



Développement d'une Application Web Interactive et Intelligente pour le Guide du Code du Travail Marocain

Travail réalisé par EL MAKHCHOUNE FAROUK

15/6/2024

date de remise

Table des matières

1	Resume	3
2	Introduction	4
2.1	Contexte et Importance du Projet	4
2.2	Objectifs et Motivations	4
2.3	Aperçu des Technologies Utilisées	4
3	Projet	6
3.1	Description de l'architecture générale de l'application	6
3.2	Fonctionnement	6
3.2.1	PyMuPDF	6
3.2.2	RecursiveCharacterTextSplitter	7
3.2.3	Jina Embeddings	7
3.2.4	FAISS (Facebook AI Similarity Search)	8
3.2.5	RetrievalQA	8
3.2.6	Mistral AI	9
4	Resultat	10
5	Conclusion	11

1 Resume

Le projet vise à développer une application web interactive et intelligente dédiée au Guide du Code du Travail Marocain, destinée à simplifier l'accès à l'information juridique pour les professionnels du droit, les entreprises et le grand public. À travers l'utilisation de technologies avancées, l'application permet aux utilisateurs de poser des questions spécifiques sur le Code du Travail et d'obtenir des réponses instantanées et contextuelles. En segmentant les documents PDF du Code du Travail en unités de traitement, l'application assure une recherche rapide et précise des informations nécessaires. L'interface utilisateur intuitive, construite avec Streamlit, offre une expérience utilisateur fluide et accessible, facilitant ainsi la navigation à travers les réglementations complexes du travail au Maroc. Ce projet combine expertise juridique et innovation technologique pour répondre aux besoins croissants de compréhension et d'application du Code du Travail Marocain dans un environnement numérique moderne et interactif.

2 Introduction

2.1 Contexte et Importance du Projet

Le Code du Travail Marocain régit les relations de travail entre employeurs et employés, définissant les droits et les obligations des deux parties. En tant que document juridique fondamental, il joue un rôle crucial dans la protection des travailleurs et la régulation du marché du travail. Cependant, sa complexité et son volume peuvent rendre son interprétation et son application difficiles, en particulier pour ceux qui ne possèdent pas une expertise juridique approfondie. Ce projet vise à combler cette lacune en développant une application web interactive et intelligente qui facilite l'accès et la compréhension du Code du Travail Marocain. En rendant l'information juridique plus accessible, ce projet contribue à la promotion de la justice sociale et à l'amélioration des relations de travail au Maroc.

2.2 Objectifs et Motivations

L'objectif principal de ce projet est de créer une application web intuitive qui permet aux utilisateurs de poser des questions spécifiques sur le Code du Travail Marocain et de recevoir des réponses précises et contextuelles en temps réel. Les motivations derrière ce projet incluent :

- Accessibilité : Rendre le Code du Travail Marocain facilement accessible à un large public, incluant les employeurs, les employés, les avocats, et les étudiants.
- Clarté : Simplifier la compréhension des textes juridiques complexes en fournissant des réponses claires et précises.
- Efficacité : Améliorer la rapidité et l'efficacité avec laquelle les utilisateurs peuvent trouver des informations pertinentes dans le Code du Travail.
- Innovation : Utiliser des technologies avancées de traitement du langage naturel et d'intelligence artificielle pour moderniser l'accès à l'information juridique.

2.3 Aperçu des Technologies Utilisées

Pour atteindre ces objectifs, le projet s'appuie sur une combinaison de technologies avancées et de méthodologies éprouvées :

- LangChain : Utilisé pour structurer les interactions entre différents composants de traitement du langage naturel, facilitant ainsi la création de chaînes de traitement complexes.
- Jina Embeddings : Utilisé pour générer des embeddings à partir du texte, permettant une représentation vectorielle efficace des segments du Code du Travail pour des recherches ultérieures.
- LLM Français (Mistralai) : Un modèle de langage avancé en français, utilisé pour la génération de réponses précises et contextuelles aux questions posées par les utilisateurs.
- PyMuPDF : Utilisé pour le chargement et l'extraction de texte à partir des documents PDF du Code du Travail, permettant de traiter les documents sources de manière efficace.
- FAISS (Facebook AI Similarity Search) : Utilisé pour créer une base de données vectorielle permettant une recherche rapide et efficace des informations pertinentes dans les documents.

PDF du Code du Travail.

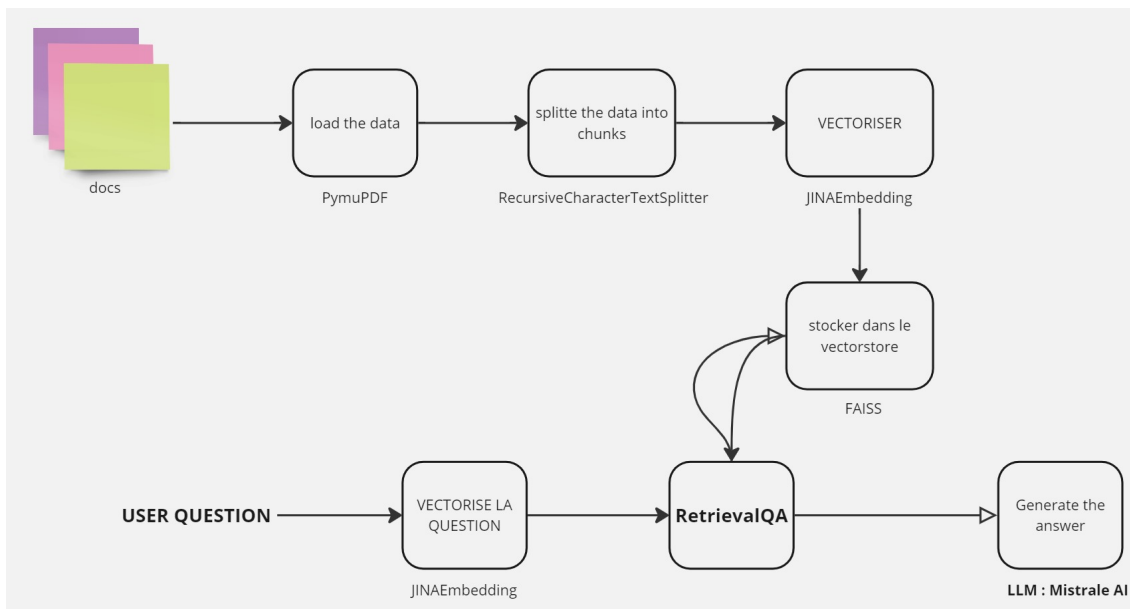
- Streamlit : Une bibliothèque de développement web permettant de créer des interfaces utilisateur interactives et intuitives. Streamlit est utilisé pour construire l'interface frontale de l'application, offrant ainsi une expérience utilisateur fluide.

En combinant ces technologies, le projet vise à offrir une solution complète et innovante pour l'accès et la compréhension du Code du Travail Marocain, répondant ainsi à un besoin crucial dans le domaine juridique au Maroc.

3 Projet

3.1 Description de l'architecture générale de l'application

L'application web interactive pour le Guide du Code du Travail Marocain est conçue pour simplifier l'accès à l'information juridique. Le processus commence par le chargement du document PDF du Code du Travail, qui est ensuite segmenté en unités de texte plus petites et gérables. Ces segments de texte sont stockés dans une base de données optimisée pour permettre des recherches rapides. Lorsqu'un utilisateur pose une question via l'interface de l'application, la question est convertie en une forme qui peut être comparée aux segments de texte stockés. Le système recherche alors les segments les plus pertinents pour la question posée. Sur la base de ces segments trouvés, l'application génère une réponse claire et contextuelle, offrant ainsi une assistance précise et rapide sur les aspects du Code du Travail Marocain.



3.2 Fonctionnement

3.2.1 PyMuPDF



Définition : PyMuPDF, aussi connu sous le nom de Fitz, est une bibliothèque Python utilisée pour accéder et manipuler les documents PDF. Elle permet de lire, écrire et modifier des fichiers PDF de manière efficace.

Rôle dans le Projet : Dans ce projet, PyMuPDF est utilisé pour charger et extraire le texte des documents PDF du Code du Travail Marocain. Le texte extrait est ensuite segmenté en portions gérables pour un traitement ultérieur. PyMuPDF permet une extraction précise et rapide du contenu textuel, facilitant ainsi les étapes suivantes du processus.

Afin de réaliser une analyse comparative, veuillez consulter le lien suivant : [PyMuPDF](#)

3.2.2 RecursiveCharacterTextSplitter

Définition : RecursiveCharacterTextSplitter est une technique ou une bibliothèque utilisée pour diviser un texte en segments plus petits et plus gérables, en se basant sur des caractères spécifiques ou des règles de segmentation.

Rôle dans le Projet : Après l'extraction du texte avec PyMuPDF, le RecursiveCharacterTextSplitter est utilisé pour diviser ce texte en chunks. Ces chunks sont des unités de traitement plus petites qui rendent la recherche et l'analyse plus efficaces. La segmentation se fait de manière récursive pour s'assurer que chaque chunk est de taille appropriée pour les traitements ultérieurs.

Pour plus d'information veuillez consulter le lien suivant : [RecursiveCharacterTextSplitter](#)

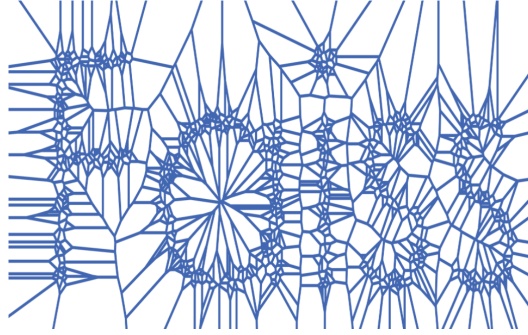
3.2.3 Jina Embeddings

Définition : Jina Embeddings est une technologie qui transforme des segments de texte en représentations vectorielles, appelées embeddings. Ces vecteurs capturent le sens sémantique du texte et permettent une comparaison efficace entre différents morceaux de texte.

Rôle dans le Projet : Dans ce projet, Jina Embeddings est utilisé pour convertir chaque chunk de texte segmenté en un vecteur numérique. Ces vecteurs sont ensuite utilisés pour indexer le texte dans une base de données vectorielle et pour effectuer des recherches de similarité.

Afin de réaliser une analyse comparative, veuillez consulter le lien suivant : [Jina Embeddings Paper](#)

3.2.4 FAISS (Facebook AI Similarity Search)



Définition : FAISS est une bibliothèque développée par Facebook AI qui permet une recherche rapide et efficace de similarité entre vecteurs. Elle est optimisée pour des opérations à grande échelle et offre des algorithmes pour l'indexation et la recherche de vecteurs.

Rôle dans le Projet : FAISS est utilisé pour indexer les embeddings générés par Jina et pour effectuer des recherches de similarité. Lorsqu'un utilisateur pose une question, l'embedding de la question est comparé aux embeddings indexés pour trouver les segments de texte les plus pertinents.

Voici un comparatif simple entre FAISS et Chroma afin d'évaluer la performance de FAISS dans la recherche sur de multiples documents.

Table 2: Comparing Vector Store FAISS vs. Chroma Retrieving 3 Documents

	context_precision	context_recall	f_measure	search_time [second]
FAISS	0.97	0.97	0.97	1.87
Chroma	0.72	0.73	0.73	2.18

Table 3: Comparing Vector Store FAISS vs. Chroma Retrieving 6 Documents

	context_precision	context_recall	f_measure	search_time [second]
FAISS	0.93	0.98	0.95	1.84
Chroma	0.76	0.77	0.76	2.33

Afin de réaliser une analyse comparative, veuillez consulter le lien suivant : [Faiss](#)

3.2.5 RetrievalQA

Définition : RetrievalQA est une technique ou un module qui combine des capacités de recherche (retrieval) avec des modèles de question-réponse (QA). Il permet de récupérer des informations pertinentes et de générer des réponses précises basées sur ces informations.

Rôle dans le Projet : Dans ce projet, RetrievalQA orchestre la recherche de similarité et la génération de réponses. Après que FAISS a identifié les segments de texte pertinents, RetrievalQA utilise ces segments pour générer une réponse contextuelle et précise à la question posée par l'utilisateur.

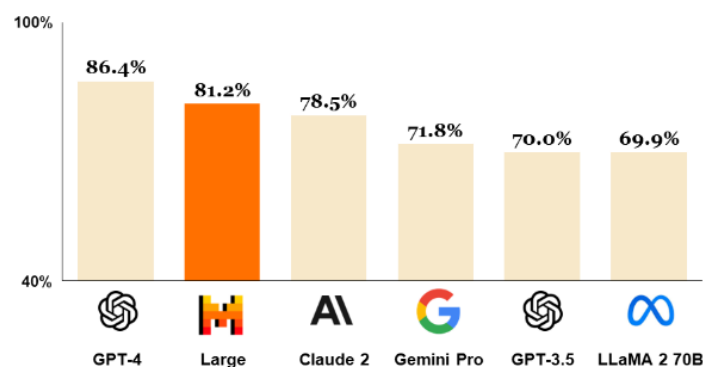
3.2.6 Mistrale AI



Définition : Mistral AI est un modèle de langage avancé en français, capable de comprendre et de générer du texte en langue française de manière fluide et contextuelle. Il est conçu pour traiter des requêtes complexes et fournir des réponses précises.

Rôle dans le Projet : Dans ce projet, Mistral AI est utilisé pour générer des réponses basées sur les segments de texte trouvés par FAISS. Il prend en compte le contexte et la précision des segments pour formuler des réponses compréhensibles et utiles pour l'utilisateur. Voici un comparatif simple

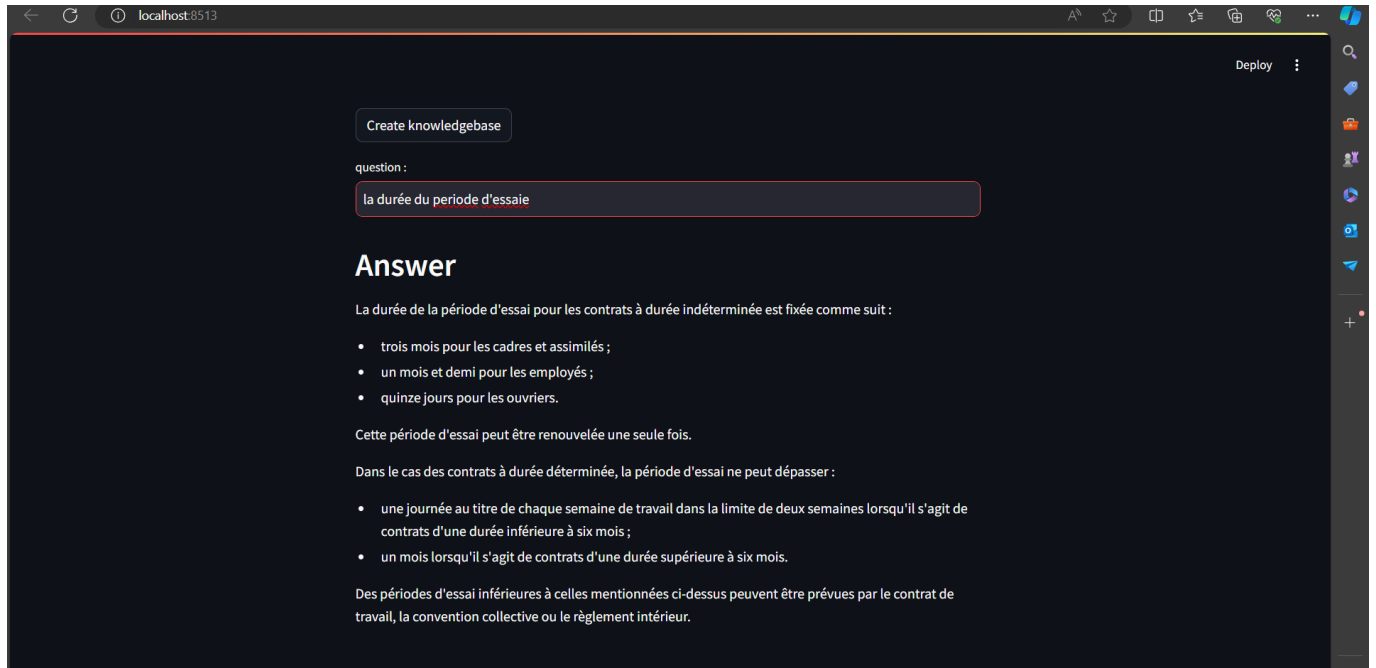
entre Mistral AI et d'autre LLMs



Afin de réaliser une analyse comparative, veuillez consulter le lien suivant : [Mistral AI](#)

4 Resultat

J'ai récemment lancé et testé avec succès mon application dédiée au Guide du Code du Travail Marocain. En utilisant l'interface intuitive que j'ai développée avec Streamlit, j'ai posé une question spécifique sur la durée de la période d'essai prévue par la loi. Grâce à l'intégration de technologies avancées comme Jina Embeddings pour la recherche de similarité et Mistral AI pour la génération de réponses, l'application a rapidement identifié et présenté une réponse claire et précise. Cette expérience démontre l'efficacité et la fonctionnalité de l'application dans la fourniture d'informations juridiques pertinentes et accessibles aux utilisateurs.



5 Conclusion

Ce projet a démontré la conception et la mise en œuvre d’une application web interactive et intelligente dédiée au Guide du Code du Travail Marocain. En utilisant une combinaison de technologies avancées telles que PyMuPDF pour l’extraction de texte, le RecursiveCharacterTextSplitter pour la segmentation du texte, Jina Embeddings pour la création de vecteurs sémantiques, FAISS pour la recherche vectorielle, et Mistral AI pour la génération de réponses, l’application offre une solution robuste pour accéder et comprendre les réglementations complexes du Code du Travail. Chaque composant joue un rôle crucial dans l’architecture globale, assurant une performance optimale et une expérience utilisateur fluide. L’application a le potentiel de transformer la manière dont les

professionnels du droit, les entreprises et le grand public interagissent avec le Code du Travail Marocain. En facilitant l’accès à des informations précises et contextuelles, elle peut considérablement réduire le temps et les efforts nécessaires pour trouver des réponses aux questions juridiques. Cela peut conduire à une meilleure compréhension et application des lois du travail, améliorant ainsi la conformité et la prise de décision informée.

felmakhchoune@gmail.com
[https ://github.com/Faroukelmak](https://github.com/Faroukelmak)