

Service cognitif Azure pour la détection de langue

Tester la possibilité d'utiliser les services cognitifs de Microsoft Azure pour identifier la langue d'un message textuel en utilisant un jeu de données open source

➔ Tester si le service cognitif Microsoft Azure peut détecter la langue d'un message à partir du jeu de données [Wikipedia Language Identification Database](#).

Exemple de réponse :

```
{  
  "isTranslationSupported": true,  
  "isTransliterationSupported": true,  
  "language": "hi",  
  "score": 0.93  
},
```

Démonstration de l'utilisation du service cognitif Azure pour la détection de langue depuis le terminal de commande.

Langues	Résultats sur 5 entrées
Anglais (en)	0,98 – 1,0 – 1,0 – 0,99 – 0,98
Chinois Mandarin (zh-Hans)	1,0 – 1,0 – 1,0 – 0,90 – 0,96
Hindi (hi)	0,93 – 1,0 – 1,0 – 1,0 – 0,98
Espagnol (es)	0,98 – 0,96 – 1,00 – 0,99 – 0,95
Arabe (ar)	1,0 – 1,0 – 1,0 – 1,0 – 0,98

Fiabilité ok pour des langues avec un très grand nombre d'entrées dans le modèle (ici les cinq langues les plus parlées dans le monde d'après <https://www.ethnologue.com/guides/ethnologue200>)

Quelques limites :

- Dialectes et langues peu usités
- Fautes d'orthographe et mélange de langues
- Insertion de noms propres (marques, noms, prénoms, noms de société)