



NUMÉRIQUE I Pour améliorer l'exploration des bases de données multimédias, sur le web notamment, les interfaces doivent se faire plus intuitives et les systèmes d'indexation être capables d'extraire automatiquement les contenus. ■ PAR AURÉLIE BARBAUX

Les défis du moteur de recherche multimédia

Ce qui doit être amélioré

➤ **La reconnaissance vocale** pour transcrire le contenu des enregistrements audio et vidéo

➤ **L'extraction automatique** d'informations contenues dans des fichiers multimédias : objets, concepts, personnages, musiques...

➤ **L'analyse sémantique** pour affiner les recherches par concept et pallier les erreurs de transcription

➤ **Les interfaces hommes-machines** pour s'affranchir du clavier

Plus de 65 000 nouvelles vidéos publiées chaque jour sur le site Youtube.com, acheté par Google pour 1,65 milliard de dollars ! 3,6 millions d'heures d'émissions télé et radio dans les archives de l'INA ! Et aucun moyen d'effectuer une recherche sur le contenu réel de ces documents. Les moteurs de recherche ne savent qu'indexer les informations textuelles associées à ces documents dans des étiquettes électroniques (on parle aussi de métadonnées ou de signature). Enrichies automatiquement pour les informations techniques comme la date de création ou le format, ces étiquettes nécessitent encore une intervention humaine pour y inclure un peu d'intelligence, via une légende ou des mots clés. Et même si les technologies du Web 2.0 permettent de solliciter les internautes pour faire le travail, on est loin d'une solution universelle. Industriels et instituts de recherche multiplient donc les programmes pour développer les technologies nécessaires à la création d'un moteur de recherche multimédia. Le plus médiatique, Quaero, attend le feu vert de Bruxelles et les financements promis pour démarrer. D'autres, comme Infomagic en Ile-de-France, sont un peu plus avancés. Mais, vue l'ampleur de la tâche, toutes les énergies sont les bienvenues. Il s'agit en effet de revisi-

ter les interfaces de recherche pour des milliers de requêtes simultanées et de développer les technologies d'extraction automatique des contenus pour des milliards de documents.

L'EXTRACTION AUTOMATIQUE DES CONTENUS

Ancré dans nos habitudes, le principe de la recherche par mots clés devrait persister, mais ne sera plus l'unique porte d'entrée. « Nous voulons proposer de nouvelles stratégies de recherche, adaptées au profil de l'utilisateur », explique Philippe Martin, responsable du projet Infomagic pour Thales Land & Joint.

Déjà, des moteurs comme celui du français Exalead proposent de fonctionner par association d'idées pour affiner une requête. Le fonctionnement par proximité est encore plus intuitif. Un prototype, développé par le chercheur viennois Andi Rauber propose une navigation sonore à partir d'une interface visuelle rapprochant des morceaux de musique suivant leur empreinte de

fréquences. Il a ainsi cartographié toute l'œuvre de Mozart.

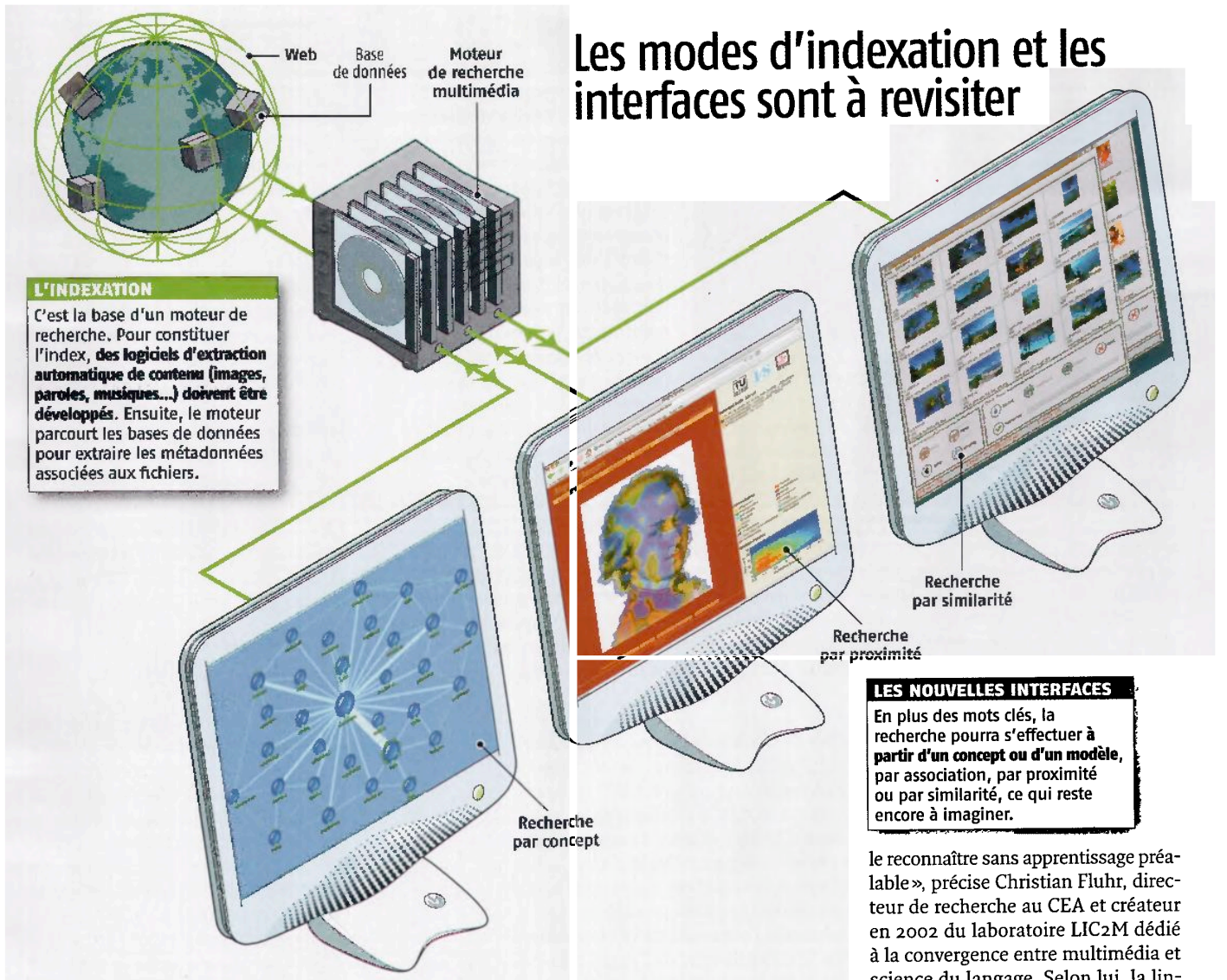
De leur côté, l'Inria et le CEA ont développé des systèmes de recherche d'images par similarité, qui ont déjà donné lieu à la création de deux entreprises innovantes, LTU et New Phenix. « Mais ces systèmes, basés sur des méthodes d'auto-apprentissage, ne fonctionnent bien que sur des corpus définis », remarque Pierre-Alain Moëllic, chercheur au CEA et responsable du programme d'évaluation ImageVal. Surtout, ces technologies se limitent à extraire des informations liées à la texture, la forme ou les couleurs, pour enrichir la signature d'une image avant l'étape de comparaison.

« Il faut maintenant dépasser le fossé sémantique », prévient Nozha Boujemaa, directrice de recherche à l'Inria et responsable du projet Imedia. Car tous les chercheurs sont d'accord : pour obtenir des réponses pertinentes, il faut nommer les éléments contenus dans les documents. Ainsi, l'indexation des documents multimédias, indispensable au fonctionnement des moteurs de

Sur une image, on reconnaît déjà une scène d'intérieur ou d'extérieur, de jour ou de nuit, en ville ou à la campagne.



Les modes d'indexation et les interfaces sont à revisiter



recherche, pourra répertorier des informations beaucoup plus complètes. Le principal défi de l'indexation multimédia relève donc de l'extraction automatique du contenu des documents pour s'affranchir au maximum de l'intervention humaine. Pour les images, on parle de CBIR, pour Content Based Image Retrieval. «Aujourd'hui, on sait reconnaître s'il s'agit d'une scène d'intérieur ou d'extérieur, de jour ou de nuit, en ville ou à la campagne. Mais on ne sait pas identifier des objets précis sans avoir, au préalable, proposé des modèles similaires», reconnaît Pierre Alain Moëllic. En revanche, la détection de zones de

Des applications très variées

- **RECHERCHE** d'informations sur internet
- **VALORISATION** des fonds patrimoniaux: archives photos, vidéo, sonores...
- **RENSEIGNEMENT** et sécurité nationale
- **VEILLE** économique
- **PROTECTION** de la propriété intellectuelle et industrielle
- **EXPLORATION** des données médicales
- **ROBOTIQUE** pour la perception de l'environnement

textes ou d'images transformées (par comparaison avec un original) est assez bien maîtrisée. On sait aussi déceler la présence d'un personnage. « Mais pas

le reconnaître sans apprentissage préalable», précise Christian Fluhr, directeur de recherche au CEA et créateur en 2002 du laboratoire LIC2M dédié à la convergence entre multimédia et science du langage. Selon lui, la linguistique diminuerait le taux d'erreur des systèmes de reconnaissance de contenu. Sur le web, l'une des techniques consiste à travailler sur la mise en contexte des œuvres en analysant le texte environnant. Cette approche ne peut être que complémentaire. «On réfléchit aussi à la propagation automatique des annotations d'une image vers une autre», prévient Henri Maître, directeur adjoint de Telecom Paris. Les technologies d'extraction automatique existantes sont en fait loin d'être au point. Ainsi, la reconnaissance vocale montre ses limites. «Le taux d'erreur n'est que de 4 % sur la voix d'un journaliste de radio, mais Suite page 82 ▶



Chercheurs et industriels européens se réveillent



F. PALLACCI/STUDIO

Ambitions. Pour Nozha Boujemaa, directrice de recherche à l'Inria, responsable du projet Imedia et coordinatrice de programmes européens : « il faut dépasser le fossé sémantique ».

4500 développeurs chez Google, environ 1 milliard de dollars investis chaque année par l'administration américaine dans les technologies d'analyse de l'information pour les besoins de la sécurité nationale... les Etats-Unis disposent de moyens colossaux pour développer les futurs moteurs de recherche multimédia. Un éditeur, Virage, vient même de sortir une solution d'indexation pour les vidéos. L'Europe, elle, ne s'est réveillée que depuis deux ou

trois ans côté multimédia... Les premières initiatives pour éviter l'éparpillement des forces ne datent que du 6^e Programme cadre de recherche (PCRD), notamment avec le réseau d'excellence Muscle, pour partager les ressources. A partir de novembre 2006, un volet du programme Corus (Coopération pour la recherche universitaire et scientifique) organisera l'échange d'expériences des trois projets européens de moteur de recherche multimédia, le franco-allemand Quaero, le norvégien IAD (avec l'éditeur Fast) et celui de l'université d'Amsterdam. Les industriels européens ont enfin pris conscience des enjeux. En France, outre Thomson, initiateur de Quaero, Thales pilote le projet Infomagic. De son côté, la DGA est partie prenante dans des programmes sur la reconnaissance vocale (Ester) et d'images (Imageval). Objectif : améliorer les techniques du renseignement.

dix ans», reconnaît Henri Sansom, responsable d'unité de recherche chez France Télécom R & D. Pour la vidéo, le traitement du son est donc utilisé en complément de l'analyse des flux, décomposés pour revenir aux technologies de l'image fixe. Les applaudissements ou les passages musicaux servent au séquençage des scènes pour les émissions. L'analyse de la voix compléterait plutôt les tentatives de reconnaissance de personnages, réalisées à partir de l'analyse de points fixes des visages (coins des yeux, de la bouche). « Mais il faut bien accepter qu'à un moment il y aura encore une intervention humaine, pour définir à qui appartient l'image de référence », rappelle Jean Carrive du service de recherche de l'INA. Dans le cadre du programme Infomagic, il réfléchit en effet à des systèmes de recherche à partir de la photographie d'une ou plusieurs personnes pour retrouver les documents sur lesquels elles sont présentes. On en est encore loin, comme de la transcription totale du script d'un film!

L'ÉTENDUE DU DOMAINE DE RECHERCHE

Le dernier grand défi technologique est le changement d'échelle. « Il s'agit de préparer des systèmes capables d'indexer des dizaines de milliards de documents d'ici à quelques années, et de répondre à des milliers de requêtes simultanées », explique François Bourdoncle, le P-DG d'Exalead, membre du consortium Quaero. L'Irisa y travaille déjà dans son projet Texmex. Mais il faut aussi étendre la recherche aux archives. « Si la numérisation est devenue triviale, le problème réside dans le volume des données », rappelle Henri Maître. Le problème est aussi financier. « Pour numériser les 25 millions d'ouvrages de la Bibliothèque nationale, il faudrait 400 millions d'euros! », prévient Benoît Drigny, directeur de l'activité numérique chez Jouve, autre membre de Quaero. Deux autres défis attendent encore les chercheurs. Pour être universel, ces nouveaux moteurs de recherche devront inclure des solutions de traduction simultanée et tenir compte du passage à la 3D. Si des travaux sont en cours pour le premier, rien n'apparaît pour le second. ■

► Suite de la page 81 en situation de micro trottoir, le taux dépasse 20 %, soit plus d'un mot sur cinq non reconnu », explique Martine Garnier Rizet, la vice-présidente de Vecsys, spin-off du CNRS.

Après le projet de moteur de recherche vocale Audiosurf, sur un corpus d'émissions de Radio France, des travaux sont donc en cours sur la transcription de la parole spontanée, à partir d'enregistrements réalisés dans ses centres d'appels d'EdF. Or, les systè-

mes de reconnaissance vocale fonctionnent par apprentissage. Pour chaque langue, il faut au préalable réaliser des transcriptions manuelles de milliers d'heures d'enregistrement pour alimenter les modèles probabilistes. De plus, pour de bons résultats, il faut au minimum quatre passes, soit quatre fois le temps réel, ce qui empêche un traitement au fil de l'eau. Un frein de taille pour indexer les émissions diffusées sur les médias. « Nous travaillons sur le sujet depuis au moins

LES REPÈRES DE «L'USINE NOUVELLE»

À RETENIR

► **La recherche actuelle** ne porte que sur des données textuelles associées aux documents : informations techniques (date, formats...) et légendes ou mots clés introduits manuellement.

► **La reconnaissance vocale** bute sur des problèmes de bruits et de temps de réponse.

► **L'extraction de sens** n'en est qu'au stade expérimental.

SUR LE NET

<http://www.usinenouvelle.com>

Complétez vos informations sur notre site :

► Les programmes de recherches Imageval (Nicephore Cité, CEA), Imedia (Inria), Infomagic (Thales, Supélec, INA)...

► Les prototypes AudioSurf (Sinequa), Marvel (IBM), Mozart (Muscle)...

► Les entreprises innovantes Advestigo, IUT, NewPhenix, Vecsys...