

|                     |  |
|---------------------|--|
| <b>Zeitschrift:</b> | Arbido   |
| <b>Herausgeber:</b> | Verein Schweizerischer Archivarinnen und Archivare; Bibliothek Information Schweiz   |
| <b>Band:</b>        | - (2008)   |
| <b>Heft:</b>        | 4: Informationswissenschaft: die Instrumente der Zukunft = Information documentaire: les outils du futur = Scienze della informazione: gli strumenti di domani |
| <b>Rubrik:</b>      | Tendances générales, recherches et projets = Allgemeine Tendenzen, Forschungen und Projekte  |

### **Nutzungsbedingungen**

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

### **Conditions d'utilisation**

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

### **Terms of use**

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

**Download PDF:** 09.01.2026

**ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>**

# I. Tendances générales, recherches et projets

## Allgemeine Tendenzen, Forschungen und Projekte

### Compression des données et archivage: le binôme du futur

Pierre Vanderghenst, professeur à l'EPFL, en collaboration avec Stéphane Gillioz, rédaction *arbido*

Les ondes radio, les circuits téléphoniques et les câbles d'ordinateurs véhiculent quotidiennement des quantités astronomiques d'informations numériques. Or, comment les référencer si les professionnels de l'information documentaire, entre autres, veulent pouvoir les archiver et les usagers les utiliser? Un double défi attend les chercheurs: la compression et l'indexation.

#### Les faits

Lorsqu'on parle de «quantités astronomiques» d'informations numériques véhiculées chaque jour par les différents modes de transmission (TV, téléphones, internet, caméras de surveillance, etc.), ce n'est pas une figure de style, loin s'en faut! Et le futur – proche! – va nous permettre de le vérifier à la puissance n.

Le livre blanc édité par l'IDC en mars 2007 (1) constate que la quantité d'informations numériques créée, saisie et transmise en 2006 était de  $1,288 \times 10$  puissance 18 bytes. Ce qui correspond à 161 exabytes ou 161 billions de gigabytes;

#### La norme JPEG2000

JPEG2000 est un nouveau système de codage d'image utilisant l'état de l'art des techniques de compression et basé sur la transformée en ondelettes. Son architecture devrait être appropriée à un grand nombre d'applications depuis les appareils photos numériques jusqu'à l'imagerie médicale et d'autres secteurs clé. Le codage comporte des informations sur le contenu ainsi qu'une indexation primaire.

autrement dit environ 3 millions de fois l'information contenue dans tous les livres écrits depuis le début des temps. Mais le meilleur est encore à venir, puisque, toujours selon le rapport de l'IDC, le volume d'informations sera multiplié par 6 d'ici à 2010...

Se pose dès lors la question suivante: lorsque l'on sait que 95% de ces données ne sont pas structurées, comment les référencer? Or, la réponse à cette question est de toute première importance pour les professionnels de l'information documentaire qui seront appelés à utiliser les nouveaux outils que les scientifiques sont en train de mettre au point dans ce domaine.

#### Compresser, puis indexer

La solution comporte deux étapes: 1) il faut d'abord compresser, puis 2) indexer. La difficulté est de taille, puisqu'il s'agit de comprimer les données tout en les structurant «sémantiquement». On connaît déjà des formats de compression comme MPEG, ZIP, JPEG et, plus récent, JPEG2000 (voir encadré), mais ils ne sont encore que des embryons de solutions face au défi que représentent les volumes de données à valoriser.

Prenons par exemple les archives du Festival de Montreux, donc pour l'essentiel des données son et image. L'EPFL se charge actuellement de la numérisation de l'archivage de ce fonds. Mais comment accéder à l'information voulue dans des délais raisonnables? La réponse est sur toutes les lèvres: par recherche «sémantique».

#### Le défi de la recherche «sémantique»

Certes, mais ici aussi le défi est de taille. Les contenus sont de toute première importance dans ce contexte. Or, l'on

sait que ces contenus comprennent du son, du texte, de l'image et de la vidéo. Il faut donc rechercher sur différents types de données. La recherche que l'on propose actuellement est indépendante d'un type de données à un autre. La solution réside donc dans l'intégration de ces données, afin qu'une recherche ciblée soit possible.

Autre exemple: les meetings virtuels, qui sont de plus en plus fréquents et qui seront certainement appelés à se multiplier à l'avenir si l'on considère l'explosion des coûts de déplacement due à la pénurie croissante des énergies non renouvelables. L'archivage de ces meetings (politiques, scientifiques, associatifs, sportifs, culturels) sera donc indispensable et nécessitera des solutions au niveau de la compression des données et de leur stockage qui n'existent pas encore. Le fameux «binôme du futur» sur lequel des milliers de chercheurs se penchent actuellement de par le monde...

#### Conclusion

La tâche est donc titanesque pour les chercheurs et il faudra encore du temps avant que les professionnels de l'information documentaire puissent disposer d'outils leur permettant de fournir à leurs clients des prestations dignes de ce nom en matière de fonds audio visuels.

#### Références:

(1) *The Expanding Digital Universe. A Forecast of Worldwide Information Growth Through 2010*, sous la direction de John F. Gantz, mars 2007

Contact: pierre.vanderghenst@epfl.ch

# Datenkompression und Archivierung: das Binom der Zukunft

Pierre Vanderghyest, Professor  
an der EPFL, in Zusammenarbeit mit  
Stéphane Gillioz, Redaktion *arbido*

**Radiowellen, Fernsprechleitungen und Computerkabel transportieren jeden Tag astronomische Mengen von digitalen Daten. Wie soll man diese Daten referenzieren, wenn u.a. Profis der Informationsdokumentation diese Daten archivieren und «gewöhnliche» Nutzer sie nutzen wollen? Die Forschungsgemeinde erwartet eine doppelte Herausforderung: einerseits die Kompression, andererseits die Indexierung.**

## Fakten

Wenn die Rede ist von «astronomischen Datenmengen», die tagtäglich mit verschiedenen Geräten und Medien (TV, Telefon, Internet, Überwachungskameras etc.) transportiert werden», so ist das keine Übertreibung – die (nahe!) Zukunft wird uns zeigen, dass «astronomisch» noch um den Faktor  $n$  zunehmen wird. Das von der IDC im März 2007 herausgegebene Weissbuch (1) hält fest, dass die Gesamtsumme von digitalen Daten, die 2006 produziert wurde,  $1,288 \times 10$  hoch 18 Bytes beträgt, das sind 161 Exabytes oder 161 Billionen Gigabytes; oder mit anderen Worten drei Millionen Mal die Information, die in sämtlichen je geschriebenen Büchern enthalten ist. Das Beste kommt aber noch: Gemäss dem Bericht der IDC wird diese Informationsmasse bis zum Jahr 2010 noch um den Faktor 6 anwachsen.

Damit drängt sich folgende Frage auf: Man weiss, dass 95% der Daten nicht strukturiert sind – wie soll man sie also referenzieren? Die Antwort auf diese Frage ist für die Berufsleute aus dem Bereich Informationsdokumentation von entscheidender Wichtigkeit: Sie werden mit unter den Ersten sein, welche die von den Forschern gegenwärtig zu diesem Zweck entwi-

ckelten neuen Instrumente anwenden werden.

## Komprimieren, dann indexieren

Die Lösung umfasst zwei Schritte: 1) Zuerst müssen die Daten komprimiert und dann 2) indexiert werden. Die damit verbundenen Schwierigkeiten haben es in sich, geht es doch darum, die Daten zu komprimieren, indem man sie «semantisch» strukturiert. Komprimierungsformate wie MPEG, ZIP, JPEG und, neueren Datums, JPEG2000 (siehe Kasten) sind bereits bekannt, sie sind aber zurzeit angesichts der gigantischen Datenmengen, die es zu verarbeiten gilt, noch nicht mehr als «Lösungsembryonen».

Werfen wir beispielsweise einen Blick auf die Archive des Jazzfestivals Montreux. Dabei handelt es sich mehrheitlich um Ton- und Bilddaten. Die EPFL ist zurzeit mit der Archivierung dieser Daten beschäftigt. Wie soll man innert nützlicher Zeit Zugriff auf exakt jene Daten erhalten, die man sucht? Die Antwort ist in aller Munde: mittels «semantischer» Abfrage.

## Die Herausforderung semantische Abfrage

Auch diese Lösung weist zahlreiche Fallstricke auf. Die Inhalte sind in diesem Zusammenhang sehr wichtig. Nun ist aber bekannt, dass die Inhalte Töne, Text, Bild und Video umfassen. Man muss also in verschiedenen Datentypen suchen. Die Suche, die gegenwärtig vorgeschlagen wird, ist unabhängig von der Art und Weise der Daten. Die Lösung heisst also Integration von Daten – erst mit integrierten Daten wird eine zielgerichtete Suche möglich.

Ein anderes Beispiel: virtuelle Sitzungen. Immer häufiger werden Sitzungen virtuell durchgeführt. Diese Tendenz wird sich angesichts der explodierenden Kosten für nicht erneuerbare Energien und damit für örtliche

Verschiebungen künftig noch akzentuieren. Die Archivierung dieser Sitzungen (Politik, Wissenschaft, Verbände, Sport, Kultur) wird damit unumgänglich und verlangt nach Lösungen im Bereich Datenkompression und Lagerung/Speicherung. Entsprechende Lösungen sind zurzeit noch nicht greifbar. Über die Lösung für dieses berückichtigte «Binom der Zukunft» beugen sich heute in der ganzen Welt Heerscharen von Forschern...

## Schlussfolgerung

Die Aufgabe der Forscher hat titanische Ausmasse. Es wird noch eine gewisse Zeit dauern, bis die Berufsleute aus dem Bereich Informationsdokumentation über Instrumente verfügen werden, die es ihnen ermöglichen, ihrer Kundschaft Dienstleistungen im Bereich audiovisuelle Bestände anbieten können, die diesen Namen auch verdienen.

## Anmerkung:

(1) The Expanding Digital Universe. A Forecast of Worldwide Information Growth Through 2010, unter der Leitung von John F. Gantz, März 2007.

Kontakt: pierre.vanderghyest@epfl.ch

### Die Norm JPEG2000

JPEG2000 ist ein neues Bildcodierungssystem, das die modernsten Komprimierungstechniken anwendet und auf der Transformation in Wavelets aufbaut. Die Systemarchitektur ist für eine Vielzahl von Anwendungen (von digitalen Fotoapparaten bis hin zu medizinischen Bildgebungsverfahren und anderen Schlüsselbereichen) geeignet. Die Codierung umfasst Informationen über den Inhalt sowie eine primäre Indexierung.

# L'initiative européenne i2010: bibliothèques numériques et réseau du futur

Un des objectifs clés de l'initiative i2010 («Une société européenne de l'information pour la croissance et l'emploi») lancée en 2005, qui forme le cadre stratégique des lignes d'actions politiques de l'Union dans le domaine des nouvelles technologies de l'information et de la communication (TIC), est de réaliser «des performances de niveau mondial dans la recherche et l'innovation dans le secteur des TIC en réduisant l'écart avec les principaux concurrents de l'Europe». Le septième programme cadre (2007–2013) de la Communauté européenne porte notamment sur les bibliothèques numériques et le réseau du futur. Extraits du programme de travail.\*

## 1. Bibliothèques et contenus numériques

Dans la société actuelle, les individus et les organisations sont confrontés à une masse et une diversité toujours croissantes d'information et de contenu, et à des demandes de plus en plus fortes de connaissances et de compétences. Répondre à ces demandes exige des progrès dans trois domaines étroitement liés. Il convient tout d'abord de rendre le contenu accessible à travers des bibliothèques numériques et d'assurer sa préservation, son accessibilité et sa facilité d'utilisation à long terme. En second lieu, nous avons besoin de technologies plus efficaces pour la création et la gestion intelligentes de contenu, pour permettre l'acquisition de connaissances ainsi que leurs partage et réutilisation. Troisièmement, les individus et les organisations doivent trouver de nouvelles méthodes pour acquérir, apporter et exploiter ces connaissances, et de ce fait apprendre.

Le défi consiste, par conséquent, à exploiter les synergies rendues possibles en associant les contenus, les

connaissances et l'apprentissage; à faire en sorte que les connaissances et les contenus soient abondants, accessibles, interactifs et utilisables dans le temps par les hommes, ainsi que par les machines. Ceci nécessitera de prendre en compte les tendances actuelles en matière de production et de consommation de contenu et en particulier la transition du modèle de service à sources rares et destinations multiples vers des modèles de service à sources et destinations multiples. L'Europe, avec son acquis culturel et son potentiel créatif uniques, est bien placée pour tirer profit de ce changement de paradigme et pour jouer un rôle majeur dans le marché de la connaissance.

La recherche devrait établir fermement les services de bibliothèques numériques comme une composante clé des infrastructures de contenu numérique, permettant de produire, stocker, gérer, personnaliser, transmettre, conserver et exploiter de façon sûre, efficace et peu coûteuse du contenu et des connaissances, selon des normes largement acceptées.

Le soutien de services davantage personnalisés et collaboratifs, en particulier au sein de communautés auto-organisatrices, débouchera sur des approches plus créatives pour la production de contenu et de connaissances.

Des améliorations sont également attendues en termes d'utilisation, d'accessibilité, d'évolutivité et de rentabilité des méthodes, technologies et applications produites, au vu du grand nombre de données et d'utilisateurs simultanés.

Le travail renforcera le lien entre le contenu, les connaissances et les processus d'éducation permanente. Il améliorera notre capacité à maîtriser et exploiter contenus et connaissances et à apprendre dans des environnements de travail de plus en plus dynamiques.

Le travail mené dans le cadre du présent défi contribuera à la mise en

œuvre de l'initiative «i2010: bibliothèques numériques».

## 1.1 Résultats visés pour les bibliothèques numériques

a) *Bibliothèques numériques européennes de grande envergure*, avec des services d'accès innovants supportant des communautés partageants des prati-

Le défi consiste, par conséquent, à exploiter les synergies rendues possibles en associant les contenus, les connaissances et l'apprentissage; à faire en sorte que les connaissances et les contenus soient abondants, accessibles, interactifs et utilisables dans le temps par les hommes, ainsi que par les machines.

ques communes dans la création, l'interprétation et l'utilisation de contenu culturel et scientifique, y compris d'objets numériques multi-formats et multisources. Elles devront être combinées à des environnements robustes et évolutifs intégrant des capacités de recherche basées sur la sémantique et des fonctions de conservation numériques essentielles. Une attention particulière est accordée aux procédés de numérisation rentables et à l'utilisation de ressources numériques dans des contextes multilingues et pluridisciplinaires.

b) *Approches radicalement nouvelles de la conservation numérique*, telles que celles inspirées de la capacité humaine à traiter l'information et les connaissances, explorant le potentiel des TIC avancées pour traiter automatiquement de gros volumes de contenu numérique dynamique et volatil, garantissant sa conservation, assurant le suivi de l'évolution du contexte sémantique et d'utilisation

\* Synthèse de la rédaction

tout en préservant son intégrité, authenticité et accessibilité à long terme.

Les efforts sont concentrés particulièrement sur la création d'un réseau de centres de compétence dans le domaine de la numérisation et de la conservation, en s'appuyant sur la mise en commun et l'actualisation des ressources existantes dans les Etats membres ou associés.

## 1.2 Résultats visés pour le contenu intelligent et sémantique

a) Environnements de création avancés pour créer de nouvelles formes de contenu interactif et expressif permettant une expérimentation multimodale et une narration non linéaire. Ces environnements faciliteront le partage et le remixage de contenu, même par des utilisateurs non experts, en étiquetant automatiquement les contenus avec des métadonnées sémantiques et en utilisant des normes ouvertes pour les stocker dans des dépôts en réseau offrant des fonctions de recherche et d'indexation symbolique et par similarité, pour tous les types de contenu.

b) Environnements de flux de travail automatisés et collaboratifs pour gérer le cycle de vie des anciens et nouveaux médias ainsi que des biens en terme de contenu de l'entreprise, depuis l'acquisition des matériels de référence jusqu'au contrôle des ver-

Des améliorations sont également attendues en termes d'utilisation, accessibilité, évolutivité et rentabilité des méthodes, technologies et applications produites, au vu du grand nombre de données et d'utilisateurs simultanés.

sions, au conditionnement et à la réaffectation de produits complexes, y compris leur adaptation linguistique et culturelle à des marchés et groupes d'utilisateurs cibles. Les résultats empiriques de la psychologie de la perception et de l'attention humaines seront utilisés pour identifier des segments multimédias marquants et pour appliquer des programmes de résumé et de codage qui

amélioreront le stockage et la transmission du contenu sans en affecter les propriétés perceptives.

c) Architectures et technologies de distribution, présentation et consommation personnalisées de contenu adaptatif et intuitif. Grâce à la détection et à l'exploitation de l'intelligence ambiante émergente, elles feront appel aux caractéristiques enfouies dans des objets de contenu et matériels de rendu, pour permettre l'adaptation

Le travail mené dans le cadre du présent défi contribuera à la mise en œuvre de l'initiative «i2010: bibliothèques numériques».

dynamique de dispositifs, les expériences multimodales immersives et un soutien contextuel aux objectifs et préférences linguistiques des utilisateurs. Des algorithmes d'apprentissage protégeant la confidentialité analyseront les interactions des utilisateurs avec des dispositifs et d'autres utilisateurs afin d'actualiser et de servir efficacement ces objectifs et préférences.

d) Actions orientées vers la création de communautés, visant à encourager des approches interdisciplinaires et un dialogue utilisateur/fournisseur plus efficace, et autres mesures, notamment la validation sur le terrain et les normes, ayant pour but une adoption plus rapide des résultats de la recherche. Etudes d'évaluation technologique et de fonctionnalité, analyses économiques et feuilles de route pour apprécier la démocratisation des outils de production et de gestion multimédias personnels et communautaires.

e) Bases sémantiques: modélisation probabiliste, temporelle et modale et raisonnement approximatif reposant sur une recherche fondée sur des objectifs allant au-delà des formalismes courants. Les résultats théoriques seront assortis de mises en œuvre de référence robustes et évolutives. La facilité d'utilisation et les performances seront testées par l'intégration web ontologique et à grande échelle de

sources de données hétérogènes, évolutives et bruitées ou contradictoires, allant de dépôts multimédias distribués à des flux de données provenant de dispositifs et capteurs ambiants, permettant la résolution en temps réel de nombres massifs de requêtes et l'induction d'hypothèses scientifiques ou d'autres formes d'apprentissage.

## 1.3 Impact attendu

Ces activités rendront les ressources numériques sur lesquelles s'appuient la créativité et la sémantique plus faciles et plus rentables à produire, organiser, rechercher, personnaliser, distribuer et (ré)utiliser dans la chaîne de valorisation.

- Les créateurs pourront concevoir des formes de contenu plus participatives et plus communicatives.
- Les éditeurs des industries créatives, des entreprises et des secteurs professionnels augmenteront leur productivité grâce à un contenu innovant d'une plus grande complexité et d'une plus grande facilité de réaffectation.
- Les organisations seront en mesure d'automatiser la collecte et la diffusion de contenus numériques et de connaissances exploitables par machine et de les partager avec des organisations partenaires dans des environnements collaboratifs de confiance.
- Les scientifiques travailleront plus efficacement en automatisant le lien entre l'analyse de données, la théorie et la validation expérimentale.

## 2. Le réseau du futur

### 2.1 Infrastructures de réseaux et de services omniprésentes et fiables

Avec ses atouts en matière d'équipements, de terminaux, de réseaux et services de communication électroniques, l'Europe est bien placée dans la course mondiale visant à définir et développer les infrastructures réseaux et services du futur. Celles-ci généreront de nouvelles possibilités économiques avec de nouvelles classes d'applications en réseau, tout en réduisant les frais d'exploitation. Les réseaux internet, mobiles, fixes et de radiodiffusion actuels et les infrastructures connexes de service logiciel doivent progresser en conséquence afin de permettre une autre va-

gue de croissance de l'économie et de la société en ligne au cours des quinze prochaines années.

Le défi consiste à produire les infrastructures de réseaux et de services omniprésentes et convergentes de la prochaine génération pour la communication, l'information et les médias.

**Une attention particulière est accordée aux procédés de numérisation rentables et à l'utilisation de ressources numériques dans des contextes multilingues et pluridisciplinaires.**

Ceci implique de surmonter les verrous liés à l'évolutivité, la flexibilité, la fiabilité et la sécurité, alors que les architectures de réseaux et de services d'aujourd'hui sont principalement statiques, ne peuvent accueillir qu'un nombre limité de terminaux et de services et n'offrent qu'un niveau de confiance limité. Ces nouvelles infrastructures permettront une grande diversité de modèles économiques capables d'associer des ressources de bout en bout et de façon continue à travers une multiplicité de terminaux, de réseaux, de fournisseurs et de domaines de service.

Les futures infrastructures envisagées devront:

- *Etre omniprésentes et fortement dynamiques.* Elles doivent offrir aux utilisateurs des possibilités quasiment illimitées, en étant compatible d'une grande variété de terminaux et de services nomades interopérables, d'une large gamme de formats de contenu et d'une multiplicité de modes de transport. Elles doivent également permettre l'intelligence du contexte et le comportement dynamique nécessaires pour des applications dont les exigences varient en fonction du temps et du contexte;
- *garantir la robustesse, la résilience, la fiabilité et la sécurité* compatibles avec des réseaux et plateformes de services logiciels dont la complexité et l'étendue sont considérablement supérieures à celles des infrastructures d'aujourd'hui;
- *permettre la convergence d'activités et de services en réseau*, ainsi que leur gestion, qui soient compatibles avec de

multiples environnements tels que l'environnement domestique, l'entreprise ou les utilisations nomades.

Ceci nécessite de prendre en compte l'évolution des grandes infrastructures existantes vers de nouvelles infrastructures, en trouvant un juste équilibre entre les exigences de compatibilité ascendante et la recherche portant sur des architectures de rupture ayant pour but de construire les infrastructures futures d'Internet, des communications mobiles, du haut débit et des services associés.

Les moteurs de l'évolution de ce défi ont trait principalement à l'évolution technologique de réseaux pervasifs de communication mobile et à haut débit, à la disponibilité de plateformes de services dynamiques, à la fiabilité et la sécurité, dans le cadre d'environnements en réseau convergents et interopérables. A cet égard, l'activité proposée porte en grande partie sur les verrous technologiques et les scénarios socio-économiques définis dans les Agendas de Recherche Stratégique des plateformes technologiques européennes eMobility, NESSI, NEM et ISI.

La participation d'organismes de pays tiers à des activités de recherche

**Les résultats empiriques de la psychologie de la perception et de l'attention humaines seront utilisés pour identifier des segments multimédias marquants et pour appliquer des programmes de résumé et de codage qui amélioreront le stockage et la transmission du contenu sans en affecter les propriétés perceptives.**

est encouragée lorsque des bénéfices mutuels peuvent être mis en évidence. Ceci concerne notamment i) la possibilité d'avancer, grâce à des partenariats stratégiques de recherche, vers un consensus global et des normes mondiales; ii) les opportunités d'étalonnage mutuel des performances; iii) l'échange des bonnes pratiques, incluant les enjeux réglementaires et socio économiques en tant que moteurs technologiques; iv) la validation à grande échelle de technologies et d'applications en réseau et leur gestion dans un contexte global. La participation de partenaires de

pays tiers et la sélection des régions cibles les plus prometteuses sont laissées à l'initiative des soumissionnaires.

## 2.2 Résultats visés

a) *Infrastructures et architectures de réseau omniprésentes* permettant: i) la convergence et l'interopérabilité de technologies réseau mobile et haut débit hétérogènes; ii) un accès radio flexible et efficace au niveau du spectre permettant un accès omniprésent à des services mobiles large bande depuis des réseaux à courte portée jusqu'à des réseaux étendus; iii) l'élimination des obstacles à l'accès large bande et à la connectivité de bout en

**Le défi consiste à produire les infrastructures de réseaux et de services omniprésentes et convergentes de la prochaine génération pour la communication, l'information et les médias.**

bout à très grande vitesse avec des protocoles et un routage optimisés; iv) l'intelligence du contexte; v) le traitement optimisé du trafic entre le cœur du réseau et les réseaux périphériques; vi) l'extensibilité permettant l'augmentation significative du nombre de terminaux connectés et permettant l'émergence d'applications de type machine à machine ou fonctionnant sur base de capteurs – au-delà de la RFID – et qui sont en mesure de fonctionner dans une multitude d'environnements d'exploitation publics ou privés.

b) *Contrôle, gestion et flexibilité optimisées de la future infrastructure réseau*, permettant une évolution vers des réseaux cognitifs et capables de: i) assurer la composition et l'exploitation continues et de bout en bout de réseaux et de services à travers une multitude d'opérateurs et de secteurs d'affaires; ii) permettre une grande diversité de caractéristiques et de besoins de services, qui seront beaucoup plus complexes que ceux des infrastructures d'aujourd'hui, grâce à des fonctions dynamiques et de programmabilité, avec reconfigurabilité de l'allocation des ressources, des protocoles et du routage, d'auto-

organisation et d'auto-administration; iii) gérer en temps réel de nouvelles formes de communications ad hoc avec des besoins de connectivité intermittente et une topologie réseau variant avec le temps; iv) permettre la distribution intelligente de services à travers des technologies d'accès multiples avec commande centralisée ou distribuée.

c) *Technologies et architectures des systèmes pour l'internet du futur*, destinées à surmonter les limites d'Internet prévues à long terme en termes de capacités, d'architecture et de protocoles, et dérivant des besoins des éléments suivants: mobilité généralisée; facteurs d'échelles liés au grand nombre de terminaux connectés, caractéristiques des services et environnements applicatifs; sécurité; domaines de confiance; nouvelles formes de routage et de distribution du contenu avec reconfiguration dy-

namique du trafic pour la distribution et le contrôle de bout en bout, de connectivité ad hoc dans un environnement sans fil généralisé. Le travail de nature exploratoire examinera dans quelle mesure différentes catégories de nouveaux besoins représentent une contrainte à l'évolution prévisible d'Internet et proposera des solutions à long terme.

d) *Actions de coordination et de soutien*:  
i) soutien au développement de feuilles de route et à des conférences;  
ii) coordination avec des initiatives ou programmes nationaux ou régionaux apparentés.

#### Impact attendu

– Normes mondiales pour une nouvelle génération d'infrastructures de réseau et de services omniprésentes à très haute capacité. Ces infrastructures devraient prendre en charge la convergence, l'interopérabilité totale,

un nombre significativement plus grand et varié de terminaux, de nouveaux services et de besoins complexes des utilisateurs.

– Renforcement du leadership industriel européen en matière de réseaux câblés et sans fil; développement de plus fortes synergies entre divers acteurs du secteur et contribution à de nouveaux modèles de gestion tirant profit de la convergence et de l'interopérabilité totale.  
– Nouvelles possibilités industrielles/de services en Europe, en particulier dans le domaine des technologies internet, où l'Europe n'a pas encore atteint une position correspondant à son potentiel technologique.

Source: [http://ec.europa.eu/information\\_society/eeurope/i2010/key\\_documents/index\\_en.htm](http://ec.europa.eu/information_society/eeurope/i2010/key_documents/index_en.htm)

## TRIALOG

*Dokumentations- und Organisationsberatung*

*In der Unternehmensberatung die Spezialisten für  
Dokumentation, Bibliothek, Archiv/Records Management*

**Ihre Zukunft planen:  
Aus Ideen eine  
Strategie entwickeln.**

*Trialog AG, Holbeinstr. 34, 8008 Zürich  
Tel. 044 261 33 44, Fax 044 261 33 77*

*E-Mail: [trialog@trialog.ch](mailto:trialog@trialog.ch)  
Homepage: <http://www.trialog.ch>*

## ArchPlus® 1.1

**Klar** in der Struktur  
**Übersichtlich** in der Darstellung  
**Effizient** in der Verwaltung der Daten

Die Archivsoftware ArchPlus® 1.1 arbeitet mit einer SQL-Datenbank. Wichtige Features sind:

- Einrichtung einer unbeschränkten Zahl von virtuellen Abteilungen mit je eigenen Archivplänen oder Klassifikationssystemen.
- Kreuzrecherchen über alle Felder und Abteilungen
- Export von Auswertungen und Verzeichnissen in verschiedene Formate (RTF, XLS, PDF, HTML).
- Einbindung von Abbildungen und elektronischen Dokumenten
- Zuweisung und Verwaltung von individuellen Anwenderrechten.

Das Basismodul ArchPlus® 1.1 ist ab Ende August 2005 lieferbar. Bestellen Sie jetzt eine Testversion!

EBERLE AG, Strälgasse 2, CH-9533 Kirchberg  
Tel: 0041(0)719312221 Fax: 0041(0)719313030  
E-Mail: [info@ebarch.ch](mailto:info@ebarch.ch) Homepage: [www.ebarch.ch](http://www.ebarch.ch)

# Die europäische Initiative i2010: digitale Bibliotheken und das Netz der Zukunft

Ein wichtiges Ziel der Initiative i2010 («Eine Europäische Informationsgesellschaft für Wachstum und Beschäftigung», 2005), die den strategischen Rahmen für die Politik im Bereich Informations- und Kommunikationstechnologie (IKT) in der EU bildet, lautet: «Wir wollen in der IKT-Forschung und -Entwicklung Spitzenleistungen erbringen und zur internationalen Konkurrenz Europas aufschliessen.» Das gegenwärtige siebte Arbeitsprogramm (2007–2013) der EU stellt v.a. digitale Bibliotheken und das Netz der Zukunft in den Mittelpunkt seines Förderprogramms. Nachstehend einige Auszüge aus dem Arbeitsprogramm der EU.\*

## 1. Digitale Bibliotheken und Inhalte

In der modernen Gesellschaft sehen sich der Einzelne wie auch Organisationen einer wachsenden Fülle und Vielfalt an Informationen und Inhalten konfrontiert, die ihnen immer mehr Wissen und Fähigkeiten abverlangen. Um den Anforderungen gerecht zu werden, müssen in drei eng zusammenhängenden Bereichen Fortschritte erzielt werden. Erstens müssen Inhalte durch digitale Bibliotheken zugänglich gemacht werden, und ihre langfristige Erhaltung, Zugänglichkeit und Nutzbarkeit müssen gewährleistet sein. Zweitens brauchen wir effektivere Technologien für die Schaffung und das Management intelligenter Inhalte, und um den Erwerb von Wissen und die gemeinsame Nutzung von Wissen zu unterstützen. Drittens müssen Einzelne und Organisationen neue Wege finden, um Wissen zu erwerben, einen Beitrag zur Nutzung zu leisten sowie sich das notwendige Know-how anzueignen.

Die eigentliche Herausforderung besteht darin, die Synergien, die durch die Verknüpfung von Inhalten, Wissen und Lernen möglich werden, zu nutzen. Es muss dafür gesorgt werden,

dass Inhalte und Wissen interaktiv und auf Dauer von Mensch und durch den Computer genutzt werden können. Dabei muss auf die gegenwärtigen Entwicklungstrends bei der Herstellung und in der Nutzung von Inhalten eingegangen werden. Ganz besonders ist die Entwicklung weg vom Wenige-an-Viele-Modell (few-to-many) hin zum Viele-an-Viele-Modell (many-to-many) zu berücksichtigen. Europa kann mit seinem einzigartigen kulturellen Erbe und kreativen Potenzial aus diesem Paradigmenwechsel Nutzen ziehen und damit zu einem wichtigen Akteur der Wissensgesellschaft werden.

Die digitalen Bibliotheken müssen dank der Forschung zu Schlüsselkomponenten der digitalen Inhaltsinfrastrukturen entwickelt werden, damit Inhalte und Wissen produziert, gespeichert, verwaltet, personalisiert, übertra-

**Die eigentliche Herausforderung besteht darin, die Synergien, die durch die Verknüpfung von Inhalten, Wissen und Lernen möglich werden, zu nutzen. Es muss dafür gesorgt werden, dass Inhalte und Wissen interaktiv und auf Dauer von Mensch und durch den Computer genutzt werden können.**

gen und erhalten werden können. Die Nutzung hat zuverlässig, effizient, preisgünstig und nach allgemein anerkannten Standards zu erfolgen.

Die Unterstützung von stärker personalisierten und im Verbund arbeitenden Dienstleistungen, insbesondere innerhalb sich selbstorganisierenden Organisationsstrukturen, führt zu kreativeren Ansätzen für die Inhalts- und Wissensproduktion. Es ist zu erwarten, dass die Forschung auch künftig Verbesserungen in punkto Anwendbarkeit, Zugänglichkeit, Skalierbarkeit und Kostengünstigkeit im Umgang mit

grossen Datenmengen bringen wird. Die sich aus der Forschung ergebenden Methoden, Technologien und Applikationen lassen nicht nur ein besseres Management von grossen Datenmengen, sondern auch eine grössere simultane Nutzung dieser Daten erwarten.

Durch die Forschung werden die Verbindungen zwischen Inhalt, Wissen und permanenten Lernprozessen gestärkt. Unsere Fähigkeit, Inhalt und Wissen zu beherrschen und zu nutzen und in immer dynamischeren Arbeitsumgebungen damit umzugehen, wird so zunehmen. Die im Rahmen dieser Problemstellung gemachten Forschungsarbeiten werden zur Umsetzung der Initiative «i2010 – digitale Bibliotheken» beitragen.

## 1.1 Erwartete Ergebnisse für digitale Bibliotheken

a) *Grosse europaweite digitale Bibliotheken* mit innovativen Zugangsdiensten, die anwenderorientierte Gruppen bei der Schaffung, Interpretation und Nutzung von kulturellen und wissenschaftlichen Wissensinhalten, gerade auch bei der Verwendung von mehrformatigen und vielschichtigen Digitalobjekten, unterstützen sollen. Sie müssen in stabilen und skalierbaren Informatikapplikationen integriert werden. Dazu gehören semantikbasierte Suchmöglichkeiten und grundlegende Aspekte der digitalen Langzeitarchivierung. Besondere Aufmerksamkeit gilt der Entwicklung kostengünstiger Digitalisierungsprozesse und der Nutzung von digitalen Ressourcen in vielsprachigen und multidisziplinären Umgebungen.

b) *Radikal neue Ansätze für die digitale Langzeitarchivierung.* Die Fähigkeit der Menschen im Umgang mit Information und Wissen verlangt auch Forschungsansätze, die sich mit dem

Potenzial innovativer Informations- und Kommunikationstechnologie befassen. So müssen automatisch digital hergestellte Inhalte und dynamische sowie flüchtige digitale Inhalte in grossem Umfange für lange Zeit zugänglich gemacht werden. Das bedingt auch, dass die sich verändernde Semantik und der jeweilige Anwendungskontext verfolgt werden können, damit ihre Integrität, Authentizität sowie die langfristige digitale Zugänglichkeit gewährleistet sein müssen.

Ein spezieller Schwerpunkt stellt die Bildung eines Netzwerkes von Kompetenzzentren für Digitalisierung und Erhaltung dar. Dieses Netzwerk stützt sich auf bereits vorhandene Ressourcen in den Mitgliedstaaten oder in den assoziierten Staaten. So gelingt durch diese Netzwerke eine Bündelung oder eine Verbesserung der Kompetenzen.

## 1.2 Gezielte Resultate in der Forschung für intelligente Inhalte und Semantik

a) Innovative *Authoring*-Umgebungen zur Schaffung *neuartiger Formen* interaktiver und ausagekräftiger Inhalte, die multimodale Experimente und nichtlineares Erzählen ermöglichen. Diese Umgebungen werden die gemeinsame Nutzung und die Neuzusammenstellung von Inhalten, auch durch fachfremde Nutzer, erleichtern, indem die Inhalte automatisch mit semantischen Metadaten gekennzeichnet und unter Verwendung offener Standards in vernetzten Speicherservern gespeichert werden, die die symbolische und ähnlichkeitsbasierte Indexierung und Suchmöglichkeiten für alle Inhaltstypen unterstützen.

b) Kollaborative und automatisierte *Workflow*-Umgebungen zur Verwaltung des *Lebenszyklus* neuer und bereits gespeicherter Medieninhalte sowie Informationsdateien von Unternehmen in all ihren Phasen des Lebenszyklus von der ersten Versionierung bis zur Umwandlung komplexer Informationsangebote einschliesslich ihrer sprachlichen und kulturellen Anpassung an die jeweiligen Zielmärkte und Nutzergruppen. Unter Heranziehung empiri-

scher Ergebnisse der Psychologie in Bezug auf die menschliche Wahrnehmung und Aufmerksamkeit sollen wesentliche Multimediasegmente ermittelt werden. Mittels Zusammenfassungs- und Kodiersysteme soll eine Verbesserung der Inhalts-speicherung und -übertragung angestrebt werden, ohne dass deren Wahrnehmungseigenschaften beeinflusst werden sollen.

c) Neue Architekturen und Technologien zur *personalisierten Verteilung, Präsentation und Nutzung* veränderbarer Medieninhalte, die ihre Meta-

Es ist zu erwarten, dass die Forschung auch künftig Verbesserungen in punkto Anwendbarkeit, Zugänglichkeit, Skalierbarkeit und Kostengünstigkeit im Umgang mit grossen Datenmengen bringen wird. Die sich aus der Forschung ergebenden Methoden, Technologien und Applikationen lassen nicht nur ein besseres Management von grossen Datenmengen, sondern auch eine grössere simultane Nutzung dieser Daten erwarten.

daten zu ihrer Beschreibung und ihrer Verarbeitung bereits in der Datei integriert haben. Sie entdecken und nutzen die technologischen Verbesserungsmöglichkeiten, die das System bietet, und verwenden die in Inhaltsobjekten und Rendering Equipment eingebetteten Funktionen, um sich dynamisch an die jeweiligen Geräte anzupassen. Lernende Algorithmen, welche die Privatsphäre wahren, analysieren den Gerätegebrauch ihrer Nutzer und den Umgang mit anderen Nutzern. Damit werden die Ziele und Präferenzen der Nutzer jeweils aktualisiert und effektiv registriert.

d) Forschungsvorhaben, die auf die *Entwicklung von Netzwerken* ausgerichtet sind, die fächerübergreifende Konzepte und einen wirksameren Nutzer-/Anbieterdialog ermöglichen sollen. Andere Massnahmen, dazu gehören auch die Feldvalidierung und gleichgelagerte Standards, sollen eine schnellere *Rezeption* von Forschungs-

ergebnissen ermöglichen. Studien zur Machbarkeit und zur Abschätzung der Technologiefolgen, Wirtschaftsanalysen und Roadmaps für den Einsatz von persönlichen und gemeinschaftsbasierten Multimedia-produktions- und -managementwerkzeuge runden dieses Forschungsvorhaben ab.

e) Neue *semantische Grundlagenerkenntnisse*: probabilistische, zeitliche und modale Modellierung und Forschung soll eine noch bestehende Lücke in den gegenwärtigen Forschungserkenntnissen zielgerichtet schliessen. Diese Ergebnisse sollen in stabilen und skalierbaren Referenzanwendungen getestet werden. Die Anwendbarkeit und Leistung müssen mittels *Webintegration* in grossen ontologiebasierten heterogenen, sich veränderten und fehlerhaften oder inkonsistenten Datenquellen getestet werden. Diese Tests reichen von verteilten Multimedia-Repositories bis zu Datenströmen aus Peripheriegeräten. Damit soll in Echtzeit eine sehr grosse Anzahl von Suchanfragen unter Ableitung wissenschaftlicher Hypothesen oder andere Formen der Resultatanalyse durchgesetzt werden.

## 1.3 Erwartete Resultate

Diese Forschungsaktivitäten machen es einfacher, kostengünstige, digitale Quellen, unter Einsatz von Kreativität und Semantik, durch die ganze Wertkette hindurch zu produzieren, sie zu verändern, sie auffindbar und profilorientiert nutzbar zu machen, zu verteilen und sie wieder für die Nutzer gebrauchsfähig zu machen.

- Informationsproduzenten können damit partizipativere und kommunikative Inhaltsformen entwickeln.
- Unternehmen in kreativen Branchen, Unternehmen und kreative Berufsleute werden ihre Produktivität erhöhen und mit innovativem Inhalt arbeiten können, der zwar komplexer, aber auch leichter veränderbar ist.
- Institutionen werden die Erfassung und Verteilung von digitalen Inhalten und maschinenlesbarem Wissen automatisieren können und es dann mit Partnerorganisationen in kolla-

borativen Umgebungen gemeinsam nutzen können.

- Wissenschaftler werden durch die automatisierte Verknüpfung von Datenanalyse, Theorie und experimenteller Validierung effizienter arbeiten können.

## 2. Das Netz der Zukunft

### 2.1 Allgegenwärtige und vertrauenswürdige Netz- und Dienstinfrastrukturen

Mit seinen Stärken in den Bereichen Kommunikationsinfrastruktur, Geräte, Netze und E-Services ist Europa im globalen Wettbewerb um die Beschreibung und Entwicklung der Netz- und Dienstinfrastrukturen der Zukunft gut aufgestellt. Hieraus werden neue wirtschaftliche Möglichkeiten mit neuen vernetzten Anwendungen entstehen. Dies geht mit einer gleichzeitigen Senkung der operationellen Kosten einher. Die bestehenden Internet-, Mobil-, Fest- und Funknetze und die dazugehörige Software müssen entsprechend entwickelt werden, um einen weiteren Wachstumsschub der Online-Wirtschaft und der Online-Gesellschaft für die nächsten 15 Jahre zu ermöglichen. Die Herausforderung besteht darin, die nächste Generation allgegenwärtiger und konvergierender Netz- und Dienstinfrastrukturen für Kommunikation, Rechenprozesse und Medien bereitzustellen. Dazu müssen die in Bezug auf Skalierbarkeit, Flexibilität, Zuverlässigkeit und Sicherheit bestehenden Engpässe überwunden werden, denn die heutigen Netz- und Dienstarchitekturen sind vorwiegend statisch ausgelegt und können nur eine begrenzte Zahl von Geräten, Dienst-

**Die Herausforderung besteht darin, die nächste Generation allgegenwärtiger und konvergierender Netz- und Dienstinfrastrukturen für Kommunikation, Rechenprozesse und Medien bereitzustellen.**

leistungen sowie nur eine eingeschränkte Privatsphäre gewährleistet werden. Mit der neuen Infrastruktur können ganz unterschiedliche Geschäftsmodelle für dynamische und lückenlose Endpunkt-zu-Endpunkt-Verbindungen verschiedenster Geräte,

Netze, Anbieter und Dienstleistungssektoren entstehen.

Die künftigen Infrastrukturen sollen:

- raumgreifend, allgegenwärtig und hochgradig dynamisch sein. Sie sollen nahezu unbegrenzte Kapazitäten für die Nutzer bereithalten und eine Vielzahl interoperabler mobiler Geräte und Dienste verschiedenster inhaltlicher Formate und eine Vielzahl unterschiedlicher Verteilmodi unterstützen. Sie sollen die Bewusstwerdung und das dynamische Verhalten der Nutzer unterstützen, das für Anwendungen benötigt wird und deren Anforderungen sich zeit- und kontextabhängig ändern können;
- Stabilität, Elastizität, Vertrauenswürdigkeit und Sicherheit gewährleisten, die mit Netzen und Softwaredienstplattformen kompatibel sein müssen, die sehr viel grösser und komplexer sind, als wir das von den heutigen Infrastrukturen her kennen;
- die Unterstützung konvergenter und vernetzter Geschäfts- und Dienstleistungsmodelle in den verschiedensten Umgebungen wie beispielsweise Wohnungen, Unternehmen oder wechselnde Arbeitsumgebungen.

Dazu muss in der Entwicklung, ausgehend von bestehenden hin zu neuen Infrastrukturen, ein Mittelweg zwischen Anforderungen wie Abwärtskompatibilität und der Notwendigkeit disruptiver Architekturen für künftige Internet-, Mobil-, Breitband- und dazugehörige Dienstinfrastrukturen gefunden werden.

Die Antriebskräfte für diese Herausforderung stecken vor allem in der technologischen Entwicklung allgegenwärtiger Mobil- und Breitbandnetze der Verfügbarkeit von dynamischen Dienstplattformen. Wichtig sind auch das Vertrauen und die Sicherheit im Umgang mit interoperablen und vernetzten Infrastrukturen. Insofern zielt die vorgeschlagene Massnahme weitgehend auf die technologischen Schranken und sozioökonomischen Annahmen, wie sie in den strategischen Forschungsagenden der Europäischen Technologieplattformen E-Mobilität, NESSI, NEM und ISI festgeschrieben sind.

Die Teilnahme an den Projektvorhaben von Institutionen aus Drittländern wird in den Forschungsbereichen unterstützt, in denen gegenseitiger Nutzen nachweisbar ist. Das betrifft vor allem i) die Möglichkeit, durch gemeinsame strategische Forschungspartnerschaften zu einem globalen Konsens und zu globalen Standards zu gelangen.

**Die Antriebskräfte für diese Herausforderung stecken vor allem in der technologischen Entwicklung allgegenwärtiger Mobil- und Breitbandnetze, der Verfügbarkeit von dynamischen Dienstplattformen. Wichtig sind auch das Vertrauen und die Sicherheit im Umgang mit interoperablen und vernetzten Infrastrukturen.**

gen; ii) die Möglichkeiten für ein gegenseitiges Benchmarking;

iii) den Austausch über die besten Verfahren einschliesslich der regulatorischen und sozioökonomischen Aspekte als technologische Antriebsmittel; iv) die breit angelegte Akzeptanz von Technologien und vernetzten Anwendungen in einem weltweiten Rahmen. Die Beteiligung von Partnern aus Drittstaaten und die Auswahl der vielversprechenden Regionen bleiben der Initiative der jeweiligen Antragsteller überlassen.

### 2.2 Erwartete Ergebnisse

a) Allgegenwärtige Netzinfrastrukturen und -architekturen sollen die nachstehenden Punkte unterstützen:

- i) Konvergenz und Interoperabilität von heterogenen Mobil- und Breitbandnetztechnologien;
- ii) flexibler und raumfüllender Funkzugang für einen Zugang zu mobilen Breitbanddiensten für Netze mit geringer Reichweite bis herunter zu WANs;
- iii) Beseitigung der Schranken für Breitbandzugang und ultraschnelle Ende-zu-Ende-Konnektivität mit optimierten Protokollen und Routing-Verfahren;
- iv) Kontextbewusstsein;
- v) optimierte Verkehrsverwaltung zwischen Kern- und Randnetzen;
- vi) Skalierbarkeit, die eine erhebliche Zunahme der Zahl verbundener Geräte sowie Maschine-zu-Maschine-

oder sensorbasierte (über RFID hinaus) Anwendungen ermöglichen soll. Damit will man künftig in der Lage sein, in unterschiedlichen öffentlichen und privaten Umgebungen funktionstüchtig zu sein.

b) Optimierte Steuerung, Management und Flexibilität der künftigen Netzinfrastruktur, die die Evolution zu kognitiven Netzen unterstützt und die in der Lage ist:

i) nahtlose Endpunkt-zu-Endpunkt-Netzinfrastruktur- und Dienstaufbau und -betrieb zwischen vielfältigen Netzbetreibern und Geschäftsbereichen zu ermöglichen; ii) eine breite Vielfalt von Dienstmerkmalen und -anforderungen zu unterstützen, die sehr viel komplexer sein werden als die heute bekannten Infrastrukturen, durch Unterstützung bei der Programmierbarkeit und der dynamischen Auslastung der Netze, mit erneuter Konfigurierbarkeit bei geänderter Verteilung der Ressourcen, von Protokollen und Routing-Verfahren, Organisation und Management der Netze; iii) neue Formen der Verwaltung von Ad-hoc-Kommunikation in Echtzeit mit intermittierenden Konnektivitätsanforderungen und zeitvariabler Netztopologie; iv) das Ermöglichen einer intelligenten Verteilung der Dienste auf unterschiedliche Zugangstechnologien mit zentraler oder verteilter Steuerung.

c) Technologien und Systemarchitekturen für ein künftiges Internet. Die Überwindung der langfristig zu erwartenden Grenzen des heutigen Internets, seiner Architektur und Protokolle. Die Orientierung an neuen Bedürfnissen: generelle Mobilität; Skalierbarkeit beim Anschluss weiterer Geräte, Dienstmerkmale und Anwendungsumgebungen; Sicherheit; vertrauenswürdige Domänen; neue Routing- und Verteilverfahren mit *dynamischem Peering*, neuen Verfahren der Endpunkt-zu-Endpunkt-Auslieferung und -Steuerung sowie neuen Verfahren mit Ad-hoc-Konnektivität in einer drahtlosen Umgebung. Bei den Untersuchungen wird es darum gehen, wie ver-

Bei den Untersuchungen wird es darum gehen, wie verschiedene Ansprüche in der vorhersehbaren Entwicklung des Internets das World Wide Web einschränken und wo entsprechende langfristige Lösungen gefunden werden können.

schiedene Ansprüche in der vorhersehbaren Entwicklung des Internets das World Wide Web einschränken und wo entsprechende langfristige Lösungen gefunden werden können.

d) Koordinierungs- und Unterstützungsmassnahmen: i) Planungs-

und Netzwerkunterstützung; ii) Koordinierung mit entsprechenden nationalen oder regionalen Programmen oder Initiativen.

### 2.3 Erwartete Auswirkungen

– Globale Standards für eine neue Generation allgegenwärtiger und extrem leistungsfähiger Netz- und Dienstinfrastrukturen. Sie sollen die Konvergenz, die vollständige Interoperabilität und die erheblich höhere Anzahl und Vielfalt von Geräten sowie neue Dienste und die komplexen Anforderungen der Nutzer unterstützen.

– Die Stärkung der industriellen Führungsrolle Europas bei verkabelten und drahtlosen Netzen; die stärkeren Synergien zwischen den Akteuren verschiedener Bereiche und ein Beitrag zu neuen Geschäftsmodellen, die sich Konvergenz und vollständige Interoperabilität zunutze machen.

– Neue Chancen für Industrie und Dienstleistungen in Europa, vor allem im Bereich der Internettechnologien. So soll Europa gemäss seinem technologischen Potenzial eine entsprechende Position in der globalisierten Wirtschaft erreichen.

Quelle: [http://ec.europa.eu/information\\_society/eeurope/j2010/index\\_en.htm](http://ec.europa.eu/information_society/eeurope/j2010/index_en.htm)

[a[r[b|i]d]o]

ANNONCES/INSERATE

– arbido newsletter

– arbido print

– arbido website

→ [inserate@staempfli.com](mailto:inserate@staempfli.com)

→ 031 300 63 89

# IDIAP: institut de recherche et pépinière de projets

**Stéphane Gillioz, en collaboration avec Jean-Albert Ferrez, directeur adjoint, et Céline Aymon Fournier, Relations publiques**

**Institut de recherche universitaire privé, indépendant, à but non lucratif, reconnu par la Confédération et le canton du Valais, l'IDIAP s'est spécialisé dans la gestion de l'information multimédia et les interactions hommes-machine multimodales. Portrait de ce centre de recherche et de développement sis à Martigny, en Valais.**

Lancé en 1991 dans le but de développer l'intelligence artificielle perceptive et d'étudier les interactions homme-machine, l'IDIAP est soutenu dès le début par la ville de Martigny, Swisscom, l'EPFL et le canton du Valais. A l'époque, le Valais, qui n'a pas d'université, décide en effet de mener une politique de promotion d'instituts para-universitaires. Cette politique se traduira notamment par la création de l'IUKB (Institut universitaire Kurt Bösch), à

Sion, et de l'IDIAP, dans le Bas-Valais. Depuis, ce dernier a connu une croissance constante; il abrite aujourd'hui, dans des locaux flambant neufs inaugurés en 2007 (voir photo), une dizaine de start-up et occupe une centaine de personnes (chercheurs, enseignants et personnel administratif). Financé à 30% par des fonds publics et à 70% par des contributions du Fonds national de la recherche et la participation à des projets européens, l'IDIAP collabore également avec des industries dont elle reçoit des mandats ponctuels.

## De solides partenariats

Institut indépendant, l'IDIAP garantit son existence en multipliant les partenariats, autant publics que privés. C'est ainsi qu'il a développé des relations étroites avec les Universités de Lausanne, Fribourg et Berne, avec lesquelles il met en œuvre des projets communs. Autre partenariat privilégié avec l'EPFL qui octroie le doctorat à ses étudiants qui mènent leurs recherches à l'IDIAP. A cela s'ajoute la participation

de l'IDIAP aux recherches internationales dans le cadre de projets de l'Union européenne, qui dispose d'un remarquable vivier de chercheurs. Les partenariats privés ne sont pas oubliés non plus, puisque l'IDIAP en a conclu avec Logitech, Nokia et Swisscom, pour n'en citer que quelques-uns.

Ces partenariats constituent la condition sine qua non pour le transfert des recherches vers la pratique, l'IDIAP jouant ici le rôle d'interface indispensable

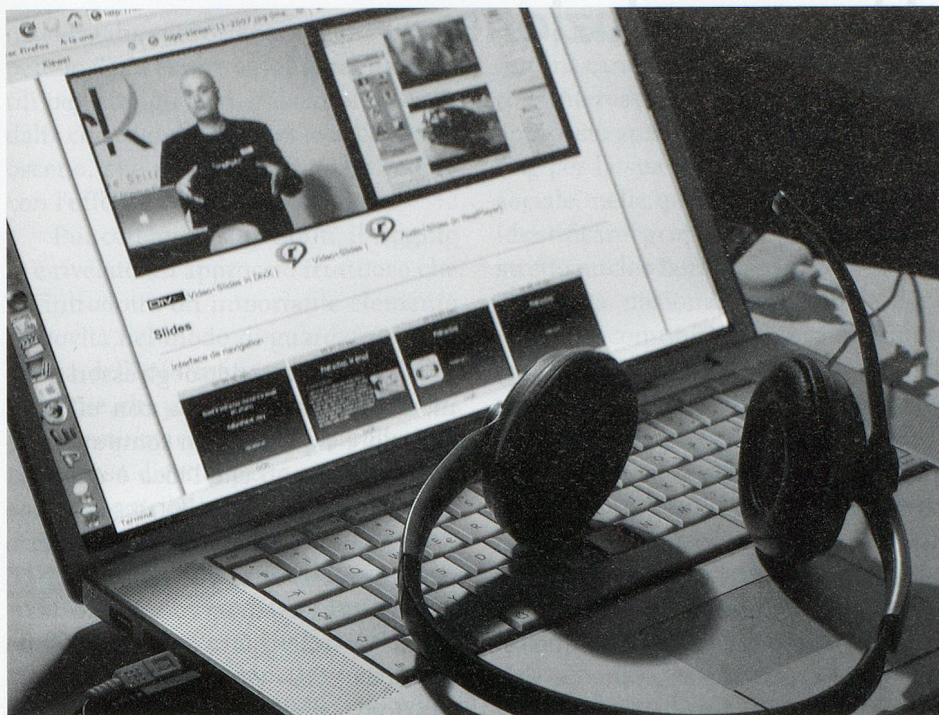
## Le cycle de vie d'un projet de recherche

De l'idée qui surgit à la création d'un nouvel outil technologique, puis à la perspective de le lancer sur le marché, il y a un monde. Le chercheur vérifie la nouveauté, la formalise, et, couchée sur le papier, l'idée se fait projet de recherche. Pour compléter ses compétences, le chercheur s'adjoint des partenaires. Reste à trouver le financement. Mis en forme, le projet est soumis à une instance de soutien. Si l'évaluation est positive, le projet peut vivre: les expériences des différents partenaires sont alors mises en commun, le programme de recherche et le planning définis, et peu à peu les résultats sont rendus publics (publications, sites web, logiciels, conférences, rapports, etc.). Périodiquement, les autorités de financement se penchent sur l'avancement des travaux. Le processus s'achève, selon les cas, par le dépôt d'un brevet ou la livraison d'un prototype.

Dans ce type de processus, l'IDIAP joue un rôle primordial. L'institut offre en effet son aide pour la rédaction et la mise en forme du projet, le suivi administratif et financier, la création de sites internet, l'encadrement du personnel, l'organisation de réunions et de conférences, la réalisation de supports de communication, la diffusion d'informations et la mise à disposition de matériel informatique ultraperformant. Autant de prestations non scientifiques qui constituent un terreau favorable à l'éclosion des projets.

(tiré du «Rapport annuel 2007», p. 7)





ble entre la recherche et le monde industriel.

### Les pôles de recherche

Si l'IDIAP observe attentivement ce qui se passe dans le secteur industriel, les industries s'intéressent elles aussi de très près aux travaux de l'institut. IdeArk, société fille de l'IDIAP, assume le rôle d'interface entre les deux mondes. Elle relaie les préoccupations des entrepreneurs auprès des scientifiques, identifie les tendances du marché et facilite la transformation des résultats de recherche en solutions innovantes. A noter que IdeArk a également pour mission de soutenir les start-up et les PME actives dans les domaines qui sont ceux de l'IDIAP.

L'IDIAP a justement décroché l'un des pôles de recherche défini par le Fonds national de la recherche, à savoir la gestion interactive d'informations multimédias. L'institut se retrouve ainsi au cœur des réseaux de recherche en Suisse dans le domaine de la gestion efficace de grandes quantités de données numériques.

### Exemples de projets

Parmi les projets qui s'inscrivent dans ce contexte, on mentionnera notamment celui de la *spin-off* Klewel (voir [www.klewel.ch](http://www.klewel.ch)) qui propose l'enregistrement de conférences et leur mise en ligne dans les 24 heures, ce qui permet

de les rendre accessibles aux internautes du monde entier.

Quant à KeyLemon, un petit logiciel d'identification biométrique né à l'IDIAP, qui permet à n'importe quel utilisateur possédant une webcam de déverrouiller son ordinateur d'un coup d'œil ou d'un simple sourire, il a été téléchargé par plus de 200 000 utilisateurs dans le monde.

### Collaboration avec les start-up

A l'IDIAP, un projet sur dix provient d'un partenariat avec une entreprise, ce qui place l'institut de Martigny parmi les plus actifs du monde sur le plan industriel. Grâce à l'appui de l'agence fédérale pour la promotion de l'innovation, la CTI, l'institut a pu entreprendre de nouveaux projets avec des start-up et des PME. Ces collaborations avec des entreprises d'avant-garde sont particulièrement intéressantes pour l'IDIAP, car elles permettent d'entamer le dialogue avec les industriels de demain.

Autre projet particulièrement important pour les systèmes de sécurité, la reconnaissance vocale d'une personne qui parle: comment distinguer l'original d'un imitateur? Ce projet a permis de développer un système d'authentification de la voix, qui a d'ores et déjà été testé avec succès par le Service de l'informatique de l'Etat du Valais.

D'autres projets de recherche menés à l'IDIAP sont présentés dans les pages qui suivent (voir les articles de A. Vinciarelli, S. Marcel et J.-P. Gehrig).

Contact: [jean-albert.ferrez@idiap.ch](mailto:jean-albert.ferrez@idiap.ch)  
[celine.aymon@idiap.ch](mailto:celine.aymon@idiap.ch)

## ABSTRACT

### IDIAP: Forschungsinstitut und Projektgenerator

Das IDIAP (Institut Dalle Molle d'Intelligence Artificielle Perceptive) ist ein privates, unabhängiges und universitäres Forschungsinstitut mit Sitz in Martigny. Im Zentrum der Forschungsaktivitäten stehen das Multimedia-Management und die Mensch-Multimodalmaschine-Interaktion.

In den 2007 neu eingeweihten Räumlichkeiten sind u.a. 10 Start-up-Unternehmen untergebracht. Das Institut beschäftigt gegen 100 Mitarbeiter. Das Institut wird zu 30% aus der öffentlichen Hand und zu 70% vom Schweizerischen Nationalfonds sowie von europaweiten Forschungsbeteiligungen finanziert.

Das IDIAP unterhält zahlreiche private (u.a. Logitech, Nokia, Swisscom) und öffentliche (u.a. Universitäten Bern, Freiburg und Lausanne, ETH Lausanne) Partnerschaften. Das IDIAP übernimmt hier die Rolle eines Interface für den Transfer von Forschungsergebnissen in die Praxis (über die Tochtergesellschaft IdeArk). Dabei werden Wünsche und Visionen aus der Industrie aufgenommen und in den Forschungssektor weitergegeben, aber es werden auch innovative Forschungsergebnisse der Industrie zugänglich gemacht. IdeArk unterstützt zudem die Start-up-Unternehmen, die aus den Aktivitäten und Partnerschaften des IDIAP hervorgegangen sind.

Die auf höchstem Niveau stattfindende Grundlagenforschung, die im IDIAP stattfindet, sichert dem Institut eine führende Rolle in seinen Kerngebieten – und zieht zahlreiche talentierte junge Forscher an, die so am Institut erste Schritte in die Welt der Forschung unternehmen können.

# L'utilizzatore: da problema a soluzione

**Dott. Alessandro Vinciarelli, Senior Researcher, IDIAP Research Institute**

**Questo articolo si propone di descrivere alcune delle novità che si osserveranno, o si stanno già osservando, nei sistemi di information retrieval. Parte di ciò che si descriverà è già oggi realtà e fa parte della vita quotidiana di milioni di utilizzatori, parte è ancora oggetto di ricerca e, per il momento, è disponibile solo in forma di prototipo nei laboratori di imprese e università.**

In entrambi i casi, l'attenzione sarà concentrata su quelle tecnologie che hanno l'utilizzatore come elemento attivo nel proprio approccio. Vale a dire su quelle tecnologie che coinvolgono gli utenti, in modo più o meno esplicito e consapevole, in operazioni normalmente svolte in modo automatico. In questo senso, gli utilizzatori si trasformano da problema a soluzione: se prima erano presenti soltanto attraverso le loro necessità da soddisfare, adesso partecipano attivamente al processo tecnologico di cui fruiscono, spesso con efficacia superiore alle macchine.

## Tagging

Da questo punto di vista, la prima importante novità è il *tagging*, una tecnica già oggi applicata da servizi popolari come Flickr e Youtube, i quali ne hanno fatto un punto di forza del proprio successo. Il tagging permette agli utilizzatori che lo desiderano di associare a dati tradizionalmente difficili da trattare, come immagini o video, delle tags, cioè delle parole chiave che descrivono il contenuto dei dati così come gli utilizzatori stessi lo vedono. Si tratta di un'innovazione importante per il motivo seguente: spesso la fase più difficile in un sistema di information retrieval è l'*indexing*, cioè la rappresentazione del contenuto dei dati in una forma adatta per il trattamento automatico. Quando

si tratta di immagini o video, tale operazione è pressoché impossibile per un computer. Infatti, nessuna tecnica automatica è attualmente in grado di capire e descrivere il contenuto di un'immagine o di un video, se non in casi talmente particolari da avere una scarsa utilità pratica. Non a caso, sistemi come Google e Yahoo, che pure sembrano capaci di descrivere il contenuto dei dati multimediali, si basano per l'*indexing* sul testo che accompagna i dati, una tecnica che, per quanto efficace, serve più ad aggirare il problema che non ad affrontarlo.

Il tagging altro non è che un'operazione di *indexing* effettuata dagli utiliz-

zatori, i quali, essendo degli esseri umani, dovrebbero svolgerla con efficacia superiore a quella di un computer. I risultati mostrano che l'idea è effettivamente corretta, ma nello stesso tempo trascura un aspetto importante: proprio perché esseri umani, i taggers, cioè gli utilizzatori che spontaneamente associano le parole chiave ai dati, non sono mossi da criteri tecnologici, ma da motivazioni sociali e personali che spesso finiscono per influenzare negativamente il processo che dovrebbero aiutare. Ecco quindi i taggers «vanitosi» che usano talmente tante tag da avere un'influenza spropositata sul funzionamento del sistema, quelli «egoisti» che usa-

## ABSTRACT

### *Der User: vom Problem zur Lösung*

User werden immer mehr von einem Bestandteil eines Problems (sie haben Bedürfnisse, die über automatisierte Suchanfragen befriedigt werden müssen) zu einem Bestandteil der Lösung (User verbessern dank ihrem interaktiven Beitrag die Qualität der Resultate).

Populäre Dienste wie Flickr oder Youtube nutzen schon heute die Technik des *Taggings*, d.h. schwer zu klassifizierender Inhalt wie etwa Videofilme kann mit Schlüsselwörtern (tags) versehen werden und sind so leichter zu finden. Sie verbessern damit die Indexierung von Inhalten, indem sie den automatischen Moment des Computers durch einen menschlichen ersetzen. Ein Problem entsteht durch ehrgeizige, egoistische oder zerstörerische Tagger, welche das an sich austarierte System durch ihre Aktivitäten verfälschen.

Ein weiterer Ansatz ist der Versuch, User in Interessengruppen oder Klassen einzuteilen («*social computing*»). Mit der Einteilung werden die eventuell verfälschenden Einflüsse eines einzelnen Users abgeschwächt. Die Schwäche des *social computing* ist allerdings genau die, dass User eben Individuen sind und sich nie gänzlich einer Gruppe zuordnen lassen.

Mit dem *social signal processing* wird versucht, User so zu begreifen, wie sich Menschen untereinander verstehen, d.h., der Computer versucht, die «sozialen Signale», die der User aussendet, zu verstehen und «sein» Angebot entsprechend auszurichten. Über Webcams und Mikrofone sollen nonverbale Äusserungen des Users herausgefiltert und interpretiert werden. Dabei entsteht zwangsläufig so etwas wie ein Verkaufsgespräch, der User ist gleichsam der Kunde, und der Computer der Verkäufer.

Das *social signal processing* ist vorderhand Zukunftsmusik und befindet sich aktuell noch im reinen Forschungsstadium. Die Debatte, wie viel Information ein System sammeln darf und wie viel Information ein Nutzer preisgeben muss, um ein derartiges System überhaupt nutzen zu können, ist bereits angelaufen. Die Zukunft wird zeigen, wie viel «Privacy» die User letztlich bereit sind preiszugeben.

no tags talmente orientate ai propri fini personali da essere assolutamente inutili per tutti gli altri, o i semplici «vandali» che usano tags, spesso di carattere osceno, che nulla hanno a che vedere con l'effettivo contenuto dei dati.

Pur con tutti i suoi limiti, il tagging si è rivelato un approccio fruttuoso che ha introdotto un importante elemento di novità nel modo di guardare al problema dell'information retrieval. L'utilizzatore, che prima era visto come semplice destinatario finale dell'applicazione, adesso diventa parte attiva del processo e contribuisce al funzionamento di un sistema tanto quanto le componenti automatiche. Questo cambiamento di visione ha dato origine a molteplici linee di ricerca che daranno frutti soltanto tra diversi anni. Il resto dell'articolo descrive quelle che sono, almeno a giudizio dell'autore, le più promettenti e interessanti.

### Social computing

Un primo tentativo, in parte già attuato in alcuni servizi on-line, è quello di rag-

gruppare gli utilizzatori in classi o comunità caratterizzate da comportamenti e interessi comuni. Si tratta di un approccio spesso definito *social computing* per la sua similitudine con la vita sociale, nella quale è sempre possibile identificare gruppi che vanno dal ristretto nucleo familiare fino alle intere comunità nazionali. Raggruppare gli utenti in comunità significa eliminare gli aspetti più individuali di ogni utilizzatore in favore di quelli più comuni e quindi più affidabili dal punto di vista statistico. Il servizio che ne risulta è meno centrato sul singolo utilizzatore, che si trova ad avere risposte mediate sul profilo del gruppo cui appartiene, ma nello stesso tempo è molto più robusto rispetto ai problemi del tagging descritti in precedenza («vanità», «egoismo» e «vandalismo»).

Il limite fondamentale dell'approccio risiede nel fatto che ogni utilizzatore resta un individuo e non è possibile identificarlo totalmente con un gruppo. Inoltre, la stessa persona può apparte-

nere a gruppi diversi in momenti diversi e ciò introduce una notevole ambiguità. Alla luce di ciò, una corrente di ricerca ancora in fase embrionale, ma in rapida espansione, mira a rendere i computer capaci di capire le persone così come lo fanno normalmente gli esseri umani. In altre parole, si tratta di insegnare ai computer a cogliere le necessità e le condizioni degli utilizzatori nello stesso modo in cui un buon venditore riesce immediatamente a capire se il cliente ha o no intenzione di comprare e quali sono i migliori argomenti per convincerlo a farlo quando è in dubbio.

### Social signal processing

Tale linea di ricerca, nota come *social signal processing*, fa leva sulla comunicazione non verbale come fonte di informazione circa l'attitudine di una persona rispetto agli altri e agli oggetti (per quanto possa sembrare sorprendente, le persone tendono a rapportarsi con le cose nello stesso modo in cui lo fanno con le persone). Nell'ambito dell'information retrieval, ciò significa che la macchina diventerebbe capace di capire l'utilizzatore, il suo stato d'animo, i suoi bisogni, e tutto quanto necessario per offrire un miglior servizio, nello stesso modo del personale di un buon locale che cerca di far sentire i clienti a proprio agio.

Il social signal processing consiste nell'utilizzare i sensori di cui molti computer sono oggi dotati (microfoni e webcam) per identificare comportamenti non verbali come la postura dell'utilizzatore, la sua espressione facciale, la direzione del suo sguardo, i suoi gesti e quant'altro può essere carpito per via automatica. Secondo le indicazioni di psicologia sociale e antropologia, tali comportamenti sono direttamente correlati con l'attitudine delle persone verso la situazione in cui si trovano, quindi possono fornire indicazioni preziose su quali sono i bisogni degli utilizzatori e su come soddisfarli. In questo senso, l'utilizzatore diventa una volta di più partecipante attivo, per quanto del tutto inconsapevole, del processo tecnologico di cui fa uso.

Laddove tagging e social computing sono già oggi parte della realtà, sebbene

## ABSTRACT

### *L'utilisateur: du problème à la solution*

D'élément du problème (il a des besoins qui doivent être satisfaits via des requêtes automatisées), l'utilisateur devient de plus en plus un élément de la solution (grâce à sa contribution interactive, il améliore la qualité des résultats).

Les services populaires tels que Flickr ou Youtube utilisent d'ores et déjà la technique du *tagging*, c'est-à-dire que les contenus difficiles à classer comme les films vidéo, par exemple, peuvent être pourvus de mots clés (tags ou étiquettes), ce qui permet de les trouver plus facilement. Ils améliorent ainsi l'indexation des contenus dans ce sens qu'ils remplacent le moment automatique de l'ordinateur par un moment humain. Un problème peut survenir lorsqu'un tagger falsifie par ses activités un système en soi équilibré, que ce soit pour des raisons égoïstes ou destructrices.

Une autre approche est d'essayer d'intégrer un utilisateur dans un groupe d'intérêt ou une classe (*social computing*). Une telle intégration permet d'affaiblir les éventuelles influences néfastes d'un utilisateur mal intentionné. La faiblesse du *social computing* réside toutefois dans le fait que les utilisateurs sont justement des individus et qu'ils ne se laissent pas complètement subsumer sous un groupe.

Le *Social Signal Processing* est quant à lui censé permettre de saisir l'utilisateur de la manière dont les individus se comprennent, c'est-à-dire que l'ordinateur essaie de comprendre les «signaux sociaux», que l'utilisateur envoie, et d'orienter «son» offre en conséquence. Les expressions non verbales de l'utilisateur sont filtrées et interprétées via les webcams et les micros. Il en résulte obligatoirement une sorte d'entretien de vente, l'utilisateur devenant en même temps le client, et l'ordinateur le vendeur.

Le *Social Signal Processing* est pour l'instant encore de la musique d'avenir et se trouve actuellement au stade de la pure recherche. Les débats sur la quantité d'informations qu'un système peut récolter et combien d'informations un utilisateur doit livrer pour pouvoir utiliser un système de ce type sont déjà lancés. L'avenir montrera combien de *privacy* les utilisateurs seront en fin de compte disposés à abandonner.

a uno stadio diverso di sviluppo e diffusione, il social signal processing è ancora pura ricerca. Tuttavia si apre già un dibattito su quanto sia opportuno rendere le tecnologie capaci di raccogliere informazioni sempre più dettagliate circa gli utilizzatori. La discussione su tale tema esula da questo articolo, ma è sempre bene ricordare che molte tecnologie sulle quali si fa quotidiano affida-

mento hanno un prezzo in termini di informazioni private che si rilasciano, più o meno consapevolmente, a terzi (carte di credito, prenotazioni on-line, ecc.).

Poche tecnologie di punta finiscono, come l'information retrieval, per avere un impatto sociale così ampio (non a caso si parla di sociotecnologie). Il loro

sviluppo sarà pertanto determinato non solo da fattori puramente tecnici e scientifici, ma anche dal più ampio dibattito che tale sviluppo determinerà nella società. I prossimi anni ci diranno se e quanto gli utilizzatori accetteranno di partecipare ai processi tecnologici a scapito della loro privacy.

Contatto: [alessandro.vinciarelli@idiap.ch](mailto:alessandro.vinciarelli@idiap.ch)

## La conservation numérique des films argentiques

### Une solution idéale pour la consultation et le catalogage du contenu

Jean-Pierre Gehrig, André Perrig et Pierre Ihmle, Cinetis SA

**Les films sur pellicule constituent un patrimoine historique, artistique et culturel inestimable. Pourtant, la majorité de ces films ont déjà disparu ou vont disparaître au cours du XXI<sup>e</sup> siècle. De plus, l'accès à leur contenu est très limité. Une solution simple et efficace pour préserver et consulter ces archives sans les détériorer: la numérisation suivie de l'indexation semi-automatique.**

Aujourd'hui, environ 90% des films muets et 50% des films tournés avant 1950 sont définitivement perdus [1]. Car, bien que la pellicule soit un excellent support de conservation, les conditions de stockage sont limitantes: la température et le taux d'humidité doivent être rigoureusement contrôlés. L'émulsion photosensible – la plus importante des trois couches qui constituent la pellicule – est très vulnérable à l'humidité. Un lieu de stockage trop humide entraîne donc la prolifération de moisissures et autres bactéries hydrophiles qui détruisent peu à peu les colorants organiques qui composent l'émulsion photosensible. Le support physique de cette dernière peut également s'altérer avec le temps et l'humidité: le triacétate de cellulose se transforme en acide acétique et une

odeur de vinaigre très caractéristique des pellicules en voie de détérioration sévère se dégage. Ce phénomène est plus connu sous le nom de «syndrome du vinaigre» [2]. La qualité de stockage des bobines de films cinématographiques demeure problématique et l'accès au contenu de ces films est aujourd'hui quasiment impossible, la projection à l'aide d'un projecteur pouvant gravement endommager la pellicule et accélérant immanquablement son vieillissement. L'utilisation de ce patrimoine à caractère unique pour la consultation ou la recherche est par conséquent fortement déconseillée.

La solution imaginée par Cinetis SA à Martigny, en collaboration avec l'Institut de recherche IDIAP, pour préserver et accéder aisément au patrimoine cinématographique sans l'endommager, est une numérisation systématique des films d'archive suivie d'une analyse semi-automatique de leur contenu. Ce qui veut dire que les films sont mis en état, si possible nettoyés, avant d'être numérisés par un laboratoire spécialisé, puis ils sont stockés sur une archive digitale consultable à souhait. Ensuite, l'analyse semi-automatique du contenu assiste le documentaliste dans l'annotation des œuvres. L'archive digitale augmentée des métadonnées permet un accès ra-

pide et structuré aux objets conservés, facilitant le travail de l'archiviste dans ses tâches de gestion, d'entretien et de développement des collections. Cette nouvelle technologie permet d'apprécier le contenu des films d'archives à un niveau de qualité suffisant pour la recherche et la documentation, sans risque de détérioration pour l'original. Elle permet également le repérage de films d'exception qui mériteraient d'être numérisés à un niveau de qualité supérieur ou même d'être restaurés manuellement.

Pour l'avenir, l'indexation des films numérisés va s'automatiser. Des recherches sont en cours pour développer de telles méthodes. L'indexation semi-automatique permettra de transformer les dialogues en texte, d'extraire un texte d'une séquence ou encore d'analyser le contenu des plans. Il est même envisagé, dans un futur proche, de reconnaître automatiquement les visages ou la voix d'une personne afin de faciliter la recherche de personnalités connues. Il serait aussi possible de repérer des plans similaires par rapport à une image de référence.

Cinetis SA est également en train de mettre au point un scanneur de films qui permettra de numériser de manière qualita-

tive, respectueuse et efficace les formats Super8, 8 mm, 9,5 mm et 16 mm, à un coût abordable. L'appareil sera accompagné d'un logiciel de segmentation automatique qui décompose les films en plans. Les plans sont représentés sous forme d'images clé permettant un aperçu rapide du contenu. Cette innovation permettra à l'opérateur de repérer immédiatement les zones sensibles à traiter manuellement.

La conservation des films argentiques sous forme numérique est une alternative simple, évolutive et d'un coût abordable. Elle permet la valorisation et l'accessibilité des films de professionnels comme de privés. Les films ainsi numérisés et indexés constitueront un catalogue des trésors filmographiques encore trop peu connus des fonds publics et privés. Sans parler des films inédits en possession des particuliers.

Contact: jp.gehrig@cinetis.ch

#### Bibliographie

- [1] M. Friend, Film/digital/film, in AMIA Conference, octobre 1994
- [2] Commission Supérieure Technique de l'Image et du Son (CST), La restauration numérique des films cinématographiques, CST, Paris, 1997

## Wertvolles Filmmaterial digitalisieren und aufbewahren

### Eine ideale Lösung für die Einsicht und die Katalogisierung des Inhalts

Jean-Pierre Gehrig, André Perrig und Pierre Ihmle, Cinetis SA

Alte Filme sind ein historisches, künstlerisches und kulturelles Erbe von unschätzbarem Wert. Der grösste Teil davon ist leider nicht mehr auffindbar oder wird im Laufe des 21. Jahrhunderts verschwinden. Ausserdem ist der Zugang zu diesen Filmen begrenzt. Eine einfache und effiziente Lösung, um diese Archiven zu schützen und einzusehen, ohne sie zu beschädigen, ist die Digitalisierung und die halbautomatische Indexierung.

Heute sind 90% der Stummfilme und 50% der vor 1950 gedrehten Filme für immer verloren [1]. Das rührt daher, dass Filme, obwohl sie ausgezeichnete Aufbewahrungsträger sind, nur eingeschränkt lagerbar sind: Es muss streng auf die Temperatur und den Feuchtigkeitsgrad geachtet werden, damit die Filme keinen Schaden nehmen. Die lichtempfindliche Emulsion – die wichtigste der drei Schichten, aus denen ein Film besteht – ist sehr feuchtigkeitsempfindlich. Eine Lagerung bei zu hoher Feuchtigkeit führt zu einer Vermehrung von Schimmelpilzen und anderen hydrophilen Bakterien, die nach und nach die organischen Farb-

mittel zerstören, aus denen die lichtempfindliche Emulsion besteht. Auch der physikalische Träger der Emulsion kann sich mit der Zeit wegen der Feuchtigkeit verschlechtern: Das Zellstoff-Triazetat verwandelt sich in Essigsäure, und es entfaltet sich ein ausgeprägter Essiggeruch (auch bekannt als «Essig-syndrom»), welcher für beschädigte Filme charakteristisch ist [2]. Die Lagerungsqualität der Filmrollen ist und bleibt ein generelles Problem. Der Zugang zu diesem Material ist jedoch weiterhin fast unmöglich; die Vorführung führt unweigerlich zu einer Beschleunigung des Alterungsprozesses. Die Verwendung dieses einmaligen Erbguts nur zur Einsicht oder zu Forschungszwecken ist deshalb nicht ratsam.

Cinetis S.A. in Martigny arbeitet zusammen mit dem Forschungsinstitut IDIAP daran, eine Lösung zu entwickeln, um alte Filme aufbewahren zu können, ohne sie zu beschädigen. Dazu werden die Archivfilme systematisch digitalisiert, und deren Inhalt wird halbautomatisch analysiert. Danach stehen die Filme wieder zur allgemeinen Besichtigung zur Verfügung.

Das bedeutet, dass die Filme wenn möglich gereinigt werden, bevor sie in einem Speziallabor digitalisiert wer-

den. Man überträgt sie dann auf einen Server, wo sie ohne weiteres angesehen werden können. Die halbautomatische Analyse des Inhalts hilft dem Dokumentalisten, die Werke mit den gewünschten Anmerkungen zu versehen. Das digitale Archiv, das dank der Metadaten fachdienliche Angaben enthält, erlaubt einen schnellen und einfachen Zugang zu den gespeicherten Objekten und vereinfacht die Verwaltung, den Unterhalt und den Ausbau der Sammlungen. Anhand dieser neuen Technologie kann man Archivfilme in einer zufriedenstellenden Qualität sehen, ohne dass das Original dabei Schaden nimmt. Darüber hinaus können mit diesem Verfahren auch aussergewöhnliche Filme entdeckt werden, die es wert sind, in einer höheren Qualität digitalisiert oder von Hand restauriert zu werden.

In Zukunft wird die Indexierung von digitalisierten Filmen automatisiert werden. Neue Methoden befinden sich in der Entwicklungsphase. Dank der halbautomatischen Indexierung wird es möglich sein, Dialoge in Texte umzuwandeln, einen Text aus einem Bildausschnitt zu entnehmen oder den Inhalt einer Aufnahme zu analysieren. Es wird sogar an einem Projekt gearbeitet, das darauf

abzielt, Gesichter oder Stimmen automatisch erkennen zu können. Dadurch wird die Suche nach berühmten Personen erleichtert werden. Anhand eines Referenzbildes wird es zudem möglich sein, ähnliche Ausschnitte aufzufinden.

*Cinetis S.A. arbeitet momentan auch an einem Filmscanner, um Super-8-, 8-mm-, 9,5-mm- und 16-mm-Filme qualitativ, originaltreu und effizient zu einem erschwinglichen Preis zu digitalisieren. Dieser Apparat wird mit einem Softwareprogramm ausgestattet sein, das die Filme dank einer automatischen Seg-*

mentierung in einzelne Sequenzen aufteilt. Diese Sequenzen – in Form von Schlüsselbildern – erlauben es dem Operator, die empfindlichen Zonen zu finden, die er manuell digitalisieren will.

Die Erhaltung von wertvollen Filmmaterial in digitalisierter Form ist eine einfache, erschwingliche Alternative. Sie erlaubt die Aufwertung von Filmen, sei es im professionellen oder im privaten Bereich. Die digitalisierten und indexierten Filme bilden einen Katalog von gefilmten Kunstwerken, die bislang verborgen waren. Und vergessen wir nicht all die unveröffentlichten

Filme, die sich noch in Privatbesitz befinden ...

Kontakt: [jp.gehrig@cinetis.ch](mailto:jp.gehrig@cinetis.ch)

#### Bibliografie

- [1] M Friend, Film / digital / Film Dans AMIA Conference AMIA Octobre 94
- [2] Commission Supérieure Technique de l'Image et du Son (CST). La restauration numérique des films cinématographiques. CST, Paris, 97

## Ergonomic Minding of Media Collections

Dr. Sébastien Marcel\* and  
Dr. Stéphane Marchand-Maillet\*\*

**Personal or professional collections of media, such as photos, music, movies and home videos tend to increase quickly in size, mainly due to the ease of collecting lots of digital content using a variety of capture devices (digital cameras, mobile phones, camcorders) or over the Internet. However, the level of management (ie organization and search) of those multimedia collections remains largely behind the ease of content creation.**

### Ergonomic Minding of Media Collections

In spite of advances in content-based retrieval and automatic multimedia indexing, multimedia content management is still difficult. For example, when it comes to personal photography,

the frequent procedure is to place items into folders, often organized by date or periods, possibly events, and loosely annotated. As a result, images and events are difficult to find and search becomes a frustrating or even painful operation because tools for browsing personal image collections are not appropriate to the user's needs. In some cases, it is possible to organize a collection by accurate date/time or location/place (thanks to the EXIF and GPS data recorded by modern capture devices). However, this is still limited and it remains desirable to search using the multimedia content itself.

The current challenge of multimedia information systems is thus to design and provide professional or non-professional users with new interactive tools that:

- improve the browsing experience to access both personal or professional collections;
- make the search easier and more natural than the folder-style layout;
- allow finding images and videos easily, rapidly, and accurately.

Media management can be significantly improved using the current results of low-level content abstraction tech-



Dr. Sébastien Marcel

niques (e.g., color, contours, texture) and high-level content abstraction techniques (e.g., image classification, face recognition) by combining those results with information obtained by minding a particular collection and developing novel browsing interfaces guided by the user's personal preferences.

The *Ergonomic Minding of Media Collections* is thus facing the following challenges:

- extract relevant information from the media content for efficient indexing, search and retrieval. More precisely, the following technologies should be

\* Idiap Research Institute, Centre du Parc,  
Rue Marconi 19, CP 592, 1920 Martigny,  
Switzerland, phone: +41 (27) 721 77 27,  
e-mail: [marcel@idiap.ch](mailto:marcel@idiap.ch)

\*\* Viper group – CUI – University of Geneva,  
Battelle Building A, 7, Route de Drize,  
1227 Carouge, Switzerland,  
phone: +41 (22) 379 01 54,  
e-mail: [marchand@cui.unige.ch](mailto:marchand@cui.unige.ch)

investigated: (i) robust extraction of invariant visual descriptors for image classification and object recognition, (ii) accurate and reliable unconstrained face detection and recognition as the human face is recognized to be an important semantic cue in visual content.

- develop novel multimedia minding strategies and techniques to optimally prepare and enrich the collection content for constructing new user interaction models.
- propose and validate in front of users new interaction models for image/video search and browsing based on clustering or topology-preserving di-

**In spite of advances in content-based retrieval and automatic multimedia indexing, multimedia content management is still difficult.**

dimensionality reduction and projection techniques. The main goal is to offer the user a global and precise access to the multimedia collection at minimal cost.

In an example scenario, the user navigates within a flat 2-dimensional arrangement of his/her media collection, organized by clusters. Each cluster corresponds to a given search criterion (query) and is represented by a statistical visual summary of the query result. One main challenge is to make available data interoperable at all levels. Thus, the construction of clusters

would use and combine EXIF information (e.g., date/time, location), meta-data (e.g., tags, events) and/or content (e.g., dominant color, people occurrence). In a complementary faceted-like approach, if the current facet combination isolates a subset of the collection,

**One main challenge is to make available data inter-operable at all levels.**

then it is possible to further refine the search by seamlessly navigating within the cluster-based representation of this subset.

### Examples of Existing Systems

As examples of relevant directions in the development of the *Ergonomic Minding of Media Collections* methodology, we detail two applications illustrating different perspectives.

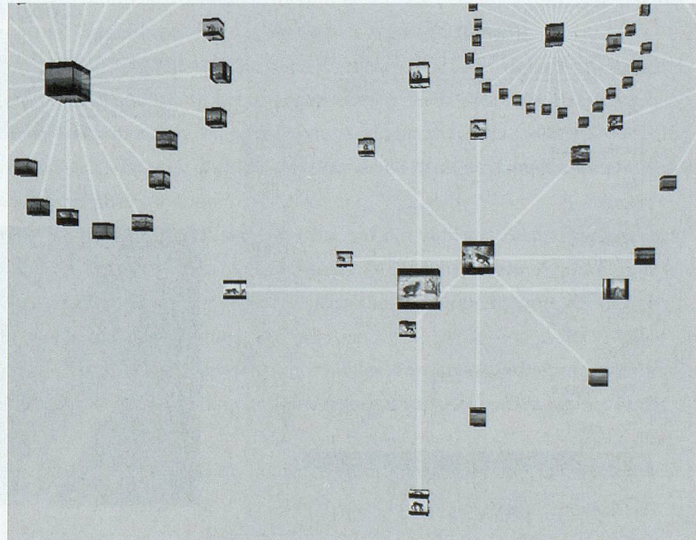
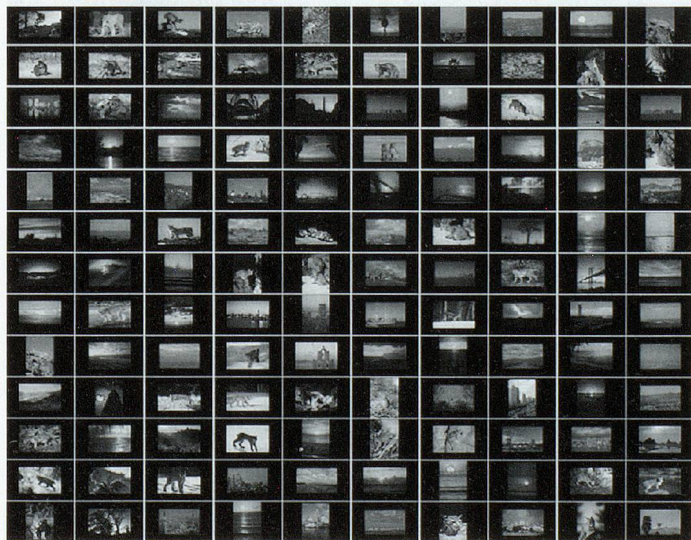
#### Collection Guiding

The Collection Guide proposes an alternative approach to many current information management systems, which are centered on the notion of a query. This is true over the Web (with all classical Web Search Engines), and for digital libraries. In the domain of multimedia, available commercial applications propose rather simple management services whereas research prototypes are also looking at responding to queries. The notion of browsing comes as a complement or as an alternative to query-based operations in several possible contexts.

In the most general case, multimedia browsing is designed to supplement search operations. This comes from the fact that the multimedia querying systems largely demonstrate their capabilities using the Query-by-Example (QBE) scenario, which hardly corresponds to any usable scenario. Multimedia search systems are mostly based on content similarity. Hence, to fulfill an information need, the user must express it with respect to relevant and non-relevant examples. The question then arises of how to find the initial examples themselves. Researchers have therefore investigated new tools and protocols for the discovery of relevant examples. These tools often take the form of browsing interfaces whose aim is to help the user exploring the information space in order to locate the sought items.

**Researchers have therefore investigated new tools and protocols for the discovery of relevant examples. These tools often take the form of browsing interfaces whose aim is to help the user exploring the information space in order to locate the sought items.**

In Marchand-Maillet, 2005, the principle of *Collection Guiding* is introduced. Given the collection of images, a path traversing the complete multimedia collection is automatically created so as to “guide” the visit of the collection. For that purpose, image intersimilarity is computed and the path is created via a



Travelling Salesman tour of the collection. The aim is to provide the user with a base exploration strategy based on a minimal variation of content at every step. This implicitly provides a dimension reduction method from a high-dimensional feature space to a linear ordering. The Collection Guide methodology provides also several multidimensional arrangements and is therefore directly related to information visualization.

The figure on page 23 shows an example opposing the classical unordered grid-based display and a 3D display

strategy exploiting content similarity and clustering. Similar images are automatically arranged around their representative image in a planet metaphor. The user may thus obtain a global overview of the collection (interplanet arrangement) or visit a specific subset of the collection (i.e. visit a specific “solar system”).

It is now clear that browsing comes as a necessity to closely adapt information inspection and retrieval to the specific user’s needs. There is no doubt that future information systems will comprise this emergent aspect, as a comple-

ment to currently dominating search operations.

### Google Portrait





In 2007, S. Marcel and al. proposed *Google Portrait* (<http://www.idiap.ch/googleportrait>) to retrieve and browse images from the Internet containing only one particular object of interest: the human face. The goal is to filter the

**The Collection Guide methodology provides also several multidimensional arrangements and is therefore directly related to information visualization.**

Search for portraits of a person. Example: celebrity portraits such as *Paris Hilton* or *Tom Cruise*

Tom Cruise      Search

**Detected faces**

| Faces   | Confidence | Size    | Original image                 | Site  |
|---|------------|---------|--------------------------------|---|
|    | 1          | 300x410 | <a href="#">Original image</a> | <a href="http://www.hitparade.ch">http://www.hitparade.ch</a>           |
|    | 2          | 375x242 | <a href="#">Original image</a> | <a href="http://www.ioblo.com">http://www.ioblo.com</a>                 |
|  | 3          | 393x590 | <a href="#">Original image</a> | <a href="http://www.kinoweb.de">http://www.kinoweb.de</a>               |
|  | 4          | 365x500 | <a href="#">Original image</a> | <a href="http://imstars.aufeminin.com">http://imstars.aufeminin.com</a> |

images provided by a standard image retrieval system with a face detector and to present portraits as a result instead of the complete image.


Image search starts with a text query. The Google Image engine is used to retrieve images matching the query. Each image URL is extracted from the Google Image result page, then images are processed in parallel. This processing includes download and face finding. Images with detected faces are presented on a new result page listing face portraits together with a confidence and direct links to the image URL and to the source page. The result page is a table with 5 columns and with as many rows than images with detected faces. The first column contains image close-ups (“portraits”), the second column contains a confidence on the likelihood of the portrait to be a face, the third column contains the size of the original

**Renaming**

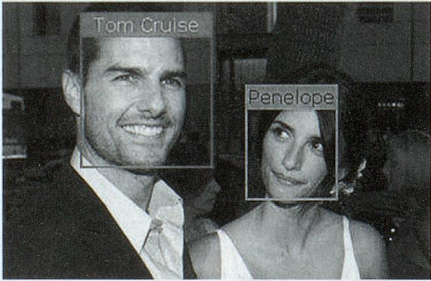
Name:

- Thomas Jüngling
- Thomas Ruland
- Thomas Sidler
- Thomas Wiegandt
- Tiam
- Tiger Wood
- Tiger Woods
- Tim Cootes
- Timberlake
- Time
- Timo Stripl
- Timothy Cootes
- Timothy Suh
- Tobey
- Tobias
- Toby Maguire
- Tom
- Tom Cruise
- Tom Cruis
- Tom Cruise

**Google Portrait**



Click on a face to tag it!



Each image URL is extracted from the Google Image result page, then images are processed in parallel.

image and the last columns contain links on the the original image and on its website.

Google Portrait includes a module for manual annotation. Indeed, portraits (detected faces) are very likely to correspond to the query, but there is no guarantee as Google Portrait uses a face detection system, not a face recognition system. It is then possible to edit a result and to change the tag of the portrait

(name of the person). Tags are saved in a database which can then be populated based on collaborative working.

Google Portrait has been released on November 27, 2006. Later in spring 2007, Google Image was providing an "unofficial" face finding search mode. Google Image face finding is now directly available in Advanced image

search at least since June 24, 2007. Nearly at the same time, probably during summer 2007, Microsoft added also face detection to Microsoft's Live Search. The comparison with Google Portrait in terms of performance is impractical since both Google and Microsoft (those companies have incomparable computing facilities) online face

detectors give results on already batch-processed images. Conversely, Google Portrait is performing live face detection on images downloaded on the fly from the Internet.

### Conclusion

The *Ergonomic Minding of Media Collections* implicitly acknowledges the limits

## ABSTRACT

### *Le maniement ergonomique des collections de médias*

Il est aujourd'hui très simple de se constituer des collections de médias (photos, musique, films, etc.). La gestion et le maniement de grandes quantités de données, qui deviennent rapidement trop volumineuses, sont néanmoins toujours aussi difficiles.

La gestion des médias peut être considérablement améliorée si l'on combine des techniques *low-level* pour abstraire des contenus (p.ex.: couleurs, contours, surfaces) avec des techniques *high-level* (p. ex.: classification des images, sélection des visages).

Tandis que le «browsing» classique dépend d'indications de recherche précises, le *Collection Guiding* renverse le processus de recherche: les collections sont en effet computérisées, les vues multidimensionnelles sont réduites à une présentation linéaire le long de laquelle le «browser» peut se mouvoir; il apprend ainsi à reconnaître le portrait simplifié de la collection complète et peut décider dans quelle direction, respectivement dans quel espace il veut poursuivre sa recherche. De tels systèmes de *browsing*, qui peuvent être adaptés aux besoins de tout un chacun, remplaceront de plus en plus à l'avenir les systèmes de recherche rigides.

*Google Portrait* procède d'une autre manière: le système délaisse toutes les autres composantes au profit d'un seul et unique critère de recherche: le visage humain. Les images sont filtrées au moyen d'un programme de détection de visages (à ne pas confondre avec un programme de reconnaissance des visages; en anglais: *to detect*, resp. *to recognize*) où uniquement le portrait et non pas l'image dans son ensemble est intégré dans les résultats de la recherche.

Microsoft a suivi entre-temps avec une détection des visages en mode «Live Search».

Globalement, on constate dans la gestion des objets multimédias un glissement du traitement classique des contenus avec indexation vers une approche qui privilégie le *knowledge management* et le *data minding* et qui implique davantage l'utilisateur dans la construction de systèmes interactifs.

En Europe et en Suisse, de nombreux projets sont actuellement en cours dans le domaine de la recherche sur les contenus multimédias et leur organisation.

## ABSTRACT

### *Ergonomischer Umgang mit Mediasammlungen*

Es ist heutzutage einfach, digitale Mediasammlungen (Fotos, Musik, Filme etc.) anzulegen. Die Verwaltung und der Umgang mit den rasch wachsenden und damit unübersehbaren Datenmengen gestalten sich aber immer schwieriger.

Das Medienmanagement kann erheblich verbessert werden, wenn einfache Techniken der Bildbeschreibung, z.B. Farben, Konturen, Oberfläche, mit entsprechenden anspruchsvollen Techniken, z.B. Bildklassifizierung, Gesichtsdetektion, gemeinsam eingesetzt werden.

Während herkömmliches «Browsen» von präzisen Sucheingaben abhängig ist, kehrt das «*Collection Guiding*» den Suchvorgang um: Sammlungen werden informatisiert, mehrdimensionale Ansichten auf eine lineare Darstellung heruntergebrochen, deren entlang sich der «Browse» bewegen kann. Der Nutzer lernt so das vereinfachte Abbild der gesamten Sammlung kennen und kann sich entscheiden, in welche Richtung resp. in welchem Raum er weitersuchen möchte. Derartige Browsersysteme, die sich den Wünschen des Einzelnen anpassen können, werden starre Suchsysteme in naher Zukunft vermutlich in den Hintergrund drängen.

Einen anderen Weg geht *Google Portrait*: Das System vernachlässigt alle übrigen Suchelemente bis auf ein einziges Suchkriterium: das menschliche Gesicht. Bilder werden mittels eines Gesichtsdetektionsprogramms (nicht zu verwechseln mit Gesichtserkennungsprogramm; engl. *to detect* resp. *to recognize*) gefiltert, anstelle des gesamten Bildes wird ausschliesslich das Portrait in das Suchergebnis einbezogen. Microsoft hat in der Zwischenzeit mit einer Gesichtsdetektion in «Live Search» nachgezogen.

Insgesamt zeigt sich beim Multimedia-Management eine Verschiebung des Ansatzes von der traditionellen Content-Verarbeitung mit Indexierung hin zu einem neuen Ansatz, der das Knowledge-Management und das Data Minding in den Vordergrund stellt und damit den Nutzer stärker als bislang in den Aufbau von interaktiven Systemen einbezieht.

Zur Zeit sind in der Schweiz und in Europa zahlreiche Projekte zum Thema Suche nach und Organisation von multimedialen Inhalten am Laufen.

of current information access systems and paves the way to new solutions and challenges in the Multimedia Information Retrieval and Management community. The emphasis is placed on shifting the focus from the traditional content-processing and indexing viewpoint to a knowledge and data minding approach complemented with a strong involvement of users in the construction of interactive systems.

We believe that such a joint data-processing and user-centric approach will demonstrate that the strong implication of users, as a source of semantic information via dedicated interfaces adapted to efficiently capture useful information, is a robust and scalable solu-

**The emphasis is placed on shifting the focus from the traditional content-processing and indexing viewpoint to a knowledge and data minding approach complemented with a strong involvement of users in the construction of interactive systems.**

tion to the problem of high-level management of multimedia information.

Various European and Swiss research projects in the area of Multimedia Information Retrieval and Management are currently active. Additionally, a project initiative on the *Ergonomic Minding of Media Collections* is currently under setup to target the development of these new generations of interactive multimedia management systems and to encourage their transfer to usable commercial solutions.

Contact: marcel@idiap.ch

stephane.marchand-maillet@unige.ch

#### References

- S. Marchand-Maillet and E. Bruno, "Collection Guiding: A new framework for handling large multimedia collection", *International Workshop on Audio-Visual Content and Information Visualization in Digital Libraries (AVIVDiLib'05)*, 2005 (<http://viper.unige.ch/collectionguide>).
- S. Marcel, P. Abbet and M. Guillemot, "Google Portrait" (<http://www.idiap.ch/googleportrait>), *Idiap Communication, Idiap-Com-07-2007*, 2007.

Katholischer Konfessionsteil des Kantons St. Gallen



STIFTSBIBLIOTHEK ST. GALLEN

Auf den 1. Februar 2009 oder nach Vereinbarung ist die Stelle eines

**wissenschaftlichen Bibliothekars (60%)**

oder einer

**wissenschaftlichen Bibliothekarin (60%)**

neu zu besetzen.

Die Stiftsbibliothek mit ihren reichen und wertvollen Beständen an mittelalterlichen Manuskripten und Inkunabeln ist heute eine moderne wissenschaftliche Fachbibliothek für Mediävistik, Handschriftenkunde sowie St. Galler Klostergeschichte und Klosterkultur. Sie zählt zu den bedeutendsten und schönsten Bibliotheken in Europa und gehört auch zu den meistbesuchten Museen der Schweiz.

Zum Aufgabenbereich des wissenschaftlichen Bibliothekars oder der wissenschaftlichen Bibliothekarin gehören die Betreuung der Forschungsbibliothek und des Lesesaals, die Benützung und Konservierung der Handschriften, das Bestellwesen und die Katalogisierung sowie die Zusammenarbeit mit anderen Bibliotheken. Nach Anweisung der Bibliotheksleitung sind weitere Aufgaben vor allem administrativer Art zu übernehmen.

Anforderungen: abgeschlossenes Universitätsstudium geisteswissenschaftlicher Richtung und Ausbildung zum wissenschaftlichen Bibliothekar oder zur wissenschaftlichen Bibliothekarin, gute EDV-Kenntnisse, organisatorische und kommunikative Fähigkeiten, Interesse an der Geschichte und Kultur des Klosters St. Gallen.

Weitere Auskünfte erteilt: Dr. Karl Schmuki,  
Tel. 071 – 227 34 20.

Bewerbungen sind bis zum 1. Januar 2009 zu richten an:  
Prof. Dr. Ernst Tremp, Stiftsbibliothekar, Stiftsbibliothek  
St. Gallen, Klosterhof 6d, 9004 St. Gallen.

[ a [ r [ b [ i [ d [ o ] ] ] ] ]

**Leserbriefe/  
Courrier des lecteurs**

*gilliozs@bluwin.ch*

# OpenURL: pour améliorer l'expérience de l'utilisateur

Sylvain Machefert, ingénieur en informatique et diplômé de l'INTD

Face à la concurrence des ressources librement accessibles en ligne, les bibliothèques doivent innover et valoriser leurs spécificités, parmi lesquelles les ressources électroniques payantes. Ces ressources électroniques sont devenues un élément majeur des fonds documentaires, dans les bibliothèques de recherche en particulier, mais restent souvent méconnues des lecteurs ou peu utilisées en dehors d'un public d'initiés. L'une des solutions se trouve dans les résolveurs de liens OpenURL.

## 1. Introduction

L'utilisation limitée de ces ressources s'explique par l'enchevêtrement des différents fournisseurs (bouquets de re-

vues, bases de notices bibliographiques, éditeurs ...) qui rend difficile la navigation entre les ressources et l'accès au texte intégral des articles. Une solution pour pallier ce problème consiste à mettre en place des outils de recherche fédérée pour permettre à l'utilisateur de retrouver l'article sans se soucier du fournisseur. Mais la navigation entre les articles reste problématique, avec des références à copier pour les rechercher dans une autre fenêtre, sans être certain de trouver le document (cf. fig. 1).

Une autre solution, que nous développerons ici, se trouve dans les résolveurs de liens OpenURL. Leur but est d'indiquer à un utilisateur qui se trouve face à une référence bibliographique si sa bibliothèque peut lui fournir un accès au document primaire cité, et les informations lui permet-

tant d'y accéder (le lien direct vers l'article si la bibliothèque y a accès).

On parlera principalement ici des articles scientifiques, cible d'origine de l'OpenURL et autour desquels le service s'est développé, mais tous les types de documents (numériques ou non)

Une solution pour pallier ce problème consiste à mettre en place des outils de recherche fédérée pour permettre à l'utilisateur de retrouver l'article sans se soucier du fournisseur.

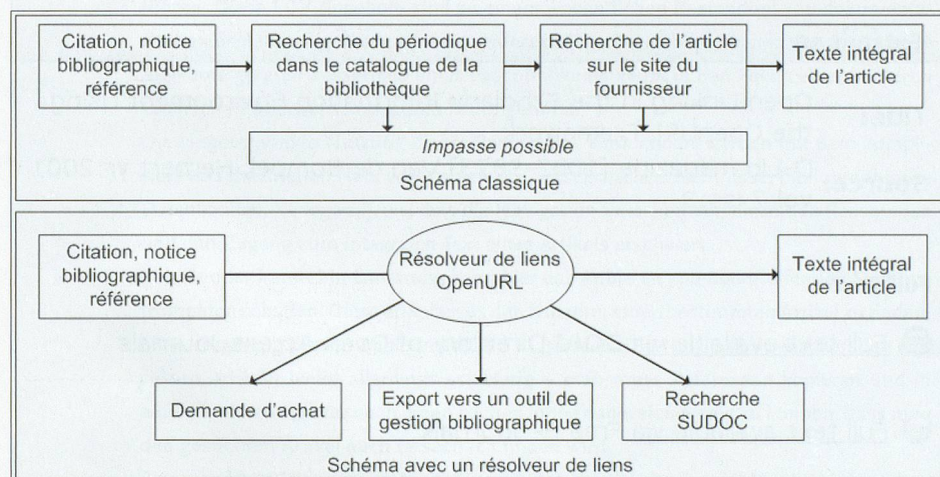


Figure 1: Recherche d'un article sans et avec un résolveur

- Apps, Ann; MacIntyre, Ross (May 2006). "Why OpenURL?" *D-Lib Magazine* 12 (5). doi:10.1045/may2006-apps. ISSN 1082-9873.
- Van de Sompel, Herbert; Beit-Arie, Oren (March 2001). "Open Linking in the Scholarly Information Environment Using the OpenURL Framework" *D-Lib Magazine* 7 (3): 34826. doi:10.1074/jbc.M004545200. ISSN 1082-9873. PMID 10942764.

Figure 2: Recherche d'un article sans et avec résolveur

sont potentiellement concernés. Certain résolveurs se connectent au SIGB pour enrichir les résultats des ouvrages papiers.

## 2. Fonctionnement

Avec l'OpenURL, lorsque l'utilisateur consulte une notice bibliographique, celle-ci est suivie d'une icône cliquable (cf. le bouton SFX de la fig. 2) ou d'un lien, qui active une fenêtre pop-up (cf. fig. 3) lui signalant les services disponibles pour la référence.

Cette fenêtre est affichée par le résolveur de liens, qui est une application autonome mise en place par la bibliothèque et disposant d'une base de connaissance décrivant les fonds de l'établissement. Lorsque l'utilisateur clique sur le lien à la suite des références, ce lien transmet les métadonnées au résolveur qui affiche alors les services disponibles, le principal étant l'accès au texte intégral. Parmi les autres services que la fenêtre peut afficher, on trouve la possibilité de lancer une recherche sur la référence dans un catalogue externe (Worldcat par exemple), d'envoyer un formulaire de demande d'achat ou de P.E.B. avec les métadonnées bibliographiques pré-remplies, la

possibilité d'exporter ces métadonnées vers un logiciel bibliographique ...

### 3. Génération des liens OpenURL

Une des principales difficultés de la mise en place d'un résolveur de liens est liée à la génération des boutons ou des liens à la suite des références. En effet, derrière ces boutons se cache un lien OpenURL sous la forme suivante: `http://www.exemple.com/resolveur?genre=article&atitle=Why%20OpenURL%3F&date=2006&volume=12&issue=5&issn=1082-9873`

Deux parties se distinguent dans cette adresse:

- `http://www.exemple.com/resolveur?`: l'adresse du résolveur de liens, propre à la bibliothèque.
- `genre=article&atitle=Why%20OpenURL%3F&date...`: les métadonnées liées au document cherché, encodées selon la norme OpenURL.

La deuxième partie est générée au moment de l'affichage de la référence, elle est identique quelle que soit la bibliothèque à laquelle le lecteur est inscrit et ne pose pas de problème. La première

Une des principales difficultés de la mise en place d'un résolveur de liens est liée à la génération des boutons ou des liens à la suite des références.

en revanche, dépend de la bibliothèque de rattachement du lecteur, elle est donc dynamique. Deux solutions existent pour l'indiquer:

- Signaler cette adresse à son fournisseur de contenus. C'est la solution proposée par les éditeurs de base payante. L'utilisateur étant connu car identifié (via un *login* ou son adresse IP), il suffit que la bibliothèque ait indiqué au fournisseur l'adresse de base de son résolveur pour que le lien puisse être généré. Cette solution a l'avantage de présenter le lien à tous les utilisateurs autorisés, mais nécessite d'indiquer l'adresse du résolveur

à tous les fournisseurs. Cette opération est parfois prise en charge par les éditeurs de résolveur.

- Utiliser le principe des COinS<sup>1</sup>: le fournisseur cache après chaque référé-

L'avantage de ce système, en plus d'affranchir du travail de signalement propre à la solution précédente, est qu'il permet de disséminer la bibliothèque sur un nombre potentiellement illimité de sites.

rence les métadonnées encodées selon la norme OpenURL. Par défaut l'utilisateur ne voit rien, mais s'il dispose d'un *plugin* spécifique<sup>2</sup>, un lien vers le résolveur sera généré dynamiquement en incluant les métadonnées. L'avantage de ce système, en plus d'affranchir du travail de signalement propre à la solution précédente, est qu'il permet de disséminer la bibliothèque sur un nombre potentiellement illimité de sites. Le principe est utilisé sur Wikipedia (cf.

fig. 2), Worldcat, Citebase... Sur les postes de la bibliothèque, on peut installer et configurer le *plugin*, mais pour les machines personnelles des utilisateurs, la question est plus complexe et nécessite de communiquer sur le service pour le voir utilisé.

La génération d'OpenURL sous l'une de ces deux formes est supportée de manière hétérogène. Chez les acteurs internationaux (type Sciencedirect, Springerlink...), la génération de liens est prise en charge. Pour les plus petits éditeurs en revanche, ou ceux qui se trouvent dans des domaines où le numérique est moins prégnant, la pénétration de l'OpenURL est moindre, mais commence à arriver, avec l'augmentation des installations de résolveurs dans les bibliothèques.

### 4. Le résolveur

L'élément central du résolveur de liens est sa base de connaissance. Elle est fournie avec l'outil, puis le plus souvent mise à jour sous la forme d'un abonnement. Cette base est un critère détermi-

**ExLibris** The bridge to knowledge  
ExLibris SFX

**Title:** Open Linking in the Scholarly Information Environment Using the OpenURL Framework  
**Source:** D-Lib magazine [1082-9873] Van de Sompel, Herbert yr:2001 vol:7 iss:3

**Full Text**

- Full text available via DOAJ Directory of Open Access Journals
- Full text available via Free E- Journals

**Holding information**

- Holdings in Local Catalog  
Ex Libris ALEPH catalog
- Holdings in this Union Catalogue  
Union Catalogue of British and Irish Libraries - COPAC

Figure 3: Exemple de fenêtre de résolveur (SFX d'Ex-Libris)

1 `http://coins.info/`

2 OpenURL referrer d'OCLC par exemple:  
`http://nj.oclc.org/openurlref/`

nant dans le choix du résolveur. Les éléments qui entrent en compte sont:

- Adéquation avec le fonds de la bibliothèque: même si la bibliothèque peut ajouter elle-même ses fournisseurs, c'est un travail fastidieux si l'on souhaite avoir une base de qualité. On s'interrogera en particulier sur la prise en compte des bouquets de périodiques. Si la bibliothèque est abon-

Une petite dizaine d'outils se partagent le marché, dominé par Ex-Libris avec son résolveur SFX, qui fait suite aux travaux de recherche ayant donné naissance à l'OpenURL en 1998.

née à un bouquet de plusieurs centaines de revues et si ce bouquet est prédéfini dans la base de connaissance, la bibliothèque n'aura pas à

sélectionner tous les titres un à un, et se contentera d'indiquer qu'elle est abonnée au bouquet. Par la suite, les titres entrant ou sortant du bouquet seront pris en compte dans les mises à jour régulières de la base de connaissance (si l'éditeur du résolveur de liens suit bien ces mises à jour).

- Qualité de la base proposée par le fournisseur, en particulier au niveau des liens vers le texte intégral. En effet, même si certains fournisseurs de texte intégral disposent de leur propre résolveur de liens, et peuvent ainsi afficher le texte intégral à partir d'informations encodées selon la norme OpenURL, ce n'est pas le cas pour tous. Pour les autres, si l'on souhaite fournir un lien direct vers l'article, il est nécessaire de savoir comment encoder les métadonnées pour construire le lien direct vers l'article si cela est possible.

A l'exception de la base de connaissance, les services supplémentaires évoqués plus haut sont disponibles sur la plupart des solutions. Une petite dizaine d'outils se partagent le marché, dominé par Ex-Libris avec son résolveur SFX, qui fait suite aux travaux de recherche ayant donné naissance à l'OpenURL en 1998.

On signalera aussi quelques tentatives de développement de logiciels libres, dont la principale difficulté est liée au travail colossal que représente le maintien de la base de connaissances. Les outils actifs dans ce domaine sont CUFTS<sup>3</sup> et Openlinker<sup>4</sup> (ce dernier est développé par la bibliothèque universitaire de médecine et le centre de documentation en santé publique de Lausanne).

Parmi les services qui peuvent faire la différence entre deux résolveurs, on citera la présence d'un connecteur interrogeant de manière dynamique le SIGB de la bibliothèque pour ajouter aux résultats les collections physiques.

## ABSTRACT

### OpenURL

Angesichts des Wettbewerbs, der unter den online frei zugänglichen Ressourcen herrscht, müssen die Bibliotheken innovative Ansätze wählen und ihre Qualitäten und Spezifitäten (dazu gehören auch kostenpflichtige EDV-Angebote) entsprechend ins Feld führen. Diese EDV-Angebote sind zu einem wesentlichen Bestandteil von dokumentarischen Beständen geworden, insbesondere in Forschungsbibliotheken. Entsprechende Angebote geraten ausserhalb eines Fachpublikums kaum in den Fokus einer breiteren Leserschaft und werden dadurch kaum genutzt.

Die eingeschränkte Nutzung dieser Ressourcen kann erklärt werden mit dem kompliziert verschachtelten Netz von Anbietern (Zeitschriftenverzeichnisse, bibliographische Datenbanken, Verleger...), welches die Navigation zwischen den einzelnen Ressourcen und den Zugang zum integralen Text eines Artikels erschwert.

Abhilfe oder immerhin Linderung kann hier das Anbieten von übergreifenden Suchinstrumenten schaffen. Diese erlauben es den Nutzern, einen bestimmten Artikel zu finden, ohne vorgängig dessen Anbieter suchen zu müssen. Die Navigation zwischen den einzelnen Artikeln bleibt allerdings schwierig – man muss Referenzen kopieren und in einem anderen Fenster nach ihnen suchen, ohne dabei sicher sein zu können, dass man den gesuchten Artikel auch tatsächlich finden wird.

Einen anderen Lösungsansatz, der vom Autor entwickelt wird, bilden die Linkresolver OpenURL. Diese zielen darauf ab, dem Nutzer, der einen bibliographischen Verweis sucht, mitzuteilen, ob «seine» Bibliothek ihm einen Zugang zum Originaldokument verschaffen kann. Gleichzeitig liefern die Linkresolver die Informationen, die benötigt werden, um auf das Dokument zuzugreifen (d.h. den direkten Link auf den Artikel, sofern die Bibliothek über einen entsprechenden Zugang verfügt).

Der Autor untersucht in erster Linie den Bereich wissenschaftliche Artikel und Dokumente, die ursprünglich den Ausgangspunkt für die Dienstleistung des OpenURL bildeten. Inzwischen verarbeitet OpenURL aber sämtliche Formen von Dokumenten (informatisierte und nichtinformatisierte). Es gibt Resolver, welche sich mit dem integrierten Bibliothekssystem verbinden, um die Trefferquote und -qualität von gedruckten Artikeln zu erhöhen.

## 5. L'impact

La question principale à se poser concerne l'impact de l'outil sur l'utilisateur. Les études disponibles indiquent des retours globalement positifs, en particulier dans les bibliothèques disposant de fonds électroniques larges. Pour les autres, le pop-up du résolveur affichera

La question principale à se poser concerne l'impact de l'outil sur l'utilisateur.

souvent que le texte intégral n'est pas disponible et proposera des services avec une valeur ajoutée moindre, à un utilisateur qui cherche en premier lieu le texte intégral, limitant ainsi l'intérêt de l'outil.

En ce qui concerne l'impact sur l'utilisation des ressources électroniques, les quelques études existantes font toute ressortir une augmentation générale des usages.

## 6. Conclusion

Le résolveur de liens est un outil dont les avantages sont indéniables pour les

3 <http://cufts.lib.sfu.ca/>

4 <http://openlinker.org/>

bibliothèques disposant de ressources électroniques. C'est dans cette situation que l'outil prend tout son intérêt, dans sa capacité à faciliter la navigation entre ressources. Après dix ans de travaux autour de cette norme, les résolveurs

En 2007 en France, plus de la moitié des établissements ayant répondu à un questionnaire sur le sujet disposaient d'un résolveur de liens ou étaient en train d'en installer un.

#### Bibliographie

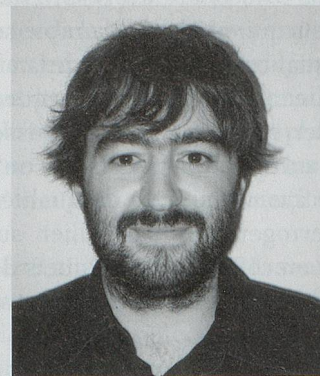
Pour une bibliographie sélective sur ce sujet, on consultera le mémoire de l'auteur *L'Open-URL dans les institutions françaises, une chance pour la valorisation des ressources électroniques?*, disponible en ligne ([http://memsic.ccsd.cnrs.fr/mem\\_00000613/fr/](http://memsic.ccsd.cnrs.fr/mem_00000613/fr/)).

sont devenus un élément central des systèmes de gestion de ressources numériques. En 2007 en France, plus de la moitié des établissements ayant répondu à un questionnaire sur le sujet disposaient d'un résolveur de liens ou étaient en train d'en installer un.

Mais la mise en place de cet outil doit s'accompagner d'un suivi de la part des professionnels sans quoi l'utilité sera très limitée. Pour qu'un projet de résolveur de liens réussisse, il faut en effet:

- Paramétrer la base de connaissance
- Signaler le résolveur de liens auprès des principaux fournisseurs de notices bibliographiques, mais aussi de ressources électroniques (texte intégral pour les références dans les bibliographies d'articles)
- Suivre les mises à jour de la base de connaissance

Contact: [smachefert@gmail.com](mailto:smachefert@gmail.com)



#### Biographie

Ingénieur en informatique et diplômé de l'INTD, Sylvain Machefert s'est spécialisé dans l'ingénierie documentaire. Il est depuis 2005 consultant au sein du cabinet doXul-ting, où il prend en charge des missions de conseil et d'assistance. Il est notamment coauteur de l'ouvrage *Mener un projet Open Source en bibliothèque, documentation et archives* paru en 2007 aux Editions du Cercle de la librairie (978-2-7654-0954-0).

## PREDATA

Bibliothekssysteme  
Hardware-Kompetenzzentrum  
Netzwerke und IP-Telefonie  
Software-Entwicklung

Predata AG – seit 1984 Ihre Partnerin für Bibliothekssoftware.

## winMedio.net

Die leistungsfähige und anwenderfreundliche Software für Bibliotheken, Ludotheken und Dokumentationsstellen bietet umfangreiche Ausleih-, Katalog-, Recherche-, Statistik- und Einstellungsfunktionen sowie einen OPAC mit Selbstbedienungsmöglichkeiten.

winMedio.net unterstützt die zukunftssträchtige Radio Frequency Identification-Technologie mit einer RFID-Schnittstelle (Medienpaket- und Stapelverarbeitung, Selbstverbuchung, Diebstahlsicherung) und ist sowohl für zentrale als auch für dezentrale Lösungen geeignet.

## BIBLIOTHECA 2000

Das innovative und benutzerfreundliche Bibliotheksmanagement-System umfasst alle Bereiche wie Katalog, Ausleihverbuchung, Recherche, Erwerbung, Statistik, Periodikaverwaltung und OPAC mit Selbstbedienungsfunktionen. Weitere Module sind zusätzlich erhältlich – z.B. Fernleihe, Inventur, Z39.50-Client, Webportal, RFID-Schnittstelle. Zur Optimierung der Betriebsabläufe kann BIBLIOTHECA2000 durch umfangreiche Einstellungsmöglichkeiten problemlos an Ihre individuellen Bedürfnisse angepasst werden.

Predata AG Burgstrasse 4 3600 Thun  
Tel. 033 225 25 55 Fax 033 221 57 22 [info@predata.ch](mailto:info@predata.ch) [www.predata.ch](http://www.predata.ch)

# Le Web 2.0 comme outil d'un nouvel individualisme

Stéphane Gillioz, rédaction *arbido*

Les agrégateurs, ces outils qui foisonnent sur le web et que la «génération numérique» s'approprie sans complexe, sont en train de transformer la relation de l'individu au monde. Ou plutôt à sa vision du monde. Car chaque usager peut désormais créer son histoire, proposer un tableau de «son» monde et le partager. Le Web 2.0 serait-il le cheval de Troie d'un nouveau narcissisme? Et comment les professionnels de l'information documentaire peuvent-ils tirer leur épingle du jeu?

## Le Robert vs Yahoo Pipes and Co

Nous avons probablement tous le souvenir de soirées au coin du feu, ou autour d'une table, en famille ou avec des amis, où les discussions parfois enflammées étaient ponctuées de pauses «vérification», le *Petit Larousse*, le *Robert* ou, pour les mieux lotis, l'*Encyclopedia Britannica* étant les références ultimes et donc des pacificateurs bienvenus.

Même scène, quelques années plus tard ou avec des «digital natives» (la génération née avec l'avènement du numérique), qui exhibent *Yahoo Pipes*, un service gratuit d'éditeur visuel en ligne qui combine et fait un «mash up», le dernier mot tendance dans le monde Web 2.0, de plusieurs sources d'informations («mashup»: littéralement «purée», il consiste à mélanger des images et des sons numériques). L'éditeur graphique est judicieux quoique déroutant pour des non-initiés: on clique, on tire et on laisse tomber les modules dans une fenêtre, on réunit avec des modules d'union, on ajoute des modules de tri, des modules filtres; on agite le tout et voilà une «gerbe d'informations» personnalisée que l'on peut partager avec d'autres.

Même possibilité de créer des interfaces personnalisées avec des agréga-

teurs comme *iGoogle* et *Netvibes* ou des sites web de réseau social comme *MySpace* ou encore *Facebook*, qui permettent de créer son propre monde.

## Les folksonomies: l'explosion des communautés

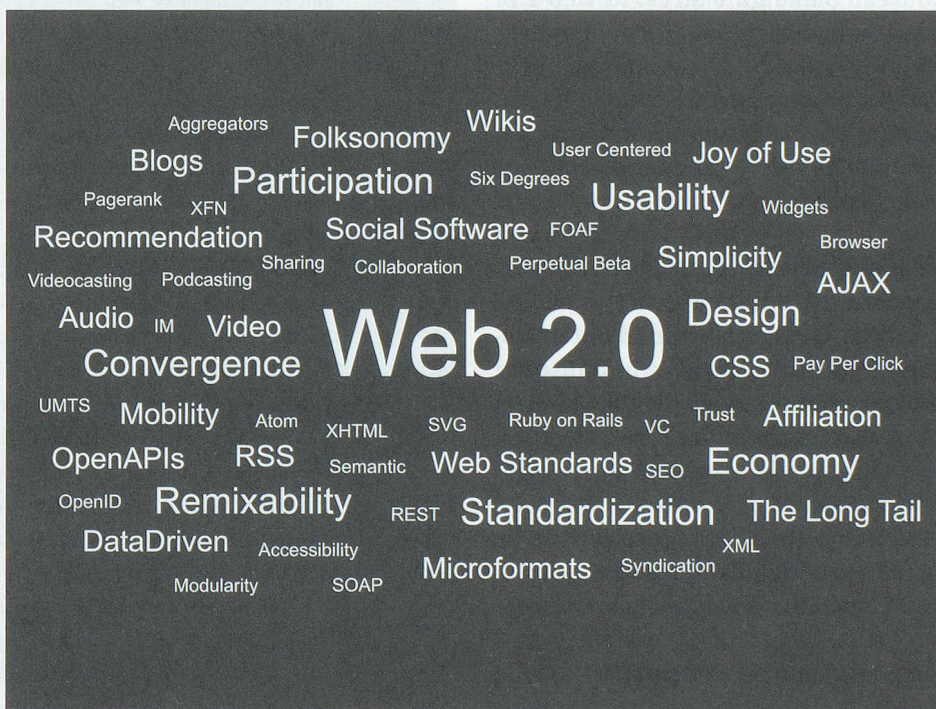
Elément le plus important du Web 2.0, les folksonomies désignent un système de classification collaborative décentralisée spontanée. Chaque utilisateur nomme en effet comme il l'entend les informations qu'il ajoute. On retrouve par exemple un tel système de classification sur *YouTube* pour les vidéos.

Parmi les nombreux sites fonctionnant selon ce système, on peut mentionner les suivants:

- Delicio.us, sites de partage de signets
- Flickr, site de partage de photos
- Wikipedia, site de partage de contenus
- rss4you et Bloglines.com, sites de partage de flux RSS

- StumbleUpon, site de classement de pages web
- Technorati, qui permet d'effectuer des recherches sur les blogs selon leurs étiquettes
- U-lik, qui permet de partager l'ensemble de ses goûts (livres, films, œuvres d'art, musées)

Contrairement aux systèmes hiérarchiques de classification, les contributeurs d'une folksonomie ne s'en tiennent pas à une terminologie prédéfinie et sont libres d'adopter les termes qu'ils souhaitent pour classifier leurs ressources. Ces termes sont souvent appelés mots clés ou *tags* (*étiquettes*). Autrement dit, il n'y a plus de règles de catalogage, ce qui peut paraître scandaleux pour les professionnels de l'information documentaire! De l'ordre logique porté par ces derniers, on glisse vers un ordre organique qui semble faire imploser toute forme de classification rationnelle. Autre conséquence non néglige-



Auteur: Luca Cremonini

able: les fournisseurs professionnels perdent leur exclusivité, puisque les usagers se réapproprient le web comme outil, qu'ils y créent leurs propres portraits et les proposent à tout un chacun.

Les folksonomies présentent pourtant plusieurs avantages. En effet, d'une part, les requêtes sont plus intuitives puisque la classification relève elle-même des idiosyncrasies du contributeur; d'autre part, l'intérêt des folksonomies est lié à l'effet communautaire: pour une ressource donnée, sa classification est l'union des classifications de cette ressource par les différents contributeurs. L'utilisateur dispose ainsi de nuages de tags – ou *tagclouds* – qui lui permettent de suivre de proche en proche les terminologies des autres contributeurs pour explorer et découvrir des ressources connexes.

#### Alors que faire des OPAC's?

Les catalogues en ligne (OPAC's: Online Public Access Catalogue) seraient-ils donc devenus superflus? On peut en effet se poser la question de savoir ce que l'on peut en faire vu que les utilisateurs ont déjà accès gratuitement aux ouvrages sur *Googlebook*, par exemple. La réponse coule de source: les rendre accessibles sur Web 2.0!

En effet, les OPAC's contiennent beaucoup d'informations utiles qui peuvent être valorisées dans le cadre du nouveau paradigme qu'est le Web 2.0. Ces catalogues seront combinés avec d'autres sources d'informations. On mentionnera ici, à titre d'exemple, *Librarything.com*, qui permet à tout un chacun de créer sa propre bibliothèque et de la partager.

Le projet allemand *MedioVis* est particulièrement intéressant dans ce contexte. Il s'agit en effet d'un catalogue multimédia de type «mashup» qui est interconnecté avec d'autres sites et sources d'informations. *MedioVis* permet la recherche dans des données bibliographiques du fonds d'une médiathèque et offre diverses possibilités de visualiser les résultats, qui sont partiellement enrichis de données externes.

Le catalogue lui-même n'est plus que le noyau à partir duquel on accède à une foule d'informations annexes, glanées dans un espace hétérogène.

#### L'adaptation des prestations des professionnels de l'information documentaire

Face à la multiplication des outils du Web 2.0 et à ce véritable tournant culturel et épistémique que constituent notamment les folksonomies, les professionnels de l'information documentaire doivent adapter leurs prestations. René Schneider, professeur en information documentaire à la Haute école de gestion de Genève, suggère plusieurs solutions. Les petites bibliothèques peuvent par exemple créer un univers de sources d'informations pour leurs clients et proposer leur propre service sur *Netvibes*. Ces mêmes bibliothèques peuvent par ailleurs devenir membres d'une communauté. Dernier exemple en date: la Bibliothèque nationale, qui va mettre les photos numérisées sur

Flickr.com. La British Library et la Library of Congress le font également...

Quant à savoir comment enrichir les catalogues en ligne avec d'autres sources ou comment les intégrer ou participer à une communauté, les pistes sont nombreuses et les solutions multiples. Celles-ci ne dépendent pas des possibilités techniques, mais bel et bien de l'intérêt que les professionnels peuvent y trouver et de la charge de travail supplémentaire que cela peut impliquer.

A moins qu'il ne s'agisse, plus radicalement, d'opérer une révolution dans leur manière de concevoir leur relation à des clients dont le comportement peut laisser penser qu'ils pourraient très bien se débrouiller seuls...

Contact: [gilliozs@bluewin.ch](mailto:gilliozs@bluewin.ch)

## ABSTRACT

*OPACs, Mash-ups, Folksonomien und Co. – Das Web 2.0 als Werkzeug eines neuen Individualismus*

Yahoo Pipes und andere Anbieter eröffnen die Möglichkeit, verschiedene Informationsquellen miteinander zu kombinieren und seine (Such-)Vorlieben zu personalisieren. Die von visuellen Editoren unterstützten Programme generieren ein «Mash-up» (von engl. to mash: mischen, verstampfen) – gemischt wird dabei, was der Nutzer vorgibt.

Mit «Folksonomie» wird das spontane, dezentrale, von bestimmten «Communities» (z.B. YouTube, Wikipedia) vorgenommene Verschlagwort benannt. Jeder User verschlagwortet die Information, die er zur Verfügung stellt, nach seinen ganz eigenen Vorstellungen. Die User halten sich nicht an eine vordefinierte, hierarchisch definierte Struktur, sondern klassifizieren die Inhalte individuell. Dabei kommen Schlüsselwörter (sog. Tags) zur Anwendung, die zu einer Schlagwortwolke (sog. Tag-clouds) aneinandergereiht werden können. Die dadurch entstehende «Ordnung» ist organisch und nicht unbedingt logisch.

Werden Online-Kataloge oder OPACs (Online Public Access Catalogue) dadurch obsolet? Angesichts der Tatsache, dass die User freien Zugang zu Diensten wie *Google Book* haben, kann man sich diese Frage durchaus stellen. Eine Lösung wäre das Zur-Verfügung-Stellen der OPACs im Web 2.0 – und damit das Nutzen der entsprechenden Vorteile (z.B. Kombinierbarkeit mit weiteren Informationsquellen).

Die Profis aus dem Informations- und Dokumentationswesen werden nicht darum herumkommen, ihre Dienstleistungen angesichts des kulturellen Paradigmenwechsels (v.a. junge, «web-native» User haben diesen Wechsel vollzogen resp. kennen nichts anderes) den neuen Gegebenheiten anzupassen. Bibliotheken etwa können ihr Angebot ausbauen und für ihre Kunden einen eigenen *Netvibes*-Service anbieten, und sie können Mitglieder von Communities werden.

Für die entsprechende Aufbereitung der Daten stehen zahlreiche Lösungen zur Verfügung. Die Technik ist kein Problem, entscheidend ist der Wille und die (nicht zuletzt zeitliche...) Bereitschaft der Betreiber, diesen Schritt zu machen.

# La sécurité informatique

Daniel Brügger, ingénieur en informatique, spécialiste de la sécurité IT

**Le développement des réseaux, le stockage et le transfert de volumes d'informations de plus en plus importants, l'explosion du nombre d'utilisateurs, voilà qui nécessitent des systèmes (hardware et software) de plus en plus sophistiqués et performants. Mais qu'en est-il de la sécurité informatique? Comment préserver ces quantités phénoménales de données qui forment la richesse d'une entreprise ou d'une institution? Rappel de quelques principes que l'on sera de plus en plus contraint d'appliquer à l'avenir.**

La sécurité informatique est un terme galvaudé, servi à toutes les sauces, quasiment un leitmotiv qu'on se lance en société comme: la sécurité, c'est très important! Ce à quoi quelqu'un répondra par: et de plus en plus! A ce moment-là, dans un mouvement solidaire, chacun hochera de la tête d'un air entendu, comme si cela coulait de source et... sans savoir du tout de quoi il retourne vraiment.

La sécurité informatique doit offrir toutes les techniques, mécanismes, architectures ou règles permettant de préserver le système d'information d'une entreprise. C'est aussi simple que ça.

## Le système d'information

Encore un terme à la mode. Le système d'information est composé d'une multitude d'informations ainsi que du système lui-même. On pouvait s'en douter. L'information d'une entreprise, ce sont tous les fichiers créés, les courriels, les messages internes, les recherches et les commandes en ligne, les mémos, les fichiers Excel, Word, les schémas, les éléments financiers, les secrets de fabrication, les banques de données, etc. En ce sens, l'information d'une entre-

prise est son âme, son historique, sa valeur première.

Préserver cette valeur est donc absolument essentiel à la bonne marche de l'entreprise. Ne pas y penser, sous-estimer sa richesse ou remettre à plus tard les tâches de préservation de l'information signifierait mettre potentiellement en péril l'entreprise. En ce sens la responsabilité ultime revient toujours au management, aucune possibilité de se rabattre sur le méchant informaticien. La direction de l'entreprise doit mettre l'infrastructure, les outils et les personnes à disposition de son système d'informations avec pour but premier d'assurer sa défense! Plus loin encore, la direction doit s'assurer de la bonne application des mesures avec l'appui d'auditeurs si nécessaires.

## Quels sont les risques?

Ils sont malheureusement innombrables. Tout d'abord, il faut impérativement se souvenir que 80% des attaques ou pertes de données sont réalisées depuis l'intérieur de l'entreprise. Le bandit masqué derrière son ordinateur existe, mais les personnages les plus dangereux vous côtoient à la cafétéria. Ou bien est-ce vous-même? Rassurez-vous, il s'agit très souvent d'erreurs de manipulations, de méconnaissances du système, de tentatives de découvertes ou encore de malveillance. Peu importe la façon, les données perdues peuvent malheureusement l'être de manière irrémédiable.

Si l'utilisateur est le risque majeur, bien d'autres événements peuvent avoir des répercussions catastrophiques. En voici une liste non exhaustive:

- Les personnes: mauvaise formation, faible prise de conscience de la sécurité, désinvolture, perte de donnée, divulgation d'informations confidentielles, non-respect des contrôles

d'accès, fraude, vol, tentative d'accès ou d'attaque etc.

- Les équipements: défaillance, destruction de données, mauvaise résilience à la panne, mécanismes de sécurité faibles, etc.
- Les contrôles d'accès: contrôles d'accès peu respectés, pas de contrôle d'accès aux bâtiments, divulgation ou échange de mot de passe.
- Les éléments naturels ou environnementaux: le feu, les inondations, le vandalisme, les pannes d'électricité, etc.
- Les applications: mal structurées, elles entraînent des erreurs, des pertes de données, un ralentissement ou un arrêt de la production, des pannes intempestives.

Les mesures permettant d'écarter ou de réduire ces risques précités ne sont pas purement techniques. Elles doivent être à parts administrative et technique égales. Cela signifie qu'une formation adéquate des personnes, une sensibilisation à l'utilisation et aux mesures de sécurité est déjà un élément de réponse très important.

Ensuite, il existe évidemment une armada d'équipements de sécurité permettant de procéder de manière quasiment systématique. L'exemple de l'antivirus est des plus clairs:

- Dissuader: l'utilisateur signe une charte de l'utilisation de son accès à internet et des équipements informatiques de l'entreprise. En ce sens, accéder à un site pour le moins équivoque ou tenter d'installer un logiciel ne respecterait pas la première barrière dissuasive. Mon patron pourrait le savoir!
- Prévenir: empêcher l'événement! Le poste est équipé d'un antivirus mis à jour et empêche ainsi l'infection.
- Détecter: un virus tente d'infecter le poste, l'antivirus le détecte et l'identifie.

– Réagir: la mesure de protection du poste de travail est lancée, le virus est effacé ou isolé si cela n'est pas possible. Le poste et donc ses données sont préservés.

### Les risques du futur?

Si l'exemple du virus a presque un côté amusant, chacun en ayant été un jour ou l'autre victime, les enjeux de la sécurité informatique ont atteint un tel niveau que le rapprochement avec la sécurité d'un pays au sens strict devient évident. A l'heure du terrorisme qui prend d'ores et déjà la forme électronique, les nations doivent prévoir des techniques de défense appropriée. Cela n'a rien d'utopique ou d'alarmiste, l'actualité nous donne raison jour après jour.

Des cas de prise d'otage d'information ont été révélés. Une entreprise ou une institution fait l'objet d'une attaque ciblée. Un dossier ou carrément un disque sensible est crypté et l'on ne peut plus du tout y accéder. Les attaquants contactent ensuite la société et lui proposent la clé permettant de décrypter les données contre une rançon. Simple et efficace. Les cibles sont des organismes financiers ou des sociétés à forte capitalisation sur leur image. Si la rançon n'est pas versée, une divulgation dans les médias de l'attaque est organisée, entraînant inmanquablement un très fort déficit d'image pour la société cible.

Plus impressionnant encore fut l'attaque en avril 2007 envers un pays, l'Estonie en l'occurrence. Le pays décide le déplacement d'une statue d'un soldat russe dans un parc en banlieue. Vainqueur des nazis pour les uns, oppresseurs pour les autres, le fait de toucher à un symbole engendre les foudres des hackers russes. En quelques heures, les sites gouvernementaux, banques, médias ou encore services d'urgence su-

bissent une attaque par déni de service. Les sites sont submergés de requête amenant à un arrêt complet de leur service respectif. L'attaque est si impressionnante et déstabilisante pour le pays qu'elle ne peut pas raisonnablement n'être que le fait d'individus isolés. D'aucuns n'y vont pas par quatre chemins, la Russie démontre ici sa force de frappe dans ce qui s'appelle dorénavant le «war game». Pour la petite histoire, un an plus tard c'était au tour de la Lituanie d'être la victime désignée pour des motifs quasiment similaires.

Au sein de l'OTAN, plusieurs pays ont signé ce printemps un accord visant à la création d'un centre de cyberdéfense. Il est reconnu que la criminalité informatique ne tient pas compte des frontières et qu'une collaboration entre pays est indispensable. L'Estonie a demandé la qualification juridique des événements susmentionnés en acte de terro-

risme. A noter également que l'Union européenne dispose depuis 2004 de sa propre agence de sécurité informatique, l'ENISA (European Network and Information Security Agency) dont le but est simple et fondamental à la fois, à savoir: «Contribuer à moderniser l'Europe et à assurer le bon fonctionnement de l'économie numérique et de la société de l'information.»

Contact: [dab@oxygen-company.com](mailto:dab@oxygen-company.com)

### Références:

- Centrale d'enregistrement et d'analyse pour la sûreté de l'information (MELANI), rapports semestriels
- Shon Harris, CISSP All-in-one, 4<sup>th</sup> edition
- Divers sites internet avec recoupement d'informations
- <http://www.enisa.europa.eu/>

## ABSTRACT

### Informatiksicherheit

Firmen produzieren viel Information. Diese ist nicht nur leeres Datenmaterial, sondern gleichsam die «Seele» und die Geschichte des Unternehmens. Wer sich über die Wichtigkeit dieses Fundus' im Klaren ist, wird die Sicherheit ganz gross schreiben.

80% der Angriffe auf Daten werden intern verursacht. Das muss nicht willentlich geschehen. Oft sind Unwissen, Fehlmanipulationen oder Unachtsamkeit der Grund für Datenverluste.

Neben dem Personal können auch fehlerhaftes Material, mangelhafte Zugangssicherheit (ins Gebäude, ins System), «natürliche» Katastrophen (z.B. Feuer, Wasser) oder schlecht strukturierte Software für Verluste verantwortlich zeichnen.

Es gibt eine Reihe von technischen und administrativen Vorkehrungen, welche die Informatiksicherheit erhöhen.

Die Vorkehrungen reichen von Antivirenprogrammen über verschiedene präventive Massnahmen bis hin zu Personalschulungen.

Künftig werden Vorkehrungen auf nationaler Ebene an Wichtigkeit gewinnen (Stichworte Terrorismus, Hackerangriffe, «Geiselnahme» von Informatikdaten, die gegen Lösegeld wieder «freigegeben» werden). Wichtig ist in diesem Zusammenhang die internationale Zusammenarbeit. Die NATO («cyber-defense») und auch die EU (ENISA – European Network and Information Security Agency) sind daran, entsprechende Strukturen aufzubauen.