

Recherche collaborative d'informations : repenser l'architecture des systèmes de recherche d'information à l'ère du numérique

The Cooperative Research of Information: An Examination of the Architecture of Information Research Systems in the Digital Era

Búsqueda de información en colaboración: replanteamiento de la arquitectura de los sistemas de búsqueda en la era digital

Abderrazak Mkadmi and Imad Saleh

Volume 50, Number 2, April–June 2004

URI: <https://id.erudit.org/iderudit/1030080ar>

DOI: <https://doi.org/10.7202/1030080ar>

[See table of contents](#)

Publisher(s)

Association pour l'avancement des sciences et des techniques de la documentation (ASTED)

ISSN

0315-2340 (print)

2291-8949 (digital)

[Explore this journal](#)

Cite this article

Mkadmi, A. & Saleh, I. (2004). Recherche collaborative d'informations : repenser l'architecture des systèmes de recherche d'information à l'ère du numérique. *Documentation et bibliothèques*, 50(2), 157–168. <https://doi.org/10.7202/1030080ar>

Article abstract

This article describes a protocol for the cooperative research of information created at the Paragraphe laboratory. The protocol provides some solutions to the issues related to the research for information in the age of digital libraries where, on the one hand, the body of documents has become too large and too often unstructured and, on the other hand, the users who are alone in front of a computer screen during the time of their search. In order to solve the first problem, we suggest the use of the technology of Xtensible Markup Language (XML) to structure and search for information and, for the second issue, we propose a cooperative work space that allows users to share results and help one another. In concluding, we elaborate on the future of this work.

Recherche collaborative d'informations : repenser l'architecture des systèmes de recherche d'information à l'ère du numérique

ABDERRAZAK MKADMI

Abderrazak.Mkadmi@univ-paris8.fr

IMAD SALEH

Imad.Saleh@univ-paris8.fr

RÉSUMÉ | ABSTRACTS | RESUMEN

Cet article présente un prototype de système de recherche collaborative d'informations créé dans le cadre des travaux du laboratoire Paragraphe. Ce prototype apporte quelques éléments de réponse à des problèmes liés à la recherche d'informations à l'ère des bibliothèques numériques où, d'une part, les corpus de documents sont devenus trop gros et souvent peu structurés, et d'autre part, les utilisateurs se retrouvent isolés devant l'écran tout au long de leurs recherches. Afin de résoudre le premier problème, nous proposons l'exploitation des apports de la technologie Xtensible Markup Language (XML) pour la structuration et la recherche d'informations et pour le problème d'isolement des chercheurs, nous proposons un espace de travail collaboratif permettant aux utilisateurs de dialoguer et de s'entraider. Enfin, nous élaborons les perspectives de ce travail.

The Cooperative Research of Information : An Examination of the Architecture of Information Research Systems in the Digital Era

This article describes a protocol for the cooperative research of information created at the Paragraphe laboratory. The protocol provides some solutions to the issues related to the research for information in the age of digital libraries where, on the one hand, the body of documents has become too large and too often unstructured and, on the other hand, the users who are alone in front of a computer screen during the time of their search. In order to solve the first problem, we suggest the use of the technology of Xtensible Markup Language (XML) to structure and search for information and, for the second issue, we propose a cooperative work space that allows users to share results and help one another. In concluding, we elaborate on the future of this work.

Búsqueda de información en colaboración : replanteamiento de la arquitectura de los sistemas de búsqueda en la era digital

Este artículo presenta un prototipo de sistema de búsqueda de información en colaboración creado en el ámbito de las actividades del laboratorio Paragraphe. Este prototipo brinda algunos elementos de respuesta a problemas relacionados con la búsqueda de información en la era de las bibliotecas digitales en la cual, por una parte, los cuerpos de documentos son demasiado grandes y con frecuencia poco estructurados y, por la otra, los usuarios hacen sus investigaciones aislados frente a pantallas. Para resolver el primer problema, proponemos aprovechar las contribuciones de la tecnología Xtensible Markup Language (XML) para la estructuración y la búsqueda de información; y para el problema de aislamiento de los investigadores, proponemos un espacio de trabajo en colaboración que permite a los usuarios dialogar y ayudarse mutuamente. Finalmente, elaboramos las perspectivas de este trabajo.

Avec la récente explosion du Web, des bases de données multimédia et des bibliothèques numériques, de grands changements ont vu le jour à la fois en ce qui concerne la nature des documents et corpus, ainsi que pour les besoins des utilisateurs. Les corpus de documents sont devenus beaucoup plus gros et sont souvent composés de documents hétérogènes aussi bien dans la forme que dans le contenu, présentant ainsi une faible structuration. Les documents, catalogues, thésaurus et outils de recherche sont représentés électroniquement et peuvent être consultés à distance à travers des réseaux informatiques. Cette réalité de recherches à distance va rendre les interactions collaboratives traditionnelles plus rares en éliminant la proximité physique avec d'autres chercheurs. Déjà, depuis un certain temps, les interactions sociales sont réduites par l'accès à domicile aux OPACs (serveurs Web de consultation de catalogues) et par le développement des bibliothèques numériques. Cette tendance devrait s'intensifier. De ce fait, le besoin d'un « intermédiaire » entre le consommateur et le producteur de l'information devient de plus en plus urgent pour maîtriser la recherche d'informations. L'utilisateur ne peut pas toujours exprimer précisément le but de sa recherche et le producteur ne sait pas toujours présenter l'information d'une manière adéquate et conviviale. Toutefois, les échanges d'expériences entre les utilisateurs et la classification des résultats de recherche semblent apporter une aide substantielle pour maîtriser l'espace informationnel actuel.

À la suite de ces grands changements, nous pouvons résumer les nouveaux besoins comme suit :

- ▶ structurer les corpus des documents de façon à en avoir une représentation facilement manipulable par machine;
- ▶ rendre ces corpus plus actifs permettant ainsi une interaction avec l'utilisateur qui peut ajouter et/ou modifier des informations, formuler et reformuler des requêtes en utilisant la structure des documents, etc.;
- ▶ créer et mettre à la disposition des utilisateurs des espaces de travail collaboratif permettant de partager à la fois les expériences et les connaissances, mais aussi les résultats et les processus de recherche.

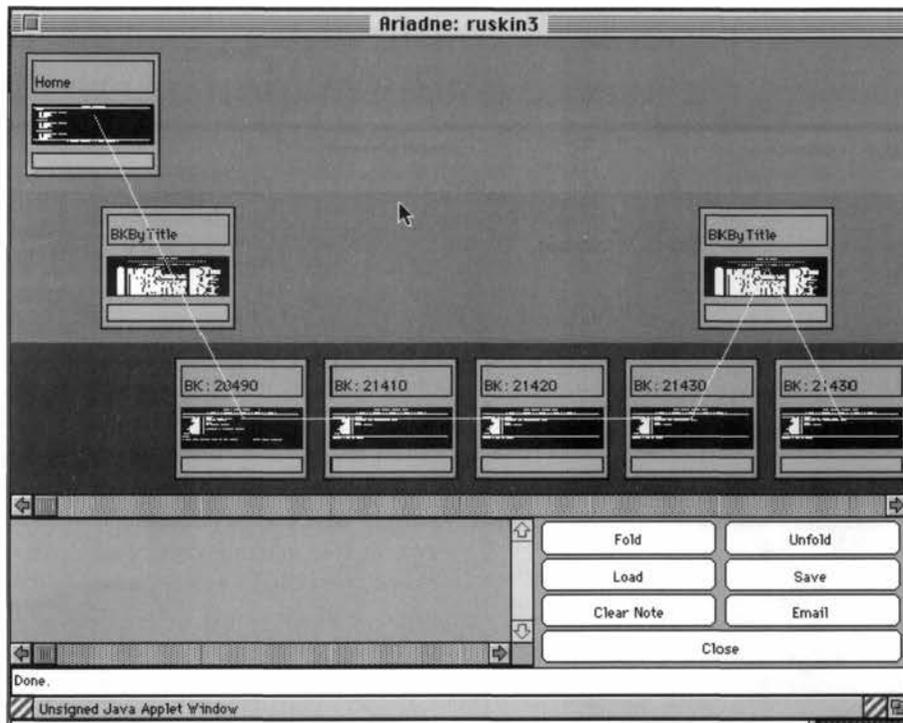


Figure 1 : Visualisation du Processus de la Recherche d'ARIADNE

PROBLÉMATIQUE

Tout système de recherche d'informations (SRI) a pour but de satisfaire les besoins des utilisateurs, mais il arrive, malgré toutes les techniques intégrées dans ces systèmes, que les utilisateurs n'obtiennent pas des résultats pertinents. Les causes de ces échecs sont multiples, allant d'une mauvaise formulation de question de la part de l'utilisateur à un manque de savoir sur le fonctionnement du système (David 1999, 1999a). De ce fait, les échecs sont moins nombreux lorsque les utilisateurs peuvent établir un dialogue et formuler adéquatement leurs besoins tout en profitant de l'expérience, des connaissances, voire des résultats de recherche d'autres chercheurs. Cet échange est rendu possible grâce aux techniques de représentation des documents (des contenus aux structures logiques). *Xtensible Markup Language* (XML), avec tous ses langages dérivés, s'impose aujourd'hui comme la solution technique la plus intéressante pour structurer des données de formats variés et des volumes très importants.

Nous proposons donc un prototype de système de recherche collaborative d'informations tenant compte du contexte social que doit avoir un SRI (ce que nous appelons « travail collaboratif ») et des nouveaux besoins liés à la structuration de l'information (en profitant de la technologie XML et des travaux liés au Web sémantique). Une recherche qui doit être comprise aussi bien par l'humain que par la machine...

ÉTAT DE L'ART

Différents systèmes ont été développés pour bien soutenir la collaboration des groupes d'individus dans différents domaines. Néanmoins, la majorité des systèmes développés ne fonctionnent avec succès que dans les laboratoires, peu d'entre eux étant commercialisés. Le Web, le courriel, les forums et les vidéoconférences sont les outils les plus utilisés actuellement pour appuyer tous styles de collaboration.

Ces types de systèmes existent dans différents domaines, allant de la rédaction à la recherche d'information, en passant par le dessin, la musique et l'apprentissage, pour n'en citer que quelques-uns.

Systèmes collaboratifs de rédaction et d'écriture collaborative

La rédaction collaborative constitue l'un des premiers domaines de recherche dans les CSCW (*Computer Supported Collaborative Work*). De nouveaux systèmes d'édition et de rédaction des documents électroniques existent déjà. Ils tiennent surtout compte des possibilités d'interactions entre les personnes qui créent les documents et celles qui les consultent.

Actuellement, les systèmes élaborés pour assurer l'écriture collaborative tiennent compte de la notion de « version du document », c'est-à-dire que pour chaque document, les traces de l'auteur des versions et les dates sont conservées. Cependant, le détail de ce qui a fait quoi, par rapport à la version précédente, devient

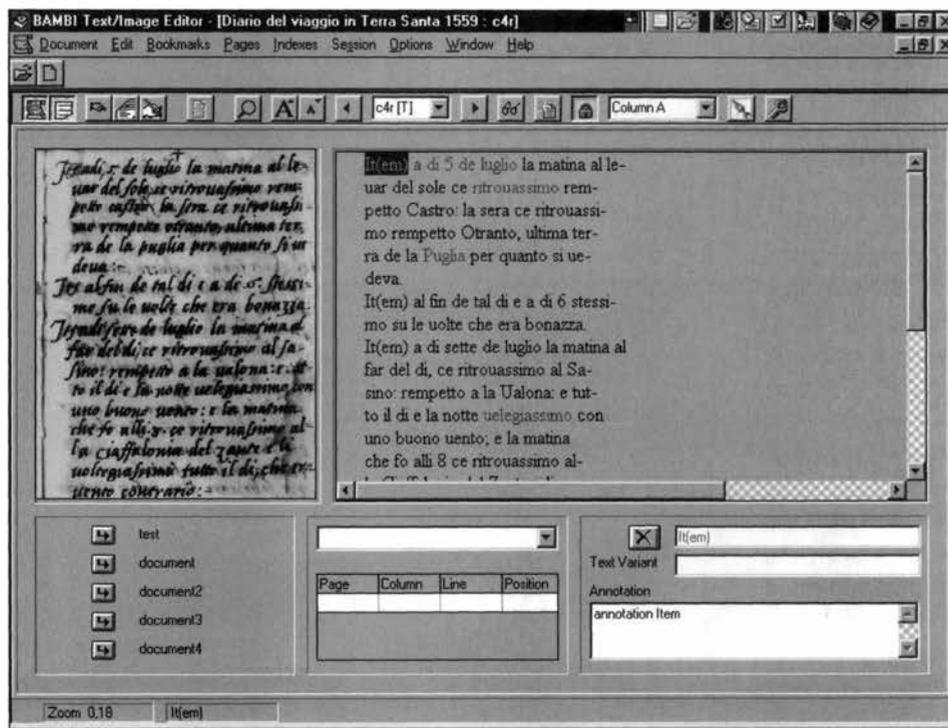


Figure 2 : Station BAMBI

moins évident lorsqu'il y a plus de deux co-auteurs (Twidale 1998).

Prenons l'exemple d'Adobe InCopy CS qui est un logiciel de rédaction et d'édition professionnelle distribué par Adobe dont le but est de « *simplifier et optimiser les flux d'édition collaborative. [...] Ce système rend le contrôle du calibrage aux mains des éditeurs, des rédacteurs et des concepteurs; [...] il leur permet de travailler simultanément sur le même document* » (Adobe 2003).

Systemes d'apprentissage collaboratif

Un système d'apprentissage collaboratif a pour rôle d'encourager le partage, la construction et la recherche d'information. Ce système peut « *se composer des outils de production (traitements de texte, outils de dessin, bases de données), des passerelles vers les autres apprenants (messagerie locale ou courrier électronique), vers les réseaux de bases de données (bibliothèques locales, réseau, Internet), et d'autres outils permettant de stocker, organiser, partager et traiter l'information dans le cadre d'une recherche commune* » (U. M. 1997). Citons ici le système Learn-Nett qui représente une expérience d'apprentissage collaboratif à distance. Il vise à expérimenter des modalités d'introduction de l'enseignement ouvert et à distance au sein de formations s'adressant à de futurs enseignants (agrégation et licence en sciences de l'éducation). Il existe sous forme de réseau européen incluant huit universités, une école supérieure, un centre de recherche et un service d'enseignement à distance. D'un

point de vue technique, un campus virtuel est mis sur pied en intégrant notamment divers outils de communication et de collaboration accessibles par Internet.

Les fonctions associées à ce campus sont les suivantes :

- ▶ « *faciliter le déplacement dans le campus (barres de menus, aides contextualisées et cartes des espaces);*
- ▶ *donner accès aux informations relatives à la communauté Learn-Nett (pages personnelles des étudiants et des tuteurs, travaux des groupes, consignes de travail, contextes des institutions partenaires, archivage de documents divers, News, etc.);*
- ▶ *favoriser la régulation et la gestion du travail collaboratif (carnet de bord pour les étudiants et tuteurs, gestionnaire de tâches);*
- ▶ *permettre la communication synchrone et asynchrone (café-forum général, forum pour chaque groupe de travail);*
- ▶ *favoriser le partage de ressources pédagogiques (ressources générales et ressources pour le groupe) »* (Charlier 2000).

Systemes de recherche collaborative d'informations

Dans le domaine de la recherche d'informations plusieurs systèmes ont aussi été développés dans le but de créer un dialogue entre les utilisateurs. Dialogue que nous avons perdu avec l'arrivée du document numérique et l'accès à domicile aux différentes bibliothèques et bases de données documentaires.

Le projet ARIADNE (*Annotable Retrieval Information And Database Navigation Environment*) représente un de ces systèmes. Il est conçu dans le but d'augmenter les occasions et l'efficacité de l'apprentissage collaboratif à distance par le biais de la « *visualisation du processus de la recherche* » [Figure 1] et de « *la navigation collaborative* » (Twidale 2002).

BAMBI (*Better Access to Manuscripts and Browsing of Images*), (Calabretto 1999), est un autre projet européen qui vise à mettre à la disposition d'un grand nombre d'utilisateurs (notamment le personnel de la bibliothèque et les chercheurs intéressés par l'histoire) un système permettant d'accéder, de manipuler et de naviguer dans une collection de manuscrits du Moyen Âge, à l'aide d'un ensemble d'outils de recherche plus ou moins performants [Figure 2]. Dans ce système, l'aspect du travail collaboratif se manifeste à travers les transcriptions, l'indexation et surtout les annotations publiques. En effet, les utilisateurs peuvent, suivant des critères et des règles de conventions accessibles, transcrire un manuscrit. Cette opération consiste à « *noter la prononciation d'une langue donnée à l'aide d'un système de signes d'une langue de conversion* » (Calabretto 1998). Quant à l'indexation, elle se fait après avoir transcrit le manuscrit et se base sur la technique plein-texte.

DEBORA (*Digital accEss to Books of the RenAissance*), (Nichols 2000), est également un projet européen qui a été conçu dans le même contexte. Il a pour but de développer des outils d'accès distants à des collections numérisées du XVI^e siècle qui appartiennent à des bibliothèques européennes (Lyon, Rome et Coimbra). Le travail collaboratif au sein de ce projet se représente essentiellement par de nouvelles fonctionnalités ajoutées à l'interface client permettant

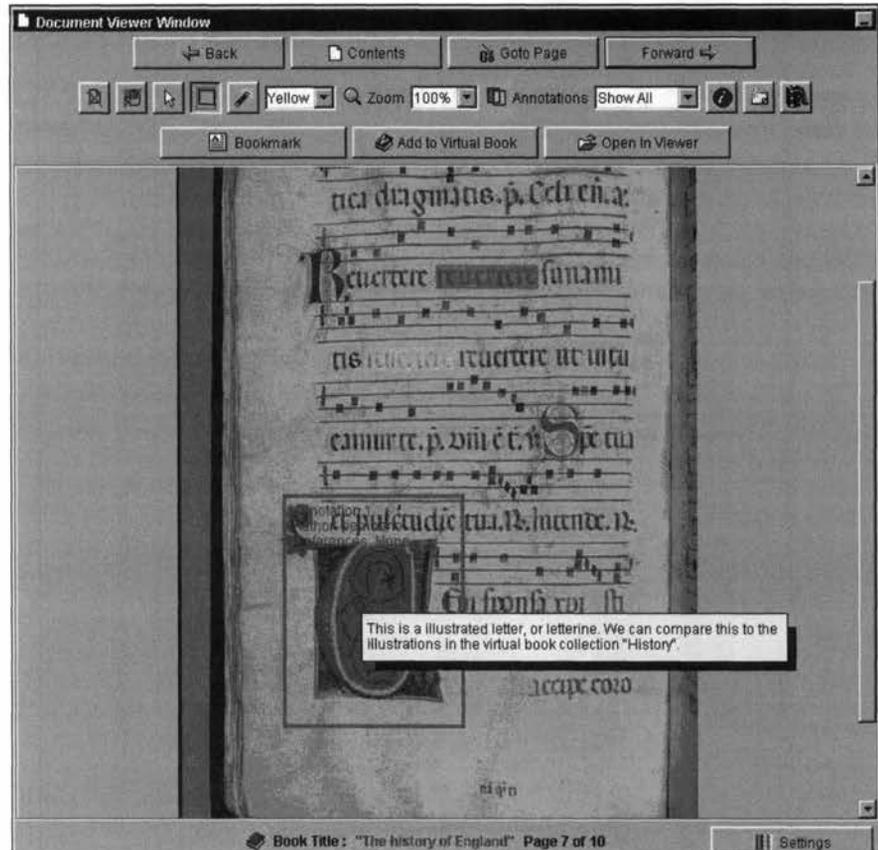


Figure 3 : Annotations (DEBORA)

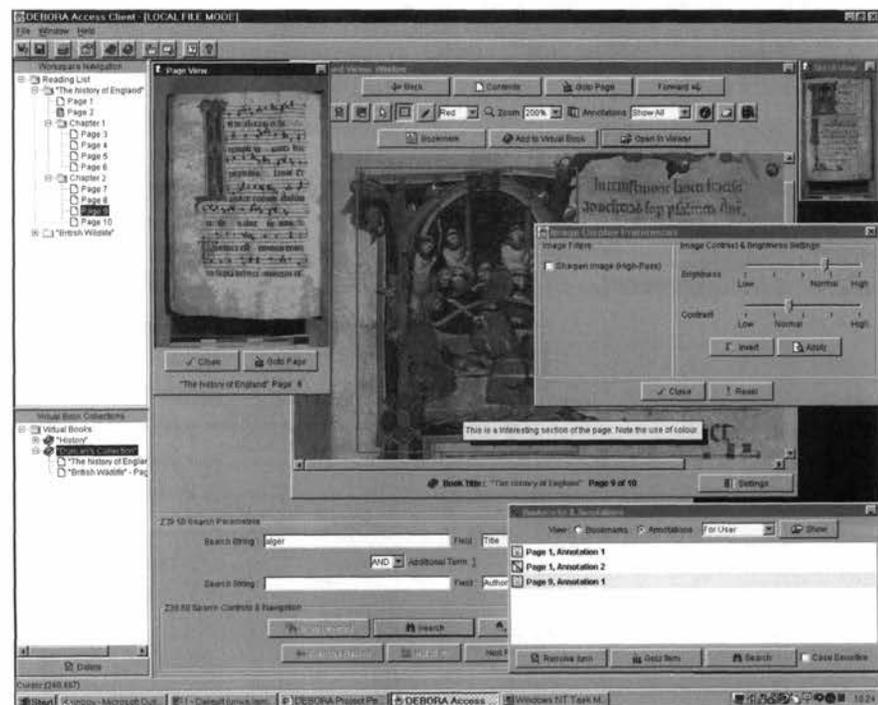


Figure 4 : Ajout des livres virtuels (DEBORA)

aux utilisateurs de coopérer. Cette coopération se traduit notamment par les annotations [Figure 3] qui peuvent être des explications de certaines idées, des liens ou des bibliographies personnelles, etc.

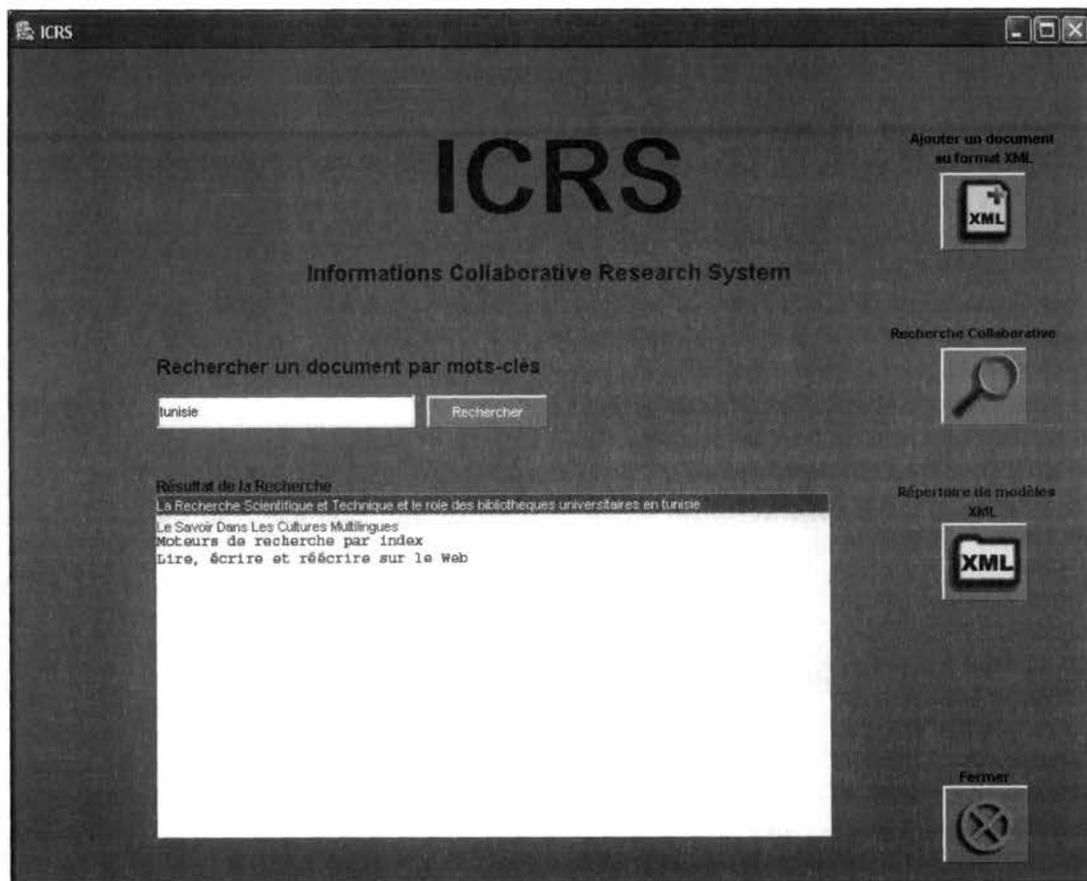


Figure 5 : Interface d'accueil d'ICRS

Une autre forme de coopération peut être représentée par l'ajout des livres virtuels [Figure 4], créés par des utilisateurs, à la collection de base à partir des fragments des documents numérisés consultés. Ces livres virtuels peuvent être consultés séparément de la collection des documents numérisés.

Ces exemples de projets, qui ne sont pas exclusifs, montrent bien l'intérêt de la collaboration entre les utilisateurs de système de recherche d'informations. En effet, le fait de partager avec d'autres utilisateurs des expériences, des annotations, des évaluations, entre autres, recrée, à l'évidence, un contexte social dans lequel vivaient les chercheurs. Plus le dialogue existe, plus la recherche d'informations est aisée et efficace. Cependant ces projets s'adressent davantage, en raison de leurs contenus (manuscrits, documents très anciens, etc.), à des spécialistes qui font des recherches pointues.

1. Cette enquête a touché les responsables du réseau de la Bibliothèque nationale de France (BnF) et ses pôles associés, ainsi que quelques autres bibliothèques numériques repérées sur Internet. Le choix de ce terrain se justifie par deux raisons essentielles : la répartition géographique et thématique et la taille moyenne de l'échantillon que représentent les bibliothèques des pôles associés.
2. La deuxième enquête s'est adressée aux utilisateurs de ressources numériques dans quatre différentes bibliothèques lyonnaises. Une bibliothèque publique : la BML (Bibliothèque Municipale de Lyon); un centre de documentation : Doc'INSA; deux autres bibliothèques universitaires de Lyon 3; la Bibliothèque de la manufacture des tabacs et la Bibliothèque de droit-gestion sur les quais. Le choix de ce terrain se justifie par la diversité des utilisateurs liée à la typologie des bibliothèques.

DESCRIPTION D'ICRS

Notre système de recherche d'information ICRS (*Information Collaborative Research System*) a été créé suite à des enquêtes menées en France auprès des utilisateurs de ressources numériques durant l'année 2002 (Mkadmi 2003). La première enquête a été effectuée auprès de quelques responsables de bibliothèques françaises¹ et concernait leurs projets de numérisation, leurs activités de collaboration, ainsi que leurs manières d'évaluer ce type d'activités. Cette enquête fut suivie d'une deuxième auprès des utilisateurs des systèmes numériques², à travers laquelle nous visons à connaître leurs habitudes, leurs manières de faire leurs recherches, et surtout leurs nouvelles attitudes face au document numérique. Des observations de quelques utilisateurs, ainsi que certains entretiens ont complété les enquêtes en déterminant les manières dont les utilisateurs procèdent dans leurs recherches, les fonctionnalités qu'ils utilisent le plus souvent, et celles qu'ils souhaitent trouver.

Notre système répond par l'ensemble de ses fonctionnalités à deux dimensions [Figure 5] :

- *une dimension auteur* en présentant les moyens nécessaires pour qu'un utilisateur puisse rédiger son document conformément à un modèle XML qui lui sera également présenté;

- ▶ *une dimension lecteur* en permettant à tout utilisateur de faire des recherches et d'accéder à un ensemble de documents répondant à son besoin.

Pour cette deuxième dimension, nous exposons la problématique du travail collaboratif au niveau de la recherche d'informations, et son importance dans un cadre d'accès distant aux ressources numériques. Ce travail collaboratif est présenté par une interface qui permet :

- ▶ de travailler à plusieurs dans un environnement de réalité virtuelle où les activités des utilisateurs et les informations peuvent être visualisées (échange des messages, partage des historiques de recherche, formulation des requêtes collaboratives, etc.);
- ▶ de passer, selon un système d'organisation de profils, des informations jugées pertinentes et utiles à d'autres utilisateurs répondant au même profil;
- ▶ d'ajouter des informations aux documents, (bibliographies personnelles pouvant être fusionnées et partagées selon le profil et le groupe, des liens entre les documents, etc.);
- ▶ d'annoter les documents en ajoutant un contenu dans le système d'informations (partage des idées);
- ▶ d'ajouter des commentaires concernant l'évaluation des documents (utilité, qualité, intérêt, etc.);
- ▶ de sauvegarder non seulement les résultats de recherche fructueux mais aussi les processus réels de recherche;
- ▶ de gérer les différentes versions d'un document.

Notre système présente à la fois une interface de recherche simple et un accès à une interface de recherche collaborative, un répertoire de modèles XML et une interface d'alimentation de la base de données de documents.

Indexation des documents

Comme dans tout système de recherche d'informations, l'indexation est l'opération centrale de notre système qui permet de représenter la teneur du document sous une forme concise, de faire la recherche et d'éviter au système d'analyser chaque texte du document à chaque interrogation. Une fois que l'utilisateur (auteur) a soumis son document en format XML dans la base de données, un agent du système se charge de l'analyse.

Cette analyse a deux objectifs :

- ▶ vérifier la grammaire XML et sa conformité par rapport au modèle utilisé, à travers des « validateurs »³ incorporés dans notre système;
- ▶ indexer des documents. Une indexation qui se fait, elle aussi, par un autre agent parcourant les métadonnées du document pour extraire les éléments qui vont servir, par la suite, de clés de recherche.

Ces éléments seront stockés dans la base de données avec des liens identifiant les contenus des documents.

Structuration de l'information avec XML

Pour répondre à de nouveaux besoins des utilisateurs et permettre des accès à l'information plus sophistiqués et plus aisés, et pour offrir des possibilités d'échange et de travail collaboratif, nous avons tiré profit de l'apparition, ces dernières années, des normes et des recommandations dans le domaine du Web au niveau de la description et de l'échange des données, notamment XML. Créé en 1998 par le W3C (World Wide Web Consortium), ce métalangage offre une représentation des documents beaucoup plus riche que les documents ordinaires, permettant de joindre au contenu brut des informations logiques, des rapports entre éléments d'information, des métadonnées, des liens vers d'autres sources d'information, etc. Pour toutes ces raisons et pour optimiser la précision et l'efficacité de la recherche d'informations, nous avons choisi de structurer nos documents en format XML. Pour l'instant l'« xmélisation » se fait à la main, faute de disposer d'outils fiables à ce jour permettant de faire des conversions d'autres types de formats vers XML. Les documents de la base sont des travaux universitaires (mémoires et articles scientifiques) qui sont structurés selon des modèles de représentations logiques des données (schémas XML) que nous avons créés et stockés dans le répertoire de modèles XML. Nous présenterons ce répertoire plus loin dans l'article.

Interface de recherche collaborative

Avant de parler des différentes fonctionnalités de cette interface [Figure 6], signalons d'abord que dans notre système de recherche collaborative d'informations, nous avons prévu deux modes de collaborations, sans exclure bien sûr le mode autonome, qui permet à l'utilisateur d'effectuer une recherche d'une manière tout à fait indépendante, tout en excluant toute intervention par un autre utilisateur. Les deux autres modes sont :

- ▶ le mode « collaborateur » : qui consiste à faire une recherche collaborative proprement dite. Ce mode permet à tous les utilisateurs connectés au système, et ayant le même profil, de dialoguer ensemble et de voir réciproquement les requêtes émises et les résultats de recherche reçus par chacun. Ce mode permet aussi d'annoter les documents, d'évaluer leur pertinence tout en accédant aux annotations des autres chercheurs;

3. Le « validator » que nous utilisons s'appelle xsdvalid, outil gratuit créé en Java. Il permet de vérifier la syntaxe XML et de valider un document par rapport à un schéma XML prédéfini. Pour plus d'information sur l'outil, consulter le site : <http://www.xmlmind.com/xsdvalid.html>

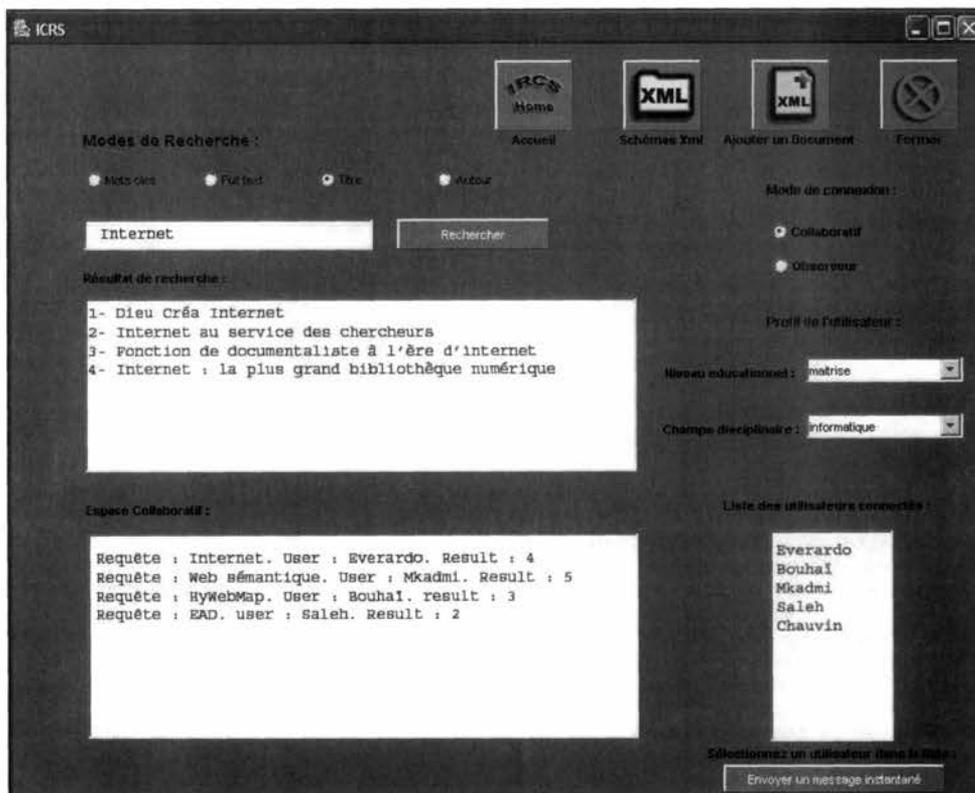


Figure 6 : Interface de recherche collaborative d'informations

- ▶ le mode « observateur » : qui permet à un utilisateur d'effectuer sa recherche tranquillement sans être dérangé par quiconque, mais qui permet quand même aux autres de suivre ses démarches de recherche et de voir les résultats qu'il a obtenus, ainsi que les annotations et les commentaires d'évaluation qu'il pourrait introduire dans un document. Ce mode de collaboration est surtout valable pour un expert d'un domaine qui veut faire profiter d'autres utilisateurs (étudiants, novices, etc.) de ses compétences et de ses connaissances.

Ces deux modes de collaboration passent par un système d'identification et de profilage des utilisateurs.

À partir de cette interface, l'utilisateur peut choisir :

- ▶ son mode de collaboration (collaborateur ou observateur);
- ▶ son profil : le système de profilage que nous avons mis en place pour l'instant est très simple, il contient deux critères qui sont le niveau éducationnel et le champs disciplinaire;
- ▶ son critère de recherche : mots clés, auteur, titre, texte intégral. Il peut bien sûr combiner les critères pour mieux affiner sa recherche.

Cette interface permet aussi à l'utilisateur d'effectuer une collaboration synchrone avec d'autres utilisateurs à travers une liste de tous les utilisateurs connectés

au système [Figure 6]. Il suffit de sélectionner le nom de la personne et de lui envoyer un message instantané. L'utilisateur pourra également visionner les requêtes des autres avec le nom de la personne, son profil, les mots clés de sa recherche, ainsi que le nombre de résultats obtenus. Cette fonctionnalité permet à n'importe quel utilisateur, s'il est intéressé par le même sujet, de cliquer directement sur la requête pour avoir le résultat sans être obligé de refaire la même opération.

Quant aux résultats de recherche, ils sont affichés par ordre de pertinence dans une fenêtre sous forme de listes sensibles (liens hypertextes) de titres de documents. En cliquant dessus, l'utilisateur peut avoir le contenu du document dans une autre interface, qui elle aussi, lui permet de collaborer avec les autres à travers les annotations et le jugement personnel des documents.

Valeur ajoutée aux documents : annotations et évaluation

Deux éléments, annotations et évaluation des documents, ont toujours été cités par les utilisateurs des ressources numériques comme pouvant être ajoutés facilement à un texte numérique. Les annotations ont été proposées fréquemment comme une technique permettant aux utilisateurs d'ajouter un contenu (et partager ainsi des idées) dans des systèmes

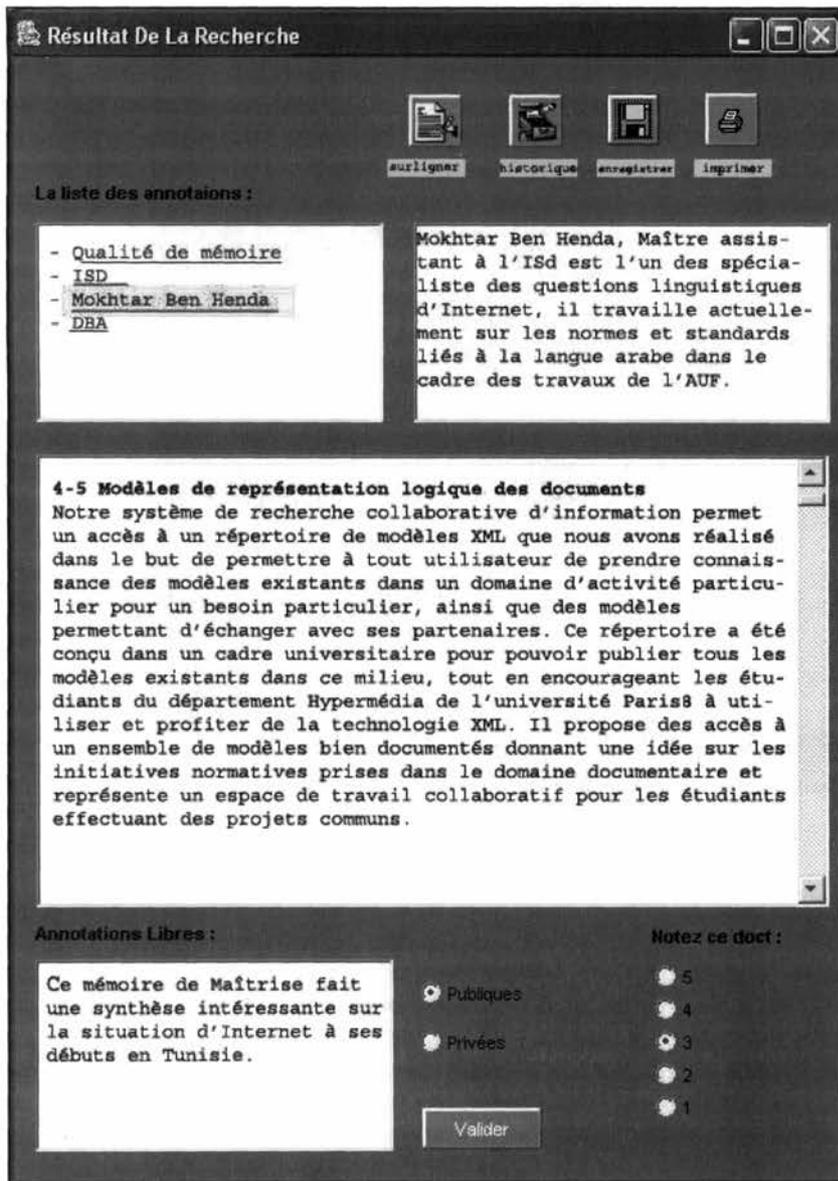


Figure 7 : Interface de valeur ajoutée au document

d'information. L'évaluation est une indication de l'utilité, de l'intérêt ou de la qualité d'un document vu par un utilisateur. Des évaluations peuvent être proposées délibérément par un utilisateur (évaluations explicites) ou peuvent être calculées par le système sur la base de l'intérêt que les utilisateurs montrent dans un document (évaluations inférées).

Dans notre système, le résultat de recherche est présenté dans une interface qui permet de voir le contenu des documents et les annotations associées faites par d'autres utilisateurs. Elle permet également d'ajouter des annotations et d'évaluer le document consulté [Figure 7].

Les annotations consistent à mettre en évidence certains passages du document au moyen de soulignements ou de surlignements de couleurs différentes selon les thèmes. Elles peuvent être plus conséquentes en donnant des éclaircissements sur le texte grâce à des connaissances acquises postérieurement à l'écri-

ture initiale ou des interprétations de lecteurs du texte à travers les « commentaires » sur le document numérique. Ces annotations sont de deux types :

- ▶ des annotations publiques que l'utilisateur émet au système pour qu'elles soient visibles pour tous les autres utilisateurs. Ces annotations sont sous forme de commentaires, de jugement de la qualité du document, de bibliographies personnelles ou autres qui peuvent donner une idée sur le contenu de document. Ces annotations sont accompagnées du nom de l'auteur;
- ▶ des annotations privées que seul le « lecteur-auteur » peut voir. Elles sont sous forme de points de repères et/ou d'indications des parties du texte sur lesquelles le chercheur aimerait bien revenir, etc.

Quant à l'évaluation, l'interface propose à l'utilisateur d'attribuer une note au document consulté portant sur sa pertinence et son utilité par rapport à la requête émise. La note va de 1 (non pertinent) à 5 (très pertinent).

À travers ces techniques, l'utilisateur peut « prendre connaissance immédiatement de l'appareil critique qui progressivement se met en place par des spécialistes répartis de par le monde » (Jacquesson 1999).

Modèles de représentation logique des documents

Notre système de recherche collaborative d'information offre accès à un répertoire de modèles XML que nous avons réalisé dans le but de permettre à tout utilisateur de prendre connaissance des divers modèles existants dans le domaine documentaire. Ce répertoire a été conçu dans un cadre universitaire pour pouvoir publier tous les modèles existants dans ce milieu, tout en encourageant les étudiants à utiliser et à profiter de la technologie XML. Il propose des accès à un ensemble de modèles bien documentés donnant une idée sur les initiatives normatives prises dans le domaine documentaire et représente un espace de travail collaboratif pour les étudiants effectuant des projets communs.

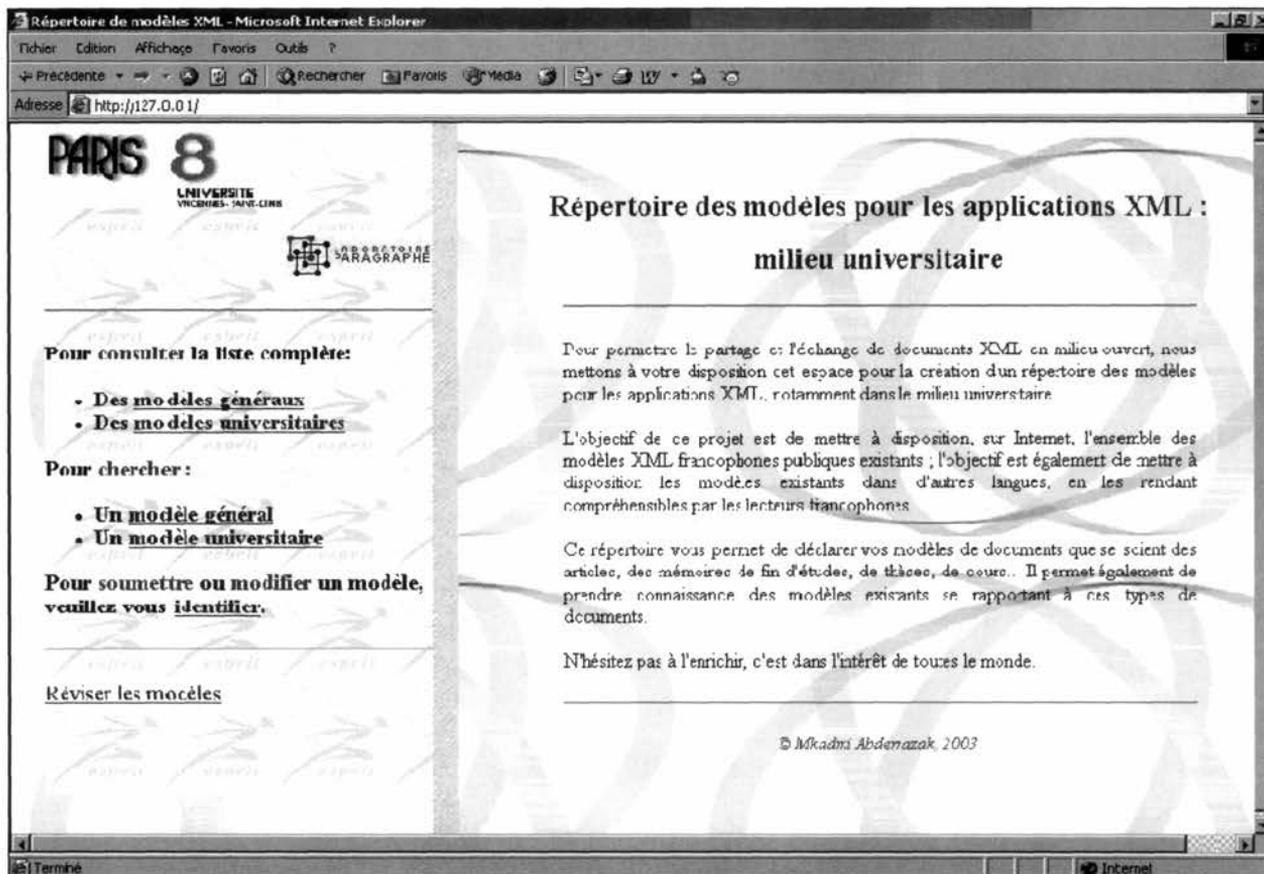


Figure 8 : Interface d'accueil du répertoire de modèles XML

RÉPERTOIRE DE MODÈLES XML

Notre répertoire de modèles est structuré en six parties contenant :

- ▶ les éléments de définition du modèle ;
- ▶ les informations associées au modèle ;
- ▶ les critères de recherche ;
- ▶ le contrôle et la révision des modèles proposés à la publication ;
- ▶ la consultation des données du répertoire ;
- ▶ la soumission des modèles et documents XML.

Cette structure a été définie, suite à une étude de besoins réalisée dans le cadre d'un projet de collaboration entre le laboratoire Paragraphe, EDIFRANCE (Association pour le développement des échanges électroniques professionnels), MUTU-XML et GFII (Groupement Français de l'Industrie de l'Information), FING qui vise à mettre en place un répertoire de modèles XML (Mkadmi 2003a).

Le répertoire de modèles est librement accessible [Figure 8] à tous sans aucune limite dans la consultation des schémas et des DTD (Définition de type de document) qui y sont stockés. Quant à la sécurité, il n'y a pour le moment aucune confidentialité concernant les accès aux données. La recherche des modèles se fait à travers plusieurs critères (secteur d'activité,

domaine d'usage, mots clés) définissant le contexte et l'appartenance de chaque modèle. Quant à la modification et à la mise à jour, elles se font par les propriétaires des modèles, seuls responsables des structures de leurs documents. Cependant, la publication de ces modèles passe par un « comité éditorial » qui vérifie la forme de ce qui est proposé à la publication, c'est-à-dire la cohérence des données contextuelles, la pertinence de la définition par rapport au contexte, la bonne syntaxe des schémas et des DTD, ainsi que la cohérence des documents d'exemples en fonction des modèles.

Notre répertoire de modèles XML est un espace ouvert à tous les utilisateurs, qu'ils soient du milieu universitaire ou autres. Chacun peut y déposer ses modèles de documents et profiter des modèles déjà publiés. C'est un espace de travail de collaboration offrant une aide à la réalisation des schémas XML tout en encourageant toutes les personnes intéressées à trouver un terrain d'entente et d'échange. Car XML s'impose aujourd'hui comme étant un format standard de description et d'échange de données.

Description des modèles XML

Nos modèles de documents XML ont été créés suite à un travail de simplification et d'adaptation de

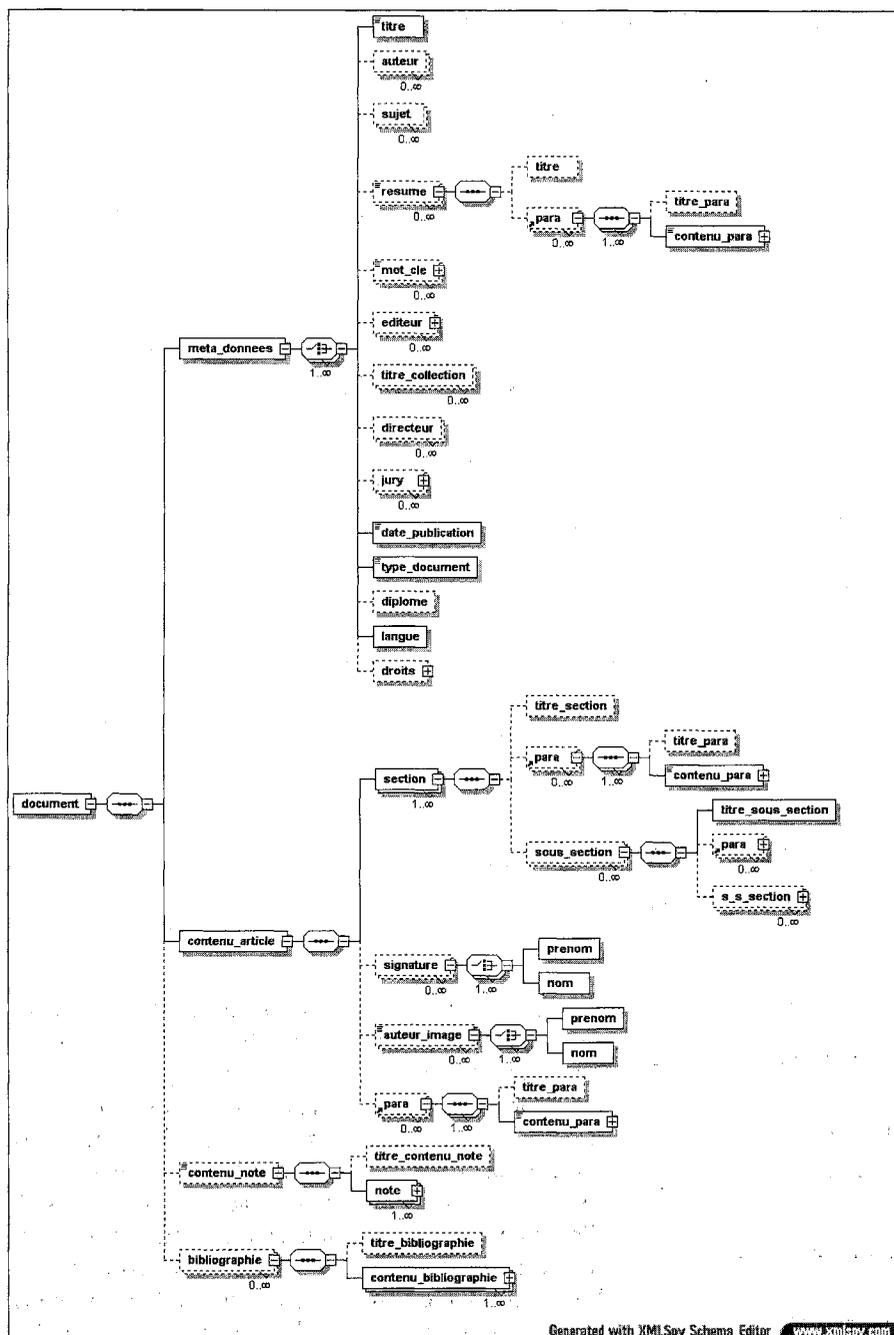


Figure 9 : Structure du modèle général

plusieurs autres initiatives dans ce domaine : DocBook (OASIS)⁴, DocUpsy (IED, Université Paris 8, Tonin 2002), Cyberthèse (Université Lyon 2)⁵. Un modèle général a été réalisé pour définir la structure de trois types de documents universitaires : un article scientifique, un mémoire de fin d'études et/ou un rapport de stage et une thèse de doctorat [Figure 9].

Ce schéma XML permet de valider tous les documents écrits dans le cadre d'une recherche scientifique au département Hypermédia de l'Université Paris 8 (articles scientifiques, mémoires, rapports de stages, thèses, etc.), mais pourra aussi être généralisé pour inclure tous les articles scientifiques. Ceci est valable pour assurer l'homogénéité des documents et, en

même temps, pour permettre à plusieurs chercheurs de travailler ensemble en se basant sur une structure prédéfinie.

Comme le montre le schéma précédent, un article est constitué d'à peu près quatre parties :

- ▶ définition des métadonnées du document. Les éléments de ces métadonnées sont inspirés de Dublin Core (DC)⁶ mais reformulés et renommés pour nos besoins spécifiques par Titre, auteur (nom, prénom, labo de recherche, courriel), revue (titre, volume, pagination...) et date de publication ;
- ▶ présentation du contenu de l'article (sections, sous-sections, chapitres, paragraphes, avec les tableaux, figures, image, etc.) ;
- ▶ présentation du contenu des notes de bas de pages ;
- ▶ présentation des références bibliographiques.

Ces modèles sont présentés dans notre système avec les interfaces utilisées pour la soumission et la modification des modèles et l'ensemble des informations associées [Figures 10 et 11].

Le modèle XML peut être présenté en texte dans

le formulaire lui-même ou comme un lien vers un fichier qui se trouve sur le serveur ou sur Internet. On peut également trouver une feuille de style utilisée par le propriétaire du modèle et à laquelle on peut accéder de la même manière qu'au modèle lui-même (en texte ou en lien hypertexte).

4. Pour plus de renseignements sur DocBook, veuillez consulter : <http://www.docbook.org/>

5. Cyberthèse est un projet commun de l'Université Lumière-Lyon 2 et Les presses de l'Université de Montréal dont le but est de permettre aux auteurs de mettre en ligne leurs thèses de façon simple. Pour plus d'informations, consultez le site : <http://mirror-fr.cybertheses.org/>

6. Il y a 15 éléments de métadonnées dans le Dublin Core : titre, créateur (auteur), sujet, description, éditeur, contributeur, date, type (de ressource), format, identifiant (URL, ISBN), source, langue, relation (URI ressource liée), couverture, droits. <http://www.dublincore.org>

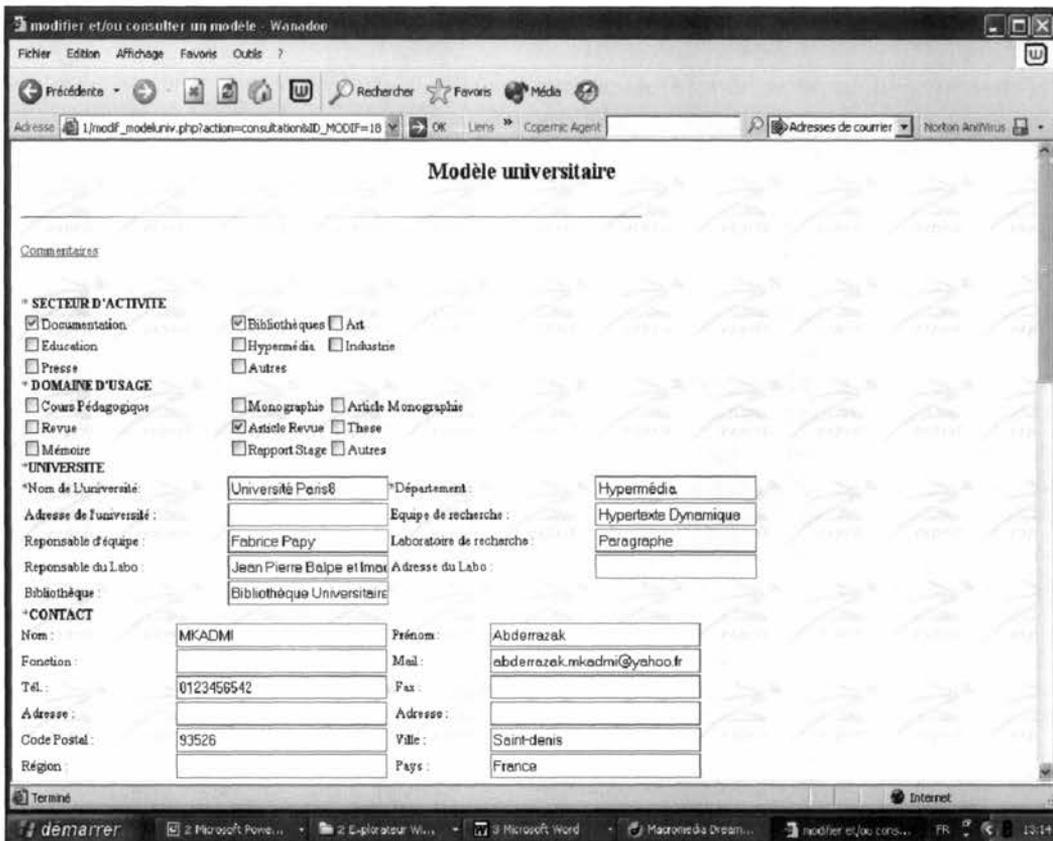


Figure 10 : Présentation des informations associées aux modèles XML

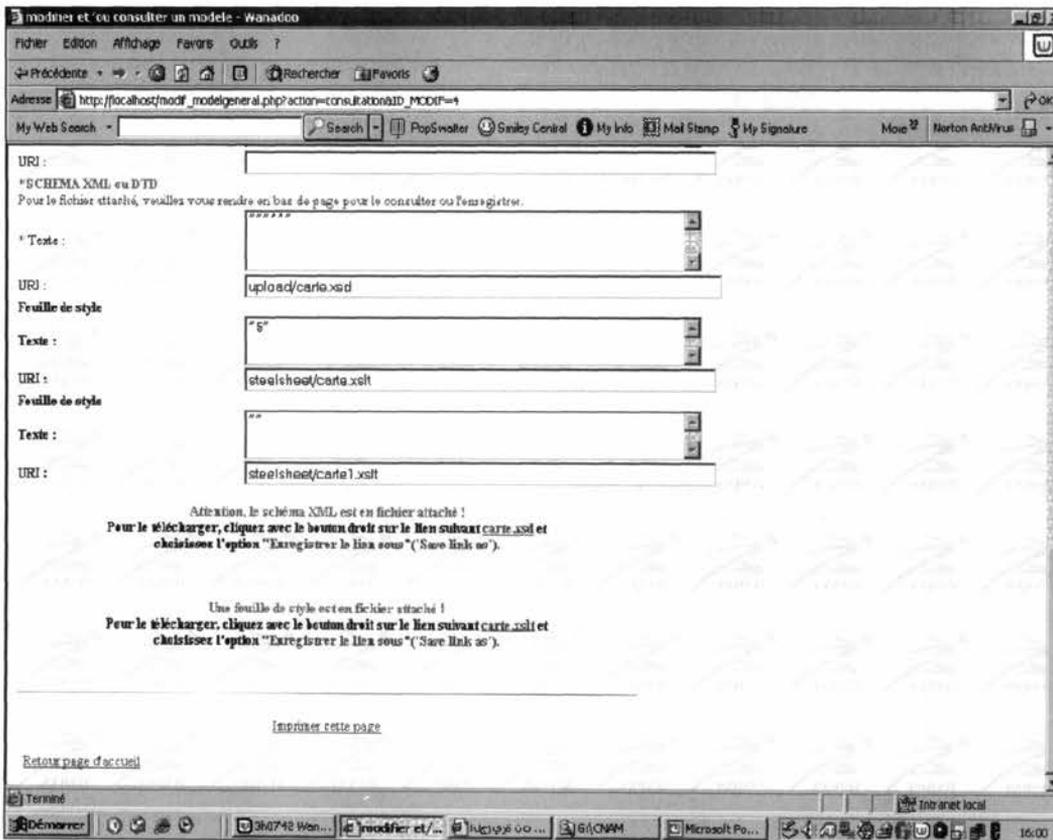


Figure 11 : Présentation des modèles en fichier attaché

CONCLUSION

La recherche collaborative des informations représente, à notre avis, une solution pour répondre aux problèmes des systèmes de recherche d'informations se rapportant notamment à la pertinence des réponses obtenues par les utilisateurs. En effet, aucun système de recherche d'informations ne peut sélectionner d'une manière précise tous les documents pertinents (et uniquement les documents pertinents) pour répondre à une requête. Ceci exige une personnalisation des réponses puisque les préférences ne sont pas les mêmes pour tout le monde.

Si, dans le Web, les fureteurs (*browsers*) ont mis depuis longtemps en valeur la notion de recherche et de diffusion de l'information, toutes les conditions maintenant incitent à revenir à l'objectif du Web, celui du travail collaboratif ou plutôt de la création collective.

Avec ce projet, qui n'est que le début d'un travail plus ambitieux voulant répondre aux problèmes de la recherche d'information par le biais de la structuration des documents, l'utilisation des apports de la technologie XML et le travail collaboratif, nous avons conçu une bonne base axée sur les besoins réels des utilisateurs et réalisée avec des outils qui ont démontré stabilité et efficacité.

À présent, nous concentrons notre travail sur :

- ▶ la création d'un système de navigation hypertextuelle à travers l'ajout d'une fonctionnalité permettant d'ajouter un lien hypertextuel à un mot ou à une expression vers d'autres documents traitant le sujet, en profitant d'un autre outil créé au sein du laboratoire Paragraphe nommée HyWebMap (Bouhaï 2002) (Mkadmi 2004);
- ▶ l'installation d'un moteur de recherche plus puissant permettant une recherche sur la structure XML des documents en utilisant le langage XQuery;
- ▶ l'intégration d'un outil permettant de générer automatiquement ou semi-automatiquement une ontologie à partir de schémas XML, plus précisément, une structure d'ontologie qui sera enrichie par les différents experts dans le cadre du travail collaboratif, et ce pour évoluer vraiment vers un Web sémantique où les informations seront comprises par les machines. ♦

SOURCES CONSULTÉES

- Adobe InCopy CS simplifie et optimise les flux d'édition collaborative. Communiqué de Presse 30 septembre 2003. Adobe Agence Brodeur SRRP. <http://www.adobe.fr/aboutadobe/pressroom/pr/sep2003/InCopyCS.pdf>, consulté le 21/12/2003.
- Bouhaï, N. 2002. Lire, réécrire et partager le savoir sur le Web : problèmes et solutions. Thèse de Doctorat en Sciences de l'Information et de la Communication, Université Paris 8.
- Calabretto, Sylvie; Pinon, Jean-Marie et Bozzi, Andrea. 1998. *BAMBI : système de gestion de manuscrits anciens pour historiens*. Document numérique, vol. 2, n°3-4 : 32-49.
- Calabretto, Sylvie; Bozzi, Andrea et Pinon, Jean-Marie. 1999. Numérisation des manuscrits médiévaux : le projet européen BAMBI. Actes du colloque *Vers une nouvelle érudition : numérisation et recherche en histoire du livre*. (décembre) Rencontres Jacques Cartier, Lyon.
- Charlier. Learn-Nett : une expérience d'apprentissage collaboratif à distance. Congrès des chercheurs en éducation, mai 2000. Rapport présenté par Bernadette Charlier, Amaury Daele, Françoise Docq, [et al.]. <http://www.det.fundp.ac.be/~ada/docs/2000cfwb.doc>. Consulté le 26/12/2003.
- David, Amos A. 1999. Modélisation de l'utilisateur et recherche coopérative dans les systèmes de recherche d'informations. In : Maniez, Jacques et Mustafa El Hadi, Widad. Organisation des connaissances en vue de leur intégration dans les systèmes de représentation et de recherche d'information.
- David, Amos A. Modélisation de l'utilisateur et recherche coopérative d'information dans les systèmes de recherche d'informations multimédia en vue de la personnalisation des réponses. Mémoire d'habilitation à diriger des recherches en Sciences de l'information et de communication, soutenu le 29 juin 1999.
- Jacquesson, Alain; Rivier, Alexis. 1999. *Bibliothèques et documents numériques : concepts, composantes, techniques et enjeux*. Electre. Paris.
- Michinov. *Un aperçu de l'apprentissage collaboratif assisté par ordinateur*. Nicolas Michinov, avril 2002. Mise à jour le 24-09-2003. — <http://www2.univ-poitiers.fr/michinov/CSCL.html>. Consulté le 19/12/2003.
- Mkadmi, A.; Bsir B. 2003. Bibliothèques numériques et travail collaboratif. *RIST (Revue d'Information Scientifique & Technique)*, vol. 13, n°1.
- Mkadmi, A.; Bouhaï, N. et Langlois, M. 2003. Partager des modèles XML : quel intérêt? *Bulletin des bibliothèques de France*, n°5, (Septembre).
- Mkadmi, A.; Reyes E. et Bouhaï, N. 2004. Semantic Network of information. *Actes de la Conférence ICTTA 2004*, Damascus, (Avril) : 19-20
- Nichols, David M.; Pemberton, Duncan et al. 2000. DEBORA : Developing an Interface to Support Collaboration in a Digital Library. *Proceedings of the Fourth European Conference on Research and Advanced Technology for Digital Libraries (ECDL2000)*, Lisbon, Portugal. (Springer-Verlag) : 239-248.
- Tonin, Cyril. 2002. DocUpsy : Les réalisations de la Bibliothèque Universitaire de Paris VIII dans le cadre du Projet Upsy. <http://www.upsy.net>.
- Twidale, Michael B.; Nichols, David M. et Paice, Chris D. 1997. *Browsing Is A Collaborative Process*. Information Processing & Management, 33(6) : 761-83.
- Université de Montréal. Introduction aux applications de la technologie éducationnelle : de 1986-2010. http://www.usagers.umontreal.ca/bordelea/fv_travaux/ETA6746-A1997/Les%20cr%20%3A%9ateurs1986-2010b/Nathalie1/mm_apcoll_msc0.htm.