

## **Projet individuel : Enoncé et attentes du projet et détails techniques**

**Enoncé :** Dans le cadre des recherches en langue des signes menées par la faculté d'informatique et la faculté de philosophie et lettres de l'UNamur, une base de données représentative de la langue des signes francophone de Belgique (LSFB) a été constituée.

L'interrogation de cette base de données demande des connaissances en SQL que n'ont pas toutes les personnes travaillant sur le projet. Il est aussi compliqué de visualiser de manière intuitive les vidéos de langue des signes capturées.

L'objectif de ce projet est de fournir un dashboard intuitif permettant de naviguer dans la base de données. L'objectif de ce projet serait les suivants :

- **Trouver une manière efficace de visualiser en 3D des signes enregistrés (voir avec Pyplot).**
- **Fournir des statistiques sur la base de données actuelles et les signes collectés.**
- **Permettre d'effectuer des recherches afin de visualiser des statistiques sur une sous-ensemble de la base de données.**

### **Attentes suite à entretien avec le client :**

De manière concrète, voici les recherches qu'il est attendu sur l'interface à concevoir :

- Chercher tous les endroits où un signe apparaît dans la base de données ainsi que des statistiques sur ce signe (ex : on cherche le signe CHIEN on aimerait bien voir combien de fois il apparaît, pouvoir voir les moments des vidéos où ce signe apparaît, savoir combien de temps on met pour le signer et si c'est un signe qui se fait à deux mains ou non).
- Chercher une expression dans les données : Pouvoir voir combien de fois apparaît une séquence de signes dans les données (ex : Combien de fois MANGER BOULETTE apparaît dans le corpus).
- Visualiser des segments de vidéos : pouvoir visualiser une vidéo et voir le squelette associé à cette vidéo si possible en 3D.
- Pouvoir faire des statistiques sur base d'information démographique du signeur (âge, région). L'objectif est de voir s'il y a des mots utilisés par les signeurs plus âgés qui n'apparaissent pas chez les signeurs plus jeunes ou si des expressions sont utilisées dans une région de Belgique et pas dans les autres.

**Ressources :** L'objectif du projet est de créer une interface permettant de visualiser des statistiques relatives à l'utilisation de la langue des signes par les sourds. Pour cela, l'on dispose d'une base de données (appelée corpus) consistant en un ensemble de fichiers video, json et numpy <https://eur02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fnumpy.org%2F&data=05%7C02%7Cloic.kamgain%40student.unamur.be%7Cdce14843464e4a4808e508dced293408%7C5f31c5b4f2e847728dd6f268037b1eca%7C0%7C0%7C638646008421997975%7CUnknown%7CTWFpbGZsb3d8eyJWljoIMC4wLjAwMDAiLCJQIjoiV2luMzliLCJBTiI6Ikl1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=6PPQFD5AFq4Zo5ZxFCIKbbyw6FOUWqWrBn2q0iY8zD0%3D&reserved=0> .

- Vidéo : vidéo d'une personne parlant en langue des signes.
- JSON : des annotations réalisées par les linguistes qui contiennent le minutage du début et de la fin d'un signe dans la vidéo (en ms) ainsi que le nom du signe.
- fichiers numpy : des fichiers contenant les informations de squelettes pour une vidéo donnée.

Un aperçu de ces données est disponible grâce à la librairie **Sign Language Tools** que nous avons développée :

<https://eur02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fpypoitier.github.io%2Fsign-language-tools%2Fvisualization%2F&data=05%7C02%7Cloic.kamgain%40student.unamur.be%7Cdce14843464e4a4808e508dced293408%7C5f31c5b4f2e847728dd6f268037b1eca%7C0%7C0%7C638646008422017939%7CUnknown%7CTWFpbGZsb3d8eyJWljoIMC4wLjAwMDAiLCJQIjoiV2luMzliLCJBTiI6Ikl1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=OKt3hb%2FBetTRsqecpmJEi8iNhqP2m%2Bf6B1bje3Ehg7o%3D&reserved=0>

**Technologies (langages) utilisés dans le projet :** SQL, JavaScript, Pyplot, Python.

**Clients :** [jerome.fink@unamur.be](mailto:jerome.fink@unamur.be), [pierre.poitier@unamur.be](mailto:pierre.poitier@unamur.be).