
12 SEPTEMBRE 202024 / 09:00 / Visio

ORDRE DU JOUR

Première réunion

1. Présentation des membres du groupes /Présentation du sujet
2. Questions – Réponses
3. Premières directives

NOTES

Données :

- Explorer les adresses mail/websites, même si elles ne sont pas toujours directement associées aux lieux (possiblement liées à des collectifs). Scraper cette information si nécessaire(À compléter plus tard, ne pas prioriser pour le moment.).
- Focus immédiat sur le nettoyage des données et l'extraction des textes existants.
- Bien que le volume de données soit faible, il est fortement connecté. Concentrez-vous sur ce que nous avons déjà en main pour avancer.

Bibliographie :

- Récupérer le code de Marathon Web (actuellement peu propre).

Visualisations :

- Intégrer D3 dans le pipeline visuel. Il y a une différence entre ce qui est fait en Python et ce qui est nécessaire pour les visualisations. Cela requiert une coordination entre les outils de traitement (Python, deep learning) et la mise en forme des données pour la visualisation.

- Site web pour les visualisations explicatives : à prévoir, mais pas une urgence immédiate.

LLM (Large Language Models) :

- Formation nécessaire. Participer à la présentation des M2 sur les LLM (tutoriels).
- Demander à M. Pasquet des précisions sur l'accès à des GPU.

Support du professeur :

- Chapitres 1 et 2 (Intro) ; importance des chapitres 3 et 4
<https://formation-cnrs-ml-dl-2023.pages.math.cnrs.fr/intro-ml-dl/>

Hugging Face :

- Travailler sur des notebooks en ligne, explorer des modèles pas trop complexes. https://huggingface.co/docs/transformers/en/model_doc/bert

BERTopic :

- Peut-être à utiliser plus tard. Regarder le code, essayer des petites variations pour s'entraîner et mieux comprendre l'outil.

BERT :

- Ne pas prioriser pour l'instant. Les métriques et la fonction de coût sont des questions importantes à aborder plus tard.

Point clé :

Fournir à l'expert un outil capable d'identifier des classes d'espaces, enrichir les textes pour proposer des types spécifiques, par exemple, des tests synthétiques sur des périodes ou lieux particuliers.

TÂCHES

1. Extraction
2. Nettoyage
3. Raccord entre les entrées et les réponses

ORDRE DU JOUR DE LA SEMAINE PROCHAINE

Rendez-vous :

Paul va réserver un rendez-vous à 8h30, vendredi prochain, avec Sandra.

Objectifs :

- Poser des questions plus spécifiques à Sandra
 - faire un premier point sur l'avancement des tâches.