



République Algérienne Démocratique et Populaire



Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université AMO de Bouira

Faculté des Sciences et des Sciences Appliquées

Département d'Informatique

Projet System De Recherche D'Information

SPÉCIALITÉ : *Master 1 ISIL*

## Thème 03

---

### Multimedia Information retrieval

---

Réalisé par :

- MAZOUZ YANIZ
- DRAIFI ABDELHAK
- RAHAL LYES

2024/2025

# Table des matières

<b>Table des matières</b>	<b>1</b>
<b>Introduction</b>	<b>3</b>
<b>1 Concepts fondamentaux de MIR</b>	<b>4</b>
1.1 Indexation multimodale . . . . .	4
1.1.1 Indexation textuelle . . . . .	4
1.1.2 Indexation visuelle . . . . .	5
1.1.3 Indexation audio . . . . .	5
1.2 Recherche d'informations multimédia . . . . .	5
1.2.1 Modèle vectoriel . . . . .	5
1.2.2 Modèle probabiliste . . . . .	6
1.3 Requête multimodale . . . . .	6
1.3.1 Processus de gestion des requêtes multimodales . . . . .	6
<b>2 techniques de récupération basée sur le contenu (CBR)</b>	<b>8</b>
2.2 Applications pratiques . . . . .	8
2.2.1 Recherche d'images . . . . .	8
2.2.2 Recherche audio . . . . .	9
2.2.3 Recherche vidéo . . . . .	9
2.2.4 Recherche textuelle . . . . .	9
2.3 Méthodologies détaillées . . . . .	10
2.3.1 Extraction de caractéristiques . . . . .	10
2.3.2 Mesures de similarité . . . . .	10

---

2.3.3	Apprentissage automatique . . . . .	10
2.4	Avantages des CBR . . . . .	11
2.5	Défis et limites . . . . .	11

]

# Introduction

Les ordinateurs, longtemps utilisés pour traiter des informations textuelles et numériques, jouent désormais un rôle central dans l'accès et la gestion d'informations multimédia. En effet, de nombreux domaines professionnels nécessitent des contenus non textuels pour répondre à des besoins spécifiques. Par exemple, les médecins consultent des radiographies, les architectes utilisent des plans de bâtiments, et les agents immobiliers montrent des photographies de propriétés.

Dans ces domaines, l'information visuelle est souvent aussi importante, voire plus, que le texte. Le besoin de recherche d'informations multimédia devient ainsi essentiel, car il permet d'accéder aux documents visuels ou audio nécessaires à ces professionnels. Il est difficile d'imaginer qu'une entreprise de construction reçoive un plan de bâtiment uniquement sous forme textuelle, ou qu'un journal soit consulté sans la disposition graphique de ses pages. La recherche d'informations multimédia devient aussi essentielle dans des contextes où les documents sont principalement textuels mais nécessitent des annotations ou illustrations visuelles, comme les formulaires d'assurance comportant des commentaires en marge.

Les progrès récents en matière de stockage et d'affichage numérique facilitent l'intégration de contenus multimédia dans les documents informatiques, et le traitement des images, vidéos, et sons devient plus accessible. La recherche d'informations multimédia permet aux utilisateurs de créer et de naviguer dans des bibliothèques de documents enrichis, où textes et médias se complètent pour offrir des perspectives variées et améliorer la qualité des informations accessibles.

# Concepts fondamentaux de MIR

Les concepts fondamentaux de la Récupération d'Information basée sur la Recherche de Médias (MIR – Multimedia Information Retrieval) englobent plusieurs notions clés qui permettent de rechercher et de récupérer des informations multimédia (textes, images, audio, vidéo, etc.) de manière efficace. Voici un aperçu des concepts essentiels dans ce domaine :

## 1.1 Indexation multimodale

L'indexation est le processus d'attribution de métadonnées aux objets multimédia afin qu'ils puissent être rapidement récupérés lors d'une requête. L'indexation multimodale fait référence à l'organisation de différents types de médias (textes, images, vidéos, audio) de manière à permettre une recherche efficace sur l'ensemble de ces données.

### 1.1.1 Indexation textuelle

L'indexation textuelle consiste à associer des mots-clés ou des métadonnées à des documents textuels pour en faciliter la recherche et l'organisation. Elle repose sur l'identification des termes significatifs ou expressions clés qui représentent le contenu, tels que *énergie solaire* ou *transition énergétique* pour un article sur les énergies renouvelables. Les métadonnées peuvent inclure des informations comme l'auteur, la date de création, les catégories thématiques ou un résumé descriptif. L'indexation peut être effectuée manuellement par des experts ou automatiquement à l'aide d'algorithmes utilisant des techniques de traitement automatique du langage naturel (TALN). Par exemple, pour un document

---

intitulé *L'intelligence artificielle dans la médecine moderne*, les mots-clés pourraient inclure *IA*, *médecine*, ou *diagnostic assisté par ordinateur*.

### 1.1.2 Indexation visuelle

L'indexation visuelle consiste à extraire et analyser les caractéristiques visuelles des images et vidéos pour les organiser, les classer et faciliter leur recherche. Ces caractéristiques peuvent inclure des informations telles que les couleurs dominantes, les textures, les formes ou encore les objets reconnus au sein des images. Par exemple, une image contenant un coucher de soleil pourra être associée à des teintes chaudes (orange, rouge), des textures de ciel et des formes naturelles comme des montagnes. L'indexation visuelle repose souvent sur des algorithmes d'apprentissage automatique ou de vision par ordinateur.

### 1.1.3 Indexation audio

L'indexation audio implique l'identification et l'extraction de caractéristiques spécifiques des fichiers audio pour en faciliter l'organisation et la recherche. Parmi ces caractéristiques, on trouve les fréquences, les rythmes, les timbres et les motifs musicaux. Cette indexation est souvent effectuée à l'aide d'algorithmes de traitement du signal et d'apprentissage automatique, facilitant ainsi la catégorisation de fichiers musicaux et les recherches par contenu.

## 1.2 Recherche d'informations multimédia

Dans MIR, les informations sont représentées sous des formats spécifiques pour faciliter leur récupération. Cela inclut la représentation de données multimédia sous forme de vecteurs ou de signatures, souvent avec des caractéristiques extraites des éléments multimédias.

### 1.2.1 Modèle vectoriel

Le modèle vectoriel pour la recherche d'informations multimédia représente les documents et les requêtes sous forme de vecteurs dans un espace multidimensionnel. Chaque dimension correspond à une caractéristique (termes, couleurs, fréquences, etc.), et la per-

---

tinence est calculée en mesurant la similarité, souvent via le cosinus de l'angle entre les vecteurs.

### 1.2.2 Modèle probabiliste

Le modèle probabiliste repose sur l'idée d'estimer la probabilité qu'un document soit pertinent pour une requête spécifique. Chaque document est représenté par un ensemble de caractéristiques (termes textuels, éléments visuels ou audio), et un score de pertinence est calculé. Ce modèle s'adapte aux incertitudes inhérentes à la recherche et est souvent amélioré par des techniques comme le feedback de pertinence ou l'apprentissage automatique.

## 1.3 Requête multimodale

Les requêtes multimodales dans la Recherche d'Informations Multimodales (MIR) sont des requêtes où plusieurs types de médias ou modalités sont utilisés ensemble pour améliorer la précision des résultats de recherche. Ces modalités peuvent inclure du texte, des images, des vidéos, des éléments audio, des gestes, ou d'autres types de données.

### 1.3.1 Processus de gestion des requêtes multimodales

**Fusion des modalités :** Lorsqu'une requête multimodale est lancée, les différents types de médias sont traités et intégrés de manière cohérente. Par exemple, dans le cas d'une requête comprenant un texte et une image, un système peut utiliser des techniques de fusion de caractéristiques, où des vecteurs de caractéristiques représentant le texte et l'image sont combinés.

**Alignement sémantique :** L'un des défis majeurs des requêtes multimodales est d'assurer un alignement sémantique entre les différentes modalités. Par exemple, une image peut contenir des objets ou des scènes qui ne sont pas directement exprimés par des mots dans une requête textuelle. Des techniques avancées comme les réseaux neuronaux multimodaux sont utilisées pour apprendre les relations entre les différentes modalités.

---

**Recherche basée sur des requêtes combinées :** Un utilisateur peut effectuer une recherche multimodale en combinant un texte descriptif avec une image ou une vidéo. Par exemple, une recherche pourrait combiner un texte comme *montagnes au coucher du soleil* et une image représentant un paysage montagneux. Le système doit analyser à la fois le texte pour en extraire les mots-clés et l'image pour en détecter les éléments visuels pertinents.

]



# techniques de récupération basée sur le contenu (CBR)

Les techniques de récupération basées sur le contenu permettent d'identifier, de rechercher et de classer des éléments en analysant leurs caractéristiques intrinsèques. Contrairement aux approches traditionnelles qui se fondent sur des mots-clés ou des métadonnées (comme les tags ou descriptions), ces méthodes analysent directement le contenu des éléments, qu'il s'agisse de données visuelles, auditives ou textuelles.

Les **principales étapes** du processus incluent :

1. **Extraction de caractéristiques** : Conversion des éléments multimédias en des représentations numériques compactes appelées descripteurs (par exemple, histogrammes de couleurs, spectrogrammes pour l'audio, etc.).
2. **Indexation** : Organisation des descripteurs dans une structure de données adaptée (par exemple, arbre de recherche).
3. **Recherche et comparaison** : Comparaison des descripteurs d'une requête avec ceux des éléments stockés en utilisant des mesures de similarité.

## 2.2 Applications pratiques

### 2.2.1 Recherche d'images

Techniques utilisées :

- 
- Histogrammes de couleurs : Mesures statistiques des couleurs dominantes d'une image.
  - Textures : Identification des motifs répétitifs dans une image (par exemple, lignes, points).
  - Formes : Analyse des contours et des structures géométriques.
  - Deep Learning : Utilisation de réseaux de neurones convolutionnels (CNN) pour extraire des caractéristiques complexes.

**Exemple :** Recherche d'images similaires à partir d'un croquis ou d'une photo.

### 2.2.2 Recherche audio

**Techniques utilisées :**

- Spectrogrammes : Représentation visuelle des fréquences dans un signal audio.
- Fingerprints audio : Création d'une empreinte unique basée sur des fréquences et des rythmes.
- Apprentissage automatique : Modèles d'apprentissage supervisé pour identifier des genres ou des instruments.

**Exemple :** Retrouver une chanson en fredonnant une mélodie (Shazam).

### 2.2.3 Recherche vidéo

**Techniques utilisées :**

- Analyse de frames : Extraction de caractéristiques à partir d'images individuelles dans une vidéo.
- Reconnaissance de scènes : Identification de contextes ou objets spécifiques dans une séquence.
- Indexation spatio-temporelle : Analyse des mouvements et changements dans le temps.

**Exemple :** Recherche de vidéos sportives contenant un type de but spécifique.

### 2.2.4 Recherche textuelle

**Techniques utilisées :**

- 
- TF-IDF (Term Frequency-Inverse Document Frequency) : Évaluation de l'importance des mots dans un document.
  - Word Embeddings : Représentation vectorielle des mots (ex. Word2Vec, GloVe).
  - Recherche sémantique : Utilisation de modèles NLP pour comprendre le contexte.

**Exemple :** Rechercher des articles similaires à un texte donné.

## 2.3 Méthodologies détaillées

### 2.3.1 Extraction de caractéristiques

Chaque type de contenu possède des descripteurs spécifiques :

- **Images** : Histogrammes, gradients orientés (HOG), descripteurs locaux comme SIFT ou SURF.
- **Audio** : Coefficients cepstraux (MFCC), analyses FFT (Fast Fourier Transform).
- **Vidéo** : Fusion des caractéristiques spatiales (images) et temporelles (mouvements).
- **Texte** : Analyse syntaxique, fréquence des mots, bigrammes/trigrammes.

### 2.3.2 Mesures de similarité

Les mesures varient en fonction du type de données :

- Distance Euclidienne : Mesure classique entre vecteurs.
- Cosine Similarity : Pour comparer des documents textuels.
- Intersection d'histogrammes : Adaptée aux données d'image.
- Distance de Levenshtein : Pour évaluer les différences entre chaînes de caractères.

### 2.3.3 Apprentissage automatique

Utilisation de techniques modernes pour améliorer la pertinence :

- Clustering : Organisation non supervisée des données (ex. K-Means).
- Classification supervisée : Modèles comme les SVM ou les réseaux neuronaux pour catégoriser les données.
- Deep Learning : Approches complexes basées sur des réseaux de neurones profonds (CNN pour les images, RNN/LSTM pour les séquences).

---

## 2.4 Avantages des CBR

1. Recherche basée sur le contenu réel plutôt que des étiquettes parfois erronées.
2. Permet la recherche d'éléments visuels, sonores ou textuels avec une précision accrue.
3. Exploitation des capacités d'intelligence artificielle pour s'adapter à des bases de données massives.

## 2.5 Défis et limites

1. Complexité calculatoire : Traitement et comparaison des descripteurs peuvent être coûteux pour de larges bases de données.
2. Qualité des caractéristiques : Une mauvaise extraction peut nuire à la pertinence des résultats.
3. Subjectivité : Les descripteurs ne capturent pas toujours la perception humaine (exemple : beauté subjective dans une image).