

noir
SUR
blanc

page 2
éditorial

pages 3, 4 & 5

dossier **Indexation multimédia**

page 7
forum

pages 8 & 9

avancées

**Indexation
et recherche
d'information à l'IRIT**

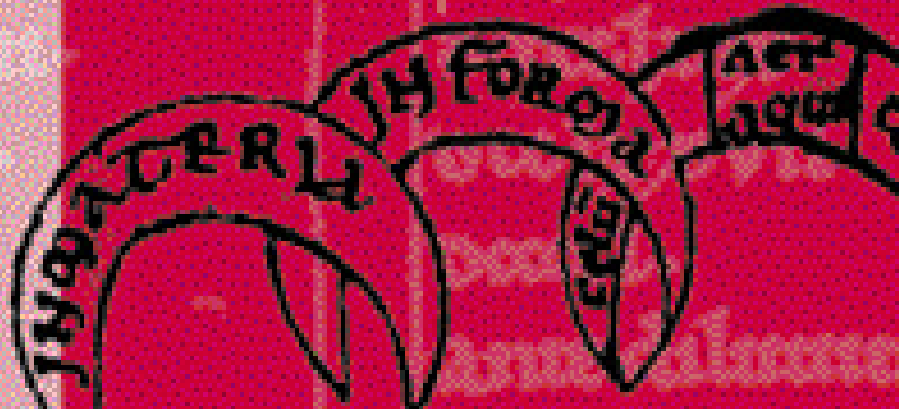
pages 10 & 11
événements

pages 6 & 12

l'invité Bruno Bachimont



Le Chronicon de Hugues de Saint-Victor, Paris, Bnf, Mss, Lat. 15009 In Tous les savoirs du monde, Bnf, Flammarion, 1996..



Bien que dans les grands gisements de documents, par exemple dans les grandes bibliothèques, nous soyons déjà en présence de quantités énormes d'information de natures très variées (et maintenant aussi numérisées), la nouveauté des bibliothèques numériques est de comprendre, sous une même communauté de formes (numériques en concurrence), plusieurs types de médias relevant chacun d'une logique de sens très différents.

Dans ce nouveau contexte, la définition de méthodes automatiques d'extraction d'information de nature diverse (textes, dessins, vidéo, etc.) devient un défi scientifique dans l'aire du document électronique.

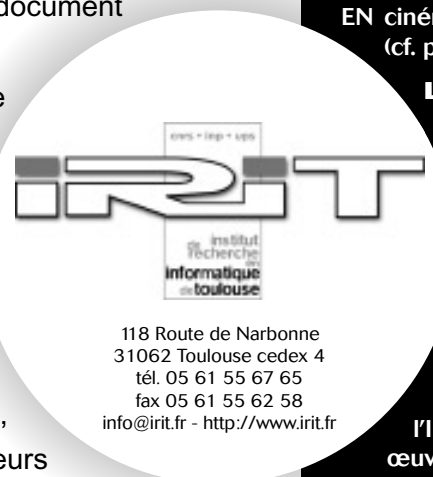
Cette thématique fait l'objet du *dossier* de ce numéro de **noir sur blanc** : elle est l'une des priorités de notre laboratoire et elle traverse les activités scientifiques de plusieurs équipes (cf. les *avancées*).

Par ailleurs, nous nous intéressons dans notre laboratoire non seulement à la mise en place de méthodes pour l'extraction des descripteurs, mais aussi à l'usage qui est fait de ces descripteurs dans l'extraction d'information de grands fonds d'archives textuels et audiovisuels. Ce qui est loin d'être un problème simple, comme l'indique notre *invité* Bruno Bachimont, directeur scientifique à l'INA.

Une « sémantique » et une « pragmatique » unifiées ou du moins articulées du multimédia ont certainement un vaste horizon de recherches théoriques et applicatives devant elles.

Luis Fariñas del Cerro

À propos de l'illustration de couverture : « *La page du Chronicon est divisée en colonnes, comme celles de la façade d'un temple ; les éléments y sont rangés de façon que l'on puisse plus aisément se les rappeler, sachant où ils se trouvent, et des couleurs sont employées pour structurer les séries qu'on doit en outre pouvoir mémoriser verticalement ou horizontalement. Encyclopédie visuelle, le Chronicon l'est ainsi jusque dans l'organisation de sa page.* » Tous les savoirs du monde, Bnf, Flammarion, 1996.



Espace ouvert d'échange et de partage, les passerelles de l'IRIT proposent des rencontres entre la recherche en informatique et le monde culturel et socio-économique. C'est dans ce cadre que s'inscrivent, entre autres, les rencontres recherche-industrie, celles avec les chercheurs d'autres disciplines scientifiques, le dialogue art-science et toutes les opérations de diffusion de la culture scientifique et technique.

Au programme des Passerelles de l'IRIT, en ce mois d'octobre 2002 :

La Fête de la Science 2002 en Midi-Pyrénées a choisi pour thème le transport.

Dans ce cadre l'IRIT a proposé : 3 animations pour le public scolaire autour du thème Le transport de l'Information (cf. p. 7)

et la 8^e édition de expérimentation EN cinéma : Transports d'images (cf. p. 7).

Le SITEF

On pourra constater, dans la rubrique *Événements*, la forte présence de l'IRIT dans l'édition 2002 du SITEF à travers des démonstrations, un colloque, ainsi que des conférences et une exposition.

À travers les Passerelles, l'IRIT poursuit la mise en œuvre de son projet de diffusion de la culture scientifique et technique, de partage des savoirs et de dialogue avec le monde socio-économique et culturel.

Le programme des Passerelles d'octobre 2002 en est une belle illustration.

La Rédaction

Directeur de la publication

Luis Fariñas del Cerro

Directrice adjointe de la publication

Colette Ravinet

Secrétariat de rédaction

Véronique Debats

Comité de rédaction

Régine André-Obrecht, Vincent Charvillat, Jérôme Lang, Mustapha Mojahid, Gérard Padiou, Pascal Sainrat, Patrick Sallé, Jacques Virbel

Maquette

Lestang Création

Contact de la rédaction

05 61 55 65 10 – nsb@irit.fr

ISSN 1620-2937

Indexation Multimédia trouver pour retrouver

Comment trouver une information lorsque celle-ci est noyée au milieu d'une gigantesque quantité de données hétérogènes ? L'indexation automatique s'attache à développer des méthodes et des systèmes qui, à défaut de se substituer à un catalogage manuel des ressources, lorsque celui-ci existe, offre des moyens substantiels pour accéder à tout type de document numérique.

Afin de retrouver un document numérique au milieu d'une très grande base d'information à l'aide d'un outil informatique, l'indexation multimédia se propose de construire un catalogue structuré autour d'un ou plusieurs index, sur un principe équivalent à celui des archives et des bibliothèques. Or la construction de ces index pose les problèmes de la définition des descripteurs : éléments pertinents pour la description des contenus et la satisfaction ultérieure de requêtes d'accès, de leur identification et de leur localisation de manière systématique et consistante dans l'ensemble des données archivées. L'automatisation de ces tâches, particulièrement coûteuse en calculs et en infrastructures techniques, doit s'accompagner de stratégies de partage des ressources, des algorithmes et des résultats malgré la nature hétérogène des contenus textuels, audio et vidéo. Par exemple, un des multiples points abordés par la recherche sur ce sujet concerne la gestion l'impact d'un descripteur en termes de fiabilité, de discriminance ou de sémantique sur les descripteurs qui en héritent dans un processus d'instanciation automatique d'index.

Construire un index nécessite dans un premier temps de prendre en compte les caractéristiques de chacun des médias, leurs modes de production, leurs transferts sur différents supports (ou plus généralement, leurs modes de transmission), et les propriétés liées à leur perception. De ces caractéristiques découlent des propositions d'indexation propres à ces médias.

Indexation et médias

Ainsi, le texte peut-il être indexé en fonction de thésaurus définis a priori lorsque le contenu est spécifique, ou bien encore à l'aide d'outils statistiques évaluant la pertinence de chacun des mots à figurer dans un index. La mise en page, le choix des polices de caractères ou encore les titres sont autant d'éléments de mise en exergue d'un contenu, susceptibles également d'être pris en compte par les processus d'indexation automatique. Etant le mode de formulation de requête privilégié, la langue naturelle est l'objet de nombreux travaux de recherche. Il s'agit non-seulement d'extraire du texte l'information sémantique, mais aussi de la représenter sous une forme symbolique.

Cette technologie est également étroitement liée à la transcription automatique de la parole présente dans les documents audio. Ce processus, encore coûteux en calcul, permet d'offrir des moyens d'accès aux contenus audiovisuels ou purement sonores particulièrement efficaces. D'autres outils exploités pour l'indexation des documents audio, effectuent la détection de mots figurant dans un lexique prédéfini, ou alignent un script sur une bande son pour retrouver, par exemple, la partie d'un document audiovisuel correspondant à un dialogue. L'identification des locuteurs est également une technologie utilisée pour retrouver les différentes interventions d'un personnage donné. Parmi les travaux les plus récents, on trouve des outils de détection de parole ou de musique, ou encore d'identification de la langue pour l'indexation des contenus multilingues. De façon complémentaire, de nombreux travaux sur l'analyse des contenus musicaux autorisent aujourd'hui

l'élaboration de systèmes d'interrogation où, par exemple, l'utilisateur chantonne l'air du document qu'il recherche.

On considère généralement que l'image et la vidéo numériques sont porteuses d'une information sémantique moins importante pour la construction d'index lorsqu'elles sont couplées aux deux précédents médias. Néanmoins, différents outils sont susceptibles d'extraire des informations précieuses pour l'indexation comme la détection et l'identification d'objet selon des méthodes le plus souvent développées pour la télésurveillance (identification de visages, localisation de personnages, suivi de véhicules), la recherche de zones de texte dans l'image et la transcription de ce texte au format ASCII, l'analyse de paramètres de production (l'éclairage, le cadrage, le montage, les mouvements de caméra), etc. La recherche a développé par ailleurs des outils très spécifiques aux contenus images, visant à identifier des signatures pour les requêtes par l'exemple reposant sur la distribution de la couleur, sur des points d'intérêts, des informations spectrales, les contours et les régions...

Citons ici, pour terminer cette liste, l'indexation des hyperdocuments, où la présence spécifique de liens renvoyant à d'autres documents, et l'usage qui en est fait, peut être utilisée pour rendre compte non-seulement de l'intérêt d'un contenu, mais aussi de la nature même de ce contenu.

Indexation et pluridisciplinarité

Reposant à la fois sur les méthodes des métiers de la production et sur les usages de consommation de contenus extrêmement diversifiés, l'indexation multimédia est une discipline éminemment pluridisciplinaire.

Si celle-ci s'appuie sur les techniques de production et les modes de perception des contenus, elle est souvent le point central de projets dans lesquels elle joue le rôle d'outil de validation dans le cadre d'une activité de création artistique, ou dans le cadre de l'élaboration de document conformément à une charte donnée. Elle est également un point d'ancrage pour l'évaluation de l'impact de certains paramètres sur la perception dans un contexte d'expérimentations liées à l'étude de la cognition.

Elle offre des éléments pour réaliser de véritables audits sur les accès aux contenus, notamment en matière de couverture des descriptions et des cheminements suivis dans une arborescence documentaire.

À travers la discrétisation temporelle des contenus audiovisuels, l'indexation multimédia permet de développer des systèmes de consultation reposant sur la notion d'objet informationnel, proche des théories structuralistes, et particulièrement utile pour développer une interaction transmodale.

D'un autre côté, il apparaît aujourd'hui que l'exploitabilité et la teneur sémantique des index produits automatiquement sont susceptibles d'être fortement améliorées par le couplage des informations issues de sources hétérogènes comme par exemple la structure de la bande son avec la structure de la vidéo dans un film.

Cette « transmédialité » nécessite non seulement le regroupement d'expertises sur les différents médias, mais aussi le développement d'outils génériques pour la combinaison des informations tels que la fusion ou l'agrégation de descripteurs, ou encore la fouille de données multimédia.

De nombreux travaux sont menés actuellement sur des outils de développement d'ontologies afin d'améliorer la qualité sémantique des index en spécialisant l'information en fonction d'un usage (d'un point de vue particulier), ou d'un type de contenu. Ces nouvelles orientations visent le développement de mécanismes d'indexation dynamique avec mise à jour en fonction de l'apparition de nouveaux outils ou de nouvelles applications sur des plate-formes d'indexation. La description devient de plus en plus symbolique et générique, c'est-à-dire non-spécialisée pour un usage particulier. La spécialisation est reléguée au moment de la présentation des résultats.

L'indexation multimédia automatique permet aujourd'hui de fournir des index de nature essentiellement numérique ou symbolique, dont la teneur sémantique est très éloignée de celle qu'on trouve dans les catalogues traditionnels. L'état de l'art actuel situe cette discipline comme porteuse d'une technologie non-directement concurrentielle des métiers de l'archivage, mais offrant une pépinière d'outils complémentaires à ceux de ce domaine, prometteuse en matière de développement de nouveaux services et de nouveaux usages. Car il ne s'agit pas seulement de promouvoir les outils d'accès au numérique, mais de légitimer la production de contenus diversifiés à l'attention du plus grand nombre et pour le plus grand éventail d'utilisations possibles. L'essor important de l'indexation multimédia se traduit aujourd'hui par le développement de nouveaux standards de description des contenus, préalable indispensable à une diffusion plus large et plus efficace des documents.

Entre l'utile et le faisable : le syndrome de l'aiguille et du lampadaire

entretien avec
Bruno Bachimont

Le numérique a profondément modifié le rapport aux contenus audiovisuels. Désormais matérialisés dans un format se prêtant au calcul, les documents audiovisuels peuvent être analysés par différents outils automatiques, mettant la puissance du calcul au service des contenus.

Parmi les outils automatiques faisant l'objet de recherche, nombreux sont ceux qui portent sur l'extraction automatique de descripteurs. Outre les difficultés inhérentes à ce type d'extraction, se pose également le problème de savoir à quoi peuvent servir ces descripteurs. Et la question n'est en rien évidente. Pour l'aborder, mettons nous dans le contexte des fonds d'archives ou des bibliothèques audiovisuelles.

Dans le domaine des bibliothèques et fonds d'archives audiovisuels, le problème essentiel est, comme souvent, d'organiser le classement pour optimiser la recherche des contenus et leur exploitation.

On décrit alors ce que contiennent les documents et où ils sont rangés pour pouvoir les retrouver : c'est la problématique de l'indexation.

Dans une bibliothèque physique, les objets manipulés sont les documents dans leur intégralité (des livres par exemple) et il n'est pas possible ni utile de considérer leurs parties car on ne sait pas les manipuler indépendamment de l'objet qui les contient : retrouver un chapitre, c'est sortir le livre qui le contient. C'est la problématique de l'indexation et de la documentation classiques. Cependant, à l'échelle d'un livre, grâce à son format permettant une pagination, des renvois et des localisations précis, des index désignant une partie particulière du document

sont progressivement intégrés aux contenus : ce sont entre autres les index, tables de matières, notes, etc. Mais, la référence à un ouvrage dans une bibliothèque, et la référence à une partie d'un ouvrage restent des tâches distinctes, la première au service du bibliothécaire, la seconde au service du lecteur. Bien qu'il ne faille pas les opposer radicalement, ces outils renvoient à des systèmes techniques distincts et non intégrés.

Avec le passage au numérique, il devient possible de désigner et manipuler n'importe quelle partie arbitraire d'un contenu indépendamment de ce qui la contient. Il devient utile et nécessaire de décrire toutes les parties présentant un intérêt et de les localiser avec précision pour être capable de les retrouver, manipuler et exploiter. C'est la problématique de l'indexation fine du contenu.

Dans ce contexte, il n'y a pas de solution de continuité entre les parties arbitraires d'un ouvrage et la collection des ouvrages : un même système d'indexation permet à la fois l'organisation de la bibliothèque (index qualifiant les ouvrages dans leur globalité) et la navigation dans les contenus (index qualifiant et référençant des contenus arbitraires d'un ouvrage). Il s'agit là d'une évolution majeure : dans les bibliothèques physiques, si la forme de codex des ouvrages a bien évidemment permis de référencer et localiser des contenus, cela n'a pas pour autant contribué à enrichir les catalogues des bibliothèques. Le numérique permet en revanche d'intégrer les deux, c'est-à-dire de pouvoir profiter, pour la gestion de la bibliothèque des outils de navigation lectoriale comme les index.

Information et intelligence : raisonner et décider

Le dossier « Raisonement et Décision » publié dans le numéro 5 de **noir sur blanc** montre bien l'importance de ce thème – en intelligence artificielle et plus généralement en informatique – et illustre de façon claire la variété des types de problèmes à résoudre.

L'importance des problématiques jointes du raisonnement et de la prise de décision dans le développement de l'informatique de demain s'est justement traduite récemment par la création, au sein du département STIC du CNRS, d'un réseau thématique pluridisciplinaire « Information et intelligence : raisonner et décider ».

Ce réseau est constitué des laboratoires de recherche français concernés par la thématique et centré autour d'un noyau d'une petite dizaine d'entre eux.

L'objectif poursuivi par les chercheurs du réseau est de concevoir et de réaliser des systèmes intelligents automatisés qui à partir d'informations disponibles de nature diverse (lois générales, observations, croyances, désirs, effets connus d'actions...)

sont capables d'effectuer des raisonnements afin d'aboutir à la prise de décisions intelligentes.

Afin d'atteindre cet objectif, plusieurs difficultés sont à surmonter. La première difficulté concerne l'imperfection des données disponibles.

En effet, les données sont souvent incomplètes, incertaines, imprécises, évolutives, bruitées, contradictoires et il s'agit d'abord de choisir un formalisme de représentation adéquat.

La deuxième difficulté réside dans le choix de l'inférence à utiliser pour exploiter au mieux les informations disponibles ; outre la déduction en logique classique, d'autres types de raisonnement peuvent être nécessaires comme, par exemple, le raisonnement par défaut, le raisonnement tolérant les exceptions ou le raisonnement paraconsistant.

Le choix d'une stratégie de prise de décision peut également se révéler problématique, en particulier, lorsque plusieurs objectifs plus ou moins incompatibles sont visés.

Par ailleurs, le traitement automatique de la prise de décision et du raisonnement conduit au respect des contraintes imposées par l'informatique, à savoir les limitations en temps et en espace. Il importe de mettre au point des algorithmes les plus efficaces possibles, d'une complexité en temps et en espace acceptable, pour traiter des instances de problèmes de taille réaliste.

Les travaux développés au sein du réseau thématique pluridisciplinaire « Information et intelligence : raisonner et décider » visent à apporter des solutions aux différentes difficultés soulevées. Ils intègrent le développement de modèles et de langages de représentation pour le traitement d'informations imparfaites mais aussi la conception de méthodes les plus efficaces possibles pour l'inférence et la prise de décision.

Pierre Marquis et Odile Papini
Animateurs du RTP 11
« Information et intelligence :
raisonner et décider »

Animations et démonstrations pour le public scolaire

Le transport de l'information 17 et 18 octobre

L'histoire des machines parlantes des années 50 aux avatars communicants (Équipe DIAMANT)

Cavalcade : un outil de prototypage virtuel collaboratif (Équipe Synthèse d'Images et Réalité Virtuelle)

Variations sur son et lumière (Équipe SYGMA de l'IRIT à l'ENSEEIH)

expérimentation EN cinéma, 8e édition

Transports d'images 16 et 17 octobre en collaboration avec Yann Beauvais, cinéaste et le Service Culture de l'UPS avec le support technique de la Cinémathèque de Toulouse

http://www.irit.fr/MANIFS/EXP_EN_CINEMA

L'IRIT à la Fête de la Science 2002

Indexation et recherche d'information multimédia à l'IRIT

Les travaux menés actuellement à l'IRIT sur divers aspects de l'indexation multimédia portent aussi bien sur l'extraction de nouveaux descripteurs, que sur leur exploitation par des outils de recherche d'information ou de navigation.

Extraction de nouveaux descripteurs

Parmi les nouveaux descripteurs extraits de la vidéo figurent des informations sur le compositing, effet de montage consistant à incruster une image dans une sous-région de l'écran. Ce type d'effet est caractéristique du type de certains documents (jeux télévisés, journaux, émissions sportives, bandes annonces). L'IRIT mène des travaux sur l'analyse de ces effets et leur usage pour la classification. > **Philippe.Joly@irit.fr**

De tous nouveaux travaux ont débuté sur l'analyse du mouvement du corps humain dans une image vidéo. Cette information, rarement exploitée, est pourtant significative pour repérer un geste sportif ou artistique particulier dans un enregistrement. Les premiers résultats permettent de produire une description à différentes échelles d'une posture quelconque sur un enregistrement numérique. > **Thomas.Foures@irit.fr**

En ce qui concerne l'indexation de la bande sonore des documents audiovisuels, un algorithme renseignant automatiquement un descripteur robuste de la présence de parole ou de musique a été développé sur la base de trois nouveaux paramètres : la modulation de l'entropie, la durée de segments obtenus automatiquement et le nombre de ces segments par seconde. Les informations issues de ces trois paramètres sont fusionnées avec celle issue de la modulation de l'énergie à 4 Hz. Le taux d'identification correcte est supérieur à 90 % et l'algorithme possède l'avantage de ne nécessiter aucun apprentissage et donc d'étiquetage manuel. > **Christine.Senac@irit.fr, Julien.Pinquier@irit.fr**

Un autre descripteur audio faisant l'objet de travaux à l'IRIT est la langue utilisée. Deux sources d'informations, l'acoustique et la prosodie, jusqu'alors peu utilisées pour l'identification des langues, sont étudiées. En acoustique, la recherche porte sur une modélisation différenciée des espaces acoustiques et en prosodie sur la définition d'une unité universelle et son exploitation en vue de la discrimination des langues. Les résultats actuels permettent d'obtenir 90 % de classification correcte dès les 5 premières secondes sur 5 langues européennes. > **Regine.Andre-Obrecht@irit.fr, Jerome.Farinas@irit.fr, Jean-Luc.Rouas@irit.fr**

Recherche d'information et navigation

Les descripteurs extraits et index ainsi constitués doivent être intégrés en une représentation unifiée multidimensionnelle. Le but en est d'obtenir un meilleur schéma d'indexation, en unifiant les différentes dimensions, afin de permettre la mise en place de mécanismes d'interrogation, via des requêtes multimédia, éventuellement flexibles. C'est sur cette représentation que repose la mise en œuvre, dans différents contextes, d'hyperbases, regroupant l'ensemble des objets documentaires multimédia et leurs descriptions (par ex., via des métadonnées), encore appelées entrepôts de documents. > **Florence.Sedes@irit.fr, Chantal.Soule-Dupuy@irit.fr**

Se posent dès lors les problèmes d'exploitation de ces données dans de très grandes bases à l'aide de systèmes de recherche et d'exploration d'information.

Le changement d'échelle dans la taille des collections et des documents eux-mêmes nous amène aujourd'hui à reconsidérer les solutions et les outils disponibles pour l'accès à l'information.

Un premier objectif est de fournir des connaissances à l'utilisateur sur le contenu des –grandes– collections qui l'intéressent et de l'aider à formuler son besoin. Les descripteurs et les index extraits fournissent une représentation multi-facettes des documents qui est mise à profit dans des activités de découverte de connaissances et de visualisation globale de collection de documents. Ces travaux sont validés en particulier au travers des systèmes Tétralogie et GeoECD (**Bernard.Dousset@irit.fr**), DocCube (**Josiane.Mothe@irit.fr**), et EasyDORView (**Christine.Julien@irit.fr**).

Nous proposons également des solutions permettant de réduire l'espace d'information via des systèmes de filtrage d'information en particulier pour des besoins à long terme représentés par des profils. Dans le contexte des documents complexes (structurés, multimédia, etc.), nous proposons des approches capables d'identifier la notion de granule de documents, de restituer et de visualiser les quelques granules pertinents parmi les centaines de millions disponibles. Ainsi, nos travaux s'intéressent à la prise en compte de l'adaptation et de la flexibilité en recherche et en filtrage d'information multilingue. > **Mohand.Boughanem@irit.fr**

L'extraction d'informations de contenu permet en particulier de concevoir des hyperdocuments comme des vidéos cliquables : l'utilisateur désigne par exemple un objet de la vidéo et des informations relatives lui sont alors fournies. L'anticipation des activités de navigation de l'utilisateur dans un hyperdocument permet de télécharger à l'avance les informations que celui-ci s'apprête à consulter. Reposant sur une modélisation statistique, ces travaux permettent de réduire les temps d'accès aux contenus.

Un outil de recherche d'information multimédia peut également reposer sur une interface de navigation. Or, un des problèmes particuliers à résoudre pour la conception d'une telle interface avec une collection de documents audiovisuels est le rendu de la dimension temporelle sous une forme statique. L'IRIT a ainsi développé des techniques dites de « mosaïcing » pour rendre compte de la globalité de l'information spatiale figurant dans un plan dans lequel la caméra effectue des mouvements de rotation. > **Vincent.Charvillat@enseeiht.fr**

Enfin, l'évaluation apparaît aujourd'hui comme une étape clé pour la validation des résultats d'indexation et de sélection d'information, mais aussi indirectement pour l'organisation des activités de recherche sur ce thème. L'IRIT s'est particulièrement impliqué dans les campagnes internationales d'évaluation « TREC ». Les résultats obtenus depuis plusieurs années sont en constante progression et classent les travaux du laboratoire sur le texte parmi les tous premiers mondiaux. > **Mohand.Boughanem@irit.fr**

Sur le stand recherche TIC (Hall 6, Allée C, Stand 6C37) : 8 démonstrations

CAMÉRA PANORAMIQUE, VIDÉOS INTERACTIVES ET APPLICATIONS MULTIMÉDIA

Une nouvelle caméra haute résolution permettant des prises de vues sur 360° sera présentée. Les images seront ensuite traitées de façon automatique pour être intégrées d'une part dans un système de navigation virtuelle utilisant une souris à accéléromètre (la Musaraigne) et d'autre part dans un système de codage de vidéo « par objets audiovisuels ». Un système de traitement d'image basé sur un processeur à reconfiguration dynamique partielle sera mis en œuvre. Les vidéos obtenues permettront la création de « vidéos cliquables » qui seront utilisables en interactivité forte avec les visiteurs.

Jean Conter, conter@enseeiht.fr

STORFACTORY : SYSTÈME DE STOCKAGE DE DONNÉES « ACTIF »

StorFactory facilite la gestion de grandes quantités de données. Des microprocesseurs internes au système de stockage traitent de façon autonome les données stockées (contrôle anti-virus, cryptage, data-mining...), permettant ainsi de limiter la quantité de données échangée avec l'extérieur. Les politiques de gestion des ressources du système de stockage s'adaptent aux besoins spécifiques des applications qui utilisent le stockage. L'administration du stockage est facilitée. Notre système est tout spécialement adapté au « stockage partagé », secteur du marché en pleine expansion. Présentation d'une maquette de StorFactory, appliquée à la « vidéo-à-la-demande ».

Christophe Guittenit, guittenit@irit.fr
Abdelaziz M'zoughi, mzoughi@irit.fr

nb événements 23-26 octobre 2002

L'IRIT au SITEF

ACCESSIBILITÉ À ET DE L'INFORMATION POUR TOUS : L'APPORT DES STIC

Cette démonstration présente une consultation multimodale distante de documents. Elle est le résultat d'une coopération interdisciplinaire (linguistique, communication homme-machine, psychologie et en informatique) pour la modélisation des stratégies de consultation. Au travers des stratégies d'oralisation qui seront démontrées, elle illustrera l'apport des technologies de l'information et la communication pour l'accès nomade à l'information pour tous (personnes handicapées et âgées, grand public). Cette présentation est le fruit d'une collaboration entre deux laboratoires (IRIT et Jacques Lordat), une PME-PMI (Elan Informatique) et des centres d'utilisateurs finaux (CCAS, notamment). Il est soutenu par la Région Midi-Pyrénées.

Nadine Vigouroux, vigourou@irit.fr
Antonio Serpa, serpa@irit.fr

TÉTRALOGIE UN OUTIL POUR L'ANALYSE STRATÉGIQUE

Présentation des nouvelles fonctionnalités de notre outil d'analyse stratégique Tétralogie : cartes géographiques, réseaux de liens, graphes statistiques. Démonstration de son utilité pour la scientométrie, l'analyse brevet, l'analyse des traces de connexions, la bibliométrie.

Bernard Dousset, dousset@irit.fr

PORTAIL POUR L'ANALYSE EN LIGNE D'INFORMATIONS STRATÉGIQUE

Après une macro analyse d'un corpus d'informations (scientifiques, techniques, commerciales, financières, juridiques, journalistiques) par notre outil de veille Tétralogie, nous proposons de mettre en ligne l'ensemble des relations obtenues afin que chaque utilisateur puisse réaliser une navigation personnalisée et produire les statistiques, les cartes et synthèses qui lui conviennent.

Bernard Dousset, dousset@irit.fr

ARIADNE : VERS UN SYSTÈME D'ENSEIGNEMENT À DISTANCE NORMALISÉ (DÉMONSTRATION COMMUNE IRIIT / CICT)

Le projet européen ARIADNE visait le développement d'outils et de méthodologies pour la production, la gestion et la mutualisation de matériaux pédagogiques médiatisés afin de diffuser des programmes de formation sur Internet. ARIADNE a grandement contribué au développement de standards définissant les règles favorisant l'indexation et la réutilisation des ressources pédagogiques. Les travaux réalisés dans ce domaine ont servi de base à la création du standard émergent IEEE/LTSC LOM. La fondation du même nom, a désormais pour objectif d'assurer la maintenance, le développement d'outils complémentaires et de former les utilisateurs à l'utilisation de ces différents outils. La démonstration permettra de présenter le système d'EAD qui a été développé dans le cadre d'ARIADNE et les outils qui permettent l'indexation des documents pédagogique ainsi que ceux qui favorisent leur réutilisation.

Philippe Vidal, vidal@irit.fr
Philippe Baqué, baque@cict.fr

LOGICIEL INDEXWEB,

PROMOTION ET RÉFÉRENCIEMENT DE SITES INTERNET

IndexWeb permet d'augmenter durablement l'audience de sites Internet et d'améliorer l'ordre d'apparition des pages dans les moteurs de recherche du web. IndexWeb est organisé en cinq étapes (Préparer/Référencer/Promouvoir/Vérifier/Planifier) autour d'une trentaine de fonctionnalités originales. Il intègre à la fois une expertise des mécanismes de référencement des moteurs de recherche et une analyse lexicale de pointe portant sur le contenu visible et invisible du site à référencer. Ce logiciel est développé et vendu par la société Synapse Développement, partenaire du projet. La collaboration entre Synapse Développement, l'IRIT et l'ERSS a porté sur l'évaluation de techniques d'extraction de terminologie développées par ces deux organismes pour l'identification et la structuration des mots-clés des champs META des pages HTML. Ces mots-clés sont un des éléments sur lesquels s'appuie IndexWeb pour suggérer des améliorations et améliorer le référencement d'un site.

Patrick Séguéla, p.seguela@synapse-fr.com
Nathalie Aussenac, aussenac@irit.fr

SIMULATION COOPÉRATIVE DISTRIBUÉE

Plusieurs utilisateurs, à distance les uns des autres, interagissent en pilotant des véhicules dans un même environnement virtuel interactif. Cette démonstration reliera un utilisateur placé sur notre stand avec un autre situé sur le stand de l'Onera, et/ou un autre poste du stand TIC, en alternance avec des utilisateurs situés à l'extérieur du SITEF. Cette démonstration est réalisée dans le cadre du thème Simulation Coopérative Distribuée de la Fédération de Recherche FÉRIA (IRIT-LAAS-ONERA).

Jean-Pierre Jessel, jessel@irit.fr
Patrice Torquet, torquet@irit.fr

Sur l'espace forum du stand recherche TIC : 4 conférences grand public

mercredi 23 octobre, 14 h 30

PREVI : UNE PLATE FORME DE RÉALITÉ VIRTUELLE POLYVALENTE DÉPLOYÉE À L'IRIT PAR Jean-Pierre Jessel, IRIIT

vendredi 25 octobre, 10 h

ARIADNE : VERS UN SYSTÈME D'ENSEIGNEMENT À DISTANCE NORMALISÉ PAR PHILIPPE VIDAL, IRIIT ET AUDREY ALIBERT, DRT MIAGE

vendredi 25 octobre, 15 h

FAIRE LIRE À HAUTE VOIX UNE MACHINE : POURQUOI ET COMMENT ? PAR Fabrice Maurel, IRIIT

samedi 26 octobre, 15 h

PRENEZ EN MAIN LA PROMOTION DE VOTRE SITE AVEC INDEXWEB PAR Patrick Séguéla, Chef de projet, consultant, Synapse Développement, Toulouse

Sur le stand CRCI (Hall 4, Allée A, Stand 4A4)

FORMATION À DISTANCE AU SERVICE DE L'INGÉNIERIE COOPÉRATIVE DANS LE SECTEUR AÉRONAUTIQUE

Un système ouvert pouvant accueillir un fonds commun partagé de formation et d'information sur l'ingénierie coopérative.

Une méthodologie de déploiement d'outils et de matériaux de formation.

Un savoir-faire autour des technologies de l'information et de communication assurant le développement de ressources pédagogiques multimédia coopératives.

Projet Européen ENHANCE 4e PCRD contrat BRPR-CT98-7001, projet CEA 97-0001 en collaboration avec Airbus, Formedia et l'IRIT.

Abdelmalek Benzekri, benzekri@irit.fr
P. Obry, Formedia

Sur le Village aéronautique et espace (Hall 3, Allée D, Stand 3D29)

ANNONCE DE LA CRÉATION DU LABORATOIRE MIDI,

laboratoire commun entre l'EADS CCR et l'IRIT sur le thème de la visualisation multimédia collective et l'ingénierie distribuée.

info@irit.fr
Eva.Dauphin@airbus.com

À l'Institut de Mécanique des fluides de Toulouse (IMFT) un colloque associé SITEF

« NOUVELLES TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION, NOUVEAUX CHAMPS POUR LES SCIENCES COGNITIVES »

Jeudi 24 octobre, de 9 h à 12 h 30 et de 14 h à 17 h 30

Journée-Phare nationale de l'ACI Cognitive, organisée par le réseau régional PRESCOT.

Journée grand-public. Entrée libre et gratuite.
http://www.irit.fr/ACTIVITES/PRESCOT/Prescot.f.html

info@irit.fr

L'IRIT est aussi :

– SUR LE STAND DU CNRS (HALL 4, ALLÉE B, STAND 4B4)

– SUR L'EXPOSITION « SCIENCES ET TECHNOLOGIES, DES MÉTIERS POUR TOUS » RÉALISÉE PAR SCIENCE ANIMATION (HALL 2, ALLÉE D, STAND 2D12)

16-18 octobre 2002

■ Fête de la science > 16 et 17 octobre expérimentation EN cinéma, 8e édition

« Transports d'images »
Salle Le Cap à 21 h
UPS, Toulouse.

> 17 et 18 octobre
Démonstrations et animations
pour le public scolaire autour
du thème « Le transport
de l'information »
IRIT, Toulouse.

23-26 octobre 2002

■ SITEF 2002 Sur le stand Recherche « Technologies de l'Information et de la Communication », présentation de 8 démonstrations Parc des Expositions Toulouse.

24 octobre 2002

■ Colloque Associé SITEF « Nouvelles technologies de l'information et de la communication, nouveaux champs pour les sciences cognitives » Journée-Phare nationale de l'ACI Cognitive, organisée par le réseau régional PRESCOT Institut de Mécanique des Fluides, Toulouse.

28-30 octobre 2002

■ DISC 2002 : 16th International Symposium on Distributed Computing ENSEEIH, Toulouse.

8-9 novembre 2002

■ First International Workshop on Isomorphisms of Types. IRIT, Toulouse.

22 novembre 2002

■ Echos-Sciences : Conférence « Les machines parlantes : les dernières avancées technologiques » par Fabrice Maurel et Nadine Vigouroux. Salle des congrès, Espace Clément Marot, Cahors.

Entre l'utile et le faisable : le syndrome de l'aiguille et du lampadaire

[suite de la page 6]

L'enjeu est de pouvoir représenter le contenu d'un document et ce qu'il signifie pour une utilisation possible.

Or, le numérique introduit de nouveaux usages et bouleverse les modes d'exploitation des contenus. Il n'est donc pas si simple de s'appuyer sur les besoins exprimés, ou des usages attestés pour proposer une indexation.

Comment en effet utiliser et exploiter une bibliothèque numérique ?

Que faire des perspectives ouvertes par la possibilité de manipuler des extraits désolidarisés de leur document d'origine ?

L'expérience manque, les usages ne sont pas encore stabilisés, l'intuition défaille. Par conséquent, comme trop souvent, au lieu de rechercher l'aiguille de ce qu'il faut faire là où elle est, on la recherche à la lumière du lampadaire de ce que l'on sait faire. Dépourvus des lumières de l'usage pour éclairer la finalité des outils, nous nous en remettons à la lueur des technologies d'extraction. Cela pose des problèmes.

En effet, les descripteurs extraits sont proches de la nature physique des contenus et sont rarement pertinents pour un usage ou une exploitation. Par ailleurs, les index utiles et pertinents mobilisent une interprétation du contenu qui demeure difficilement automatisable : elle reste le fait d'une

indexation manuelle, coûteuse et variable. Autrement dit, ce qui est utile est impossible, ce qui est possible est inutile. L'aiguille n'est définitivement pas sous le lampadaire.

Pour poursuivre la métaphore, quelles peuvent être les lampes de poche qui vont nous permettre d'élargir notre champ de vision ?

Une manière de formuler la question est de dire que l'on recherche l'information physique, donc extractible, qui puisse être une information sémantique, donc exploitable.

On peut citer des exemples : la couleur bleu d'un fond d'écran dénote une séquence « plateau » dans un journal de télévision, du fait des conditions de production. Un grain grossier d'une photographie renvoie au genre « people ».

En analysant quelles sont les effets de mise en forme propres à chaque genre de document, on sera capable d'associer le physique au sémantique. C'est la voie d'une philologie numérique, qui étudie les genres, associée à une rhétorique audiovisuelle et multimédia, qui explore les manières de dire en fonction de ce qu'elles disent.

L'enjeu est de pouvoir représenter le contenu d'un document et ce qu'il signifie pour une utilisation possible.

Or, le numérique introduit de nouveaux usages et bouleverse les modes d'exploitation des contenus.

Il n'est donc pas si simple de s'appuyer sur les besoins exprimés, ou des usages attestés pour proposer une indexation.

Comment en effet utiliser et exploiter une bibliothèque numérique ?

Que faire des perspectives ouvertes par la possibilité de manipuler des extraits désolidarisés de leur document d'origine ?