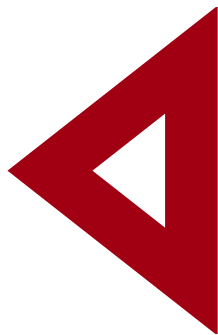


DESQL Corporation

DOSSIER DE SPÉCIFICATION

Application : UpssiDataBase



PRÉPARÉ PAR : BERNAT LOAN, GAUDILLAT ELIOTT, MISSIER
SAMUEL, SOUBIRANT DAMIEN, PHAN QUENTIN



À propos du logiciel

Notre logiciel sera en mesure d'indexer 3 types de documents : Texte (.xml), Image (.jpg, .bmp) et Audio (.wav .bin). Deux types d'acteurs pourront utiliser le logiciel. Les utilisateurs "standards", qui pourront via un moteur de recherche, récupérer un ou plusieurs documents en relation avec leur recherche. L'administrateur, qui pourra configurer et lancer l'indexation de nouveaux documents.

Périmètre fonctionnel

Voici ce qu'il va être nécessaire de réaliser au sein du logiciel pour qu'il puisse assurer sa fonction :

- **Indexation des fichiers par modalités (3 types)**
 - L'indexation correspond à la création de descripteurs pour chaque fichiers (1 descripteur = 1 fichier) et de les stocker dans un fichier dédié.
 - Un descripteur est un ensemble d'attributs communs aux mêmes types de fichiers, permettant de caractériser efficacement le contenu d'un fichier.
 - Texte :
 - Identification du fichier (Nom / Chemin)
 - Liste des termes significatifs (mots-clés) et leur nombre d'occurrences
 - Le nombre de mots-clés retenus
 - Taille du texte (après filtrage et épuration du texte)
 - Image :
 - Identification du fichier
 - Tableau donnant le nombre d'occurrences (pixels) de chaque intensité de couleur
 - Audio :
 - Identification du fichier
 - Histogramme des fréquences (par intervalles de fréquences)
 - Un descripteur est un ensemble d'attributs communs aux mêmes types de fichiers, permettant de caractériser efficacement le contenu d'un fichier.
- **Comparaison de fichiers via leur descripteur.**
 - La comparaison est paramétrée selon un taux de similarité (ou de dissemblance) explicité ci-après et compare les attributs des fichiers.

- **Utilisation en 2 modes (Admin, User)**
 - Le mode administrateur permet de lancer et configurer l'indexation des fichiers.
 - Le mode utilisateur permet de lancer le moteur de recherche via un mot-clé ou un fichier.
- **Moteur de recherche**
 - Le moteur de recherche comparera le descripteur du fichier de l'utilisateur avec ceux des fichiers déjà indexés afin de renvoyer les plus similaires. Ou, exclusivement pour les fichiers textes, le moteur de recherche pourra fonctionner avec un mot clé saisi par l'utilisateur et renverra les fichiers où le mot clé apparaît le plus de fois

Cadre technique

L'application sera développée en C sous le système d'exploitation Linux. L'application doit pouvoir fonctionner sur les machines des développeurs ainsi que sur celle du client ou celle de l'U3.

Gestion de l'accès

Le mode utilisateur sera accessible à toute personne disposant du logiciel.

Le mode administrateur quant à lui sera accessible uniquement au client, une authentification avec un mot de passe sera nécessaire pour y accéder

ZOOM SUR LES FONCTIONNALITÉS

Indexation des documents

- **Entrée** : Un fichier à indexer, fourni par son chemin d'accès
- **Sortie** : Un descripteur créé pour le document, structure synthétisant ses caractéristiques, dépendantes de son type (audio, image, texte)

Moteur de recherche

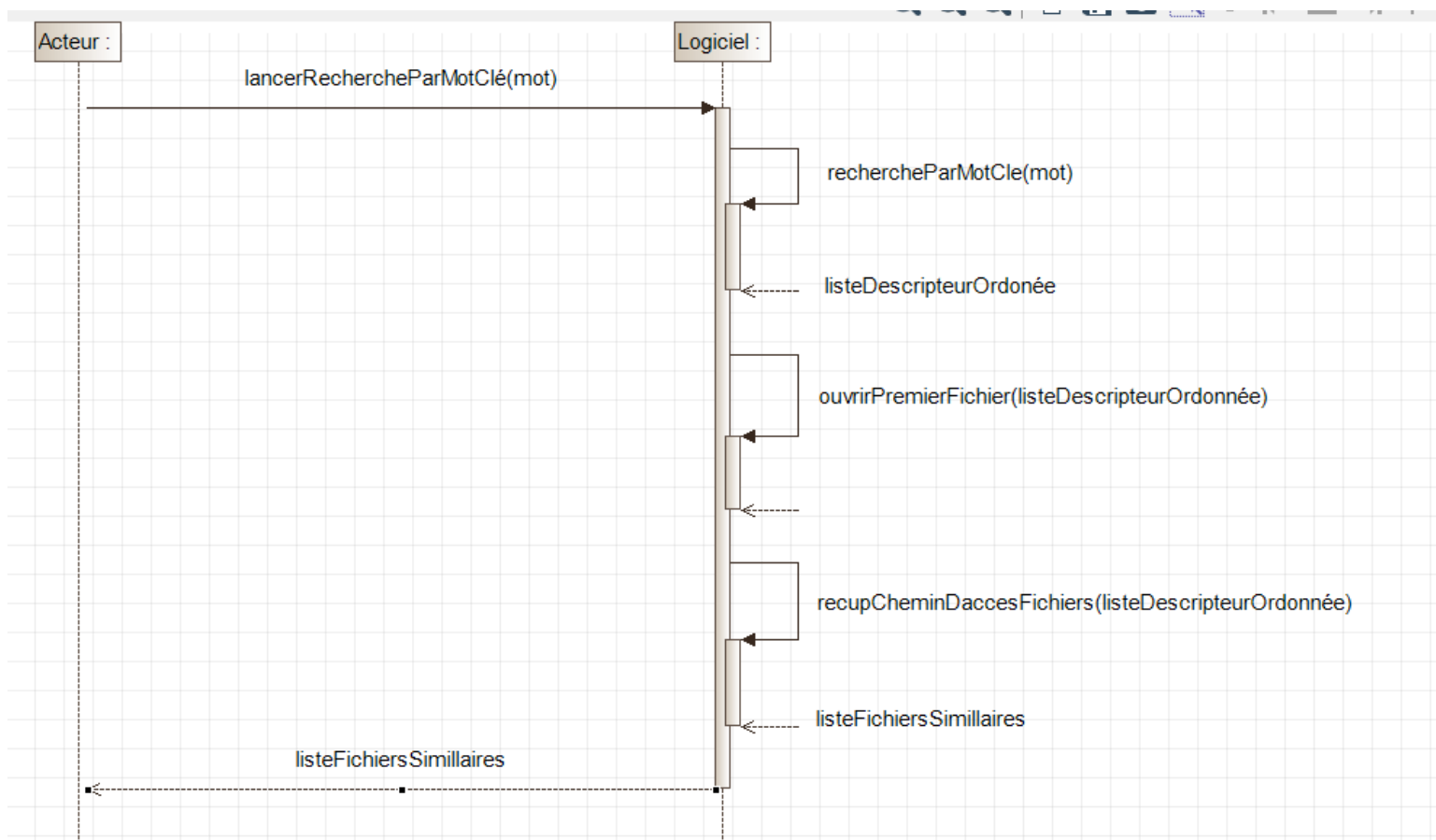
- **Entrée** : Un mot clé, mot explicitant l'idée générale recherchée par l'utilisateur (nom commun) (Mode de recherche exclusif aux fichiers textes) / Un fichier, similaire à ceux sur lequel l'utilisateur porte son intérêt
- **Sortie** : Ouverture du fichier le plus similaire possible à celui fourni en entrée, suivi d'une liste de fichier similaires, répondant à la requête.

Comparaison de Documents

- **Entrée** : 2 descripteurs, l'un correspondant au fichier donné en entrée, l'autre à un fichier déjà indexé, et ce pour tous les fichiers déjà indexés.
- **Sortie** : Un taux de dissemblance (ou de similarité), représenté par une distance numérique comprise entre 0 et 100 (plus elle est faible, plus les fichiers sont "proches" et par conséquent similaires)

DIAGRAMME DE SÉQUENCE SIMPLIFIÉE DU CAS D'UTILISATION 2

L'utilisateur (Acteur) lance une recherche via un mot clé
A l'issue, le logiciel renvoie une liste des fichiers les plus similaires et ouvre le premier de cette liste



EXEMPLES DE CAS D'UTILISATION :

Lancer une recherche via un fichier

- **Auteur déclencheur** : Utilisateur
- **Préconditions** : Le fichier est .xml , .wav , .bin, .jpg ou .bmp. Le fichier doit faire partie de la base l'indexation ou bien être indexé en amont.
- **Postconditions** : Aucune
- **Scénario Normal** :
 - **1** - L'utilisateur transmet son fichier au programme, par son chemin d'accès, pour pouvoir l'indexer et/ou lancer la recherche selon ci celui-ci est déjà indexé ou non.
 - **2**- Le programme demande au préalable le type de fichier manipulé et vérifie si le fichier est déjà présent dans la base, sinon l'indexe.
 - **3**- Le programme lance une recherche parmi les autres fichiers indexés pour trouver les fichiers les plus similaires possible, par comparaison des descripteurs.
 - **4**- Le programme retourne une liste de fichiers similaires, et ouvre celui qui l'est le plus.
- **Cas particuliers** :
 - Le fichier en entrée n'est pas d'un type de fichier reconnu ou bien introuvable : on affiche un message d'erreur "Fichier invalide"
 - Aucun fichier n'est estimé suffisamment similaire, autrement dit, aucun fichier ne dépasse le taux de similarité défini en configuration. Un message type "Pas de fichier similaire" est retourné.
 - Le programme n'a pas la permission pour ouvrir le fichier fourni en entrée par l'utilisateur. Un message d'erreur est affiché, type "Accès refusé".

Lancer une recherche via un mot clé

- **Auteur déclencheur** : Utilisateur
- **Préconditions** : Liste de descripteurs des fichiers de type texte
- **Postconditions** : Aucune
 - **1** - L'utilisateur saisi son mot-clé dans le terminal
 - **2**- Le programme vérifie la validité du mot-clé
 - **3**- Le programme lance une recherche à partir du mot-clé.
 - **4**- Le programme ouvre le fichier contenant le plus de fois le mot-clé et transmet la liste des autres fichiers possédant le mot-clé un nombre de fois minimum (défini dans le .config).
- **Cas particuliers** :
 - Le mot-clé n'est pas valide (Exemple : "Le"). Un message type "Mot-clé invalide" est retourné.
 - Le mot-clé ne donne aucun résultat. Un message type "Pas de correspondance" est retourné.

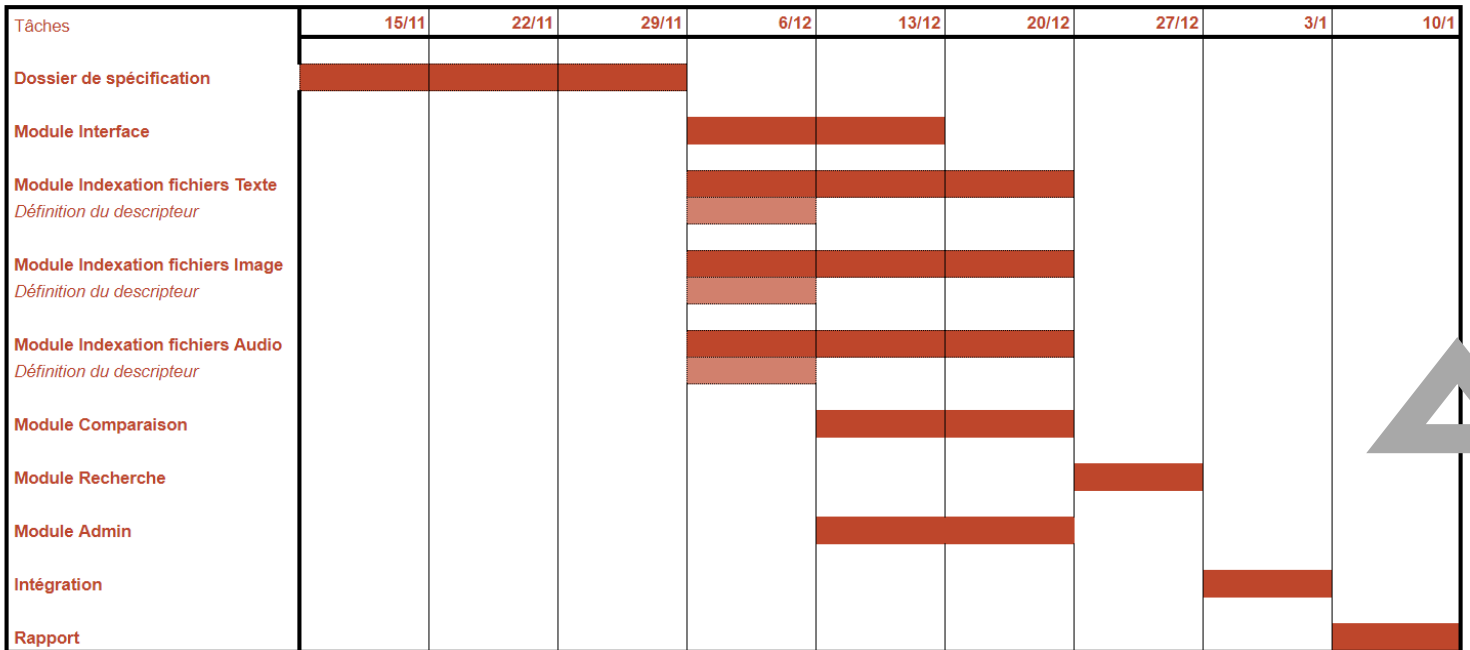
Modifier paramètres d'indexation

- **Auteur déclencheur** : Admin
- **Préconditions** : Aucune
- **Postconditions** : Aucune
- **Scénario Normal** :
 - **1** - L'admin modifie les paramètres d'indexation (taux de dissemblance / seuil d'occurrence des mots-clés / valeur bit couleur)
 - **2**- Le programme sauvegarde les modifications dans le fichier .config contenu dans la base.
 - **3**- Le programme relance une indexation de tous les fichiers avec les nouveaux paramètres d'indexations.
- **Cas particuliers** :
 - Vérification de la valeur de chaque paramètre (taux de dissemblance, le seuil d'occurrence et la valeur RGB). Si il y a un problème (valeur incohérente) on vient alors afficher un message demandant une valeur correcte et on propose à l'administrateur de saisir une nouvelle valeur.

Lancer une indexation manuelle

- **Auteur déclencheur** : Administrateur
- **Préconditions** : Aucune
- **Postconditions** : Le fichier est indexé
- **Scénario Normal** :
 - **1** - L'administrateur lance une indexation manuelle, sur un fichier unique ou sur un ensemble de fichiers à indexer.
 - **2**- Le programme parcourt l'ensemble des fichiers fourni à la recherche de ceux qui n'ont pas déjà été indexés et repère leurs chemins d'accès.
 - **3**- Le programme indexe chaque fichier selon son type.
 - **4**- Un descripteur regroupant les attributs propres à chaque fichier est créé pour chacun des fichiers. Ces attributs diffèrent selon le type du fichier.
 - **5**- Chaque descripteur est ajouté à un fichier regroupant les descripteurs déjà existants. Un fichier "liste" de descripteurs existe pour chaque type de document (Audio, Image, Texte).
- **Cas particuliers** :
 - Aucun nouveau fichier n'est détecté, on renvoie un message type "Pas de nouveau fichier à indexer".

Diagramme de Gantt



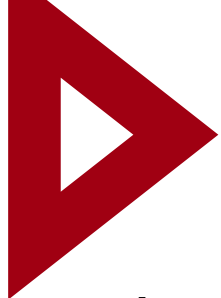
Interface : Fonctions de navigation et d'appel entre les deux modes, les différents menus et fonctionnalités possibles.

Indexation des fichiers (variant selon le type fichier) : Détermination des attributs, mise en place de la pile de descripteurs (sauvegarde/manipulation)

Comparaison : Utilisation des descripteurs de fichiers (une comparaison différentes par type). Caractérisée par le taux de similitude défini.

Module "Recherche" : Mise en œuvre des fonctions élémentaires (comparaison, indexation) afin d'obtenir une liste de fichiers similaires pour répondre à la requête de l'utilisateur saisie via l'interface.

Mode Admin : Gestion de la connexion sécurisée, de la modification du .config, de la définition des attributs et de la mise à jour de la base de descripteurs.



L'ÉQUIPE DE DÉVELOPPEMENT

loan.bernat@univ-tlse3.fr

Bernat Loan - Module "Recherche" + Comparaison

eliott.gaudillat@univ-tlse3.fr

Gaudillat Eliott - Indexation des fichiers audios

samuel.missier@univ-tlse3.fr

Missier Samuel - Indexation des fichiers textes

damien.soubirant@univ-tlse3.fr

Soubirant Damien - Indexation des fichiers images

quentin.phan@univ-tlse3.fr

Phan Quentin - Interface + Mode Admin



CLIENT

isabelle.ferrane@univ-tlse3.fr

Ferrané Isabelle