

Dossier de spécification

Projet Fil Rouge

Auteurs

Tobias BRISSAUD
Bernard PARFAITE
Pierre PINCON
Baptiste RAMONDA
Dorian SAURAT

Sommaire

I. Introduction.....	3
II. Fonctionnalités.....	5
II.1/ Mode Administrateur.....	5
II.2/ Mode Utilisateur.....	7
III. Répartition du travail.....	9
IV. Glossaire.....	10
V. Annexes.....	11

I. Introduction

Ce dossier a pour objectif de décrire les différentes fonctionnalités du moteur de recherche développé dans le cadre du Projet Fil Rouge (PFR). Ce projet s'inscrit au sein de la première année de la formation d'ingénieur « Systèmes Robotiques et Interactifs » (SRI) à l'UPSSITECH.

Ce projet a pour objectif la spécification, la conception et le développement en C d'un moteur de recherche qui permettra à son utilisateur de pouvoir rechercher différents types de documents au sein d'une base de données. Ce dossier se concentre ici sur la partie spécification de ce moteur de recherche. Si la partie utilisateur de ce projet vient d'être vaguement évoquée, le programme développé contient aussi une partie administrateur qui permet à celui-ci d'effectuer diverses actions sur la base de données. Nous pouvons donc résumer ce moteur de recherche comme ceci :

Une partie **administrateur** qui permet de :

- Configurer les processus d'indexation
- Lancer une indexation
- Visualiser les valeurs des descripteurs et de l'indexation

Une partie **utilisateur** qui permet de :

- Effectuer une recherche (via une requête spécifique en fonction du type de document recherché, cf. tableau ci-dessous)
- Visualiser les résultats de cette recherche

Comme évoqué ci-dessus, la base de données contient différents documents de plusieurs types et leur traitement est fait en fonction de leurs spécificités (voir tableau 1).

Tableau 1 : Types des documents, extensions et moyens de recherches

Types de Documents	Extension(s) Visualisation	Extension (Format) Traitement	Moyen(s) de recherche
Texte	.xml	.xml (Texte)	Mots-clés / Doc. Texte
Image	.jpg .bmp	.txt (Texte)	Couleur / Doc. Image
Audio	.wav	.bin (Binaire)	Doc. Audio

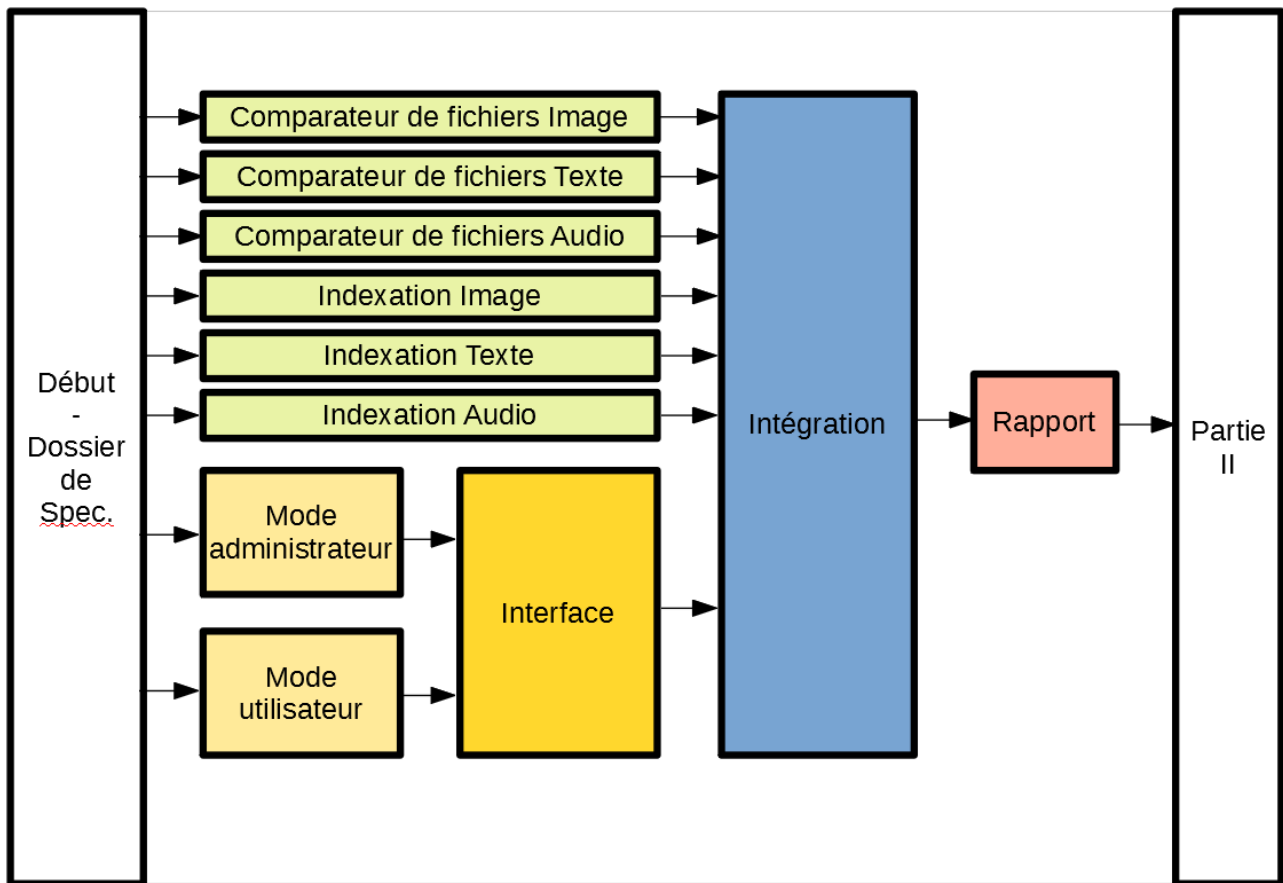


Figure 1 : Première partie du projet

La Figure 1 représente l'organisation des tâches. Si une tâche à gauche est reliée à une autre à droite par une flèche, alors la tâche de droite doit nécessairement être terminée pour pouvoir commencer la tâche de gauche. Nous estimons qu'il y a deux grandes tâches parallèles, les modules par type de fichier et les types d'utilisateurs pour préparer l'interface, en vue de l'intégration pour la version qui sera livrée au client. La rédaction du rapport interviendra après avoir livré le moteur de recherche.

II. Fonctionnalités

II.1 – Mode Administrateur

L'administrateur doit s'identifier pour avoir accès à ses fonctionnalités. Il pourra ainsi indexer de nouveaux documents en passant en paramètres soit le chemin vers ledit fichier soit un chemin vers un dossier contenant les fichiers à indexer et ainsi ajouter des descripteurs qui seront associés aux fichiers. Les descripteurs générés dépendront du type de fichier : pour les images les descripteurs seront la couleur prédominante ou le niveau de gris prédominant ; pour les textes les descripteurs seront les principales occurrences des mots du texte ; pour les audio les descripteurs seront déterminés par les histogrammes tirés des audios. L'administrateur peut également visualiser les valeurs des descripteurs en passant le nom de ceux-ci en paramètre.

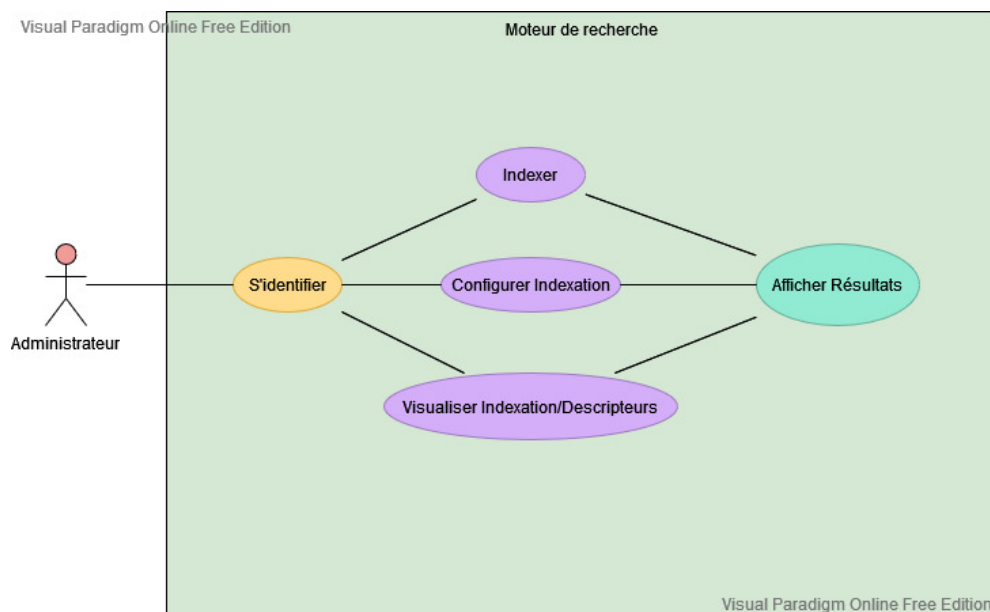


Figure 2 : Diagramme Use Case Administrateur

ATTENTION : Note importante concernant ce diagramme. Ici toutes les fonctionnalités arrivent à la même case « Afficher Résultat ». Il est donc important de spécifier que dans la grande majorité des cas, la nature de ce résultat n'est absolument pas la même (liste de descripteurs, booléen, entiers...etc). Cette modélisation a été simplifiée ainsi pour le bien de la lecture et de la compréhension du diagramme.

Ci-dessous (tableau 2), les différentes présentations des fonctionnalités évoquées :
(Code Couleur : fonctionnalités **Obligatoires** ; fonctionnalités **Facultatives**).

Tableau 2 : Fonctionnalités Administrateur

Fonctions	Entrées	Sorties	Description
Indexer	Chemin vers un dossier/fichier	Descripteurs liés à chaque doc. passé en paramètre	Génère des descripteurs correspondant au(x) fichier(s) passé(s) en paramètre(s) Chaque type de fichier (texte, audio, image) aura son propre type de descripteur
Tester	Chemin vers un dossier/fichier et un descripteur	Comparaison entre les descripteurs passés en paramètres et ceux associés aux fichiers	Vérifie que les descripteurs passés en paramètres sont identiques à ceux trouvés par la fonction <i>Indexer</i> .
Configurer Indexation	Processus d'indexation à modifier	Processus d'indexation modifié	Permet la modification de différents paramètres d'indexation en fonction du type de recherche (Ex : nombre d'occurrences minimales, taille de la fenêtre d'analyse...)
Visualiser Indexation	Fichier .config (En fonction des types de fichier qui ont été indexer)	Visualisation Du contenu du fichier. Config à l'écran	Visualise le processus d'indexation pour voir si l'indexation c'est bien passer
Visualiser Descripteur	Descripteur	Valeur du descripteur passé en paramètre	Donne la valeur du descripteur passé en paramètre.
Identifier	Identifiant ou Mot de Passe	Booléen donnant l'accès ou non à la suite du programme	Permet à l'administrateur de sécuriser ses fonctionnalités propres.

Pour mieux comprendre et visualiser les différentes actions possibles pour l'administrateur, se référer à **la Figure 2: Diagramme Use Case Administrateur**.

II.2 – Mode Utilisateur

Les fonctionnalités du moteur de recherche rattachées à l'utilisateur lui offre la possibilité de rechercher depuis un fichier texte, audio ou image, ou bien à partir d'une des spécificités intrinsèques de ces fichiers (mot-clé, couleur), les documents qui conviennent le plus à la recherche formulée. (cf. Figure 3).

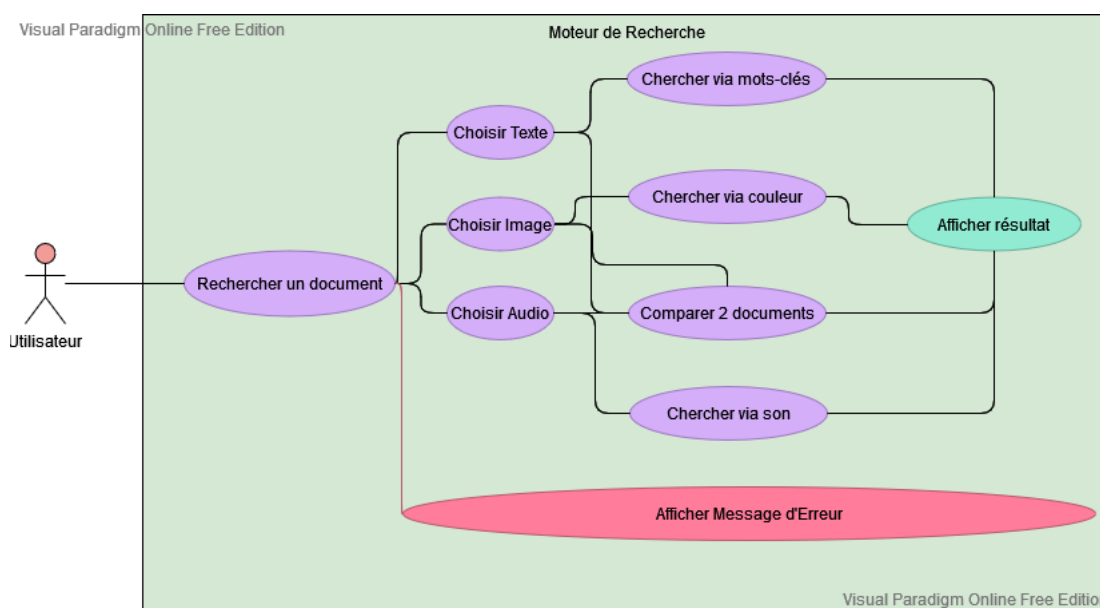


Figure 3 : Diagramme Use Case Utilisateur

ATTENTION : Note importante concernant les diagrammes Use Case. Toutes les bulles « use case » sont censé être reliées à la bulle « Afficher Erreur » (à l'exception de la case « Afficher Résultat ») car c'est ce qui se passerait dans le cas où les entrées utilisateur seraient incorrectes. Ces associations ont volontairement été oubliées pour permettre une meilleure lisibilité du diagramme.

L'utilisateur pourra aussi passer en paramètre de recherche un document et ainsi visualiser les documents du même type qui lui sont similaires en fonction justement d'un seuil de similarité calculé en interne. Ce seuil de similarité sera calculé grâce à une fonction de comparaison qui renverra en sortie ce seuil entre les deux descripteurs des documents comparés. Cette fonctionnalité pourra peut-être être rendue accessible à l'utilisateur si celui-ci souhaite voir plus précisément les seuils de similarités avec le document en entrée des différents documents en sortie. La recherche par une image se fera par comparaison entre image en calculant l'intersection des histogrammes représentant les images. L'image la plus proche sera ouverte automatiquement à l'aide d'un logiciel de visualisation. Pour la recherche à partir d'un fichier son, un calcul des intersections des histogrammes représentant le fichier sonore et ceux représentant le jingle sera fait. Le moteur de recherche localisera temporellement (en secondes) chaque

apparition du jingle dans le fichier sonore. La recherche par mot-clé pour les fichiers texte affichera par ordre décroissant le nombre d'occurrences du mot-clé dans ces fichiers.

Une autre fonctionnalité qui pourra être mise en place si celle-ci rentre dans les contraintes de temps, c'est la généralisation de la recherche utilisateur. En effet cette fonctionnalité aurait pour objectif de permettre une recherche de n'importe quel type de documents à partir de n'importe quel critère. Par exemple la recherche d'une image à partir d'un texte, la recherche d'un texte à partir d'un son, etc. Cette fonctionnalité est ambitieuse car elle est plus puissante et générale que celle présentée ci-dessus.

Ci-dessous (voir tableau 3), les différentes présentations des fonctionnalités évoquées :
(Code Couleur : fonctionnalités **Obligatoires** ; fonctionnalités **Facultatives**)

Tableau 3 : Fonctionnalités Utilisateur

Fonctions	Entrées	Sorties	Description
Recherche spécifique	Mots-clés (ou) Couleur (ou) Document	Documents correspondant le mieux à la recherche selon un seuil de correspondance	Recherche les documents similaires par rapport à un critère de recherche ou à un document du même type.
Comparaison	2 documents du même type	Seuil de similarité (en-dessous de ce seuil les documents ne sont pas similaires et au-dessus on considérera qu'ils le sont)	Compare deux documents grâce à leurs descripteurs et affiche un résultat en fonction d'un seuil de similarité.
Afficher Résultats	Résultat de la fonction Recherche	Affiche les résultats en les classant par ordre de similarité. Il sera également possible d'afficher les scores de similarité	Affiche les résultats en les classant par ordre de similarité décroissante avec en plus la possibilité d'ouvrir tous les documents trouvés.
Recherche générale	Tous critères de recherche (ou) N'importe quel type de document	Documents correspondant le mieux à la recherche	Recherche les documents similaires par rapport à n'importe quel critère de recherche ou à un document quelconque.

Pour mieux comprendre les différentes actions de base possibles pour l'utilisateur, se référer à *Figure 3 : Diagramme Use Case Utilisateur*.

III. Répartition du travail

La répartition qui suit a été établie à l'unanimité du groupe en prenant en compte les différents intérêts et compétences de chaque membre dans les domaines étudiés (voir tableau 4). Nous rappelons aussi ici que les rôles au sein de la communication visent à gérer la relation avec le client et à gérer les outils de travail : Discord, [Github](#), [Trello](#), [Drive](#), etc.

Tableau 4 : Répartition des modules/rôles

	Membres
Chef de Projet	Baptiste RAMONDA
Communication / Gestion	Pierre PINCON / Dorian SAURAT
Module Texte	Dorian SAURAT/ Pierre PINCON
Module Image	Dorian SAURAT / Tobias BRISSAUD
Module Audio	Bernard PARFAITE / Tobias BRISSAUD
Programme Principal	Baptiste RAMONDA / Pierre PINCON

Ci-dessous (figures 3 et 4), le diagramme de Gantt donnant les échéances temporaires pour finir chaque module/conception avant la deadline finale (ces dates ne sont que des indications visant à visualiser des objectifs dans le temps).

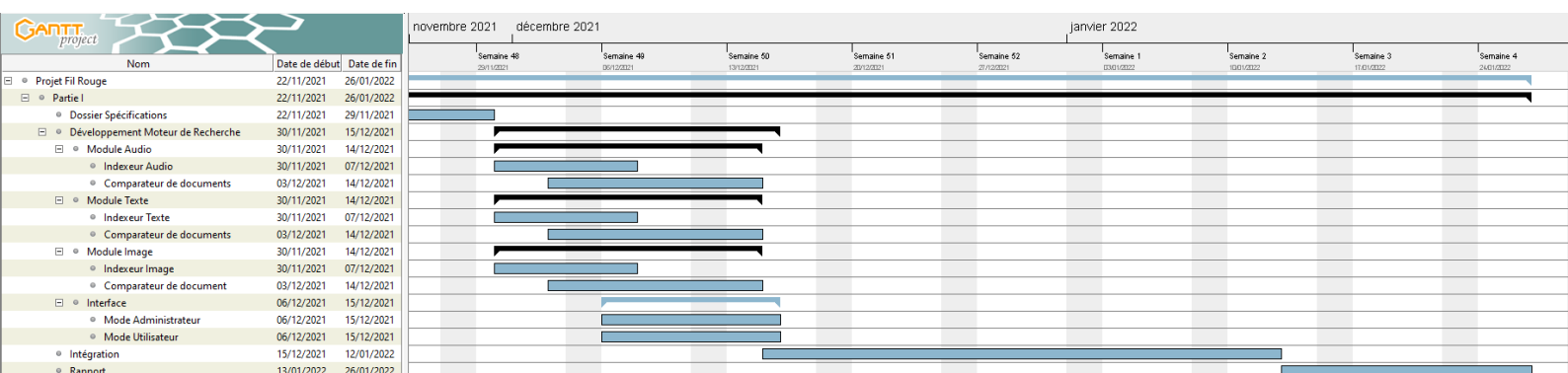


Figure 3 : Diagramme de Gantt

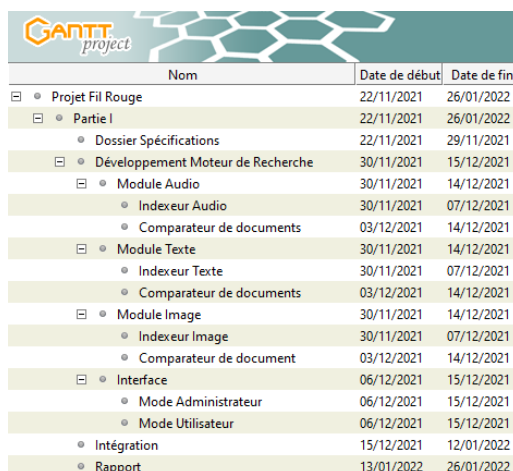


Figure 4 : Tableau récapitulatif Diagramme de Gantt

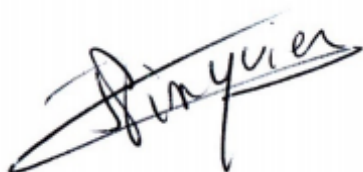
IV. Glossaire

Cette partie donne les définitions de certains mots de vocabulaire dans le cadre spécifique de ce projet portant sur la réalisation d'un moteur de recherche.

- **Administrateur** : Personne chargée de la gestion du programme. Il aura plus de droits que l'utilisateur (accès au « fonctionnement » du programme) et l'accès à son menu d'actions sera sécurisé.
- **Document** : Fichier de données, il correspond à un nom + une extension. On utilisera dans ce projet plusieurs types de documents (Texte, Audio, Image).
- **Descripteur** : Représentation simplifiée d'un document. Un descripteur possède un type (en fonction de la nature du document) et un ensemble de paramètres.
- **Diagramme de Gantt** : Diagramme permettant de visualiser dans le temps les diverses tâches composant un projet.
- **Diagramme de Séquence UML** : Diagramme représentant graphiquement des interactions entre les acteurs et le système selon un ordre chronologique.
- **Diagramme Use Case** : Diagrammes UML utilisés pour une représentation du comportement fonctionnel d'un système logiciel.
- **Indexation** : Analyse, identification et renseignement du contenu d'un document dans sa notice bibliographique (descripteur).
- **Utilisateur** : Personne qui utilise le programme. Son menu d'actions sera en libre accès et il ne pourra interagir qu'avec la barre de recherche du moteur.

Accord du maître d'ouvrage

Bon pour accord le 1/12/2021

A handwritten signature in black ink, appearing to read 'J. J. J. J.', is written over a horizontal line.

V. Annexes

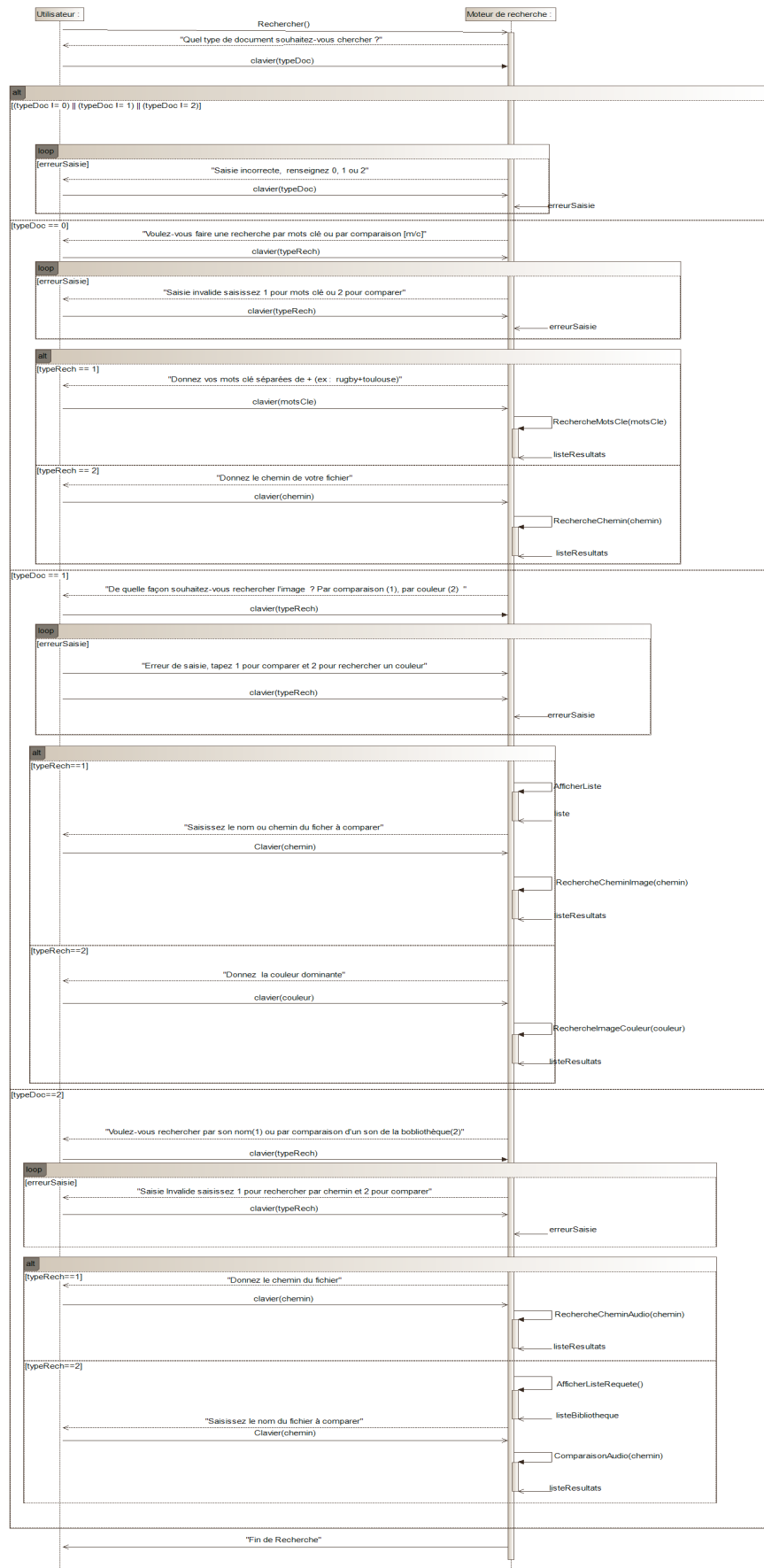


Figure 5 : Diagramme de Séquence UML sur le mode utilisateur

Dans un premier temps, l'utilisateur se verra demander le type de document qu'il souhaite rechercher. Une boucle bloquante vérifie la saisie.

Cas du document texte :

S'il choisit de rechercher un texte, il a le choix entre comparer un document ou par mots clé comme l'indique un exemple affiché à l'écran. Là aussi une structure de contrôle vérifie la saisie comme pour chaque saisie au clavier de l'utilisateur tout au long de ce diagramme.

S'il choisit par exemple de rentrer des mots clés, le programme dresse une liste des documents indexés en fonction des mots correspondants à la recherche et contenus dans les descripteurs. Pour finir, dans la recherche par mot clé comme dans la recherche par chemin, le premier document de la liste dressée est ouvert automatiquement.

Cas du document de l'image :

Si lors du choix de document l'utilisateur choisit de rechercher une image (ici typeDoc == 0), il a le choix la réaliser via une couleur prédominante saisie au clavier ou par son nom. Dans le premier cas, le programme compare le contenu des descripteurs à la couleur saisie pour dresser une liste et ouvre automatiquement le premier document de la liste. Dans le second cas, la liste des documents indexés est affichée et l'utilisateur rentre le nom du fichier qu'il souhaite ouvrir dans cette liste.

Cas du document de l'audio :

Dans le cas où l'utilisateur choisit de rechercher un document audio, il lui est donné de choisir de rechercher soit par le nom du fichier, soit selon une liste de fichiers à comparer pour dresser une liste de documents indexés similaires au fichier choisi pour la comparaison. Là aussi, le premier document de la liste est ouvert automatiquement.