



DÉTECTION DE LA LANGUE AVEC LES SERVICES COGNITIFS MS AZURE

Voahangy Joan ALEONARD – 23/07/2020

AGENDA DU JOUR



PRÉSENTATION DU
PROJET



AZURE : SERVICES
COGNITIFS



SCRIPT ET RÉSULTAT



DÉMO : CURL ET
PYTHON



PRÉSENTATION DU PROJET

BANQUE CRÉDIT : RÉINVENTER L'EXPÉRIENCE CLIENT

Credit



Projet

Concevoir un **chatbot**



Enjeu / Objectifs

Fluidifier l'expérience client lors des interactions :

- Réduire le temps de réponse;
- Améliorer la **pertinence de la réponse**.



Mission

Programmer la fonction « polyglotte » du chatbot : **détecter la langue** afin d'apporter une réponse dans la bonne langue



Ressources

Utiliser un **modèle pré-entraîné de Microsoft Azure** afin de détecter la langue, un jeu de données publiques [ICI](#)



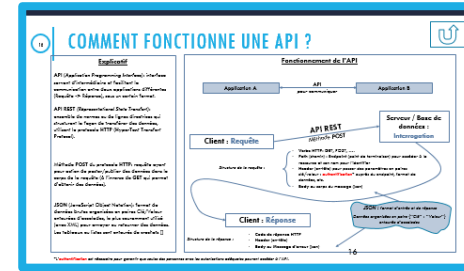
Résultat attendu

Livrer un **code complet, sécurisé et présentable**



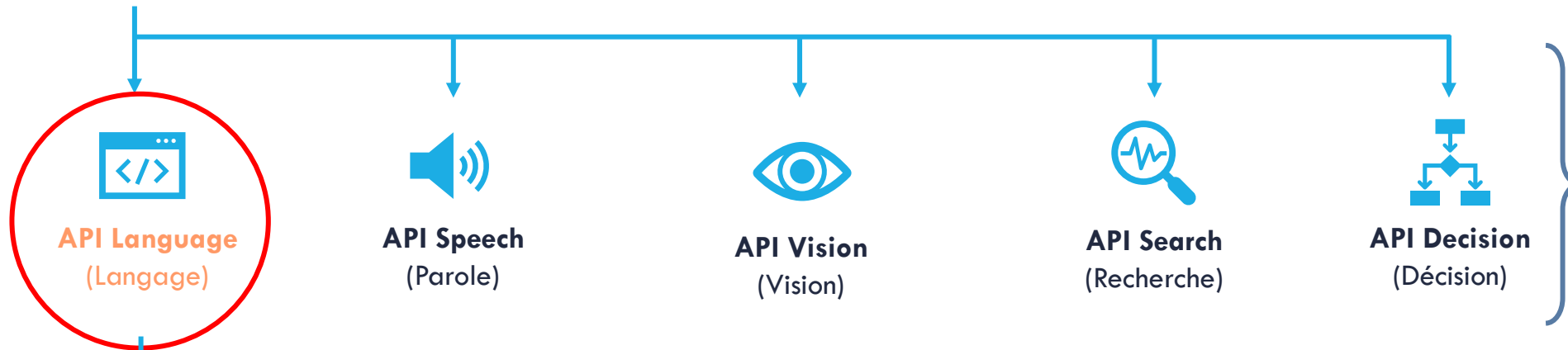
AZURE : SERVICES COGNITIFS

MICROSOFT AZURE : LES SERVICES COGNITIFS



Azure Cognitive Services

La solution **Cognitive Services** dans le cloud Azure est une famille d'**API** et de services permettant aux développeurs d'ajouter facilement des **fonctionnalités cognitives** à leurs applications, c'est-à-dire, **voir, entendre, parler, comprendre, commencer à raisonner**.



Le **catalogue des services** peut être classé en 5 grands piliers : **Langage, Parole, Vision, Recherche et Décision**.

Parmi les services disponibles en termes de **détection de langue**, l'API Language propose 2 choix :

Text Analytics
Analyse de texte

Ici, la méthode **Detect** évalue un texte et retourne les identificateurs suivants pour la langue détectée : **son nom, son code ISO et un score de puissance de l'analyse** (1=100% de niveau de confiance – soit un niveau dit haut, lorsque 0 est le niveau bas).

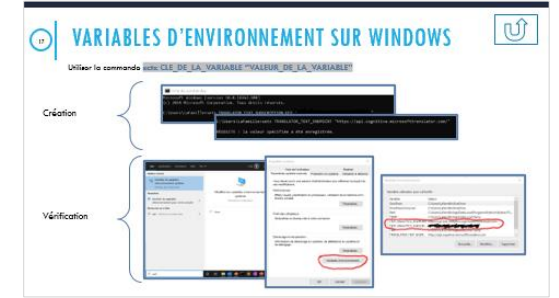
Translator
Traduction de texte

Ici, la méthode **Detect** évalue un texte et retourne les indicateurs suivants concernant le texte évalué : le **code de la langue, le score de niveau de confiance**, si les fonctionnalités « translation » et « transliteration » sont supportées (True ou False),... , tout cela **pour une langue identifiée comme principale, plus les alternatives**.



SCRIPT ET RÉSULTAT

cURL SUR L'INVITE DE COMMANDE



Ressources utilisées: l'éditeur de texte *Sublime Text* et le *Prompt Windows*

Configuration préalable: définition de la clé d'authentification de l'API dans les *variables d'environnement utilisateur*

Verbe HTTP POST

Endpoint complet de l'API pour définir le chemin

```

detect_curl.txt
1 curl -X POST "https://via-text-analytics.cognitiveservices.azure.com/text/analytics/v2.1/languages" ^
2 -H "Ocp-Apim-Subscription-Key:%TEXT_ANALYTICS_SUBSCRIPTION_KEY%" ^
3 -H "Ocp-Apim-Subscription-Region:westeurope" ^
4 -H "Content-Type: application/json" ^
5 -d "{\"documents\": [{ \"id\": \"1\", \"Text\": \" رلوی لآ امده قطخ ي دلاغل انل ع أ 21/7/2012 خي رطب وطبض لب م ل ب ق 21/7/2013\" } ]}"
  
```

Dans les en-têtes :

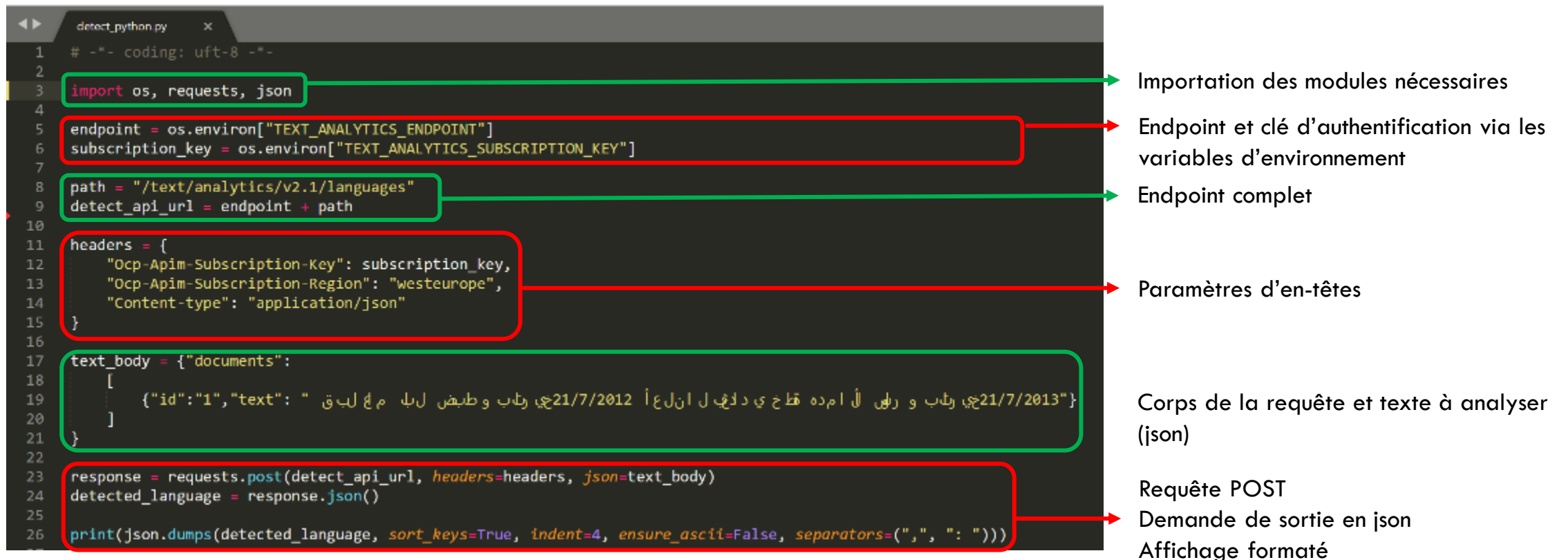
- Clé d'authentification
- Paramètre régional de l'API
- Spécification du format d'entrée

Le corps de la requête (json) et le texte à analyser pour détecter la langue

PYTHON SUR L'INVITE DE COMMANDE

Ressources utilisées : l'éditeur de texte *Sublime Text* et le *Prompt Windows*

Configurations préalables: spécification du chemin de l'exécutable Python dans les **variables d'environnement système** afin d'accéder directement à **l'interpréteur Python** sur le Prompt Windows, définition de la clé d'authentification de l'API dans les **variables d'environnement utilisateur**



The image shows a Python script named `detect_python.py` in the Sublime Text editor. The script is annotated with red and green boxes and arrows pointing to explanatory text on the right.

```

1  # -*- coding: utf-8 -*-
2
3  import os, requests, json
4
5  endpoint = os.environ["TEXT_ANALYTICS_ENDPOINT"]
6  subscription_key = os.environ["TEXT_ANALYTICS_SUBSCRIPTION_KEY"]
7
8  path = "/text/analytics/v2.1/languages"
9  detect_api_url = endpoint + path
10
11  headers = {
12      "Ocp-Apim-Subscription-Key": subscription_key,
13      "Ocp-Apim-Subscription-Region": "westeurope",
14      "Content-type": "application/json"
15  }
16
17  text_body = {"documents":
18      [
19          {"id": "1", "text": "21/7/2012 جي رٻڙ و طبيض لب م لڀ ق 21/7/2013 جي رٻڙ و رٻڙ ال امدہ قطح ي دلڙي ل انل ع ا"}
20      ]
21  }
22
23  response = requests.post(detect_api_url, headers=headers, json=text_body)
24  detected_language = response.json()
25
26  print(json.dumps(detected_language, sort_keys=True, indent=4, ensure_ascii=False, separators=(",", ": ")))
  
```

- Importation des modules nécessaires**: `import os, requests, json`
- Endpoint et clé d'authentification via les variables d'environnement**: `endpoint = os.environ["TEXT_ANALYTICS_ENDPOINT"]` and `subscription_key = os.environ["TEXT_ANALYTICS_SUBSCRIPTION_KEY"]`
- Endpoint complet**: `path = "/text/analytics/v2.1/languages"` and `detect_api_url = endpoint + path`
- Paramètres d'en-têtes**: `headers = { "Ocp-Apim-Subscription-Key": subscription_key, "Ocp-Apim-Subscription-Region": "westeurope", "Content-type": "application/json" }`
- Corps de la requête et texte à analyser (json)**: `text_body = {"documents": [{"id": "1", "text": "21/7/2012 جي رٻڙ و طبيض لب م لڀ ق 21/7/2013 جي رٻڙ و رٻڙ ال امدہ قطح ي دلڙي ل انل ع ا"}] }`
- Requête POST**: `response = requests.post(detect_api_url, headers=headers, json=text_body)`
- Demande de sortie en json**: `detected_language = response.json()`
- Affichage formaté**: `print(json.dumps(detected_language, sort_keys=True, indent=4, ensure_ascii=False, separators=(",", ": ")))`

DÉMONSTRATION





SYNTHÈSE

2 MÉTHODES, MÊME RÉSULTAT → CHOIX PYTHON

curl://

```
{
  "documents": [
    {
      "id": "1",
      "detectedLanguages": [
        {
          "name": "Arabic",
          "iso6391Name": "ar",
          "score": 1.0
        }
      ]
    }
  ],
  "errors": []
}
```

++:

- Script court et simple
- Directement interprétable sur le prompt

-- :

- Peu documentable
- Format peu présentable



```
{
  "documents": [
    {
      "id": "1",
      "detectedLanguages": [
        {
          "name": "Arabic",
          "iso6391Name": "ar",
          "score": 1.0
        }
      ]
    }
  ],
  "errors": []
}
```

++:

- Langage simple
- Facilement documentable
- Format très lisible

-- :

- Interprétable sur le prompt uniquement si le path est défini
- Requiert un peu plus de maîtrise de l'environnement de travail et du langage de programmation

CONCLUSION

Le modèle pré-entraîné, ici l'API REST « **Text Analytics** » de Microsoft Azure permet de détecter efficacement la langue d'un texte donné.

La requête est **complète** car :

- Elle récupère la prédiction du modèle;
- Elle permet de modifier facilement le texte pour lequel on veut prédire la langue.

La requête est **sécurisée** car la clé de connexion (de même que le chemin de l'endpoint) n'est pas visible dans le script.

La requête est **présentable** car le script de la requête est contenu dans un seul fichier qui permet de lancer en une fois toutes les instructions permettant la prédiction pour une entrée donnée.

The background features a complex network diagram with numerous nodes and connecting lines. The nodes are represented by circles of varying sizes and colors, including dark blue, light blue, and grey. Some nodes are highlighted with larger, concentric circles. The lines are thin and grey, creating a web-like structure across the entire image.

QUESTIONS / RÉPONSES





ANNEXES

COMMENT FONCTIONNE UNE API ?



Explicatif

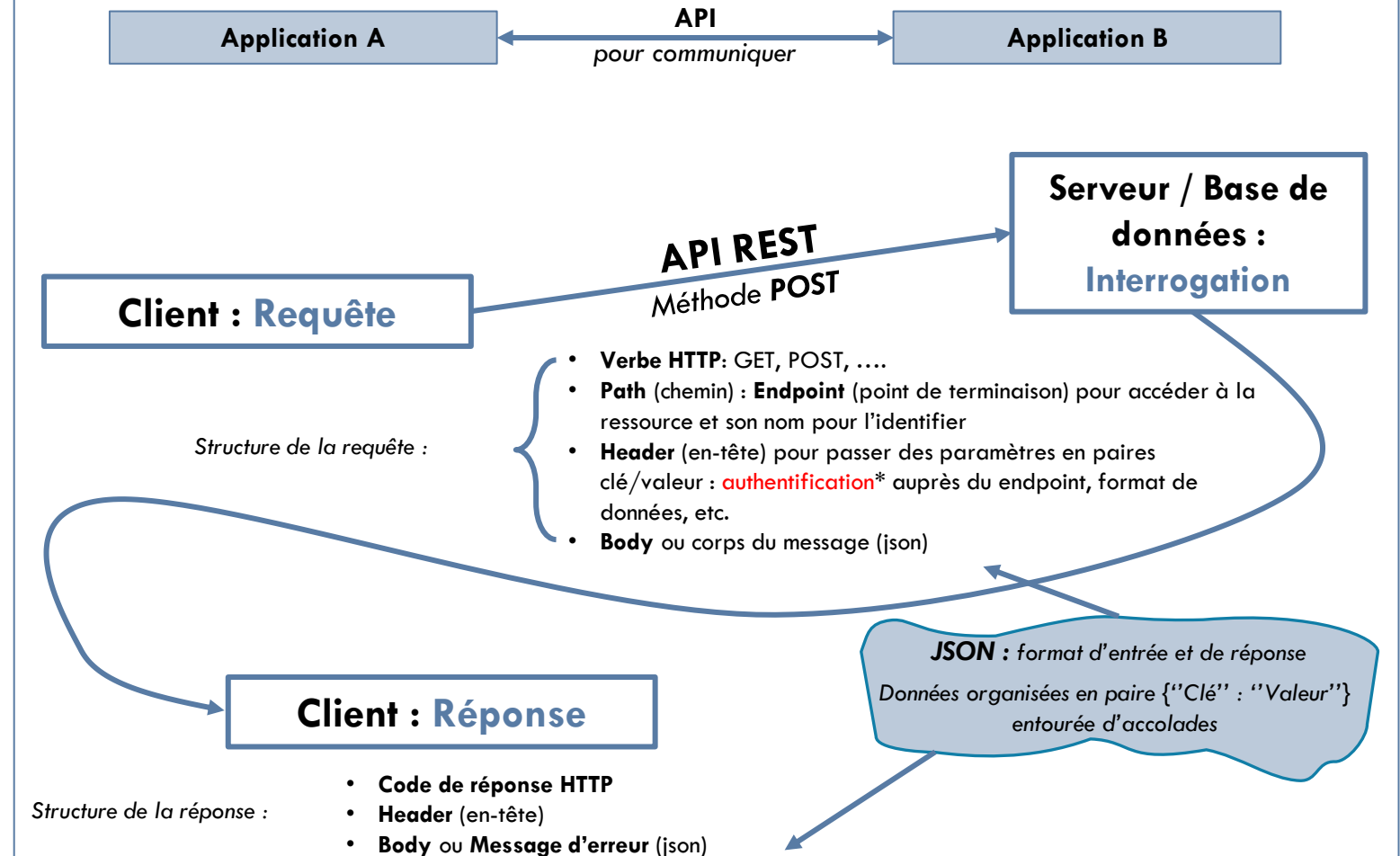
API (Application Programming Interface): interface servant d'intermédiaire et facilitant la communication entre deux applications différentes (Requête -> Réponse), sous un certain format.

API REST (Representational State Transfert): ensemble de normes ou de lignes directrices qui structurent la façon de transférer des données, utilisant le protocole **HTTP** (HyperText Transfert Protocol).

Méthode **POST** du protocole **HTTP**: requête ayant pour action de poster/publier des données dans le corps de la requête (à l'inverse de **GET** qui permet d'obtenir des données).

JSON (JavaScript Object Notation): format de données brutes organisées en paires Clé/Valeur entourées d'accolades, le plus couramment utilisé (avec XML) pour envoyer ou retourner des données. Les tableaux ou listes sont entourés de crochets []

Fonctionnement de l'API



*L'**authentification** est nécessaire pour garantir que seules des personnes avec les autorisations adéquates peuvent accéder à l'API.



VARIABLES D'ENVIRONNEMENT SUR WINDOWS

Utiliser la commande `setx CLE_DE_LA_VARIABLE "VALEUR_DE_LA_VARIABLE"`

Création

```
Invite de commandes
Microsoft Windows [version 10.0.18362.900]
(c) 2019 Microsoft Corporation. Tous droits réservés.

C:\Users\LaFamille>setx TRANSLATOR_TEXT_SUBSCRIPTION_KEY " "

C:\Users\LaFamille>setx TRANSLATOR_TEXT_ENDPOINT "https://api.cognitive.microsofttranslator.com/"

RÉUSSITE : la valeur spécifiée a été enregistrée.
```

Vérification

The screenshot shows the Windows Settings application with the 'System' section open. The 'Advanced system settings' link is selected, leading to the 'System properties' dialog box. In this dialog, the 'Advanced' tab is active, and the 'Environment variables...' button at the bottom is circled in red. To the right, the 'Environment variables' dialog box is open, showing a list of user variables for 'LaFamille'. The variables listed are:

Variable	Valeur
OneDrive	C:\Users\LaFamille\OneDrive
OneDriveConsumer	C:\Users\LaFamille\OneDrive
Path	C:\Users\LaFamille\AppData\Local\Programs\Python\Python37\...
TEMP	C:\Users\LaFamille\AppData\Local\Temp
TEXT_ANALYTICS_ENDPOINT	https://vja-text-analytics.cognitiveservices.azure.com
TEXT_ANALYTICS_SUBSCRI...	[REDACTED]
TMP	C:\Users\LaFamille\AppData\Local\Temp
TRANSLATOR_TEXT_ENDP...	https://api.cognitive.microsofttranslator.com

The 'TEXT_ANALYTICS_ENDPOINT' and 'TEXT_ANALYTICS_SUBSCRI...' variables are circled in red. At the bottom of the dialog, there are buttons for 'Nouvelle...', 'Modifier...', and 'Supprimer'.



Ce document a été produit dans le cadre de la soutenance du projet n°1 du parcours Ingénieur IA d'OpenClassrooms :
« Découvrez le métier d'Ingénieur IA »

Mentor : Thierno DIOP
Evalueur : Thierno DIOP

