



Le futur entre aux Archives

Communiqué de presse

Paris, le 16 juin 2005

Des outils numériques pour une meilleure exploitation des Archives

L'INRIA, l'INSA de Rennes et le Conseil général des Yvelines présentent aux Archives des Yvelines un nouveau système informatique révolutionnant l'accès aux documents manuscrits numérisés : accès automatique, mise à disposition favorisée pour le public des documents d'archives, base de données pouvant s'enrichir des travaux de chaque utilisateur... Une application qui sera bientôt déployée au Centre historique des archives nationales et à la ville de Lannion.

S'intéressant de longue date à la thématique du document numérique, l'INRIA investit depuis quelques années, en partenariat avec les acteurs régionaux (archives départementales), le domaine et le métier de l'archive numérique. C'est à l'INRIA Rennes, au sein de l'équipe IMADOC de l'IRISA et de l'INSA de Rennes que les compétences scientifiques les plus proches du domaine ont été identifiées.

Accéder automatiquement à des documents d'archives manuscrits numérisés

Mis en service pour la première fois aux Archives départementales des Yvelines, un nouveau système informatique est désormais capable de sélectionner automatiquement, c'est-à-dire sans aucune indexation manuelle préalable, les documents manuscrits qui intéressent un lecteur. Combinant des performances de reconnaissance de formes manuscrites, de détection de structures ou de champs propres aux documents, ce nouvel outil produit des annotations automatiques renseignant chaque type de document archivé. Grâce à ces annotations extrayant de multiples informations des images numériques des documents, l'utilisateur retrouve automatiquement, par le patronyme, les pages qui l'intéressent.

Faciliter les recherches sur les patronymes

Lorsqu'un lecteur demande au système de visualiser les documents qui contiennent le nom d'une personne, l'ordinateur va alors lire automatiquement l'ensemble des documents manuscrits numérisés et présenter ceux qui sont les plus proches de la requête. Le lecteur n'a plus qu'à sélectionner le nom qu'il souhaite pour voir l'ensemble de la page correspondant à ce nom. L'accès au bon document se fait en quelques secondes de manière complètement automatique. Cette application concerne actuellement les registres de matricules militaires des XIX^e et XX^e siècles. Les recherches se poursuivent afin de permettre un accès automatique par le nom et le prénom aux listes nominatives de recensement de population des XIX^e et XX^e siècles, et pour permettre ce type d'accès sur des documents encore plus complexes à traiter automatiquement.

Rendre accessibles des documents jusqu'ici non communicables au public

Certains documents ne sont pas consultables car ils contiennent des informations relevant du secret médical ou du secret judiciaire. C'est notamment le cas des registres matricules de moins de 130 ans. Or ces documents recèlent d'autres informations très recherchées par les historiens ou les généalogistes.

Le nouveau système développé par les chercheurs de l'INRIA et de l'INSA de Rennes donne accès à ces données communicables grâce à un découpage virtuel des pages. En effet, il masque les zones protégées par la loi pour rendre lisible ce qui reste du domaine public. Ce découpage virtuel libère donc toute une masse d'informations.

Permettre à chacun de bénéficier des recherches de tous

Cependant, parvenir à un système d'annotation automatique complet est techniquement impossible, tous les documents manuscrits numérisés ne peuvent pas être reconnus à 100 %. En effet, certains n'ont pas une organisation suffisamment bien établie pour qu'elle puisse être déchiffrée automatiquement, pour d'autres l'écriture manuscrite est trop dégradée pour être lisible. Aussi pour pallier cette perte d'informations, le nouveau système permet de tirer parti des recherches menées par tout utilisateur.

Lorsqu'un lecteur, en salle ou sur Internet, accède au document qu'il souhaite, il peut, grâce au nouveau système, compléter les annotations automatiques par des annotations manuelles, dites collectives. Ces annotations collectives sont gérées sur une plate-forme et permettent un travail en réseau. Chaque lecteur peut ainsi enregistrer les informations qu'il trouve pertinentes au fur et à mesure de ses consultations. Par son apport, il permet à d'autres lecteurs de sélectionner automatiquement des documents ainsi associés. Par exemple, sur les registres paroissiaux et d'état-civil, un utilisateur peut être amené à préciser le nom de la mère sur un acte de naissance. Il permet ainsi, grâce à cet ajout, de rendre cet acte de naissance accessible à toute personne intéressée par le patronyme de la mère. Ces renseignements supplémentaires peuvent être d'un précieux concours dans le cadre de recherches en généalogie, mais aussi en histoire (par exemple le taux d'alphabétisation d'après la présence ou non de signatures dans les actes).

La plate-forme d'annotation collective est mise en place aux Archives départementales des Yvelines sur des registres paroissiaux et d'état-civil (1 450 000 images) et des registres matricules (430 000 images). Elle sera également mise en place au Centre historique des archives nationales sur des décrets de naturalisation (150 000 images). Son utilisation peut être envisagée sur n'importe quel type de documents numérisés, comme des listes nominatives de recensement de population, des délibérations de conseils municipaux...

Des postes de lecture dits du futur pour consulter les documents manuscrits numérisés comme de vrais livres, le stylo à la main...

Chaque poste de lecture dit du futur est constitué d'une tablette graphique à retour visuel, c'est-à-dire d'un écran plat posé sur un bureau à la manière d'un sous-main, où l'interaction entre la machine et le lecteur s'effectue à l'aide d'une interface stylo. Même si un public de plus en plus large sait utiliser à vitesse opérationnelle un clavier d'ordinateur, il n'en reste pas moins que l'utilisation d'un stylo est bien plus intuitive. L'interface stylo permet de réaliser des annotations, de sélectionner des zones et de visualiser des documents sans avoir recours au clavier ni à la souris. Son utilisation correspond aux habitudes de lecture des documents papiers et permet de mélanger facilement la consultation de documents papiers et numériques.



L'application développée par l'équipe IMADOC pour les Archives des Yvelines permet de profiter de toutes les capacités offertes par ce stylo numérique. La tablette PC connue du grand public est généralement fournie avec des logiciels où l'intégration du stylo graphique fonctionne uniquement pour la bureautique et non pour des logiciels de consultation des images. Deux postes de lecture utilisant cette technologie sont mis en place dans la salle de lecture des Archives des Yvelines. Ils préfigurent ce que pourront être dans un proche avenir les outils de travail des lecteurs aux archives.

A PROPOS DE L'IRISA

L'INRIA, le CNRS, l'université de Rennes 1 et l'INSA de Rennes sont partenaires au sein d'une unité mixte de recherche appelée IRISA : Institut de recherche en informatique et systèmes aléatoires. Plus de 450 personnes, dont près de 200 scientifiques, travaillent dans ses locaux situés sur l'un des campus universitaires de Rennes.

L'IRISA s'implique fortement dans des coopérations avec les centres de recherche et a su établir des liens étroits avec le tissu économique tant régional que français et étranger.

Pour en savoir plus: <http://www.irisa.fr>

A PROPOS DES ARCHIVES DEPARTEMENTALES DES YVELINES

En octobre 2003, les Archives départementales des Yvelines ont quitté les locaux prestigieux mais inadaptés de la Grande Ecurie du Roy à Versailles pour s'installer dans un nouveau bâtiment à Saint-Quentin-en-Yvelines. Outre une augmentation de la capacité de stockage, la priorité dans ce bâtiment a été donnée à l'accessibilité des documents grâce aux nouvelles technologies : numérisation des fonds d'archives les plus consultés (2 millions de fichiers), consultation gratuite dans la salle de lecture des Archives départementales équipée de 54 postes informatiques ou sur Internet (www.archives.yvelines.fr). De plus, grâce à un partenariat avec l'INRIA et l'INSA, les Archives départementales des Yvelines mettent à la disposition des lecteurs des outils de recherche innovants (plate-forme d'annotations collectives, reconnaissance de l'écriture manuscrite...), permettant de faciliter la communication des archives numérisées.

CONTACTS PRESSE

IRISA

Bertrand Couasnon

Tél. : 02 99 84 74 11 – mail : couasnon@irisa.fr

INRIA

Laurence Hermant – Stratéus pour l'INRIA

Tél. : 01 40 41 56 11 – mail : laurence.hermant@strateus.com

Vincent Coronini

Tél : 01 39 63 57 29 – mail : vincent.coronini@inria.fr

INSA

Claire Antony

Tél. : 02 23 23 86 29 – mail : claire.antony@insa-rennes.fr

Conseil général des Yvelines / Archives départementales des Yvelines

Laurence Volf

Tél. : 01 39 07 71 53 – mail : LVolf@cg78.fr