

Test de traduction de document

Ceci est un document de test créé pour vérifier la fonctionnalité de traduction du pipeline de traitement de document intégré PyMuPDF-YOLO. Le document contient divers types de contenu qui doivent être extraits et traduits correctement.

Contenu pour texte

Section 1 : Aperçu technique

Le système de traitement de documents amélioré combine PyMuPDF pour une extraction de texte haute fidélité avec YOLO pour l'analyse de la mise en page. Cette intégration fournit un mappage intelligent du contenu et un routage du traitement basé sur les caractéristiques du document. Les documents purement textuels utilisent un traitement rapide uniquement avec PyMuPDF, tandis que les documents à contenu mixte utilisent une extraction basée sur les coordonnées.

Section 2 : Fonctionnalités de traduction

Le service de traduction prend en charge plusieurs langues et inclut des fonctionnalités avancées telles que la traduction contextuelle, la gestion du glossaire et le traitement compatible Markdown. Le système préserve la structure et la mise en forme du document tout en fournissant des traductions précises.

Fonctionnalités clés :

- Extraction de texte haute fidélité à l'aide de PyMuPDF
- Analyse de la mise en page avec les réseaux neuronaux YOLO
- Routage intelligent de la stratégie de traitement
- Services de traduction contextuels
- Préservation de la structure Markdown
- Génération de sortie multi-format

Conclusion

Ce document démontre les capacités du système de traduction intégré. Une fois traité, il doit générer une sortie traduite dans la langue cible tout en préservant la structure et la mise en forme d'origine. Le système doit créer un document Word, un rapport de traitement et extraire tous les éléments non textuels dans un dossier d'images séparé.