TP Techniques IA 2 - BUT3

Agent conversationnel - IUT-RAG

2024-2025

L'objectif de ce TP est de mettre en oeuvre un agent conversationnel dédié à l'IUT. Il devra intégrer des connaissances spécifiques à ce domaine. Il s'agit d'un TP expérimental dans lequel vous allez utiliser des technologies récentes. Le but est donc d'explorer différentes technologies et de voir leurs performances.

La chaîne de traitement globale à construire doit réaliser la séquence suivante :
1. Enregistrer une requête audio de l'utilisateur
2. Effectuer la reconnaissance automatique de ce qui est prononcé
3. Interroger le système pour obtenir une réponse
4. Générer l'audio qui correspond à la réponse (en utilisant une voix personnalisée)

L'étape 3. est celle qui doit être réalisée en priorité. Pour cela, nous allons utiliser un système de type RAG (Retrieval Augmented Generation). Les étapes 1., 2. et 4. seront réalisées ensuite.

De manière générale, il est possible de construire le prototype en utilisant Google Colab (https://colab.research.google.com/).

1. Agent conversationnel

Le système RAG proposé est LightRAG. Il s'agit d'un système qui utilise un LLM et qui permet d'intégrer une base de connaissances. Sans rentrer dans les détails, cela permet à l'utilisateur d'ajouter des documents à la base de connaissances. Dans notre cas, l'objectif est d'intégrer le contenu des pages web de l'IUT afin de pouvoir répondre à des questions spécifiques sur les formations, la structure de l'IUT, etc.

Il est possible d'utiliser différents LLM pour gérer la génération de texte. Ici, le LLM que l'on utilisera de manière couplée à LightRAG sera Gemma2 (Gemma2:2b ou Gemma2:9b).

Les liens donnés ci-après permettent de mettre en place LightRAG. Pour construire la base de connaissances, plusieurs solutions sont possibles : (1) extraction des informations depuis les pages de l'IUT (crawling) ou plus simplement pour démarrer (2) écrire d'un ensemble de textes (phrases) décrivant les connaissances à intégrer.

Liens utiles :
- https://github.com/HKUDS/LightRAG
- hhttps://huggingface.co/google/gemma-2-2b

- hhttps://huggingface.co/google/gemma-2-9b
- https://stable-learn.com/en/lightrag-introduction/

2. Interaction Orale avec l'utilisateur

Afin de rendre le système d'interaction plus naturel, nous souhaitons intégrer la modalité orale pour exprimer des requêtes et avoir le résultat. À cette fin, il est nécessaire d'ajouter une brique de reconnaissance automatique de la parole et une brique de synthèse de la parole à partir du texte :

- Whisper: il s'agit d'un modèle de reconnaissance automatique de la parole. Plusieurs versions sont possibles. Chacune offrant un niveau de précision et un niveau de performances différent. À vous de choisir le modèle le plus adapté et celui que vous pouvez faire fonctionner.
- XTTS-v2 : il s'agit d'un modèle de clonage de voix. Grâce à un échantillon de voix (une sorte d'empreinte vocale), il est possible de reproduire la voix de l'échantillon pour des textes différents.

Liens utiles :

- https://github.com/openai/whisper
- https://huggingface.co/coqui/XTTS-v2

3. Rendus et évaluation

L'objectif est d'expérimenter et de voir quels résultats on peut obtenir en utilisant de tels systèmes. Notamment, vous pourrez comparer les réponses entre un LLM classique et le LLM (LightRAG) que vous aurez mis en place. Vous pourrez utiliser différentes configurations et voir l'impact sur le résultat final (qualité du moteur de reconnaissance de parole, quantité de connaissances injectée dans le RAG, qualité de la voix de sortie, etc.)

Le rendu attendu pour ce TP est le notebook de votre code accompagné d'un compte-rendu de quelques pages décrivant vos expériences et vos résultats.