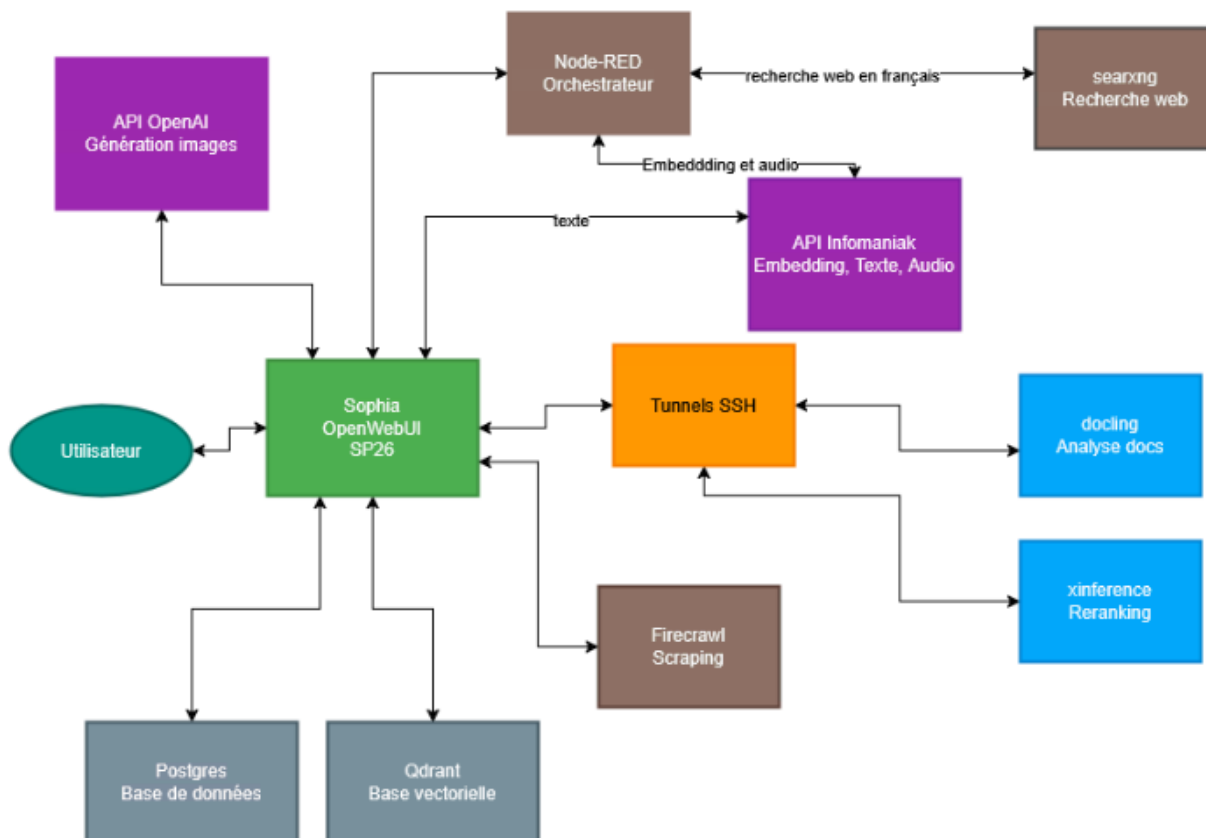


Architecture générale

Schéma d'architecture générale de Sophia

profile



Pour modifier l'image : <https://onlyoffice.vincennes.fr/Products/Files/DocEditor.aspx?fileid=40912>

L'architecture Today at 19:45

L'architecture est basée sur la documentation d'openwebui : <https://docs.openwebui.com/>.

Tableau récapitulatif des conteneurs - visualisation par Portainer

Conteneur / Service	Stack / Origine	Rôle	Port hôte (local)	Port conteneur	URL d'accès depuis Sophia
openwebui	Sophia	Interface principale	3002	8080	https://sophia.vincennes.fr
qdrant	Sophia	Base vectorielle (RAG)	6333, 6334	6333, 6334	Interne (via API locale)
postgres	Sophia	Base de données d'application	5437	5432	Interne
sophia-nodered-openwebui-1	Sophia	Automatisation, intégration des services	1886	1880	http://sp26.vincennes.fr:1886
infomaniak-xinference-1	Infomaniak (tunnel SSH)	Tunnel vers xinference (reranking)	9995	9998	http://sp26.vincennes.fr:9995
infomaniak-docling-1	Infomaniak (tunnel SSH)	Tunnel vers docling (analyse de documents)	5002	5001	http://sp26.vincennes.fr:5002
infomaniak-portainer_agent-1	Infomaniak (tunnel SSH)	Agent Portainer pour supervision distante	9002	9001	Interne (gestion Docker)
searxng	ia-app	Moteur de recherche web	8056	8080	Via Node-RED (recherche)
firecrawl-api-1	firecrawl	Analyse de pages web (scraping intelligent)	3003	3002	http://sp26.vincennes.fr:3003
docling (VM Infomaniak)	-	Analyse de documents (PDF, etc.)	5001	5001	Accès distant (via tunnel SSH depuis SP26:5002)
xinference (VM Infomaniak)	-	Modèle de reranking (bge-reranker-v2-m3)	9997	9997	Accès distant (via tunnel SSH depuis SP26:9995)
portainer_agent (VM Infomaniak)	-	Agent Docker pour supervision	9001	9001	Accès distant (via tunnel SSH depuis SP26:9002)

Réglages dans l'interface administrateur - visualisation par Openwebui

1. Menu Connexions

- **URL** : `https://sophia.vincennes.fr/admin/settings/connections`
- **Fonction** : Gestion des connexions aux modèles IA hébergés sur Infomaniak.
- **Détails** :
 - Affiche l'état de disponibilité des endpoints IA (texte, audio, image).
 - Aucune action de configuration possible directement : en cas de problème, contact avec le support requis.
 - Vérification de la connectivité aux API distantes.

2. Menu Documents

- **URL** : `https://sophia.vincennes.fr/admin/settings/documents`
- **Fonction** : Configuration du traitement des documents (analyse, embedding, reranking).
- **Paramètres** :
 - **Analyse de documents** :
 - URL : `http://sp26.vincennes.fr:5002/`
 - Connecté au service **docling** via tunnel SSH.
 - **Embedding** :
 - URL : `http://sp26.vincennes.fr:1886/`
 - Passerelle via **Node-RED**, qui relaye vers le modèle `bge_multilingual_gemma2` sur Infomaniak.
 - Utilisation de Node-RED pour limiter et réguler les appels API.
 - **Reranking (RAG)** :
 - URL : `http://sp26.vincennes.fr:9995/v1/rerank`
 - Connecté au modèle `bge-reranker-v2-m3` via **xinference** sur Infomaniak (tunnel SSH).

3. Menu Recherche web

- **URL** : `https://sophia.vincennes.fr/admin/settings/web`
- **Fonction** : Configuration de la recherche web et de l'analyse des résultats.
- **Paramètres** :
 - **Recherche** :
 - URL : `http://sp26.vincennes.fr:1886/search?q=<query>`
 - Passerelle via **Node-RED**, qui interroge **searxng**.
 - Node-RED assure la traduction de la requête en français si nécessaire.
 - **Analyse des pages** :
 - URL : `http://sp26.vincennes.fr:3003/`
 - Envoi des URLs récupérées vers **Firecrawl** pour extraction de contenu structuré.

4. Menu Interface utilisateur

- **Fonction** : Génération automatique de métadonnées pour améliorer l'expérience utilisateur.
- **Utilisation** :
 - Appels aux **API Infomaniak** pour :
 - Génération de **titres** pertinents.
 - Création de **tags** et de **mots-clés** pour la recherche.
 - Ces fonctionnalités enrichissent les réponses et les contenus générés.

5. Menu Audio

- **URL** : `https://sophia.vincennes.fr/admin/settings/audio`
- **Fonction** : Transcription de fichiers audio en texte.
- **Paramètres** :
 - **Transcription** :
 - URL : `http://sp26.vincennes.fr:1886/`
 - Utilisation de **Node-RED** pour gérer l'appel à l'API Infomaniak.
 - Le processus est asynchrone : Node-RED surveille la tâche et renvoie le texte une fois disponible.
 - Permet de traiter des fichiers audio longs via une file d'attente.

6. Menu Images

- **Fonction** : Génération d'images à partir de descriptions textuelles.
- **Paramètres** :
 - Appel direct aux **endpoints OpenAI** (ex: DALL·E).
 - Aucun traitement local ou intermédiaire : la requête est envoyée directement à OpenAI.
 - Les images générées sont intégrées dans les réponses de l'interface.

Révision #10

Créé 2025-09-26 08:53:32 UTC par Testelin Didier

Mis à jour 2025-09-28 09:57:07 UTC par Testelin Didier