



Partie I

Contexte

Création d'un moteur de recherche de document de nature diverse dans un format standard.

L'objectif de cette première partie est de spécifier, concevoir et développer des fonctionnalités d'indexation et de comparaison entre documents.

Indexation

Afin d'indexer et de comparer, nous allons produire des descripteurs étant la représentation simplifiée du document indexé.

La nature de chaque descripteur dépend du type du document indexé ainsi qu'un ensemble de paramètres donnés dans un fichier de configuration. L'ensemble des descripteurs du même type seront stockés dans une base de descripteurs.

On dit d'un document qu'il a été indexé lorsqu'il a produit la génération d'un descripteur.

Un document ne peut être indexé qu'une seule et même fois.

Comparaison

La comparaison entre deux documents de même type revient à comparer leurs descripteurs associés.

Ils sont identiques si leurs descripteurs sont strictement identiques.

Pour ce moteur de recherche, nous allons introduire la notion de similarité entre deux descripteurs qui va reposer sur le nombre de caractéristiques qu'on en commun deux descripteurs.

Cette notion de similarité va alors nécessiter de fixer une valeur seuil à partir de laquelle on va considérer que deux documents sont proches.

Le moteur de recherche

L'interface n'étant pas l'objet de cette première phase de projet, le premier prototype du moteur de recherche **INSERER UN NOM** sera sommaire et ne se composera donc que d'une entrée clavier et d'une sortie textuelle.

L'utilisation du moteur de recherche pourra se faire selon deux modes distincts :

Le mode **administrateur** qui pourra configurer les processus d'indexations, lancer l'indexation ainsi que visualiser les descripteurs.

Le mode **utilisateur** qui permettra d'effectuer une recherche dans notre panel de descripteurs et d'afficher les documents les plus proches du document en entrée.



Equipe

Dans une logique d'améliorer nos mindset, nous avons choisi d'ouvrir l'équipe à diverses filiales.

Etienne : Celui qui a vu pire comme équipe mais quand même, c'est des beaux branquignols

Raphaël : Celui qui finance le projet mais en fait il est pauvre

Gaël : Celui qui va essayer d'être sérieux le jour de l'oral alors qu'il sera entouré de 4 crétins

Oualid : Celui qui va mettre des memes dans le rapport

Clément : Celui qui va finir le projet au prix de 3L de café/heure

Bilan d'ouverture

	ACTIF			PASSIF	
	Brut	Amort. Dépréc	Net		
Actif Immobilisé				Capitaux Propres	
Salle X-free	10 000,00 €	8 000,00 €	2 000,00 €	Capital	8 000,00 €
Restaurant Universitaire	5 000,00 €	1 366,47 €	3 633,53 €	Reserves	- €
Ordinateurs	3 259,63 €	1 583,78 €	1 675,85 €	Resultat	- €
Machine à Café	500,00 €	38,00 €	462,00 €	Subventions	1 500,00 €
Actif Circulant				Dettes	
Stock Matiere Premiere	- €		- €	Emprunts	3 190,03 €
Stock Produit Fini	- €		- €	Dettes Fournisseurs	2 977,78 €
Créances	7 896,43 €		7 896,43 €		
Banque	- €		- €		
TOTAL	26 656,06 €	10 988,25 €	15 667,81 €		15 667,81 €



Annexes

Document Texte (XML)

Document Image (JPG)

Document Audio (WAV, BIN)

Résultat de la recherche

Pour une recherche de texte,

Par mot-clé, affichage de la liste triée des fichiers par ordre décroissant des occurrences du mot-clé et ouverture du fichier ayant le meilleur score

Par nombre de mots-clés en commun, affichage de la liste triée des fichiers par ordre décroissant des fichiers ayant le plus de mots-clés en commun et ouverture du fichier ayant le meilleur score

Pour une recherche image, le moteur de recherche prend en entrée le chemin d'une image et sort les images les plus correspondantes et ouvre la meilleure dans un logiciel de visualisation.

Pour une recherche audio, l'utilisateur devra saisir le chemin du jingle choisi et le moteur de recherche renverra la liste des corpus les plus semblables (et lire le plus semblable).