

compte rendu

mars 2024

1 Objectifs

- Développer une solution innovante pour étendre la collecte d'informations sur les produits de Made In Tracker
- Concevoir une technologie de parsing intelligent automatisant des processus clés
- Répondre aux besoins de l'entreprise en renforçant sa crédibilité et en offrant une solution évolutive
- Jeter les bases pour des avancées dans le domaine du parsing intelligent

2 Méthodologie

- Nous avons décidé d'utiliser la bibliothèque Spacy pour extraire les domaines de chaque produit, afin d'obtenir une meilleure compréhension de la structure de nos données.
- Malgré nos efforts, nous n'avons pas encore réussi à entraîner un modèle satisfaisant pour répondre à nos besoins spécifiques, mais cette expérience nous a permis de mieux cerner les défis et les limites de notre approche.
- En parallèle, nous avons développé un code dédié au prétraitement des données, permettant d'automatiser l'organisation d'un fichier sans nécessiter une saisie manuelle ligne par ligne.
- Cette automatisation s'est avérée cruciale pour maintenir notre avancée malgré les difficultés rencontrées dans l'entraînement du modèle.

3 Difficultés Rencontrées

- La bibliothèque Spacy ne réussit pas toujours à bien classer chaque produit.
- Parfois, nous avons rencontré des difficultés à comprendre les modèles de Langage de Modèles de Longueurs (LLM) ainsi que les méthodes de fine-tuning associées.
- Il n'y a pas suffisamment d'API gratuites avec des fichiers JSON adaptés à

nos besoins.

- Nous ne savons pas encore comment sauvegarder un modèle pour l'utiliser dans un autre notebook.
- Le code que nous avons développé pour l'extraction des informations à partir des API ne fonctionne actuellement qu'avec Open Food Facts.

4 Réalisations

Voici la section des réalisations basée sur nos échanges :

- Nous avons réussi à utiliser la bibliothèque Spacy pour extraire les domaines de chaque produit, ce qui nous a fourni une compréhension approfondie de la structure de nos données.
- Bien que nous n'ayons pas encore atteint notre objectif d'entraîner un modèle satisfaisant, nous avons acquis une meilleure compréhension des défis liés à cette tâche, ce qui nous guide vers de nouvelles approches à explorer.
- Nous avons développé un code de prétraitement des données qui automatise l'organisation des fichiers, facilitant ainsi la manipulation de nos données sans avoir besoin de saisie manuelle ligne par ligne.
- Malgré les obstacles rencontrés, nous avons maintenu notre avancée grâce à cette automatisation, ce qui nous a permis de progresser même en l'absence de résultats concrets dans l'entraînement du modèle.

5 Bilan

- Malgré les défis rencontrés, notre utilisation de Spacy a renforcé notre compréhension des données.
- Bien que le modèle n'ait pas été entraîné avec succès, nous avons développé des compétences précieuses en prétraitement des données.
- Notre automatisation du prétraitement a été une réussite, facilitant la manipulation des données malgré les difficultés rencontrées dans l'entraînement du modèle.