Recherche et affichage de documents en temps réel à partir de conversations en parole spontanées



Mots-clés : Reconnaissance de la parole en streaming, conversations multi-locuteurs, compréhension de la parole, traitement automatique des langues, recherche d'information, systèmes temps réel.

Description

Au sein du Pôle Recherche de l'AMIAD (Agence Ministérielle pour l'Intelligence Artificielle de Défense), vous développerez un prototype capable d'écouter une conversation entre plusieurs personnes, d'en identifier le sujet en cours et de rechercher automatiquement des documents pertinents à afficher sur un écran.

L'objectif est d'enrichir les échanges oraux par un accès immédiat à l'information, sans nécessiter de recherche manuelle ni d'interruption de la conversation. Par exemple, lorsque les participants évoquent les déplacements en métro à Paris et le prix des titres de transport, le système pourra afficher un plan du réseau ainsi que des tableaux présentant les abonnements et les tarifs des tickets individuels.

Le prototype reposera sur une architecture en cascade articulant quatre composantes :

1. **Reconnaissance de la parole en temps réel :** transcrire progressivement les échanges à l'aide d'un système de reconnaissance automatique de la parole (ASR) en streaming, adapté aux conversations multi-locuteurs. La diarisation permettra, si nécessaire, d'associer les segments de parole aux différents intervenants.
2. **Compréhension de la parole :** exploiter les transcriptions successives pour identifier le sujet courant, les entités mentionnées et les besoins d'information implicites. Cette composante devra suivre les changements de thème tout en conservant le contexte utile des échanges précédents.
3. **Recherche d'information :** construire et actualiser des requêtes à partir du contexte, retrouver les documents ou passages pertinents dans un jeu de données indexées
4. **Interface graphique :** présenter automatiquement les documents sur un écran, avec leur source, dans une organisation lisible.

Les principaux enjeux concernent le compromis entre latence et pertinence, la propagation des erreurs des différents systèmes, ainsi que la stabilité de l'affichage. Le système devra déterminer quand lancer une nouvelle recherche et quand actualiser les résultats, afin d'éviter les documents hors sujet ou les changements d'écran trop fréquents.

Objectifs

L'objectif principal du stage sera de concevoir, implémenter et évaluer un démonstrateur fonctionnel intégrant ces trois composantes. Le travail s'appuiera prioritairement sur des modèles et des outils existants, adaptés et assemblés pour un fonctionnement en continu.

Les principales étapes seront :

— sélectionner et intégrer une solution d'ASR en streaming, puis évaluer son comportement sur des échanges multi-locuteurs ;
— développer un module de compréhension de conversation et de génération de requêtes ;
— constituer et indexer un corpus documentaire de démonstration, puis mettre en place une recherche lexicale, sémantique (ou hybride) et éventuellement multimodale ;
— concevoir une interface d'affichage et une stratégie de sélection, de maintien et de renouvellement des documents ;
— évaluer le prototype sur des scénarios de conversation représentatifs, en mesurant la pertinence, la latence des résultats, et la stabilité de l'affichage de bout en bout.

Le livrable attendu sera un prototype reproductible (open-source), accompagné de sa documentation et d'une analyse expérimentale des performances et des limites. Le périmètre initial portera sur un corpus maîtrisé ; une extension à des sources externes pourra être envisagée selon l'avancement du stage.

Le stagiaire bénéficiera du soutien de l'équipe de support technique du Pôle Recherche de l'AMIAD, notamment pour les choix technologiques et l'intégration du démonstrateur.

Profil recherché

Vous suivez une formation d'ingénieur ou de master spécialisée en intelligence artificielle, traitement de la parole, traitement automatique des langues ou science des données, et vous disposez :

— de connaissances en apprentissage automatique et profond ;
— de solides compétences en programmation Python et d'une familiarité avec un framework d'IA, idéalement PyTorch ;
— d'un intérêt pour le traitement de la parole, la recherche d'information et le développement de prototypes interactifs ;
— d'autonomie, de créativité et d'un goût pour l'expérimentation ;

— d'une aisance en anglais pour la lecture de publications scientifiques et de documentation technique.

Une première expérience en traitement de la parole, en recherche d'information ou en développement d'applications temps réel serait un atout.

Attention : l'AMIAD, faisant partie du Ministère des Armées et des Anciens Combattants, privilégie fortement les nationalités européennes pour cette offre de stage.

Informations générales

— Durée : 6 mois.
— Niveau : Master 2 ou équivalent en école d'ingénieur.
— Début : dès que possible.
— Lieu : AMIAD – X-Novation Center, École polytechnique, Palaiseau.
— Ressources mises à disposition : serveurs et GPU dédiés, ainsi qu'un accès au supercalculateur Jean Zay.
— Encadrement :
  — Thomas Thebaud : *thomas.thebaud@polytechnique.edu*
  — Christophe Servan : *christophe.servan@polytechnique.edu*