

Recherche collaborative d'informations : repenser l'architecture des SRI à l'ère du numérique

A. MKADMI, I. SALEH

Laboratoire Paragraphe (Université Paris8)
2, Rue La Liberté, 93526 Saint-Denis cedex

Abderrazak.Mkadmi@univ-paris8.fr , Imad.Saleh@univ-paris8.fr

Résumé

Cet article présente un prototype d'un 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'information 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, l'utilisateur, dans sa recherche, se trouve tout seul face à un écran. Au premier problème, nous proposons l'exploitation des apports de la technologie XML au niveau de la structuration et la recherche d'informations et au deuxième, nous proposons un espace de travail collaboratif qui permet aux utilisateurs de dialoguer et de s'entraider au niveau de leurs recherches. À la fin, nous présentons les conclusions et les perspectives de ce travail.

Mots clés

Recherche collaborative d'informations, schémas XML, travail collaboratif, interaction homme-machine-homme.

1 Introduction

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 dans la nature des documents et des corpus et dans 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, possédant ainsi une faible structuration. Les documents, catalogues, thesaurus et outils de recherche sont représentés électroniquement et peuvent être accédés à distance à travers des réseaux informatiques. Cette réalité de recherche à distance va rendre les interactions collaboratives traditionnelles plus rares en perdant la proximité physique à d'autres chercheurs. Déjà, depuis un bon moment, les occasions pour les interactions sociales sont réduites par l'accès de chez soi aux OPACs, et avec le développement des bibliothèques numériques, cette tendance peut 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 toujours pas exprimer précisément son but de recherche, et le producteur ne sait toujours pas présenter l'information d'une manière adéquate. En effet, les échanges d'expériences entre les utilisateurs, la classification des résultats de recherche semblent apporter une aide substantielle pour maîtriser l'espace informationnel actuel.

Suite à 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.

2 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 peuvent être moins nombreux quand il existe d'une part, un dialogue entre les utilisateurs qui les aide à bien formuler leurs besoins et à profiter, les uns des autres, de leurs expériences et de leurs connaissances et pourquoi pas même de leurs résultats de recherche et d'autre part, des techniques de représentation des documents aussi bien par leurs contenus que par leurs structures logiques. XML, avec tous les langages dérivés, s'impose aujourd'hui comme la solution technique la plus intéressante pour structurer de telles données de formats variés et de volumes très importants.

Au travers notre travail, nous essayons de répondre à ces questions à travers la proposition d'un prototype d'un système de recherche collaborative d'informations qui tient compte à la fois du contexte social que doit avoir un SRI (ce que nous appelons travail collaboratif), et des nouveaux besoins quant à la structuration de l'information (en profitant de la technologie XML et des travaux liés au Web sémantique) qui devra être comprise aussi bien par l'humain, que par la machine.

3 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 sont commercialisés. Le Web, le mail, les forums et les vidéoconférences sont les outils les plus utilisés actuellement pour appuyer tous genres de collaboration.

Ces types de systèmes ont existé dans différents domaines, allant de celui de la rédaction à celui de la recherche d'information, en passant par le dessin, la musique et l'apprentissage pour ne citer que ceux-ci.

3-1 Systèmes collaboratifs de rédaction et d'écriture collaboratives

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, ainsi que celles qui les consultent.

Les systèmes élaborés actuellement pour assurer l'écriture collaborative tiennent compte de la notion de « version du document », c'est-à-dire ils gardent les traces sur qui a travaillé sur telle version de tel document et à quelle date. Cependant, le détail de ce qui a fait quoi, par rapport à la version précédente, devient moins évident lorsque les personnes co-écrivains dépassent deux rédacteurs. (Twidale 1998)

Nous citons l'exemple de « ADOBE INCOPY CS » qui est un logiciel de rédaction et d'édition professionnelle, distribué par Adobe dans le but 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).

3-2 Systèmes 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 (traitement 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). Nous 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 formation s'adressant à de futurs enseignants (agrégation et licence en sciences de l'éducation). Il est sous forme de réseau européen, de 8 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, 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 et aide 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) ;
- 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)

3-3 Systèmes de recherche collaborative d'informations

Dans le domaine de recherche d'informations aussi, plusieurs systèmes ont été développés dans le but de créer un dialogue entre les utilisateurs, ce dialogue que nous avons perdu avec l'arrivée du document numérique et l'accès de chez soi aux différentes bibliothèques et bases de données documentaires. Le projet **ARIADNE** (Annotatable 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, d'une part la « *visualisation du processus de la recherche* » [Figure 1] et d'autre part, par « *la navigation collaborative* ». (Twidale 2002)

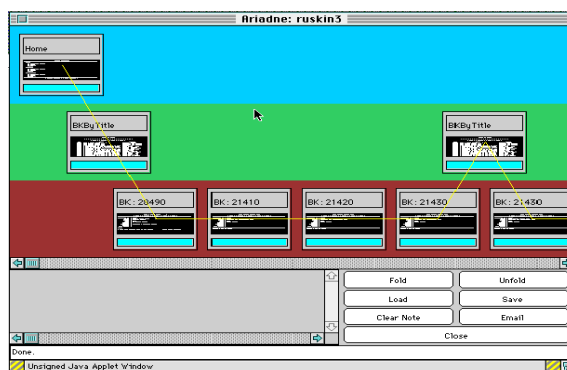


Figure 1 : Visualisation du Processus de la Recherche d'Ariadne

BAMBI (**B**etter **A**ccess to **M**anuscripts and **B**rowsing of **I**mages) (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 qui leur permet 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 transcrire le manuscrit et se base sur la technique full-text.

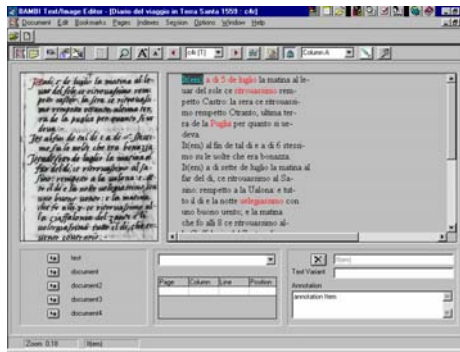


Figure2 : Station BAMBI

DEBORA - Digital accEss to Books Of RenAissance (Nichols 2000) est aussi un projet européen qui a été développé dans ce même contexte. Il a pour but de développer des outils d'accès distants à des collections numérisées du 16^{ème} siècle qui appartiennent à des bibliothèques européennes (Lyon, Rome et Coimbra). Le travail collaboratif au sein de ce projet se représente essentiellement par des nouvelles fonctionnalités ajoutées à l'interface client et qui permettent aux utilisateurs de coopérer entre eux. Cette coopération se traduit particulièrement par les annotations [figure 3] qui peuvent être des explications de certaines idées, des liens et des bibliographies personnels, etc.



Figure 3 : Annotations (DEBORA)

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.

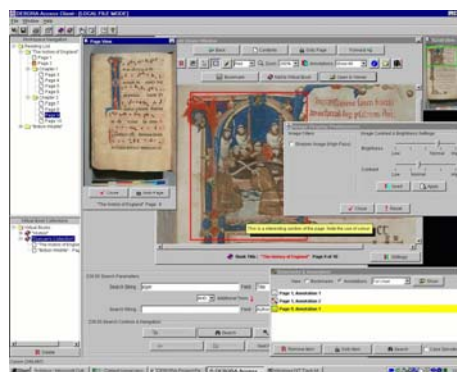


Figure 4 : Ajout des livres virtuels (DEBORA)

Ces exemples de projets (qui ne sont pas exclusifs) montrent bien l'intérêt des aspects de collaboration entre les utilisateurs au niveau de recherche d'informations. En effet, le fait de partager avec d'autres utilisateurs des expériences, des annotations, des évaluations, etc. participe, à l'évidence, à la création d'un contexte social, dans lequel vivaient les chercheurs. Plus le dialogue est là, plus la recherche d'informations est aisée et efficace. Cependant ces projets sont destinés, par leurs contenus (manuscrits, documents très anciens, ...), plutôt à des spécialistes qui ont des recherches plus pointues.

4 ICRS – Information Collaborative Research System: description

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 des bibliothèques françaises¹ portant sur leurs projets de numérisation, leurs activités de collaboration, ainsi que leurs manières d'évaluer ce type d'activités. Cette enquête était suivie d'une deuxième auprès des utilisateurs des systèmes numériques², à travers laquelle nous visons 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 [figure5] à deux dimensions :

- 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 permettant à tout utilisateur de faire des recherches et d'accéder à un ensemble de documents répondant à son besoin.

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

- à plusieurs personnes de travailler dans un environnement de réalité virtuelle, où les activités des utilisateurs et les informations peuvent être visualisées : échanger des messages, partager des historiques de recherche, formuler des requêtes collaboratives ...
- selon un système d'organisation de profils, de passer des informations jugées pertinentes et utiles par un utilisateur à d'autres utilisateurs ayant le même profil ;
- d'ajouter des informations aux documents, à savoir des bibliographies personnelles (qui peuvent être fusionnées et partagées dans un groupe (même profil)), des liens entre les documents ;
- d'annoter les documents en ajoutant un contenu dans le système d'informations (et partager ainsi des idées) ;
- d'ajouter des commentaires concernant l'évaluation des documents (utilité, qualité, intérêt, etc.) ;
- de sauvegarder outre les résultats de recherche fructueux, les processus réels de recherche ;
- de gérer les différentes versions d'un document.

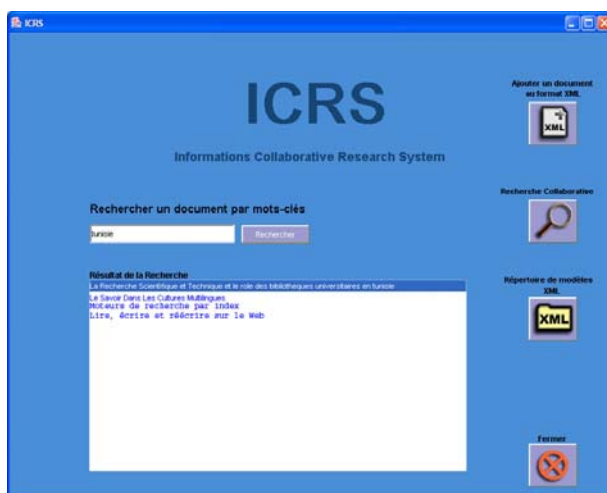


Figure 5 : Interface d'accueil de ICRS

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 à savoir : 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.

L'interface principale de 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.

4-1 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) soumette son document en format XML dans la base de données, un agent du système se charge de l'analyse. Cette analyse a deux objectifs : Le premier est de vérifier la grammaire XML et sa conformité par rapport au modèle utilisé, à travers des « validateurs »³ incorporés dans notre système. Le deuxième objectif le plus important est l'indexation des documents qui se fait, elle aussi, par un autre agent qui parcourt les métadonnées du document pour extraire les éléments qui vont servir après à des clés de recherche. Ces éléments seront stockés dans la base de données avec des liens identifiant les contenus des documents.

4-2 XML : structuration de l'information

Pour répondre à de nouveaux besoins des utilisateurs et permettre des accès à l'information plus sophistiqués et plus aisés, ainsi que pour offrir des possibilités d'échange et de travail collaboratif, nous avons profité de l'apparition, ces dernières années, des normes et des recommandations dans le domaine du Web au niveau de la description et l'échange des données notamment XML ou eXtensible Markup Language. Créé en 1998 par le W3C (World Wide Web Consortium), ce métalangage offre une représentation des documents beaucoup plus riches 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 d'autres qui ont surtout un lien avec 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 disponibilité des 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 que nous allons présenter plus loin dans cet article.

4-3 Interface de recherche collaborative

Avant de parler des différentes fonctionnalités de cette interface [figure 6], nous 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 autre intervention par un autre utilisateur. Les deux autres modes sont :

- 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 et d'évaluer leur pertinence, ainsi que d'accéder aux annotations des autres chercheurs ;
- mode « observateur » : permet à un utilisateur d'effectuer sa recherche tranquillement sans qu'il soit dérangé par quiconque, mais il 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.

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, veuillez consulter le site : <http://www.xmlmind.com/xsdvalid.html>

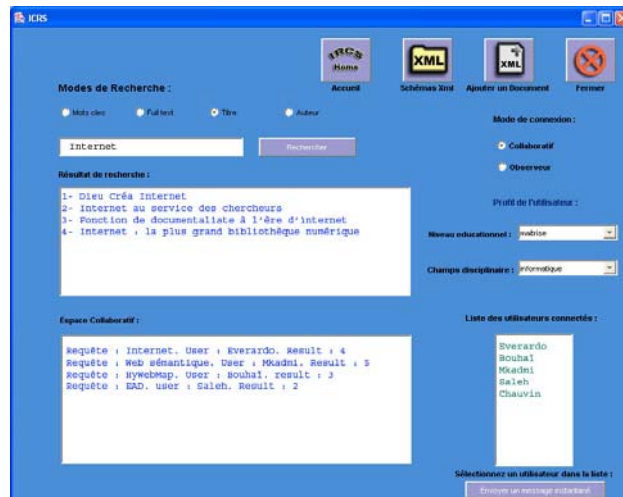


Figure 6 : Interface de recherche collaborative d'informations

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 qu'il soit obligé de refaire la même opération.

Quant aux résultats de recherche, ils sont affichés par ordre de pertinence dans une fenêtre sous forme de liste sensible (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.

4-4 Valeur ajoutée aux documents : Annotations et évaluation

Deux éléments ont été toujours cités par les utilisateurs des ressources numériques pouvant être ajoutés facilement à un texte numérique sont les annotations et l'évaluation des documents. Les annotations ont été proposées fréquemment comme une technique permettant aux utilisateurs d'ajouter un contenu (et partager ainsi des idées) dans des systèmes d'informations. L'évaluation est une indication de l'utilité, de l'intérêt ou de la qualité d'un document vus par un utilisateur. Des évaluations peuvent être approvisionnées 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 à la fois de voir le contenu des documents et s'il y a des 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'écriture initiale ou des propres interprétations des lecteurs du texte lu à travers des « commentaires » mis en regard du document numérique. Ces annotations sont de deux types :

- des annotations publiques que l'utilisateur émette 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, des bibliographies personnelles ou autres qui peuvent donner une idée sur le contenu de document. Ces annotations sont accompagnées par le nom de l'auteur ;

- des annotations privées que seul le « lecteur-auteur » puisse les voir. Elles sont sous forme de points de repères et/ou des indications des parties que le chercheur aimerait bien y revenir, etc.

Quant à l'emploi d'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).

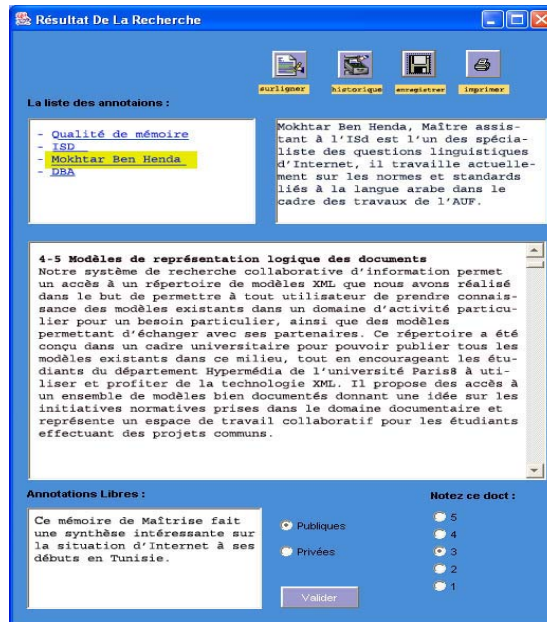


Figure 7: interface de valeur ajoutée au document

4-5 Modèles de représentation logique des documents

Notre système de recherche collaborative d'information permet un accès à un répertoire de modèles XML que nous avons réalisé dans le but de permettre à tout utilisateur de prendre connaissance des 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 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.

4-5-1 Répertoire de Modèles XML

Notre répertoire de modèle est structuré en six parties :

- une partie contenant les éléments de définition du modèle ;
- une partie définissant toutes les informations associées au modèle ;
- un module définissant les critères de recherche ;
- un module pour gérer le contrôle et la révision des modèles proposés à publication ;
- un module pour la consultation des données du répertoire ;
- et un module pour 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 accessible librement [figure 8] à tout le monde et il n'y a aucune limite dans la consultation des schémas et des DTD qui y sont stockés. Quant à la sécurité, il n'y a pour le moment aucune notion de confidentialité au niveau d'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 la mise à jour, elles se font par le propriétaire 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 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 au regard des modèles.

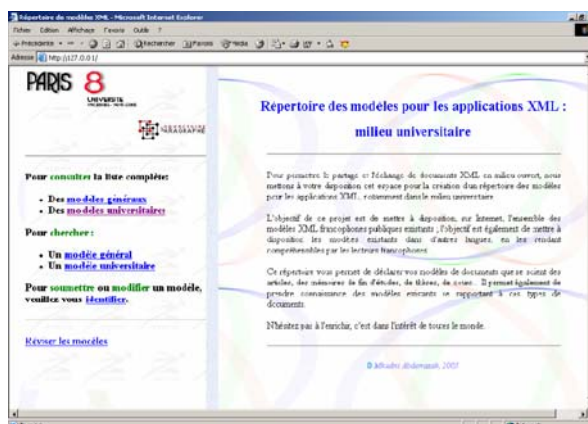


Figure 8 :Interface d'accueil du répertoire de modèles XML

Notre répertoire de modèles XML est un espace ouvert à tous les utilisateurs que ce soit du milieu universitaire ou autres pour déposer leurs modèles de documents et d'en profiter aussi des modèles déjà publiés. Il représente lui aussi un espace de travail de collaboration qui offre une aide à la réalisation des schémas XML tout en encourageant toutes les personnes intéressées à trouver un terrain d'entente et d'échange à travers XML qui s'impose aujourd'hui en tant que format standard de description et d'échange de données.

4-5-1 Modèles XML : description

Nos modèles de documents XML ont été créés suite à un travail de simplification et d'adaptation de plusieurs autres initiatives dans ce domaine (DocBook (OASIS)⁴, DocUpsy (IED, Université Paris8) (Tonin 2002), Cyberthèse (Université Lyon2)⁵). Actuellement un modèle un peu 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].

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

⁵ Cyberthèse est un projet commun de l'université Lumière-Lyon2 et les presses de l'Université de Montréal, elle a pour but de permettre aux auteurs de thèses de mettre en ligne leur pavé de façon simple. Pour plus d'Informations, consultez le site : <http://mirror-fr.cybertheses.org/>

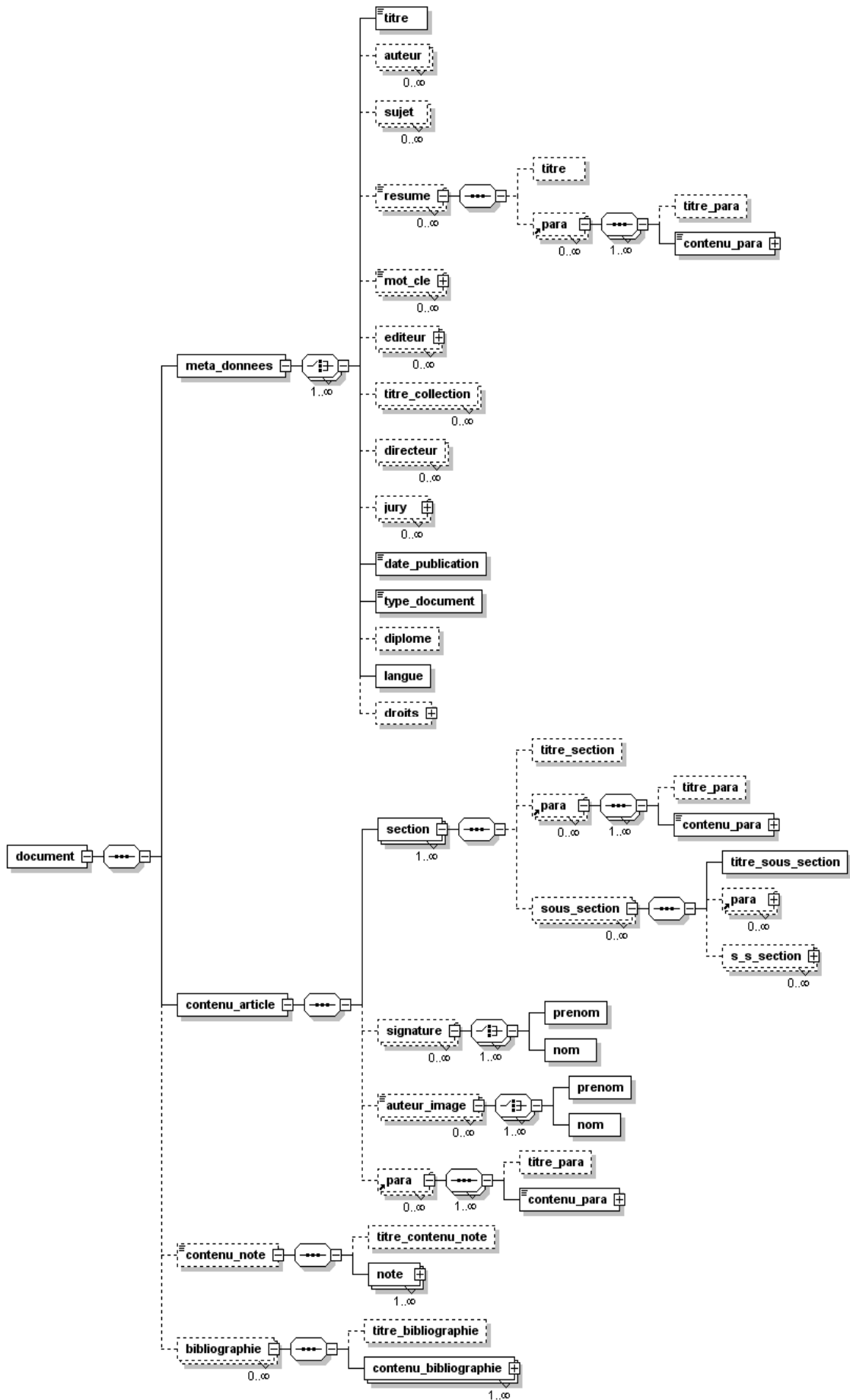


Figure 9 : Modèle de données XML

Ce schéma XML permet de valider tous les documents écrits dans le cadre d'une recherche scientifique au département Hypermédia de l'université Paris8 à savoir des articles scientifiques, des mémoires, des rapports de stages, des thèses, etc., mais bien sûr, il pourra être plus généralisé pour toucher 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 à peu près de 4 parties :

- la première permet de définir les métadonnées du document. Les éléments de ces métadonnées sont inspirés de Dublin Core (DC)⁶ mais reformulés et renommés pour nos besoins spécifiques : (Titre, auteur (nom, prénom, labo de recherche, courriel), revue (titre, volume, pagination...) et date de publication) ;
- la deuxième partie permet de présenter le contenu de l'article (sections, sous sections, chapitres, paragraphes, avec les tableaux, figures, images...) ;
- la troisième partie présente le contenu des notes de bas de pages ;
- la quatrième présente les références bibliographiques.

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

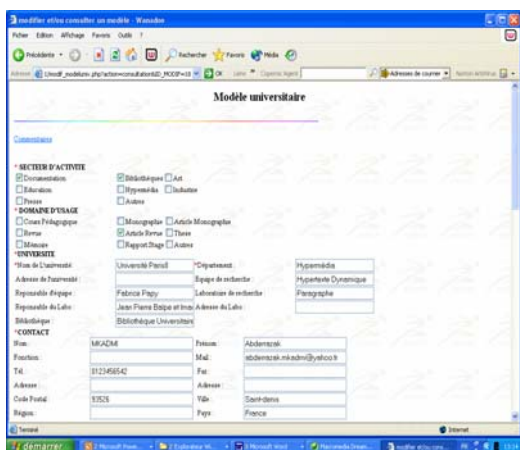


Figure 10 : Présentation des informations associées aux modèles XML

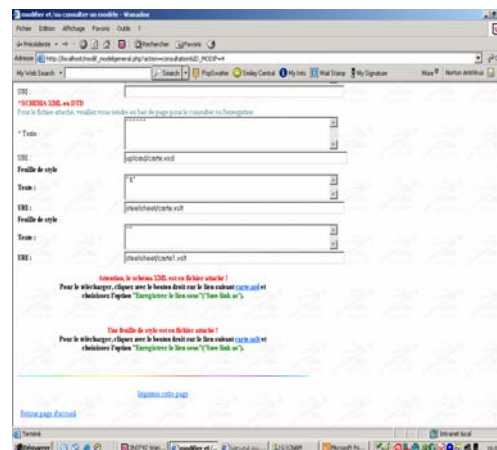


Figure 11 : Présentation des modèles en fichier attaché

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 qu'on peut y accéder de la même manière que le modèle lui-même (en texte ou en lien hypertexte).

5 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 récupérées par les utilisateurs. En effet, tout système de recherche d'information ne peut pas 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 à chaque utilisateur, puisque les préférences de chacun ne sont pas les mêmes pour les autres.

Si dans le Web, les 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.

Dans ce projet, qui n'est que le début d'un travail ambitieux qui veut répondre aux problèmes de la recherche d'information par deux techniques, celle de la structuration des documents et l'utilisation des apports de la technologie XML et celle du travail collaboratif, nous avons réalisé une bonne base du projet à l'aide d'une

⁶ les éléments de métadonnées Dublic Core sont au nombre de 15 : Titre, créateur (auteur), sujet, description, éditeur, contributeur, date, type (de ressource), format, identifiant (URL, ISBN), source, langue, relation (URI ressource liée), couverture, droits.

modélisation basée sur des besoins réels des utilisateurs et réalisée avec des outils qui ont montré leur stabilité et leur efficacité. Nous travaillons à présent sur la :

- création d'un système de navigation hypertextuelle à travers l'ajout d'une fonctionnalité permettant la possibilité d'ajouter un lien hypertextuel sur un mot ou une expression vers d'autres documents traitant le sujet, en profitant d'un autre outil créé au sein du laboratoire Paragraphe et qui s'appelle 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, ou plutôt une structure d'ontologie qui sera enrichie par les différents experts dans la cadre du travail collaboratif, et ce pour évoluer vraiment vers un Web sémantique où les informations seront comprises par les machines.

6 Références Bibliographiques

- (Adobe 2003) 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 2002) BOUHAÏ, N. – 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 Paris8, 2002.
- (Calabretto 1998) CALABRETTO, Sylvie ; PINON, Jean-Marie ; BOZZI, Andrea. – BAMBI : système de gestion de manuscrits anciens pour historiens. In Document numérique, vol. 2, n°3-4, 1998, PP. 32-49.
- (Calabretto 1999) CALABRETTO, Sylvie ; BOZZI, Andrea ; PINON, Jean-Marie. - Numérisation des manuscrits médiévaux : le projet européen BAMB, in : Actes du colloque Vers une nouvelle érudition : numérisation et recherche en histoire du livre, Rencontres Jacques Cartier, Lyon, décembre 1999.
- (Charlier 2000) 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 1999) DAVID, Amos A. - 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, 1999.
- (David 1999a) 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 1999) JACQUESSON, Alain et RIVIER, Alexis. - Bibliothèques et documents numériques : concepts, composantes, techniques et enjeux. Paris : Electre, 1999.
- (Michinov 2003) 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 2003) MKADMI, A. ; BSIR, B. – "Bibliothèques numériques et Travail collaboratif". – in : RIST (Revue d'Information Scientifique & Technique), vol. 13, N°1, 2003.
- (Mkadmi 2003a) MKADMI, A. ; BOUHAÏ, N. et LANGLOIS, M. - Partager des modèles XML : Quel intérêt ?". – in : BBF, N°5, Septembre, 2003.
- (mkadmi 2004) Mkadmi, A. ; REYES, E. ; Bouhaï, N. - "Semantic Network of information". – in : *Actes de la Conférence ICTTA 2004*, Damascus, 19-20 Avril 2004.
- (Nichols 2000) NICHOLS, David M ; PEMBERTON, Duncan ; et al. - DEBORA: Developing an Interface to Support Collaboration in a Digital Library. – in : *Proceedings of the Fourth European*

Conference on Research and Advanced Technology for Digital Libraries (ECDL2000), Lisbon, Portugal. Springer-Verlag. 239-248.

- (Tonin 2002) **TONIN, Cyril**. – DocUpsy : Les réalisations de la Bibliothèque Universitaire de Paris VIII dans le cadre du Projet Upsy, 2002. – <http://www.upsy.net>.
- (Twidale 1997) **TWIDALE, Michael B. ; NICHOLS, David M. and PAICE, Chris D.** - Browsing Is A Collaborative Process. - In : *Information Processing & Management*, 1997, 33(6), 761-83.
- (U. M 1997) **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%C3%A9ateurs1986-2010b/Nathalie1/mm_apcoll_msco.htm.