

Rencontres médias - CNAM



Bibliothèque numérique et Recherche collaborative d'informations

Abderrazak MKADMI

Département Hypermédia (Univ-Paris8)

12 Janvier 2004



1

Cursus Universitaire

- **Maîtrise** en Documentation, Bibliothéconomie et Archivistique, ISD, Tunis, 1996
- **DEA SIC**, ENSSIB (Lyon), 1997
- **Mastère Informatique**, ECP, 2002
- **Doctorant** Université paris8 (Département Hypermédia)
- Actuellement, en poste **PRAG** au département Hypermédia.

2

Publications

- A. Mkadmi, N. Bouhai, I. Saleh. – "Vers des réseaux sémantiques d'informations ". – In : Conférence CoPSTIC, Rabat, 11-13 décembre 2003, pp. 204-209
- A. Mkadmi, N. Bouhaï, M. Langlois. - "Partage des modèles XML : une solution pour les échanges électroniques professionnels". – In JFT'2003, Tours, 30 juin, 1 & 2 juillet 2003, pp.37-46
- A. Mkadmi, N. Bouhaï, M. Langlois. - "Partager des modèles XML : Quel intérêt ?" – In: BBF, N°5, 2003, pp.68-73
- A. Mkadmi. - "Interface utilisateur d'une bibliothèque numérique". Journée d'études "Système d'Informations et Interface", Lyon, 10-11 février 2000, pp. 64-78.
- Autres travaux internes sans comité de lecture portant notamment sur l'usage et les usages des bibliothèques numériques.
- <http://h2ptm.univ-paris8.fr/> ; <http://perso.wanadoo.fr/mkadmi>

3

Plan de l'exposé

- Introduction générale
- Problématique
- Méthodologie de Recherche
- Résultats primaires des enquêtes de terrain
- SRCI
 - Répertoire de modèles XML
 - Documents XML
 - Recherche collaborative d'information
- évaluation des interfaces des B.N.s
- Conclusion et suite du travail

4

État d'avancement (1)

- Après avoir défini une problématique de départ, un travail théorique autour de :
 - Systèmes de recherche d'informations : définition, rôle, typologie et problèmes d'évaluation ;
 - Bibliothèque numérique : objectifs et impacts sur les manières de s'informer ;
 - Interface utilisateur : rôles, objectifs et méthodes d'évaluation.

l'intérêt s'est porté sur l'usage de ces bibliothèques, leur capacité de maintenir un contexte social des utilisateurs (Travail collaboratif) et l'évaluation de leurs interfaces utilisateur.

5

État d'avancement (2)

Travaux réalisés

- État de l'art sur le sujet
- Définition de la problématique et de la méthodologie
- Enquêtes auprès des responsables et des utilisateurs des bibliothèques françaises sur les projets de numérisation, les ressources numériques et les activités de collaboration des chercheurs dans le cadre de leurs recherches.
- Réalisation d'un répertoire de modèles XML
- Conception d'une maquette d'un SRCI

Travaux en cours

- Finalisation de la maquette du SRCI
- Réalisation d'une grille d'évaluation des nouveaux SRIs

6

Constats (1)

Avec la récente explosion du Web, des bases de données multimédias, 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 peu structurés.

7

Constats (2)

Malgré le développement des SRIs et les technologies liées à la recherche d'informations, l'utilisateur ne peut toujours pas exprimer précisément son but de recherche et le producteur ne peut toujours pas présenter l'information d'une manière adéquate.

8

Constats (3)

L'arrivée du document numérique a changé au moins deux constantes :

Le lieu de recherche de documents :

il n'est plus l'endroit géographique où le document est localisé, mais plutôt le réseau

La manière d'accès aux documents :

l'utilisateur se trouve seul face à un écran :
- Un avantage : il se situe tout seul dans le niveau de connaissances qu'il a dans le domaine, etc.
- Un inconvénient : *pas d'assistance humaine*

9

Constats (4)

Deux usages au moins peuvent être saisis par l'ensemble des établissements :

- les services pouvant être proposés en ligne à tout un ensemble de publics «*empêchés*» par un handicap physique, par la maladie ou par toute autre forme d'éloignement social ou géographique ;
- les possibilités de dialogue en ligne avec les usagers qui peuvent ainsi s'associer à la vie de leur établissement, par exemple en partageant leurs impressions de lecture ou leurs informations documentaires avec l'ensemble des lecteurs. Il y a là une bonne occasion de renouveler et d'améliorer le lien entre un service public et ses usagers ; C'est ce deuxième type d'usage qui nous intéresse dans notre travail et que nous appellen le travail collaboratif.

10

Propositions

Pour maîtriser la recherche d'informations, il faut :
- une **structuration** et/ou restructuration de l'information

- un « **intermédiaire** » entre le consommateur et le producteur de l'information permet de mieux insérer la bibliothèque dans son environnement scientifique, intellectuel et social.

→ le besoin de développer des moyens qui permettent aux différents utilisateurs de **collaborer** au niveau de leurs recherches

11

Maîtriser la RI : par quels moyens ?

Ce contexte suscite l'exploitation :

- de l'apport des langages **XML** pour évoluer vers le Web sémantique (XML : description des documents structurés sur le Web, RDF pour la description des ressources accessibles sur le Web, XML-Query pour des recherches basées sur la structure des documents XML, Xlink pour la description de liens hypertextes sophistiqués, etc.);
- des ressources terminologiques et conceptuelles (**ontologies**) et leur structuration pour l'usage du Web ;
- de l'**annotation sémantique** des ressources documentaires ;
- des **profils utilisateurs** ;
- des outils et moyens du **travail collaboratif** entre les utilisateurs.

12

Problématique

- (1) Comment peut-on organiser les ressources numériques pour pouvoir en profiter au maximum ?
- (2) Jusqu'où les bibliothèques numériques développent et créent des occasions de collaborations entre les utilisateurs et par conséquence fallait-il repenser l'architecture des SRIs ?
- (3) Quels types de mesures sont plus appropriés pour évaluer la performance de ces nouveaux SRIs ?

13

Méthodologie de Recherche

- Pour répondre à notre problématique, nous avons opté pour la méthodologie suivante :
1. Mener des enquêtes auprès des utilisateurs des ressources numériques portant sur leurs activités collaboratives au niveau de la recherche et leurs attentes au niveau de nouvelles fonctionnalités des SRIs
 2. Réaliser une maquette d'un SRI répondant à la fois aux problèmes de structuration de l'information et aux besoins des utilisateurs en matière de collaboration.
 3. Établir un modèle d'évaluation de ces types d'activités et des interfaces utilisateurs de bibliothèques numériques.

14

Bibliothèque numérique ?

- Plusieurs définitions qui renvoient toutes à la notion de virtualité : mise en place d'une nouvelle conception de l'encyclopédisme et d'un projet de rassemblement et de mise en circulation des savoirs.
- Pour les bibliothécaires : il s'agit d'une bibliothèque qui a les mêmes buts et les mêmes fonctions que la bibliothèque traditionnelle, sauf qu'elle a besoin de nouveaux outils pour accomplir sa mission.

15

Fondements de la BN

- L'informatisation en bibliothèque
- Les services de télécommunication
- Les outils à disposition du lecteur (outils de recherche, d'accès, de transfert, de visualisation, etc.)

16

Terrain d'enquêtes

Deux enquêtes :

- **La première enquête** touche, les responsables du réseau de la BnF et ses pôles associés, ainsi que quelques autres bibliothèques numériques repérées sur Internet (48 Bibliothèques).
- **La deuxième enquête** s'adresse aux utilisateurs de ressources numériques dans quatre différentes bibliothèques lyonnaises à savoir : la BML, 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 (une centaine de réponses).

17

Résultats primaires des enquêtes(1)

- Presque toutes les bibliothèques ont confirmé leur présence sur Internet et le document numérique y est devenu omniprésent ;
- La collaboration entre les bibliothécaires est devenue plus riche (locale et distante) ;
- Tous les utilisateurs des documents numériques confirment leurs besoins de collaboration avec les autres ;
- Le moyen le plus utilisé actuellement pour la collaboration à distance est la messagerie électronique ;

18

Résultats primaires (2)

- Les utilisateurs ont besoin d'autres outils qui soutiennent leur collaboration
- Confirmation de la plupart des utilisateurs de leurs besoins d'un système qui permet :
 - l'annotation des documents ;
 - le partage de ces annotations ;
 - le partage et la réutilisation de la connaissance du processus d'information ;
 - l'accès facile à l'auteur de l'information ou au bibliothécaire qui l'a déposée.

19

Résultats primaires (3)

La collaboration dans une B.N. peut être :

- avant et pendant la recherche (courriel et aides en ligne...)
- au niveau du produit de recherche (les listes de diffusion, l'enregistrement automatique des résultats de recherche...);
- au niveau du processus de recherche (enregistrement des processus de recherche...);
- au niveau de la valeur ajoutée aux documents (bibliographies personnelles, liens, annotations, notation, etc.).

20

Partie pratique : développement

- Répertoire de schémas XML :
 - Structuration de l'information
 - Modèles XML
 - Pourquoi un tel répertoire ?
 - Structuration de l'application et des données
 - Principes et fonctionnalités
- SRCI
 - Fonctionnalités
 - Recherche collaborative : comment et pourquoi ?

21

Structuration de l'information : historique

- 1986
 - SGML (*Standard Generalized Markup Language*)
 - Norme ISO:8879:1986
- 1987
 - TEI (*Text Encoding Initiative*)
- 1990
 - HTML 1.0 (*HyperText Markup Language*)
- 1997/1998
 - XML 1.0 (*eXtensible Markup Language*)

22

XML : qu'est ce que c'est ?

- Une Recommandation du W3C en date du 10 février 1998, référencée à :
<http://www.W3.org/TR/1998/REC-xml-19980210>
- Un Langage de Balisage Extensible
- Un méta-langage
 - Permet de définir d'autres langages
- Une Simplification de la Norme SGML
 - ISO 8879:1986
 - Structure logique des documents électroniques

23

XML : Qu'est ce que ce n'est pas ?

- Surtout pas une nouvelle version de HTML
 - HTML est un ensemble de balises figé, pour afficher du texte avec des liens hypertextes et des images
 - XML ne rend pas HTML obsolète (ni SGML !)
 - HTML devrait être compatible XML (XHTML)
- XML n'a pas de balise pré-définie et permet aux créateurs de spécifier leur propre jeu de balise pour structurer leurs données

24

XML : Apports

- La **lisibilité** : aucune connaissance ne doit théoriquement être nécessaire pour comprendre un contenu d'un document XML
- **Autodescriptif et extensible** ;
- **Une structure arborescente** : permettant de modéliser la majorité des problèmes informatiques ;
- **Universalité et portabilité** : les différents jeux de caractères sont pris en compte ;
- **Déployable** : il peut être facilement distribué par n'importe quel protocole à même de transporter du texte, comme HTTP ;
- **Intégrabilité** : un document XML est utilisable par toute application pourvue d'un parser (c'est-à-dire un logiciel permettant d'analyser un code XML) ;
- **Extensibilité** : un document XML doit pouvoir être utilisable dans tous les domaines d'applications...

25

Structuration de l'information : XML(1)

XML : eXtensible Markup Language

- Format d'échange de documents ;
- Répond au besoin d'interaction et de coopération entre des S. Informatiques hétérogènes ;
- Utilisé par tous : fournisseurs d'ERP, de bases de données, éditeurs de middleware, etc.
- ? **Nécessité de partage des Docts. Structurés et des Infos. associées**

26

Structuration de l'information : XML(2)

XML permet :

- affichage de données en langage humain,
- conversion d'une structure de données en une autre structure...

Mais,

- Accord entre les acteurs sur un formalisme de structuration des Infos. selon un modèle.
- Nécessité de partage de ces modèles

• ? **Notre projet : Création du Répertoire de Modèles XML**

27

Modèle de document XML

- est une structure permettant de donner les règles d'assemblage des données.
- définit tous les éléments utilisés dans un doct.
- définit les relations et l'ordonnancement entre ces éléments.
- peut être créé selon les deux recommandations (**DTD & schémas XML**).

28

DTD vs XML Schema

DTD

- N'est pas écrite en langage XML
- Typage de données très limité
- Ne supporte pas les espaces de noms
- Ne permet de définir un nombre précis d'occurrences (sauf *, ?, +)

Schéma XML

- écrit en langage XML
- Typage très riche de données (Booléens, entiers, intervalles de temps...)
- Possibilité de créer de nouveaux types de données
- Supporte des espaces de noms
- Notion d'héritage : les éléments peuvent hériter du contenu et des attributs d'un autre élément.

29

Répertoire de modèles XML : Pourquoi ?

- Permet à tout utilisateur de prendre connaissance des modèles existants dans un domaine particulier pour un besoin particulier;
- Permet de donner une idée sur l'organisation et le contenu des modèles ;
- Favorise l'échange ouvert;
- Offre un service d'aide à l'élaboration des schémas XML.

30

Répertoire de modèles XML : état de l'art (1)

OASIS registry/repository :

- Forum indépendant des éditeurs pour soumettre publiquement, publier et échanger des spécifications et vocabulaires XML. Il contient des objets enregistrés et entrées de registre
- Un pas très intéressant dans l'orientation vers l'utilisation du langage XML.

Mais : Anglophone, très complexe : utilisation difficile pour les non spécialistes

Le répertoire de l'ATICA

- Concerne le domaine de l'administration
- Publie les schémas et DTD issus des domaines documentaires et juridique
- **Intérêts :** Dématerrialisation et interopérabilité des échanges au sein des administrations

Mais : Non précision de comment les modèles sont présentés dans le répertoire, ni quels sont les documents qui doivent accompagner ces modèles.

31

Répertoire de modèles XML : état de l'art (2)

DoCBook

- modèle de documentation technique
- maintenu par « [DocBook Technical Committee](#) » (comité technique du consortium OASIS) :
- créer et d'échanger des documents techniques ou des articles grâce à une sémantique plutôt généraliste, qui la rend utilisable dans toute l'industrie de l'électronique et de l'informatique.

Mais

- étant conçu au départ pour gérer du SGML, il est jugé trop difficile à paramétriser et utiliser : Il y a aujourd'hui une version simplifiée de ce modèle dénommée « *Simplified DocBook* », mais ...

32

Répertoire de modèles XML : état de l'art (3)

Ces exemples de projets (qui ne sont pas exclusifs) montrent bien l'intérêt de la création d'un répertoire de modèles XML. En effet, le fait de partager des modèles participe à l'évidence à l'acceptation de ceux-ci et représente, de plus, un facteur de montée en compétences des organisations confrontées à l'explosion des applications XML. Plus les modèles sont accessibles de manière facile, plus l'échange entre partenaires devient aisé.

33

Répertoire de modèles : principes de base

- Accès libre aux modèles XML
- La publication des modèles passe par un comité de révision
- Seul le propriétaire du modèle peut le modifier
- Application ouverte et évolutive
- ni le logiciel serveur, ni le logiciel client ne doit être d'une technologie propriétaire

34

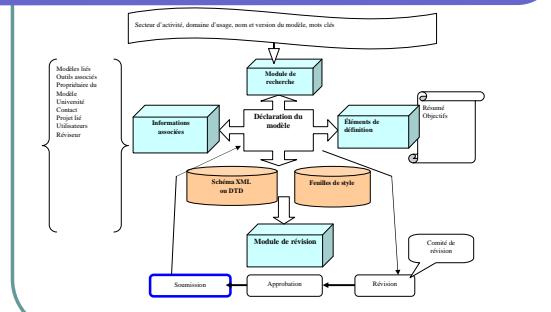
Répertoire de modèles : structure de l'application (1)

Notre répertoire de modèle est structuré en six parties :

- les éléments de définition du modèle ;
- toutes les informations associées au modèle ;
- 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.

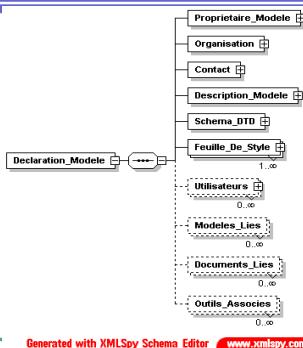
35

Répertoire de modèles : structure de l'application (2)



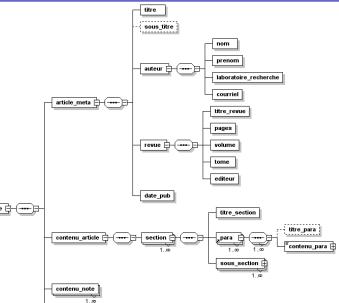
36

Structure de l'application



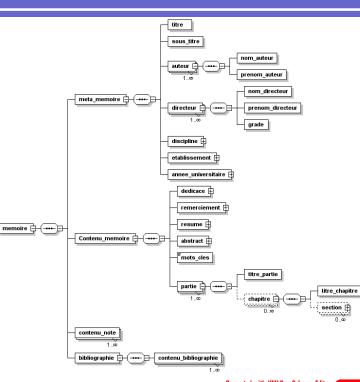
37

Structuration des données : schéma d'un article scientifique



38

Structuration des données : schéma d'un mémoire



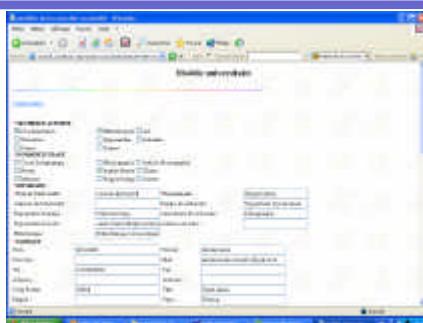
39

Répertoire de modèles : interface et fonctionnalités



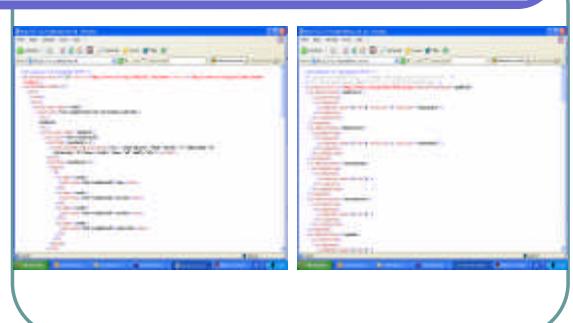
40

Répertoire de modèles : saisie et consultation



41

Consultation : Schéma XML et Feuille de style



42

Systèmes de recherche collaborative

- **ARIADNE** (Annotatable Retrieval Information And Database Navigation Environment) :
 - apprentissage collaboratif (catalogue des bibliothèques)
- **BAMBI** (Better Access to Manuscripts and Browsing of Images) :
 - accès à des manuscrits (annotations, transcription, indexation)
- **DEBORA** (Digital accEss to Books Of RenAissance) :
 - accès à des documents du 16ème siècle : annotations, livres virtuels)

43

ICRS : *Information Collaborative Research System*

Notre système répond par l'ensemble de ses fonctionnalités 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 qui permettra à tout utilisateur de faire des recherches et d'accéder à un ensemble de documents répondant à son besoin, ainsi qu'aux différents commentaires associés à ces documents.

44

Outils utilisés

- XML : langage de description des données
- Apache : serveur Web
- Tomcat : moteur de servlet
- Cocoon : transformateur XSLT
- MySQL : base de données relationnelle
- Java : langage de programmation

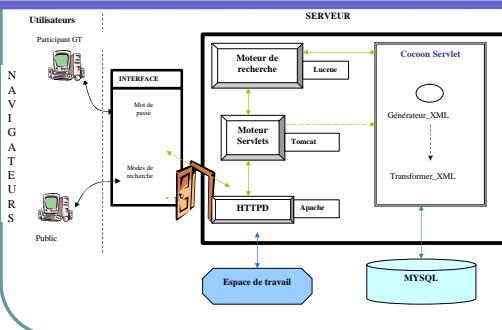
45

ICRS : Principes

- les utilisateurs et les informations peuvent être visualisés :
- 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 ;
- ajouter des informations aux documents (bibliographies personnelles, liens entre documents) ;
- annoter les documents en ajoutant un contenu dans le système d'informations (et partager ainsi des idées) ;
- ajouter des commentaires concernant l'évaluation des documents (utilité, qualité, intérêt, etc.) ;
- sauvegarder outre les résultats de recherche fructueux, les processus réels de recherche ;
- gérer les différentes versions d'un document.

46

ICRS : Architecture



47

ICRS : Fonctionnalités

Interface d'accueil du système :

- Recherche simple pour les utilisateurs qui ne veulent pas être en collaboration
- Accès au RM_XML
- Accès à une interface de recherche collaborative
- Proposer des documents pour la base de données (édition des documents en format XML)



48

Recherche collaborative

- Définir un profil utilisateur (niveau et discipline)
- Définir un mode de collaboration (collaboration, observation)
- Recherche par mots clés, auteur, titre, full-text
- Avoir la liste des personnes connectées au système
- Avoir la liste des requêtes proposées au système avec le Nb de résultats (avec possibilités de récupérer la requête et ses résultats)
- Contacter un utilisateur par un message instantané
- Avoir accès au RM_XML



49

Valeur ajoutée au document

Cette interface permet de visualiser le contenu d'un document sélectionné, mais elle permet aussi de :

- Consulter les annotations publiques associées à ce document
- Annoter un document
- Rendre les annotations publiques ou privées
- Noter un document (de 1 à 5)



50

Évaluation des interfaces utilisateurs : évaluation heuristique

- L'apprentissage
- L'efficacité
- La mémorisation
- Le taux d'erreurs
- La satisfaction subjective de l'utilisateur

(Nielsen et Molich, 1990; Nielsen 1994)

51

Évaluation d'une interface d'une bibliothèque numérique (1)

L'évaluation d'une interface d'une BN doit prendre en compte le contexte social de la recherche d'information (notamment la collaboration avec les autres) :

- la sensibilisation d'autres utilisateurs (travail sur le même document, etc.) ;
- La communication entre les différents acteurs du SRI et ses moyens (messagerie, FTP, annotations, liens personnels, etc.)
- L'appropriation des documents (copier, coller, imprimer, constituer des bibliothèques personnelles, etc.)

52

Évaluation d'une interface d'une bibliothèque numérique (2)

- Apprentissage collaboratif
- Efficacité collaborative
- Mémorisation collaborative
- Erreurs collaboratives
- Satisfaction subjective collaborative (degré de satisfaction des différents utilisateurs lors de la réalisation des tâches collaboratives) ;
- Coordination (capacité offerte par l'interface pour que les utilisateurs puissent se coordonner ensemble afin de réaliser leurs tâches communes) ;
- Support de communication (capacité offerte par l'interface pour supporter les communications entre les différents utilisateurs du groupe. Ceci est très important car la communication est l'élément de base de tout travail collaboratif).

53

Conclusion (1)

- Les bibliothèques numériques représentent des nouveaux SRI qui doivent prendre en compte les activités collaboratives des utilisateurs :

C'est ce qui devrait naturellement permettre la numérisation, qui va au-delà de la seule valorisation de fonds, et qu'il est essentiel qu'elle soit associée à la communauté des chercheurs et permet de mieux insérer la bibliothèque dans son environnement scientifique, intellectuel et social. Plusieurs opportunités peuvent être saisies par les bibliothèques au bénéfice de leurs différents publics.

54

Conclusion (2)

- La recherche collaborative des informations représente une solution pour répondre aux problèmes des SRLs
- Notre projet essaie de répondre aux problèmes de la recherche d'information par deux techniques :
 - Une structuration des documents : XML ;
 - Des moyens pour soutenir les activités collaboratives des utilisateurs

55

Perspectives

- Création d'un système de navigation hypertextuelle entre les documents ;
- l'ajout des outils graphiques qui permettent de surligner un terme dans le document, de l'entourer selon son importance et/ou d'ajouter éventuellement des commentaires sur les marges ;
- l'installation d'un moteur de recherche plus puissant qui permet une recherche full-text;
- Création d'une base de connaissances (ontologies)

56

Des questions ?

Amkadmi.Abderrazak@univ-paris8.fr

Département Hypermédia
Université paris8

57