrence des banques en ligne 100% digitales (+20% de nouveaux clients par an sur 3 ans)



Jeux et objectifs :



Fluidité de l'expérience client



Croissance des nouveaux marchés

- Fluidifier les interactions clients (temps de réponse et pertinence de la réponse)
- Nouvelles générations habituées au digital

- Potentiel de croissance dans les pays émergents (en Afrique notamment) important.
- Utilisation des téléphones portables pour le paiement rend la banque digitale indispensable pour gagner des parts de marché dans ces pays.

Préparation et lancement du projet :

Découvrir le métier d'ingénieur IA

Projet 1

Entreprise : Banque " Credit"

Contact principal : ——— Manager : David Lemoine

Support : ——— Mail

Travail demandé : Développer un chatbot

Buts

réduire drastiquement le temps de réponse
aux questions posées par les utilisateurs

fluidifier l'expérience utilisateur

Fonctions

1-Polyglotte : être capable de donner la bonne réponse dans la
bonne langue.

2-Fiable : être capable de donner les bonnes réponses pour
les questions les plus fréquemment posées par les clients.

3-Pertinent : être capable d'orienter le client vers le bon
interlocuteur humain lorsqu'il ne connaît pas la réponse à la
question posée

Ma tâche assignée

Développer la fonction " Polyglotte"

Main tools :

Microsoft Azure ——— modèles pré-entraînés

API REST ——— Curl

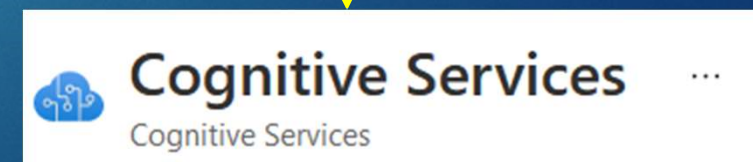
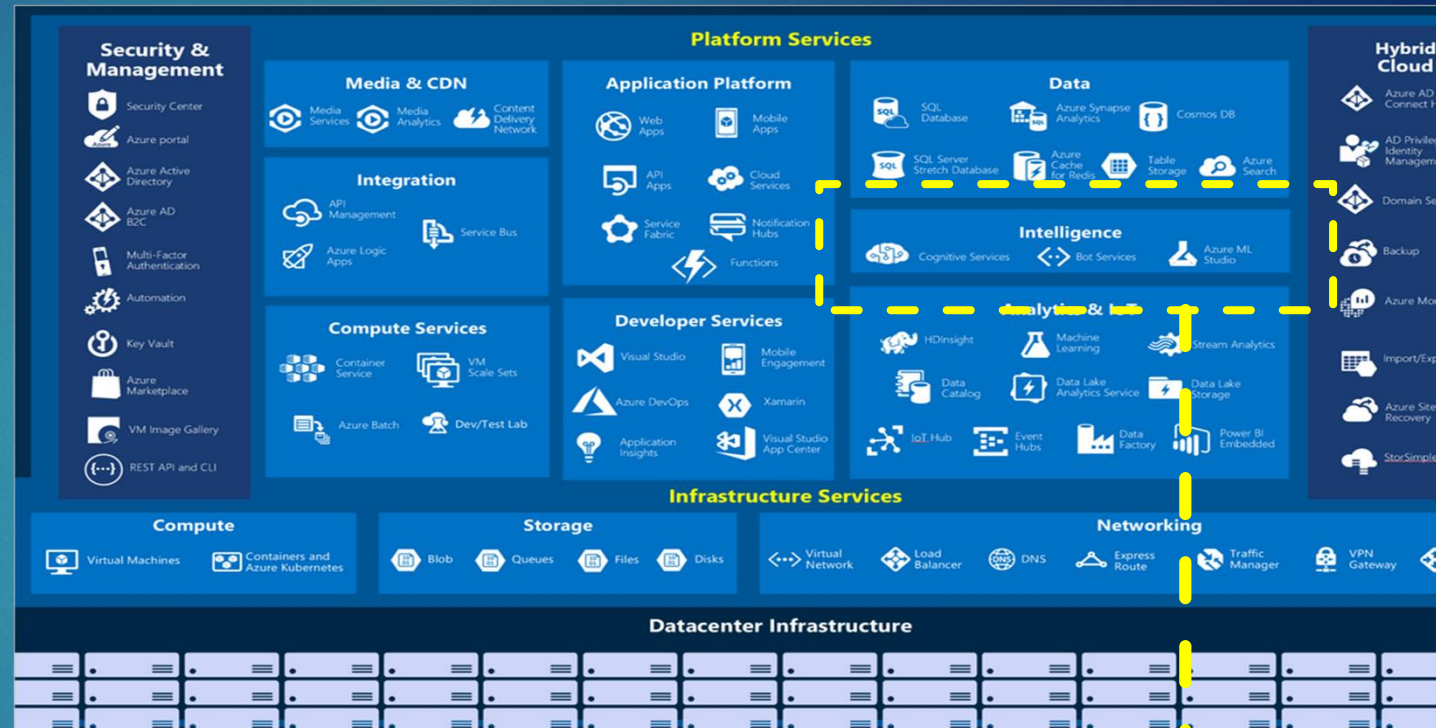
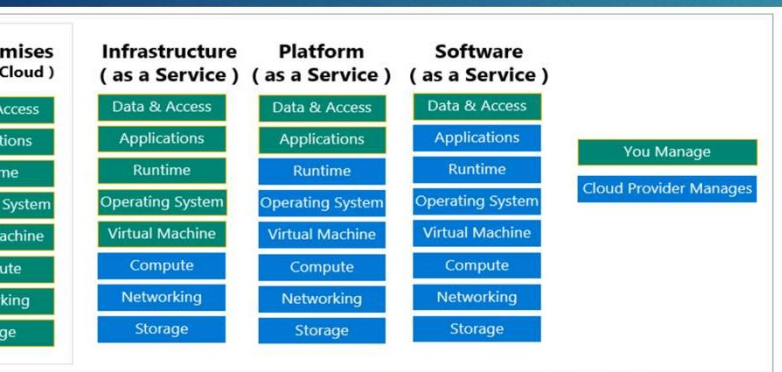
Autres...

Code Language ——— Python

PRESENTATION DU PORTAIL AZURE :

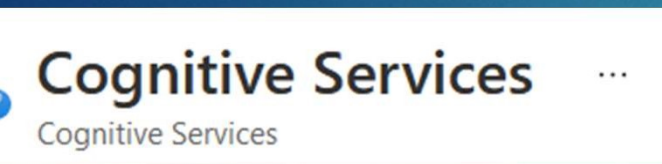
est une plateforme de services cloud qui propose une multitude de services numériques facilement paramétrables et extrêmement flexibles.

Le diagramme ci-dessus illustre les modes de services auxquels peuvent être dispensés les services Azure.




PRESENTATION DU PORTAIL AZURE :

Cognitive services est un lot d'algorithmes pour machine learning développés par Microsoft résoudre des problèmes dans le domaine intelligence artificielle.



Service name	Description
Computer Vision	Use image-processing algorithms to smartly identify, caption, index, and moderate your pictures and videos.
Speech	Convert spoken audio into text, use voice for verification, or add speaker recognition to your app.
Knowledge mapping	Map complex information and data to solve tasks such as intelligent recommendations and semantic search.
Search	Add Bing Search APIs to your apps and harness the ability to comb billions of webpages, images, videos, and news with a single API call.
Natural Language processing	Allow your apps to process natural language with prebuilt scripts, evaluate sentiment, and learn how to recognize what users want.

L'outil qu'on se propose d'utiliser dans ce projet est le module "Text Analytics" pour la détection de la langue.

Text Analytics
AI + MACHINE LEARNING

5,000
transactions per month

Extract information such as sentiment, key phrases, named entities, and language from your text. [Learn more](#)

Create

PRESENTATION DU SET DE DONNES :

La tâche sera donc de tester si le service cognitif Microsoft Azure peut détecter le langage d'un message à partir du jeu de données [Wikipedia Language Identification Database](#) (téléchargement direct à [ce lien](#)).

Le jeu de données contient plus de 200 000 paragraphes dans 235 langues. Sélectionne quelques paragraphes des cinq langues les plus parlées dans le monde et vérifie si tu envoies un paragraphe dans une langue au service cognitif, il te renvoie bien la langue identifiée.

2018 - Wikipedia Language Identification Database

The Wikipedia language identification benchmark dataset, contains 235000 paragraphs of 235 languages. The dataset is split into a training and a test set.

For more information, see the "Wikipedia language identification benchmark dataset for written language identification" paper (soon on arXiv) for more information.

File	Size
Wikipedia_Language_Identification_Database.zip	954 Bytes
x_train.txt	12.0 kB
y_train.txt	4.5 MB
x_test.txt	65.2 MB
y_test.txt	64.1 MB
urls.txt	480.5 kB
labels.csv	480.5 kB

Contents

```
* `x_train.txt`: 175000 lines of text. Each line belongs to one language.
* `y_train.txt`: 175000 lines. Each line denotes the language of the same line in `x_train.txt`
* `x_test.txt`: See `x_train.txt`
* `y_test.txt`: See `y_train.txt`
* `urls.txt`: A list of permanent URLs to all pages used for paragraph extraction
* `labels.csv`: A header line plus one line per language
* `README.txt`: This file
```

Set de paragraphes a utiliser pour le test du module Azure

The screenshot displays two text editors side-by-side. The left editor, titled 'x_train.txt', shows a list of text samples in various languages, including English, Hindi, and French. The right editor, titled 'y_train.txt', shows a list of language codes corresponding to the samples in the left editor, such as 'est', 'swe', 'mai', 'oci', 'tha', 'orm', 'lim', 'guj', 'pnb', 'zea', 'krc', 'hat', 'pcd', 'tam', 'vie', 'pan', 'szl', 'ckb', 'fur', 'wu', 'arz', 'ton', 'eus', 'map-bms', 'glk', 'map-bms', 'nld', 'arz', 'bod', and 'jpn'.

DEPLOIEMENT DU MODULE 'TEXT ANALYTICS' :

Passons le workflow relative au deployment du module 'text analytics'

Créer une ressource Azure Language, qui vous accorde l'accès aux fonctionnalités offertes par le service Azure Cognitive Service for Language. Il génère un mot de passe (appelé une clé) et une URL d'endpoint que vous utiliserez pour authentifier les requêtes API.

Créer une requête en utilisant soit l'API REST ou la bibliothèque de client pour C#, Java, JavaScript, et Python. Vous pouvez également envoyer des appels asynchrones avec une requête de batch pour combiner des requêtes API pour plusieurs fonctionnalités dans un seul appel.

Envoyer la requête contenant vos données sous forme de texte non structuré brut. Votre clé et votre endpoint seront utilisés pour l'authentification.

Récupérer ou stocker la réponse localement.



Microsoft.CognitiveServicesTextAnalyticsWithConfigurations | Overview

Deployment

Search (Ctrl+/) << Delete Cancel Redeploy Refresh

Overview

Inputs

Outputs

Template

We'd love your feedback →

✓ Your deployment is complete

Deployment name: Microsoft.CognitiveServicesTextAnalyticsWithC... Start time: 11/9/20...
Subscription: Azure subscription 1 Correlation ID: 7c...
Resource group: project01textanalytics

Deployment details (Download)

Resource	Type
✓ project01textanalytics	Microsoft.CognitiveServices/account

Next steps

Go to resource



- Extraire les "key" et le "endpoint" à partir du module déployé sur Azure
- Utiliser les codes mis à disposition pour utiliser le module et vérifier les résultats

SULTATS :

methods ont été utilisées pour vérifier ce que renvoie le module en terme de detection :

REST : Curl

phon.

base de données de Wikipedia a été utilisée comme source de text.

The screenshot displays a terminal window and a text editor window. The terminal window shows a curl command being executed to test the Azure AI Text Analytics API. The output indicates that the detected language is Estonian. The text editor window shows a list of language codes, with 'est' (Estonian) highlighted. A green arrow points from the 'est' entry in the text editor to the 'detectedLanguage' field in the terminal output.

```
AMC@DESKTOP-D4UUDU6 MINGW64 ~/Desktop
$ curl -X POST https://project01textanalytics.cognitiveservices.azure.com/text/analytics/v3.1/languages/ \
-H "Content-Type: application/json" \
-H "Ocp-Apim-Subscription-Key: 3456f41e057f4010995f8573981c3fbf" \
-d '{ "documents": [ { "id": "1", "text": "Klement Gottwaldi surnukeha palsameeriti ning paigutati mausoleumi." } ] }'
```

% Total	% Received	% Xferd	Average Speed	Time	Time	Time	Current
			Dload	Upload	Total	Spent	Left
100	263	0	160	100	103	363	234

```
errors": [], "modelVersion": "2021-01-05"}
599{"documents":[{"id":"1","detectedLanguage":{"name":"Estonian","iso6391": "est"}]}
```

est
swe
mai
oci
tha
orm
lim
guj
pnb
zea
krc
hat
pcd
tam
vie
pan
szl
ckb
fur
wuu
arz
ton
eus
map-bms
glk
map-bms
nld
arz
bod
jpn

```
In [10]: key = "3456f41e057f4010995f8573981c3fbf"
endpoint = "https://project01textanalytics.cognitiveservices.azure.com/"

from azure.ai.textanalytics import TextAnalyticsClient
from azure.core.credentials import AzureKeyCredential

# Authenticate the client using your key and endpoint
def authenticate_client():
    ta_credential = AzureKeyCredential(key)
    text_analytics_client = TextAnalyticsClient(
        endpoint=endpoint,
        credential=ta_credential)
    return text_analytics_client

client = authenticate_client()

# Example method for detecting the language of text
def language_detection_example(client):
    try:
        documents = ["Klement Gottwaldi surnukeha palsameeriti ning paigutati mausoleumi."]
        response = client.detect_language(documents = documents, country_hint = 'us')[0]
        print("Language: ", response.primary_language.name)

    except Exception as err:
        print("Encountered exception. {}".format(err))
    language_detection_example(client)
```

Language: Estonian

CONCLUSION :

1- L'EXERCICE PRESENTE UNE BONNE INTRO A L'IA ET MONTRE A TRAVERS LES MODULES AZURE LA CAPACITE DES METHODES IA A SOLUTIONNER DES PROBLEMATIQUES CONCRETES.

2- LA POSSIBILITE DE COMBINER DES MODULES EXTERNES DANS UN PROJET IA, ESSENTIELLEMENT SI CEUX LA SONT BIEN ENTRAINES.



MERCI POUR VOTRE ATTENTION