

Récapitulatif du Projet - Base de Connaissances CUS

Tâches Accomplies

1. Extraction du Document Word ✓

- **Fichier source:** `/home/ubuntu/Uploads/CUS-documents.docx`
- **Contenu extrait:** 170 paragraphes incluant :
 - Mission, Vision, Objectifs
 - Thématiques de recherche
 - Outcome (Résultats attendus)
 - Liste complète de l'équipe (15 membres)
 - Contexte mondial et défis

2. Extraction du Logo ✓

- **Format:** PNG
- **Dimensions:** 645 x 175 pixels
- **Taille:** 111 KB
- **Emplacement:** `/home/ubuntu/cus_logo.png`

3. Scraping du Site Web ✓

- **URL:** `https://cus.um6p.ma`
- **Sections explorées:**
 - Page d'accueil et About
 - Équipe (Team)
 - Projets de recherche (5 projets identifiés)
 - Publications (4 publications récentes + collection complète)
 - Collaboration et partenariats
 - Éducation et Formation

4. Consolidation des Données ✓

- **Fichier final:** `/home/ubuntu/cus_knowledge_base.md`
- **Taille:** 18 KB (~17 000 caractères)
- **Format:** Markdown structuré et optimisé pour RAG

Structure de la Base de Connaissances

La base de connaissances consolidée comprend **13 sections principales** :

1. Contact et Informations Générales

- Adresse, email, site web

2. **Contexte Mondial et Défis**

- Urbanisation en Afrique (860M nouveaux citoyens)
- Révolution numérique urbaine
- Paradoxes et défis actuels

3. **Mission**

- 4 piliers : Recherche, Éducation, Expertise, Innovation
- Rôle des Living Labs

4. **Vision**

- Objectifs stratégiques
- Approche intégrée
- Conditions de réussite

5. **Objectifs**

- Objectifs généraux (3)
- Objectifs spécifiques (3)

6. **Thématiques de Recherche**

- 9 thématiques principales
- Digital development, Smart cities, Mobilité, Énergie, etc.

7. **Outcome (Résultats Attendus)**

- 7 résultats clés attendus
- Leadership continental
- Innovation mondiale

8. **Équipe**

- 15 membres détaillés
- Profils complets avec spécialisations

9. **Projets de Recherche**

- 5 projets en cours
- Descriptions et focus

10. **Publications Scientifiques**

- 4 publications récentes (2025)
- Collection complète disponible

11. **Éducation et Formation**

- Approche pédagogique
- Programmes : Masters, Bachelors, MOOCs
- Executive Master's

12. **Collaboration et Partenariats**

- Initiatives récentes
- Masterclass pour maires
- Opportunités de recrutement

13. **Points Forts et Expertise**

- Approche unique

- Domaines d'expertise (8 domaines)
- Public cible

Optimisations pour le Système RAG

Caractéristiques Favorables

- ✓ **Structure hiérarchique claire** avec titres H2, H3, H4
- ✓ **Formatage riche** : listes, tableaux, emphases
- ✓ **Mots-clés stratégiques** identifiés et intégrés
- ✓ **Contenu bilingue** (Français prioritaire + termes anglais techniques)
- ✓ **Sections thématiques** bien délimitées pour le chunking
- ✓ **Métadonnées incluses** : dates, auteurs, contacts
- ✓ **FAQ potentielles** anticipées
- ✓ **Concepts clés** identifiés pour l'indexation

Densité de l'Information

- **17 292 caractères** de contenu riche
- **15 profils d'équipe** détaillés
- **9 thématiques** de recherche
- **5 projets** documentés
- **4 publications** récentes référencées
- **7 résultats attendus** clairement définis

Livrables Finaux

1. Base de Connaissances

```
Fichier : /home/ubuntu/cus_knowledge_base.md
Format : Markdown
Taille : 18 KB
Encodage: UTF-8
```

Contenu:

- Documentation complète du CUS
- Structure optimisée pour RAG
- Bilingue (FR/EN)
- Métadonnées riches

2. Logo

```
Fichier : /home/ubuntu/cus_logo.png
Format : PNG (8-bit RGB)
Taille : 111 KB
Résolution: 645 x 175 pixels
```

Recommandations pour l'Implémentation RAG

1. Chunking Stratégique

- Utiliser les sections H2/H3 comme limites de chunks
- Maintenir le contexte des titres dans chaque chunk
- Taille recommandée : 500-1000 tokens par chunk

2. Indexation

- Créer des embeddings pour chaque section
- Indexer séparément : équipe, projets, publications
- Utiliser les mots-clés fournis pour l'enrichissement

3. Métadonnées à Extraire

- Noms des chercheurs
- Spécialisations
- Dates de publication
- Titres de projets
- Thématiques de recherche

4. Requêtes Typiques à Optimiser

- "Qui travaille sur [thématique]?"
- "Quels sont les projets en cours?"
- "Quelle est la mission du CUS?"
- "Comment rejoindre l'équipe?"
- "Quelles sont les publications récentes?"

5. Sources Multiples

- Combiner les informations du Word + Web
- Mentionner les sources quand pertinent
- Mettre à jour régulièrement depuis le site web

Points Forts de la Livraison

- ✓ **Complétude** : Toutes les informations disponibles ont été capturées
 - ✓ **Structure** : Organisation logique et hiérarchique
 - ✓ **Richesse** : Détails complets sur l'équipe, projets, publications
 - ✓ **Bilingue** : Français (principal) + termes anglais techniques
 - ✓ **Optimisé RAG** : Format et structure adaptés pour l'indexation
 - ✓ **Maintenable** : Facile à mettre à jour et enrichir
-

Informations de Contact CUS

Email: contact.cus@um6p.ma

Site web: <https://cus.um6p.ma>

Adresse: Lot 660, Hay Moulay Rachid, Ben Guerir 43150, Maroc

Document généré le 16 octobre 2025

Projet : Base de connaissances pour Chatbot IA - Centre CUS