

T-ESP-700

Big Brain

Lucas FIXARI
Pierre ROCHETTE
William WOZIWODA

IA d'Analyse Documentaire

1. Description du Projet	2
2. Contexte du Projet	2
3. Objectifs du Projet	2
Objectifs principaux	3
Objectifs techniques	3
4. Structure de Décomposition des Produits (PBS)	3
Niveau 1 : Solution IA d'Analyse Documentaire Locale	3
Niveau 2 : Modules Principaux	3
Niveau 3 : Tâches Spécifiques	4
5. Description des Technologies	4
6. Plan de Gestion du Projet	4
7. Prototypage et Tests	4
Prototypes à Développer	4
Plan de Test	4

1. Description du Projet

Titre : IA d'Analyse Documentaire Locale

Description: Ce projet est une IA qui lit, organise et recherche des informations dans différents documents (PDF, Word, Excel, images). Elle est pensée pour être utilisée en toute confidentialité, en répondant à des questions simples en langage naturel (comme dans une conversation). Cette IA donne des réponses claires avec des liens vers les documents d'origine.

2. Contexte du Projet

Avec de plus en plus de documents en format numérique, il devient difficile de retrouver des informations précises. Cette IA aide les gens (entreprises, administrations, ou particuliers) à trouver facilement ce dont ils ont besoin, sans connaître précisément le document ou son emplacement.

Exemple : Vous voulez retrouver une facture d'il y a 3 ans sans fouiller dans vos dossiers. Cette IA vous aide en recherchant pour vous.

3. Objectifs du Projet

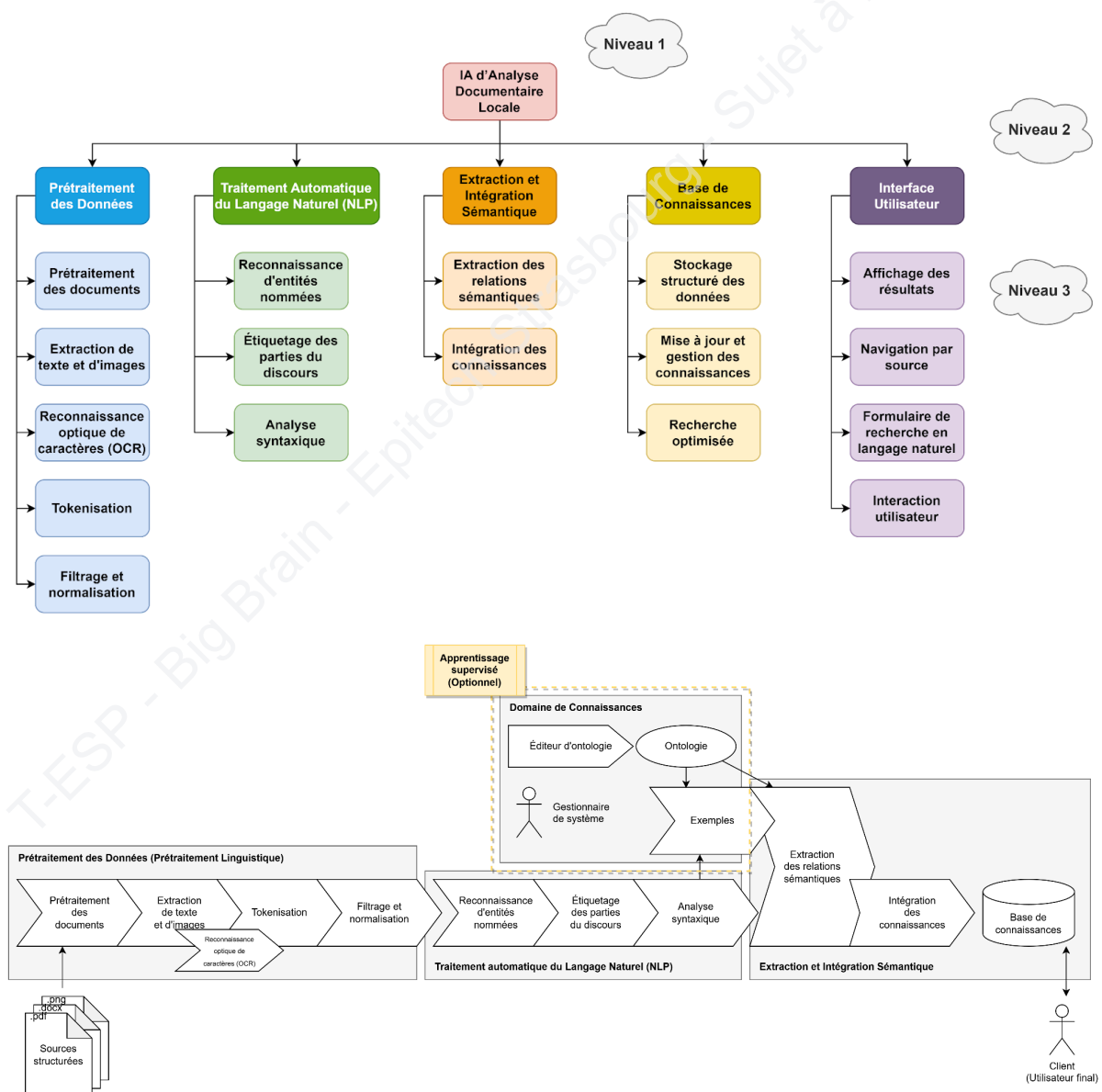
Objectifs principaux

- **Confidentialité et Sécurité** : Toute la recherche se fait localement, donc aucune donnée n'est partagée en ligne.
- **Facilité d'Utilisation** : Permet de poser des questions simples pour obtenir des réponses rapides.
- **Efficacité et Précision** : Donne des réponses claires avec des résumés et les sources (documents) associées.

Objectifs techniques

- Développer l'IA en Python pour analyser et organiser les documents.
- Créer une interface simple pour poser des questions et voir les résultats.
- Optimiser l'IA pour qu'elle soit rapide à utiliser localement.

4. Structure de Décomposition des Produits (PBS)



Niveau 1 : Solution IA d'Analyse Documentaire Locale

- Développement d'une IA qui analyse les documents pour trouver des informations rapidement.

Niveau 2 : Modules Principaux

1. **Prétraitement des Données** : Prépare les documents pour qu'ils soient bien organisés.
2. **Traitement Automatique du Langage Naturel (NLP)** : Comprend les documents en extrayant des informations clés.
3. **Extraction et Intégration Sémantique** : Organise les relations entre les informations extraites pour faciliter la recherche.
4. **Base de Connaissances** : Stocke les informations pour des recherches rapides.
5. **Interface Utilisateur** : Une interface simple pour poser des questions et voir les réponses.

Niveau 3 : Tâches Spécifiques

1. Préparer et nettoyer les documents.
2. Extraire le texte et les images des documents (PDF, Word, Excel).
3. Organiser les informations pour qu'elles soient faciles à retrouver.

5. Description des Technologies

Chaque module utilise des technologies spécifiques pour réaliser ses fonctions :

1. **Prétraitement des Données** : Techniques pour nettoyer et organiser les textes (utilisation de Python, SpaCy, etc.)
2. **Traitement du Langage Naturel (NLP)** : Comprendre les documents et les questions posées (utilisation de SpaCy, transformers).
3. **Base de Connaissances** : Utilise une base de données locale pour stocker les informations (comme SQLite).
4. **Interface Utilisateur** : Affichage des résultats (Flask, FastAPI).

6. Plan de Gestion du Projet

1. **Planification Initiale** : Définir le temps et les ressources nécessaires.
2. **Contrôle Qualité** : Vérifier que les résultats sont précis.
3. **Gestion des Risques** : Identifier les problèmes possibles et trouver des solutions.
4. **Suivi et Reporting** : Rapports mensuels sur l'avancement, les défis rencontrés, et les améliorations possibles.

7. Prototypage et Tests

Prototypes à Développer

- Test de l'extraction et de l'organisation du texte.
- Test de l'interface pour s'assurer qu'elle est facile à utiliser.

Plan de Test

- **Qualité des Résultats** : Mesurer la précision des réponses.
- **Performance** : Vérifier la rapidité de l'IA.
- **Expérience Utilisateur** : Tester l'interface pour qu'elle soit intuitive et agréable à utiliser.