Compte rendu : Solution Open Source Performante pour la Transcription Automatique de la Parole en Texte(SpeechToText) en Français

Introduction

Dans le cadre du projet "Générateur Automatique de Résumés de Cours et de Notes avec IA", un des composants clés est la transcription de la parole en texte (Speech-to-Text, STT). L'objectif de la première étape est d'identifier la solution open-source la plus performante et adaptée à la langue française. Pour cela, nous avons suivi une approche méthodique en plusieurs étapes : d'abord, la définition des critères de performance, suivie de l'analyse des solutions open-source, puis l'expérimentation et la comparaison, et enfin le choix du modèle final.

1. La définition des critères de performance

Afin de choisir la meilleure solution de reconnaissance vocale, nous avons défini plusieurs critères de performance essentiels. Ces critères nous permettent d'évaluer objectivement chaque modèle et de comparer leurs performances selon nos besoins.

- **Précision :** Taux d'erreurs dans la transcription.
- **Vitesse :** Temps d'exécution sur un même fichier audio.
- **Compatibilité :** Facilité d'intégration avec notre projet.
- **Ressources requises :** GPU, CPU, RAM nécessaires.
- **Support de la langue française :** Optimisation pour le français (accents, vocabulaire technique)**.**

2. Analyse des solutions open-source

Nous avons étudié plusieurs solutions STT disponibles en open-source, notamment :

- **Whisper (OpenAI) : https://github.com/openai/whisper**
- **DeepSpeech (Mozilla): https://github.com/mozilla/DeepSpeech**

3. Expérimentation et comparaison des modèles

Critères	Whisper (OpenAI)	Deepseech (Mozilla)
Précision	Très élevée (supporte le français nativement)	Moyenne (peut être entraîné sur des données personnalisées)
Ressources	GPU recommandé	Fonctionne sur CPU
Mode hors ligne	Simple avec whisper Python	Plus complexe (entraînement nécessaire pour de bons résultats)
Mode hors ligne	Oui	Oui

4. Justification du choix de Whisper

Après une analyse approfondie, **Whisper d'OpenAI** s'avère être la meilleure solution pour notre projet :

- **Précision exceptionnelle**, y compris pour des accents et un langage informel.
- **Support natif du français**, sans nécessité d'entraînement supplémentaire.
- **Facilité d'intégration** avec Python.
- **Disponible en open-source** et peut être utilisé localement sans API.

5. Versions de Whisper

Whisper est disponible en plusieurs versions de tailles différentes, chacune offrant un compromis entre précision et performance computationnelle. Les principales versions sont :

- **Whisper tiny** : Modèle léger, rapide mais moins précis.
- **Whisper base** : Un bon équilibre entre rapidité et précision pour les tâches simples.
- **Whisper small** : Version plus avancée avec de meilleures performances en transcription.
- **Whisper medium** : Modèle plus précis mais nécessitant plus de ressources.
- **Whisper large** : La version la plus performante avec la meilleure précision, au prix d'un temps de traitement plus long.

Dans le cadre de notre projet, nous avons opté pour **Whisper Small**, car il représente un bon compromis entre **qualité de transcription et rapidité d'exécution**. Il offre des performances satisfaisantes tout en restant accessible en termes de ressources informatiques (CPU, RAM), ce qui est essentiel pour garantir une exécution fluide de notre système

6. Intégration et mise en œuvre

Un notebook Jupyter sera fourni avec l’installation et l’utilisation de Whisper pour une démonstration pratique.

Conclusion

Pour notre projet d'application web visant à transcrire des cours en français et à générer des résumés automatiquement, nous avons choisi **OpenAI Whisper**. La raison principale de ce choix est que Whisper supporte très bien le français, ce qui est essentiel pour garantir une transcription précise et de qualité des cours. De plus, sa robustesse face à différents accents et à des environnements sonores variés nous permet d'être confiants quant à sa capacité à traiter des enregistrements audio de qualité variable, typiques des cours.

En comparaison, **Mozilla DeepSpeech**, bien que performant pour l'anglais, est moins adapté aux langues autres que l'anglais, ce qui aurait limité son efficacité pour notre projet. Grâce à sa prise en charge multilingue et à sa précision, Whisper s'avère être la solution la plus adaptée pour notre besoin spécifique de transcrire des cours en français.