

Bases de données multimédia I - Introduction

ENSIMAG
2014-2015

Matthijs Douze & Karteek Alahari



Bases de données multimedia

- Deux intervenants:
 - ▶ Matthijs Douze, matthijs.douze@inria.fr
 - ▶ Karteek Alahari karteek.alahari@inria.fr
- 6 x 3 h = 18 h de cours
- Évaluation : examen écrit + quizz sur des articles + présentations articles (bonus!)



Au programme

- Introduction
- Gestion des bases de données multimédia
- Espaces de représentation
- Mesures de similarité
- Évaluation
- Description globale
- Description locale
- Techniques d'indexation
- Recherche dans des espaces de grande dimension
- Indexation Vidéo
- Indexation Audio
- Classification

Inria

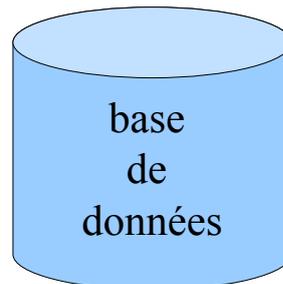


Base de données



Wikipedia

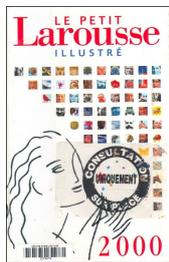
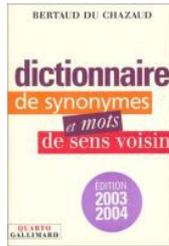
Une **base de données** est un ensemble structuré et organisé permettant le stockage de grandes quantités d'informations afin d'en faciliter l'exploitation (ajout, mise à jour, recherche de données).



Inria



Multimédia



Inria



Qui utilise plusieurs moyens de diffusion

Qui concerne plusieurs média

Diffusé par plusieurs média

Adj: Qui utilise ou concerne plusieurs médias

Media: Tout support de diffusion de l'information constituant à la fois un moyen d'expression et un intermédiaire transmettant un message à l'attention d'un groupe

Multimédia



Wikipedia

Le mot multimédia est apparu vers la fin des années 1980, lorsque les CD-ROM se sont développés. Il désignait alors les applications qui, grâce à la mémoire du CD et aux capacités de l'ordinateur, pouvaient générer, utiliser ou piloter différents médias simultanément [...]

Aujourd'hui on utilise le mot multimédia pour désigner toute application utilisant ou servant à travailler sur au moins un média spécifique. [...]

Par ailleurs, en recherche en informatique, on nomme multimédia l'étude des médias non textuels, principalement les images, les vidéo et les sons.

→ on perd souvent le côté "multi"

Inria



Multimédia : définition de travail

- Des contenus pouvant être de différents types, en particulier
 - ▶ Texte
 - ▶ Image
 - ▶ Son
 - ▶ Vidéo
- Qui peuvent être combinés
- Caractérisés par des besoins de stockage très importants

Inria



Ordres de grandeur

- **Volume unitaire** de stockage. Valeurs typiques :

	Stockage brut	Compression sans perte	Compression avec perte
Texte (Descartes, discours de la méthode, http://abu.cnam.fr)	125 Ko	Gzip: 42 Ko	N/A
Photo (6 Mpixels)	18 Mo	Gzip*: 12 Mo	1 Mo
Album de musique	750 Mo	FLAC: 406 Mo	MP3: 75 Mo
Un film de 100 minutes (720x576)	180 Go	X	700 Mo

Inria



Ordres de grandeur

- Nombre d'unités
 - ▶ Internet visible : nécessite "seulement" 5-10 To de capacité de stockage
 - Google: 3.5 milliards de recherches / jour
 - ▶ Films : <http://www.imdb.org> recense plus de 400 000 films
 - ▶ Images (semi-)pro : Corbis, Getty, Fotolia = 10-100 M images
- Croissance très importante, en raison de l'accumulation des contenus numériques auto-produits par le grand public
 - ▶ Images : par exemple, Flickr, Picasa, Facebook
 - ordre de grandeur: 10-100 milliards d'images
 - Facebook + 100 M d'images / jour
 - ▶ <http://www.flickr.com> : "9600 images uploaded in the last minute"
 - extrapolation : 15M d'images images par jour, 5.2G d'images par an
 - ▶ <http://www.youtube.com> : 2G de vidéos délivrées chaque jour + 300 h de vidéo / minute
 - ▶ croissance > loi de Moore → "datacenters"
- Images/vidéo spécifiques : photos satellites (SPOT: 3M km²/jour), vidéo-surveillance (4 M en GB), etc

Inria



Chaîne du multimédia : besoins

- Génération : outils de production et de création
- Représentation : utilisation de formats de représentations différentes
- Stockage
- Transmission : problème de réseaux, architecture
- **Recherche d'information : recherche basée sur le contenu**
- Distribution : conception de serveur de streaming

Inria



Applications types de BD multimedia

- News/Films à la demande
- Commerce électronique
- Informations médicales
- Systèmes d'informations géographiques
- Architecture/Design
- **Protection du copyright / traçage de contenu**
- Géo-localisation
- Enquêtes policières
- Militaire
- Expérimentations scientifiques
- Enseignement
- **Archivage, gestion des bases de données de contenu (personnelles ou professionnelles)**
- **Moteur de recherche (Internet, collections personnelles)**
- Applications industrielles

Dans les cours suivants :

focus sur la partie recherche et indexation dans de grandes bases d'images et de vidéos

Inria



Base de données pour le multimédia

- Initialement traitées comme des bases standards
 - ▶ Objets multimédia traités comme un seul item comme champs au sein d'une base de données relationnelle (ex: Oracle, LOB –Large Object)
 - ▶ objet opaque
- Recherche sur mots clés introduits manuellement dans le système par la personne : nécessité d'un système d'annotation, par exemple
 - Tag ID3 dans le MP3
 - Information Exif
 - ▶ Utilisation des relations entre objets
 - ▶ Recherche sur les mots présents dans les pages Web conjointement avec l'objet multimédia
 - <http://images.google.com>
 - <http://www.exalead.fr/image>
 - <http://images.search.yahoo.com>

Inria



Base de données “standard”

- Une base de données (DB) relationnelle est un ensemble de données mises en relation.
- Un système de gestion de base de données (DBMS) est un logiciel facilitant la création, la maintenance et la manipulation d'une base de données. Il garantit sécurité et intégrité de la base.
 - ▶ Oracle, DB2, Sybase, PostgreSQL, Mysql
- Système de base de données = la base + le système de gestion
- Succès important de ce type de système
 - ▶ répond à la plupart des besoins impliquant des données bien structurées
 - ▶ accroît l'indépendance des données
- Requêtes approximatives:
 - ▶ ex: noms dans les bases policières, requêtes avec correction orthographique
 - ▶ nécessite la présence d'annotations (ou méta-données) et/ou d'outils d'analyse

Inria



Spécificités des bases de données multimédia (MMDB)

- Ex de requête impossible avec un SGDB “classique”:
 - ▶ récupérer tous les images “qui ressemblent” à une image requête
 - Grosky et Fotouhi (1) caractérisent le multimédia à partir du type d'activité impliquée dans la création des données
- L'information portée par le multimédia est tout ce qui peut venir du monde réel, alors que l'information portée par une base de données classique ne peut être qu'une représentation symbolique de faits limités à l'univers de la base de données.*
- Le développeur d'une MMDB ne peut expliciter tous les aspects des données qui seront importants pour l'utilisateur
 - ▶ les informations ne peuvent se limiter à un ensemble de champs prédéfinis
 - ▶ sauf pour applications spécifiques → ici on suppose qu'on ne sait pas a priori

[1]. W.I. Grosky, F. Fotouhi and Z. Ziang. *Multimedia data Management. Using metadata to integrate and apply digital media*, chapter *Using metadata for the intelligent browsing of structured media objets*, pages 123-148. In Sheth and Klas, 1998

Inria



Richesse des bases de données multimédia

- Apprendre un concept (ici les couleurs) à partir “du monde réel” [2]



- Conclusion de l'article : utiliser Internet (c'est-à-dire Google ou tout autre base texte+images) pour définir des couleurs de nouvelles images
→ meilleurs résultats que de demander à un groupe de personnes

[2]. J. Van-der-Weijer, C. Schmid and J. Verbeek, *Learning Color Names from Real-world Images*, CVPR 2007



MMDb avec contexte d'utilisation limité

- Certaines applications correspondent à un besoin spécifique (type de requête unique ou limité)
- Pour de telles applications spécifiques :
 - ▶ bases de données classiques
+ outil d'analyse simple adapté à l'application
- Exemples :
 - ▶ base de données sur le remplissage des stades à partir d'images de retransmission
 - ▶ images en sortie de chaîne de production industrielle
 - ▶ imagerie médicale, satellite



Contenu et meta-données

- Les données “brutes” (fichier image, fichier son) provenant d’objets multimédia contiennent des informations sémantiques (= compréhensibles pour l'utilisateur).
- Ces meta-données proviennent
 - ▶ soit de propriétés de descripteurs des objets (couleur moyenne d'une image: métadonnée sur-exposé)
 - ▶ ou d'annotations manuelles.

Inria



Annotations / Meta-données

- Une annotation textuelle sera toujours trop restrictive, même si elle prend en compte à la fois des informations syntaxiques et symboliques.
- Malgré cela, l'approche la plus utilisée reste l'annotation textuelle et manuelle
- Avantage :
 - ▶ recherche indépendante du type de media
 - ▶ techniques de base de données classiques
- Inconvénients
 - ▶ Le coût d'annotation manuel est très important
 - ▶ Différentes personnes utilisent un vocabulaire différent pour signifier la même chose (ex: clair, lumineux). Connue sous le nom de “problème du vocabulaire”. L'utilisation d'un thésaurus ne règle pas tout.
 - ▶ Différentes personnes peuvent décrire des aspects différents du média (polysémie du contenu), la même personne décrira différents aspects en fonction de la situation.

Inria



Exemple d'annotation

- Charlotte et Médor pendant les vacances de Pâques
- footing femme chien pré
- mon programme minceur pour l'été



Inria



Systèmes d'annotation : musique/image

- Musique
 - ▶ pas de tag prévu dans le format originel CD
 - ▶ Mais utilisation d'un identifiant unique par CD (CDID)
 - ▶ couplé avec un serveur qui associe des tags à chaque CDID
<http://www.freedb.org> → annotation par les utilisateurs finaux
 - ▶ MP3 : utilisation de ID3
- Images
 - ▶ annotation Exif, IPTC

Inria



ID3 Tags

- Conçu en 1996 et devenu *de facto* un standard
- 128 octets

Position	Taille	Contenu
0	3	Marqueur 'TAG'
3	30	Titre
33	30	Interprète
63	30	Album
93	4	Année
97	30	Commentaire
127	1	Genre musical

Valeur numérique	Genre
0	Blues
1	Classic rock
2	Country
3	Dance
4	Disco
5	Funk
6	Grunge
7	Hip-Hop
8	Jazz
9	Metal
10	New Age
11	Oldies (rétro)
12	Autre
13	pop

- Manque de souplesse → de nombreuses extensions (ID3v1.1, ID3v2) ...
- ... mais qui ne pourront jamais répondre à tous les besoins

Inria



Exchangeable image file format (Exif)

- Spécification pour les formats d'images des appareils numériques
 - ▶ non géré par une organisation internationale, mais largement utilisé
- Pour JPEG, TIFF, RIFF. Ne supporte pas JPEG2000, PNG ou GIF
- Le format supporte
 - ▶ Date et heure, enregistrés par l'appareil
 - ▶ Les paramètres de l'appareil
 - Dépendent du modèle : inclus la marque et des informations diverses telles que le temps d'ouverture, l'orientation, la focale, l'ISO, etc.
 - ▶ Une vignette de pré-visualisation
 - ▶ La description et les informations de copyright
 - ▶ coordonnées GPS
 - ▶ ...
- Supporté par de nombreuses applications

Inria



Ambiguïté de l'annotation d'images

- Difficile de se mettre d'accord sur les annotations
- Instructions (Pascal 2009)



What to label	<i>All objects of the defined categories, unless: you are unsure what the object is. the object is very small (at your discretion). less than 10-20% of the object is visible. If this is not possible because too many objects, mark image as bad.</i>
Viewpoint	Record the viewpoint of the 'bulk' of the object e.g. the body rather than the head. Allow viewpoints within 10-20 degrees. If ambiguous, leave as 'Unspecified'. Unusually rotated objects e.g. upside-down people should be left as 'Unspecified'.
Bounding box	Mark the bounding box of the visible area of the object (<i>not</i> the estimated total extent of the object). Bounding box should contain all visible pixels, except where the bounding box would have to be made excessively large to include a few additional pixels (<5%) e.g. a car aerial.
Truncation	If more than 15-20% of the object lies outside the bounding box mark as Truncated. The flag indicates that the bounding box does not cover the total extent of the object.
Occlusion	If more than 5% of the object is occluded within the bounding box, mark as Occluded. The flag indicates that the object is not totally visible within the bounding box.

Inria



Ambiguïté de l'annotation de vidéos...

- Instructions (Trecvid MED 2012)



Event Name	Giving directions to a location
Definition:	One or more people give directions to one or more other people, either in person or over the phone, by explaining verbally and/or with gestures how to get to a particular location.
Explication:	People may give directions in response to being asked for them, or they may give them without being asked as a part of a normal conversation if the topic of conversation is a location (e.g. telling a friend how to get to a new restaurant that just opened that she may be interested in eating at). People may ask for directions from strangers they see on the street or in a store, or call an information service or someone they know. People may ask for directions over the phone. Or, the person giving directions and the one getting directions may be together, and one person is serving as the navigator while the other(s) follow the directions, such as commonly happens when the person getting directions is driving and the person giving directions is reading them from a map, printout, or smart phone. Note that a GPS giving directions is not relevant for this event, and that the person giving directions must be visible. If people are visiting a new city or country, they will often refer to a guidebook with them to reference when asking for directions from a person on the street. Depending on whether the people asking for directions are walking, driving, taking a subway, train, or bus, the directions given may reference city blocks, highways, or subway/train/bus directions often gesture along with their directions, for example pointing their arm to the right and turning their head to the right as they speak, and this could be done even if giving them on the phone. People giving directions on the phone may stay on the line with the person they are helping while they reach their destination, giving them the directions step by step in real time.
scene:	outdoors, indoors
objects/people:	map, driver, car, bicycle, subway, train, bus, pedestrian, passenger, guidebook, portable telephone
activities:	gestures indicating directions (e.g. pointing or extending arm straight/to right or left of speaker), person pointing out location on a map, head movement indicating direction
audio:	narration of directions



Inria

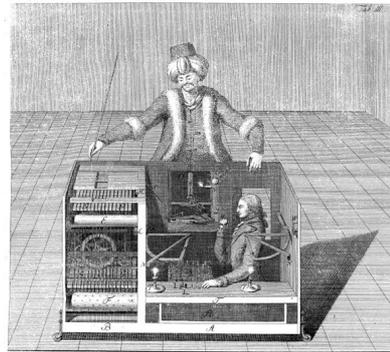
Grenoble INP ENSIMAG

Label Me

<http://labelme.csail.mit.edu> (MIT)



- Annotation précise/fastidieuse...
- Amazon Mechanical Turk



Inria

Grenoble INP
ensimag

Recherche approximative/par similarité

- L'idée clé dans la recherche de média est la recherche approximative
- Utilise la notion de proximité, de similarité, de distance entre objets multimedia
- Généralement, la requête est exprimée sous la forme d'un ou de plusieurs vecteurs dans un espace multi-dimensionnel.
 - ▶ définition d'une distance (ou mesure de dissimilarité) sur cet espace
 - ▶ recherche des objets dont la distance est minimale
- Les vecteurs d'attributs sont extraits de contenu de l'image

Recherche par le contenu (CBIR: content based information retrieval)

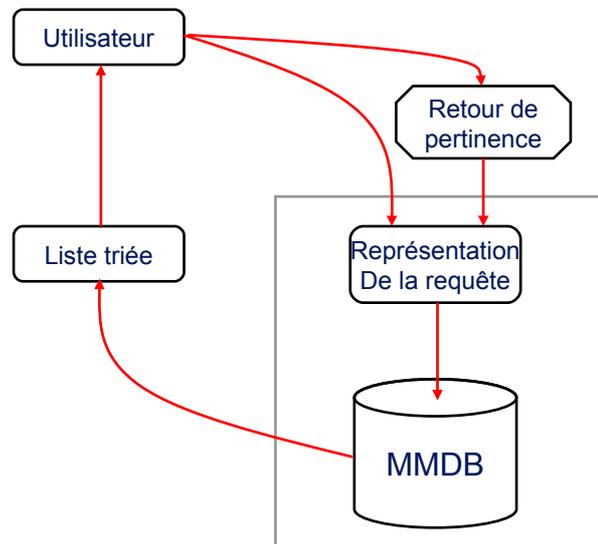
- Possibilité d'utiliser un retour de pertinence

Inria

Grenoble INP
ensimag

Retour de pertinence

- Possibilité de mettre l'interaction au coeur du système
 - ▶ parmi les réponses, le système identifie ce qu'il trouve intéressant
 - ▶ puis affinage de la requête par l'utilisateur



Inria

Grenoble INP
ensimag

Indexation

- Structurer la base pour ne pas avoir à la parcourir entièrement à chaque recherche, comme dans un dictionnaire
- Nécessite de décrire le contenu avec des index portant sur les médias
- Difficulté : les médias en question sont représentés par des vecteurs de grande dimension

The dimension curse

La malédiction de la dimension

- Nous reviendrons largement sur cet aspect dans la suite de ce cours

Inria

Grenoble INP
ensimag

Moteurs de recherche basés sur le contenu

-
- Critère de similarité global
 - ▶ QBIC (Query By Image content) – IBM 93
 - ▶ Virage 93
 - ▶ Photobook – MIT'93
- <http://pixolu.de> : organisation/recherche par couleur + forme globale



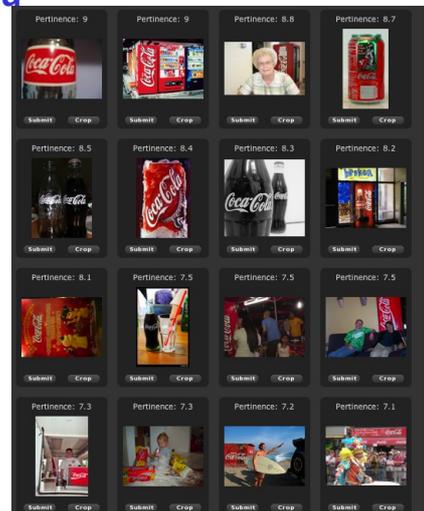
- <http://tineye.com> → recherche d'images sur le web

Inria



Moteurs de recherche basés sur le contenu

- <http://bigimbaz.inrialpes.fr>
recherche par description locale



- <http://www.google.com/mobile/goggles>
 - ▶ OCR
 - ▶ bâtiments
 - ▶ peintures
 - ▶ logos...

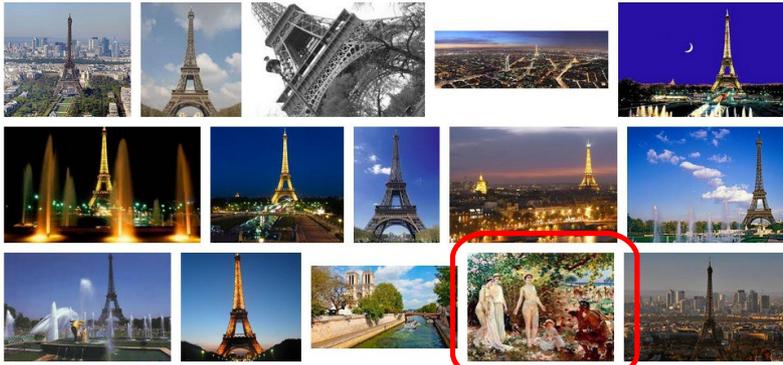


Inria



Moteurs de recherche basés sur le contenu

- Google similar image search (texte + retour de pertinence image)



- Recherche: Paris

- Similarité visuelle



- Shazam reconnaît la musique

Inria



Fonctionnalités d'un système de gestion de MMDB idéal

- Ce qu'il doit permettre
 - ▶ Ne pas reposer (uniquement) sur des annotations manuelles
 - ▶ Le processus de requête est interactif
 - ▶ Le processus de requête peut utiliser plusieurs modes de représentation (ex: texte, puis image)
- Il doit posséder deux modes
 - ▶ Gestion du contenu (ajout, modification des objets présents, suppression)
 - ▶ Interrogation de la base
- Nécessité d'une interface utilisateur permettant la manipulation des objets existants, la présentation de résultats, le retour de pertinence

Inria

