
Un modèle d'une bibliothèque numérique collaborative – ARMARIUS

Reim Doumat, Elöd Egyed-Zsigmond, Jean-Marie Pinon

*LIRIS – INSA de Lyon
7avenue Jean Capelle
69100 Villeurbanne FRANCE
{reim.doumat, elod.egyed-zsigmond,jean-marie.pinon}@liris-cnrs.fr*

RÉSUMÉ. Le nombre de bibliothèques numériques augmente rapidement sur internet et en conséquence le volume de collections et documents disponibles en ligne devient énorme. Les manuscrits anciens numérisés constituent un contenu spécifique pour ces bibliothèques. Les internautes et les utilisateurs professionnels travaillant sur ce type de documents ont besoin des systèmes d'assistance et des espaces de travail collectif pour interpréter, annoter et transcrire ces manuscrits.

Dans cet article, nous présenterons un modèle de bibliothèque numérique spécialement conçu pour des manuscrits anciens numérisés : Armarius. Celui-ci fournit des interfaces d'annotation manuelle et semi-automatique et propose un système d'assistance pour aider l'utilisateur à annoter et exploiter les manuscrits. De plus, il contient un espace de travail collaboratif qui permet à un groupe d'utilisateurs de travailler sur un ensemble de documents.

ABSTRACT. The number of digital libraries increases rapidly on internet; consequently the volume of the collections and the documents that are available on line is enormous. The digitized ancient manuscripts are specific content for these libraries. Users working on this type of documents need assistant systems and a collective work spaces to interpret, annotate and transcript these manuscripts.

In this paper, we present a digital library model specially designed for digitized ancients' manuscripts: Armarius. It provides manual and semi-automatic annotation interfaces and proposes an assistant system to help users in annotating and consulting the manuscripts. Moreover, it contains a collaborative work space that permits user groups to work on a set of documents.

MOTS-CLÉS : modèle collaboratif, bibliothèque numérique, assistant d'annotation, manuscrits numérisés.

KEYWORDS: collaborative model, digital library, annotation assistant, digitized manuscripts.

1. Introduction

Internet regroupe, aujourd'hui, une grande quantité d'informations représentées sous diverses formes (documents texte, journaux, images, vidéos...). Ces informations sont publiées par des bibliothèques numériques qui proposent la consultation à distance des versions électroniques répondant à la demande des chercheurs, des étudiants et des internautes. Il existe actuellement plusieurs types de bibliothèques numériques telles que : les bibliothèques d'héritage culturel comme Gallica¹ de la bibliothèque nationale de France, Toyo Bunko² Rare Books (Kiamoto, 2006), l'université de Southern California³ (Hunt *et al.*, 2005), les bibliothèques scientifiques (BND⁴ (Bibliothèque Numérique Nationale), BVH⁵ bibliothèque développée par CESR (Centre d'Etudes Supérieures de la Renaissance) (Ramel, 2006), PERSEE⁶ (un site web pour les journaux scientifiques dans les sciences sociales et humaines) (Morandiere *et Bouletreau*, 2006), etc.), pour ne parler que de quelques initiatives.

Avec l'augmentation de ces sources d'information, la gestion et l'exploitation des collections volumineuses est devenue fastidieuse. En effet, les utilisateurs ont besoin des systèmes d'assistance pour consulter et extraire des informations intéressantes de ces collections (Ignat *et al.*, 2006) (Mothe *et al.*, 2003) . D'où la nécessité de concevoir une bibliothèque numérique intégrant des outils d'assistance, de traçage (Tamine *et al.*, 2006) et de personnalisation (Chevalier *et al.*, 2007) afin de répondre aux besoins des utilisateurs. En effet, ces outils facilitent la récupération, la modification, l'organisation et l'enrichissement du contenu (Amous *et al.*, 2005). Avec le développement des systèmes d'information collaboratifs (Nguyen, 2006), les utilisateurs de la bibliothèque numérique sollicitent un environnement de travail collaboratif, parce que la recherche scientifique nécessite le travail de plusieurs personnes avec des compétences différentes qui doivent coopérer pour partager leurs connaissances. Nous avons réalisé une étude des bibliothèques numériques de 45 pays Européens (Doumat *et al.*, 2007) pour synthétiser les besoins d'une architecture idéale du site web des bibliothèques numériques.

Dans cet article, nous présentons un modèle de bibliothèque numérique conçu pour des manuscrits anciens. Notre objectif est de créer un modèle d'archive vivante en ligne. Cette archive intègre différents rôles d'utilisateurs pour développer et exploiter le contenu et offre un environnement de travail personnalisé, assisté et collaboratif qui facilite l'échange d'informations entre les utilisateurs.

1. <http://gallica.bnf.fr>

2. <http://dsr.nii.ac.jp/toyobunko/>

3. <http://www.inscriptifact.com>

4. <http://bnd.bn.pt>

5. <http://www.cesr.univ-tours.fr>

6. <http://www.persee.fr>

Le présent article est organisé comme suit : dans le chapitre suivant, nous exposerons un état de l'art des modèles des bibliothèques numériques et les travaux existants concernant des bibliothèques numériques des manuscrits et des objets anciens, nous montrerons aussi les limites de ces bibliothèques. Dans la partie 3, nous proposons notre modèle *Armarius* qui représente une archive vivante pour les manuscrits numérisés. Nous terminerons finalement par une conclusion et quelques perspectives.

2. Etat de l'art

Les bibliothèques numériques sont des systèmes d'information qui stockent et gèrent les documents numériques et les documents numérisés. Ces bibliothèques sont classées parmi les systèmes d'informations les plus complexes et avancés, elles doivent offrir des services de bonne qualité (recherche, personnalisation, ...). D'un point de vu informatique, la bibliothèque numérique est constituée de composants simples comme les objets numériques. L'objet numérique peut être un document ou un méta donné.

Cependant les informations dans la bibliothèque numérique ne sont pas simples. En effet, un document peut avoir plusieurs parties, une structure interne complexe et des relations avec d'autres documents. De plus, l'organisation flexible de ces informations (documents et méta données) est un des challenges de la conception pour la création d'une bibliothèque numérique.

2.1. Modèles de bibliothèques numériques

Les modèles de bibliothèques numériques concernent la structure qui rassemble les documents numérisés et leurs annotations le plus souvent dans une base de données. Ce modèle fournit des services en ligne pour la recherche et l'annotation (automatique, assistée, manuelle...) (Pouliquen *et al.*, 2006)(Savoy, 2005), des services de personnalisation qui facilite l'expression des besoins des utilisateurs (Bouzeghoub *et Kostadinov*, 2005). Cette structure contrôle des droits d'utilisation des documents et gère les fonctionnalités du site web qui représente la bibliothèque numérique. La Figure 1 représente un schéma classique de gestion d'une bibliothèque numérique (Le Bourgeois *et al.*).

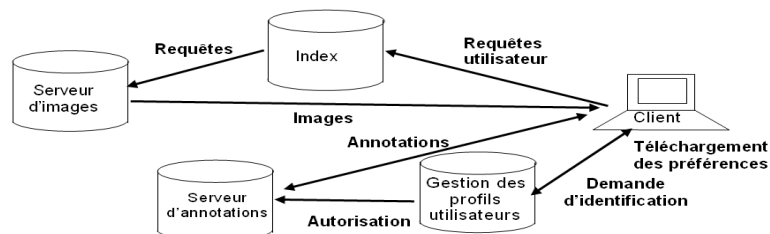


Figure 1. Schéma classique de gestion d'une bibliothèque numérique avec une base de données relationnelle

Un modèle de stockage des images, des manuscrits et leurs annotations ou leurs transcriptions est nécessaire pour la gestion de données. Malheureusement, il n'y a pas de standard pour la représentation de la structure de la base de données correspondante à une bibliothèque numérique centrée images numérisés, malgré le grand nombre des bibliothèques numériques existantes. Les modèles varient selon l'objectif de la bibliothèque, les types de documents gérés et le type d'utilisateurs (scientifique, éducatif, historique, archéologique...). Le modèle contient le plus souvent des répertoires pour stocker les objets numériques, des interfaces d'accès aux représentations des documents, souvent un système de recherche d'informations, un système de gestion (stockage, traitement et indexation des nouveaux objets), un système de sécurité et des fois un système de gestion des interfaces utilisateurs.

Un modèle de bibliothèque numérique est proposé par le 5S (Streams, Structures, Spaces, Scenarios, Societies) qui représentent Flux, Structures, Espaces, Scenarios, Sociétés (Gonçalves *et al.*, 2004). Ces cinq dimensions sont cruciales pour toutes les bibliothèques numériques. Les flux représentent la séquence des éléments (ex. le flot des caractères), les structures concernent l'organisation des informations (ex. les relations entre les acteurs, les connections du système), les espaces figurent le contenu de la bibliothèque numérique d'une façon utilisable et récupérable (ex. une interface d'une base de donnée ou un navigateur pour accéder aux objets), les scénarios sont les comportements des services dans la bibliothèque numérique et ils expliquent les fonctionnalités des structures et des espaces (ex. chercher des objets), les sociétés se concentrent sur les acteurs qui sont concernés par les fonctionnalités de la bibliothèque numérique (ex. des utilisateurs, des fournisseurs...). Selon le modèle 5S, la bibliothèque numérique (DL) est composée d'un 4-uplet : le répertoire (R), un groupe des catalogues de méta donnée (Cat), un groupe des services (Serv) et la société (Soc) : DL= (R, Cat, Serv, Soc).

Le projet DELOS est un réseau d'excellence sur les bibliothèques numériques supporté par la commission européenne (DELOS). Un des objectifs du projet est de moduler une architecture d'une bibliothèque numérique au niveau supérieur (un modèle conceptuel). Le projet propose un modèle pour les éléments fondamentaux

(qui ne changent pas) dans la bibliothèque numérique. Il considère la donnée comme unité essentielle. Il existe huit types de données dans la bibliothèque numérique (utilisateurs, fournisseurs, documents, méta données descriptives, services, commandes, livraisons qui fournissent aux utilisateurs des services ou des documents, comptes) (Formalizing the Design of Digital Libraries Based on UML, 2006). Le modèle de DELOS est basé sur le modèle de 5S mais il détaille les sociétés en utilisateurs et fournisseurs. Dans le cadre de notre travail, les fournisseurs des documents sont toujours les administrateurs du système et pas les utilisateurs. Dans DELOS les comptes remplacent les espaces du modèle 5S.

Nous avons basé notre modèle sur le modèle 5S, cependant nous l'avons modifié selon notre besoin pour permettre à un groupe d'utilisateurs de travailler en collaboration sur un ensemble de documents qui les intéressent.

2.2. Projets de bibliothèques numériques pour des manuscrits

Il y a plusieurs projets qui ont proposé des modèles de bibliothèque numérique. Dans ce chapitre nous citons quelques uns (des projets de consultation des manuscrits anciens et des objets historiques) :

2.2.1. BAMBI

BAMBI (Better Access to Manuscripts and Browsing of Images)(Calabretto *et al.*, 1998). C'est un projet européen sur la transcription et les annotations manuelles des anciens textes scannés, il traite les textes latin et grec. BAMBI contient un espace de partage des manuscrits entre plusieurs utilisateurs. Cependant la limite essentielle de BAMBI est qu'il est locale et que son architecture ne fonctionne pas sur l'internet.

2.2.2. DEBORA

Le projet DEBORA (Digital AccEss to BOoks of theRenAissance)(Le Bourgeois *et al.*, 2001), est un projet européen qui intéresse à numériser les livres rares du XVI siècle et les mettre sur internet. Ce projet permet aux différents types d'utilisateurs de participer au développement du système d'exploitation des livres numérisés, avec un accès à distance. Le système a des outils de traitement d'images développés qui permettent d'extraire des métadonnées, de compresser les images numérisés, et de séparer le texte et les figures du fond du document, de plus le système assiste la transcription du document. Néanmoins, ce projet n'a pas un système de personnalisation.

2.2.3. ETANA-DL

ETANA-DL est une bibliothèque numérique archéologique (Ravindranathan *et al.*, 2004) qui traite les données archéologiques hétérogènes. Cette bibliothèque est

un modèle de base extensible qui gère les ressources complexes des informations en utilisant le modèle S5. Le modèle d'ETANA-DL facilite la recherche des objets archéologiques aux différents types d'utilisateurs. Ce projet a des interfaces simples et conviviales. De plus il offre un espace personnel à l'utilisateur pour marquer ces collections. Il permet à l'utilisateur de sauvegarder le chemin et les résultats de sa recherche, cependant le système ne garde pas des traces du travail de l'utilisateur sur ces collections.

Les applications existantes de bibliothèques numériques sont différentes et elles ne sont pas adaptables à tous les types des documents. Dans le cadre de notre travail, les documents sont principalement des images scannées des manuscrits écrits à la main, en plusieurs langues, rendant difficile l'utilisation des applications de reconnaissance de caractères (OCR). De plus les documents ne sont pas structurés en XML. Pour cette raison les manuscrits doivent être transcrits et annotés manuellement. Ce travail est fastidieux pour une personne, nous avons donc besoin d'un espace de travail collaboratif, cet espace permet à un groupe d'utilisateurs qui ont des expériences différentes de travailler sur le même document.

2.3. Limites des modèles et projets des bibliothèques de manuscrits numérisés

Pendant notre étude sur les bibliothèques numériques, nous avons dégagé les problèmes suivants de gestion des profils utilisateurs :

- la plupart des projets d'annotation de manuscrits anciens ne gardent pas les traces des actions des utilisateurs. Le stockage de ces traces dans la base de données avec les manuscrits facilite le travail des utilisateurs, car ces traces permettent à l'utilisateur de continuer ces activités dans la bibliothèque (recherche, annotation,...) à partir du travail qu'il a déjà fait. De plus, les traces d'utilisateurs peuvent être utilisés dans un système d'assistance;
- beaucoup de bibliothèques numériques n'ont pas d'espace personnel pour les utilisateurs réguliers de la bibliothèque. Cet espace permet à l'utilisateur de garder les collections sélectionnées et leurs recherches;
- la majorité des projets des bibliothèques numériques ne contiennent pas d'espace du travail collaboratif sur les documents, cet espace collaboratif faciliterait la communication entre différents utilisateurs pour la réalisation d'un travail difficile à faire par une seule personne : i.e. confrontation des points de vues, des corrections faites par d'autres utilisateurs sur ces annotations;

Nous décrivons dans la section suivante, un système de bibliothèque numérique pour la gestion de documents anciens numérisés. Ce système permet de gérer aussi bien les annotations, les transcriptions, les utilisateurs, les groupes de la bibliothèque que leur droits d'accès, leurs interactions, leurs préférences. Notre modèle contient un système de traçage, un système d'assistance et un système de travail collaboratif pour les utilisateurs professionnels et les experts.

3. Le modèle d'une bibliothèque numérique collaborative pour des manuscrits anciens (Armarius)

Le projet Armarius est un système de gestion de documents en ligne en vue de fournir une « archive vivante ». Il vise la mise en ligne d'une base de manuscrits anciens. Une première collection de documents constitués de manuscrits d'un mathématicien du XIX^{ème} siècle contient environ 20,000 pages.

Nous décrivons dans cette section, le modèle abstrait du contenu du système Armarius ainsi que son architecture. Ensuite nous présenterons les interfaces de l'application de ce système et leurs fonctionnalités et nous détaillerons le modèle conceptuel de la base de données Armarius.

3.1. Modèle abstrait d'Armarius

Nous présenterons un modèle abstrait (Figure 2) d'une base de données pour une bibliothèque numérique collaborative. Ce modèle est composé : des collections des images des manuscrits, des métadonnées (méta contenus, annotations, transcriptions...), des utilisateurs, d'un système d'assistance, d'un système de sécurité qui gère les droits d'accès aux collections, d'un système de traçage des actions des utilisateurs. Ce système de traçage va être utilisé pour aider l'utilisateur à personnaliser sa recherche de documents, aussi pour l'assister pendant son travail d'annotation et de transcription des documents, et finalement d'un système de collaboration pour permettre la transmission des messages entre les utilisateurs. Nous, nous appuyons sur les traces des groupes et des utilisateurs pour montrer les actions réalisées par rapport à une collection d'un groupe.

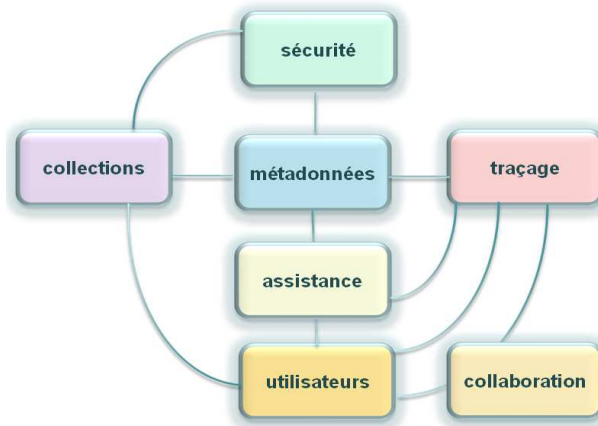


Figure 2. Un modèle abstrait de la bibliothèque numérique collaborative

3.2. L'architecture d'Armarius

L'architecture est une représentation plus détaillée du modèle abstrait d'Armarius. Cette architecture montre le contenu de notre système (Figure 3) :

- *collections* : images numérisées des manuscrits qui sont stockées en différentes formes (PDF, JPEG, PNG, Tiff,...);
- *métadonnées* : des mots-clés, commentaires, transcriptions, signatures numériques, métadonnées administratifs ou descriptifs...) sont attachées aux fragments des images numériques. Les annotations sont créées aux images de manière *indépendante des pixels*. Cette manière permet d'avoir la même annotation pour des versions différentes de l'image (des résolutions différentes).
- *application* : une application Web;
- *fonctionnalités* : un ensemble de services en ligne, comme la recherche, la visualisation, l'annotation, la transcription manuelle, l'ajout des commentaires, etc.;
- *base de données* : base de données pour la gestion des images, des collections, des annotations, des utilisateurs, des groupes d'utilisateurs et des droits d'accès. Nous allons détailler la structure de cette base de donnée dans la section (3.4.);
- *interface* : interface basée sur des services web pour intégrer des outils de traitement automatique d'image (word-spotting, ...).

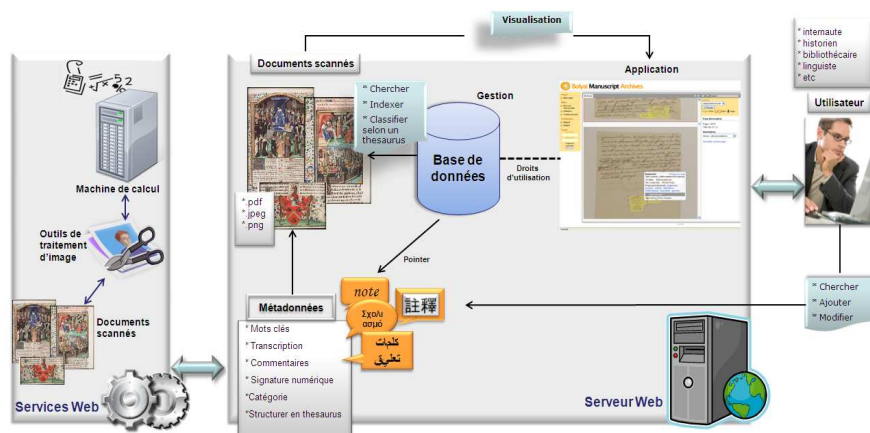
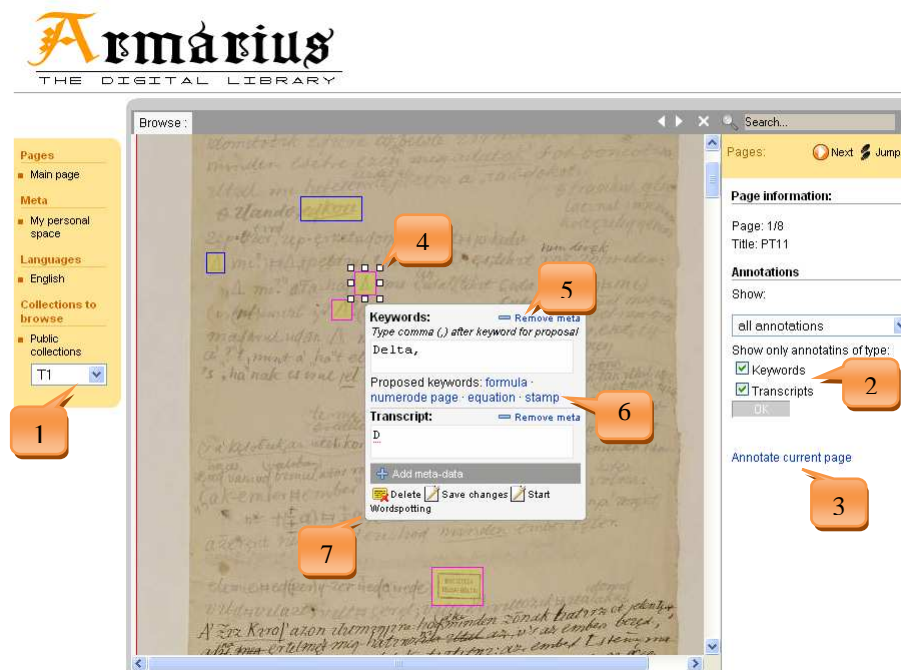


Figure 3. Vue globale du système Armarius

Dans le système Armarius, les utilisateurs se connectent au serveur de l'application pour visualiser les images des manuscrits ou des collections des documents numérisés. Les utilisateurs connectés peuvent chercher une certaine

image selon des annotations ou des transcriptions associés. Les annotations dans les documents sont de trois portées : des annotations *privées* qui sont accessibles uniquement par l'auteur lui-même, soit des annotations *collectives* partagées par un groupe d'utilisateurs ou des annotations *publiques* accessibles par tous les utilisateurs de la bibliothèque numérique. Ensuite, les utilisateurs peuvent ajouter (manuellement ou avec l'assistant de l'application) de nouveaux mots clés ou des commentaires enrichissant le document avec plus d'information, la Figure 4 .



1- sélectionner la collection à visualiser; 2- filtre sur les annotations et les transcriptions visibles; 3- créer une nouvelle annotation sur la page courante; 4- dessiner un rectangle autour d'un fragment; 5- ajouter des diverses métadonnées; 6- utiliser les mots clés suggérés par le système d'assistance; 7- lancer une session wordspotting avec le fragment sélectionné.

Figure 4. Annoter une unité documentaire dans Armarius

Nous montrons dans la Figure 5 les résultats de l'utilisation du *wordspotting* dans Armarius, le wordspotting cherche dans le document tous les fragments qui ressemblent au fragment précisé par l'utilisateur. Dans Armarius, le wordspotting est géré sous la forme de sessions : un utilisateur sélectionne un fragment d'une page, puis il lance la session qui peut prendre plusieurs heures. Sur son espace personnel, il peut ensuite visualiser l'état de ses sessions et valider les résultats de celles qui sont terminées.

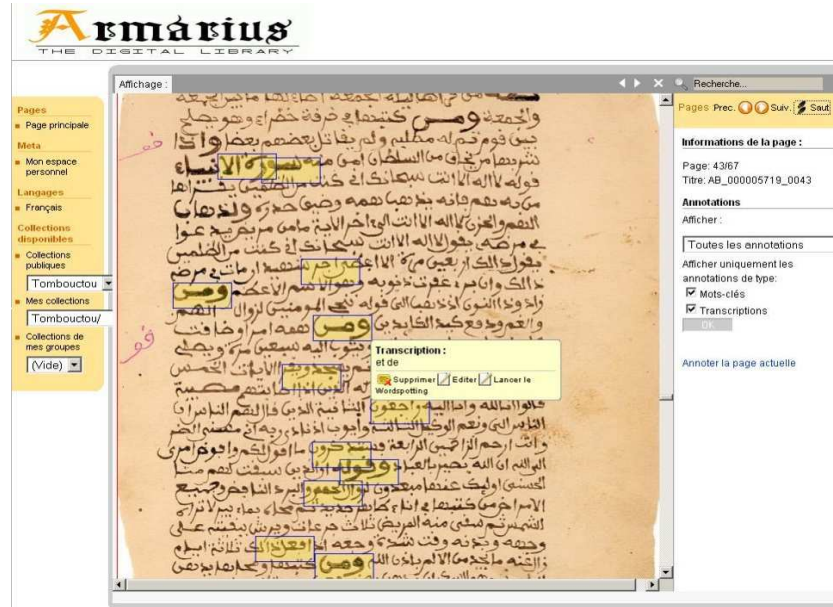


Figure 5. Résultats de Wordspotting dans Armarius

3.3. Interfaces et fonctionnalités d'Armarius

Une partie importante de la conception du projet Armarius sont les interfaces et ses fonctionnalités. Dans la Figure 6, nous présentons le schéma des interfaces de la bibliothèque numérique avec des tâches complexes et le système de collaboration (Armarius).

Depuis la fenêtre principale l'utilisateur a le choix entre trois langues. Ensuite, les utilisateurs non identifiés peuvent voir une démo des collections. Les autres utilisateurs se connectent pour accéder à leur espace personnel.

Presque toutes les fonctionnalités du système sont fournies dans l'espace personnel de l'utilisateur (chercher certaines collections ou pages puis les visualiser, faire une liste des mots clés favoris pour obtenir des informations sur ces mots, tracer les actions d'un certain utilisateur ou groupe, ajouter des marques pages aux collections/pages favorites pour créer une collection personnelle, gérer son profil utilisateur et son espace personnel,...).

L'étape la plus intéressante dans le projet Armarius arrive après la sélection et la visualisation d'une page de collection. L'utilisateur peut choisir de visualiser les annotations sur la page ou le document selon ses droits d'accès (privacy) : voir ses propres annotations (les annotations privées) ou les annotations collectives de ses groupes d'utilisateurs ou les annotations publiques qui sont disponible à tous le

monde. En plus de la visualisation des annotations, l'utilisateur peut annoter de nouveaux fragments d'image/document en créant un rectangle qui représente une unité documentaire (Figure 4). Une fois que l'utilisateur a précisé son unité documentaire, une boîte de dialogue est fournie par le système pour permettre à l'utilisateur de saisir ses annotations ou transcriptions du texte contenu dans l'unité documentaire.

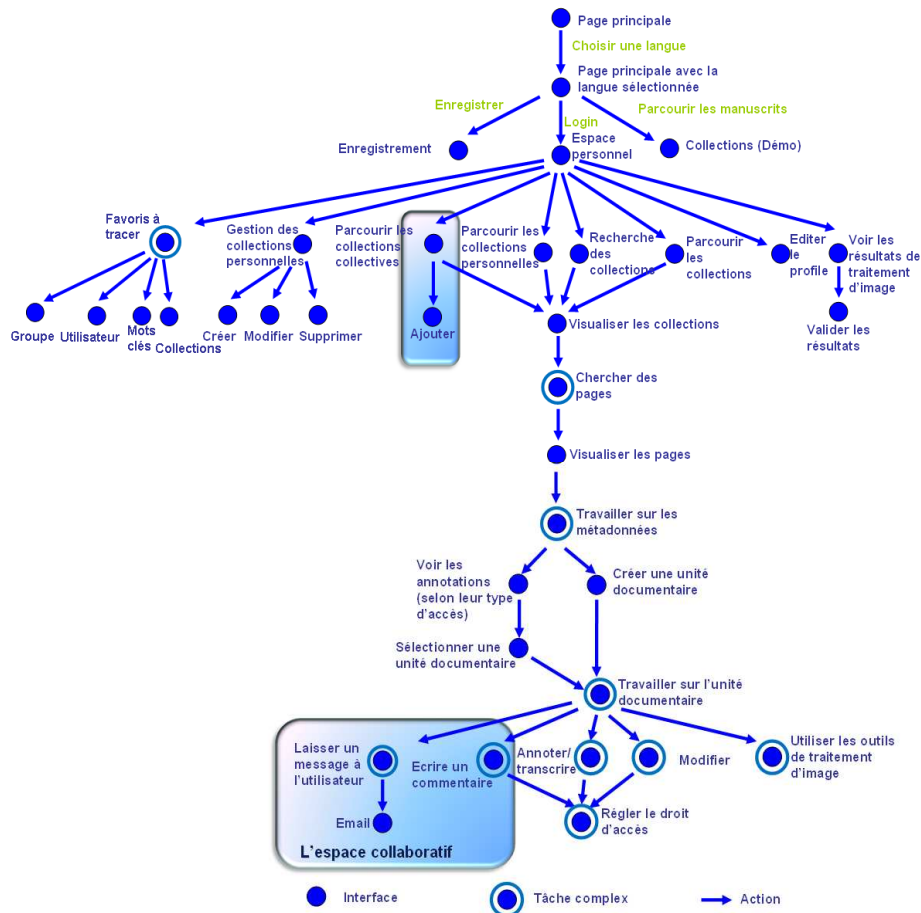


Figure 6. Schéma d'interfaces et fonctionnalités- Armarius

Le système Armarius propose à l'utilisateur des mots clés qui sont déjà utilisés par lui-même ou par d'autres utilisateurs afin de l'aider à mieux créer ses annotations. Un autre système assistant propose l'utilisation des outils de traitement d'image pour chercher d'autres fragments du document (des unités) qui ressemblent

à un fragment créé par l'utilisateur, puis les résultats de cet outil restent à valider par l'utilisateur. Ensuite, le *système de sécurité* permet à l'utilisateur de déterminer le droit d'accès (Privacy) de ses annotations/transcriptions/commentaires en définissant sa portée (privé, collective, publique). Une autre option offerte à l'utilisateur (s'il a le droit, selon son groupe) est de choisir une unité documentaire déjà annotée et de modifier les mots clés.

Notre système collaboratif fournit aux utilisateurs la possibilité d'ajouter des commentaires sur le travail des autres personnes ou d'ajouter des commentaires personnels aussi.

Dans Armarius le *système de traçage* vise à tracer les démarches des utilisateurs pendant leurs recherches de documents ou pendant leur travail sur les unités documentaires afin d'intégrer ces traces dans le système d'assistance. Le traçage s'effectue selon le schéma d'interfaces (Figure 6), le système va tracer les actions des utilisateurs identifiés, par rapport la recherche et la visualisation des collections et des pages, ainsi que les créations et les modifications des unités documentaires.

3.4. Modèle conceptuel d'Armarius

Dans cette section nous présentons le modèle conceptuel de la base de données de la bibliothèque numérique collaborative Armarius (Figure 7).

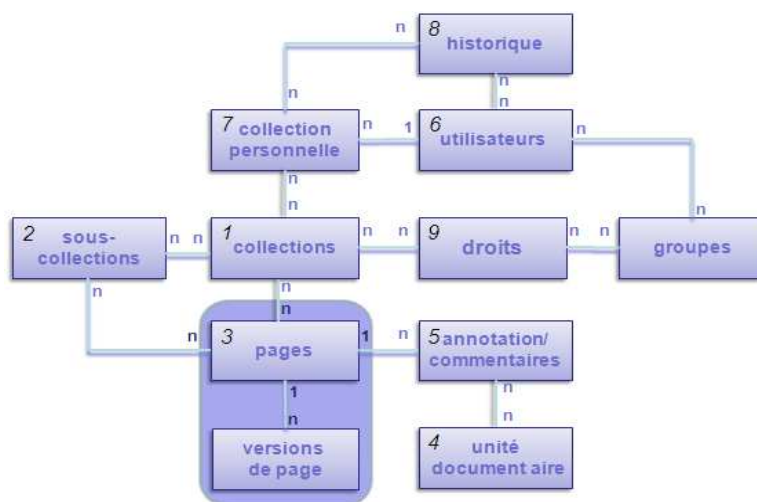


Figure 7. Le modèle conceptuel de bibliothèque numérique collaborative- Armarius

Ce modèle contient les modules suivants :

1. les images des documents manuscrits sont structurées en collections entreposées dans une base de données relationnelle;
2. les collections peuvent être organisées en sous-collections selon des critères différents (thème, date,...);
3. les collections/sous-collections sont composées de pages. Chaque page est l'image numérisée d'une page manuscrite. Dans notre base de données, nous sauvegardons plusieurs versions de documents ou d'images. Quand une image est introduite dans le système, trois versions sont produites : *vignette* pour la résolution basse, *accès* pour une résolution intermédiaire et *réel* pour une résolution haute. L'avantage de ce système est de fournir des vignettes quand l'utilisateur demande un résumé d'images, et de fournir des images en résolution intermédiaire « accès » si l'utilisateur demande un accès à l'image sans préciser sa version (résolution);
4. chaque page peut contenir plusieurs unités documentaires représentant des fragments d'image définis par l'utilisateur. Chaque unité possède des coordonnées liées à la page. Ces coordonnées changent suivant la taille de l'image en gardant l'unité au même endroit pour chaque version d'image;
5. les métadonnées sont des annotations, des transcriptions ou des commentaires associées à une page ou une unité documentaire (un fragment d'image);
6. nous avons deux types différents d'utilisateurs : premièrement, les internautes non identifiés qui peuvent voir une démo des collections choisies par l'administrateur, et sans être enregistrés dans la base de données; deuxièmement les utilisateurs enregistrés qui sont classés en groupes, et chaque utilisateur peut appartenir à un ou plusieurs groupes;
7. les utilisateurs du système Armarius peuvent créer un espace personnel qui contient des collections ou des pages choisies par l'utilisateur. cet espace fournit aussi toutes les fonctionnalités qui intéressent l'utilisateur pour réaliser son travail, annoter et transcrire les images des manuscrits;
8. le travail des utilisateurs et leurs actions accomplies sont enregistrés dans l'historique. Cet historique est utilisé dans le système de traçage et le système d'assistance pour aider les utilisateurs pendant la recherche et l'annotation;
9. les droits d'accès sont imposés d'un côté sur les collections et leurs contenu et de l'autre côté sur les groupes d'utilisateurs. Ces droits sont représentés comme (visualiser une page, visualiser les unités de documents avec leurs annotations, ajouter des nouvelles métadonnées, ajouter des commentaires, créer un espace personnel).

4. Conclusion et perspectives

Nous avons présenté dans cet article, un modèle d'archive numérique « vivante » pour des manuscrits anciens : Armarius, ainsi qu'un prototype d'application Web qui l'implante. Notre modèle proposé peut être utilisé dans autres domaines (documents scientifiques, documents médicaux...). Armarius propose également un cadre de travail collaboratif et un système d'assistance basé sur des traces d'utilisation.

Dans cet article nous avons traité les problématiques suivantes :

- comment représenter les documents numérisés dans une archive vivante ? Ça concerne la création des annotations, la structuration des documents, le stockage dans une base de données et le modèle d'accès aux documents;
- la personnalisation, création de profils, de groupes et leur évolution dynamique;
- comment intégrer une assistance utilisateurs, qui propose différents types d'aide pendant l'annotation, la recherche d'un document, ou la création de l'espace personnel. L'assistance, la collaboration, la confrontation et la discussion sont particulièrement importants lors de l'annotation des manuscrits. Pour ce type de documents, les interprétations sont multiples et les outils de traitement d'image peu efficaces;
- le besoin d'un espace de travail collaboratif qui permet la création, l'instrumentalisation et l'émergence dynamique de groupes de travail autour des collections.

Dans nos recherches à venir, nous visons à intégrer des technologies type « push » et RSS pour suivre l'évolution de certains documents, thèmes, collection, etc., à rendre possible et à assister la confrontation de points de vues et des discussions autour des documents. Nous, nous intéressons aussi au développement du module de collaboration, pour permettre aux utilisateurs d'échanger des messages pour discuter sur le contenu (collections, documents, pages, mots clés...). Un autre axe de recherche concerne l'enrichissement des outils de traitement d'image (en plus du wordspotting).

5. Bibliographie

- Borbinha J., Gil J., Pedrosa G., Penas K., «The case of the digitized works at a national digital library», *Second International Conference on Document Image Analysis for Libraries, DIAL '06*, 2006, Lyon France, p. 116-125.
- Bouzeghoub M., Kostadinov D., «Personnalisation de l'information: aperçu de l'état de l'art et définition d'un modèle flexible de profils», *CORIA 2005*, p. 201-218.
- Calabretto S., Pinon J.M., Bozzi A., «BAMBI : système de manuscrits anciens pour historiens », *Les bibliothèques numériques*, vol. 02, n° 3-4, 1998, p. 31-50.

- Chevalier M., Julien C., Soulé-Dupuy C., Vallès-Paralangeau N., « Personalized information access through flexible and interoperable profiles », *Web Information Systems Engineering – WISE 2007 Workshops, Springer Berlin*, 2007, p. 374-385.
- DELOS Network of Excellence on Digital Libraries. <http://www.delos.info/> (accédé Décembre, 2007).
- Doumat R., Egyed-Zsigmond E., Pinon J.M., « Online ancient documents in European national libraries, a survey », *Colloque International sur le Document Electronique (CIDE10)*, 2007, P. 151-162.
- Formalizing the design of Digital Libraries based on UML, présentation de DELOS, 2006, http://old.hki.uni-koeln.de/people/herrmann/forschung/Formalizing_Design.ppt.
- Gonçalves M.A., Edward Fox A., Watson L., Lipp N., « Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries », *ACM Transactions on Information Systems (TOIS)*, vol. 22, n° 2, 2004, p. 270-312.
- Hunt L., Lundberg M., Zuckerman B., « InscriptiFact: A virtual archive of ancient inscriptions from the Near East », *International Journal on Digital Libraries*, vol. 5, n° 3, 2005, p. 153-166.
- Ignat C., Pouliquen B., Steinberger R., Erjavec T., « A tool set for the quick and efficient exploration of large document collections ». *CoRR*, 2006
- Ikram A., Jedidi A., Sèdes F., « A contribution to multimedia document modeling and querying », *Multimedia Tools and Applications, SpringerLink*, vol. 25, n° 3, 2005, p.391-404.
- Kiamoto A., Onishi M., Ikezaki T., Deuff D., Meyer E.; Sato S., Muramatsu T., Kamida R., Yamamoto T., Ono K., « Digital bleaching and content extraction for the digital archive of rare books », *Second International Conference on Document Image Analysis for Libraries, DIAL '06*, 2006, Lyon France, p. 133-144.
- Lebourgeois F., Emptoz H., Trinh E., Duong J., « Networking Digital Document Images », *Sixth International Conference on Document Analysis and Recognition (ICDAR'01), IEEE Computer Society*, Seattle, WA, USA, 2001, p. 379-383.
- Morandiere B., Bouletreau V., « Persee: Addressing the needs of the digitalisation and online accessibility of back collections through robust and integrated tools », *Proceedings of the Second International Conference on Document Image Analysis for Libraries (DIAL'06), IEEE Computer Society*, Lyon France, 2006, p. 156-163.
- Mothe J., Chrisment C., Dousset B., Alaux J., « DocCube: Multi-dimensional visualisation and exploration of large document sets », *Journal of the American Society for Information Science and Technology*, vol. 54, n° 7, 2003, p. 650-659.
- Nguyen A., COCoFil2 : Un nouveau système de filtrage collaboratif basé sur le modèle des espaces de communautés. Université Joseph Fourier – Grenoble I, 2006.
- Pouliquen B., Steinberger R., Ignat C., « Automatic annotation of multilingual text collections with a conceptual thesaurus », *CoRR*, 2006.

Ramel J.Y., Busson S., Demonet M.L., «AGORA: the interactive document image analysis tool of the BVH project», *Second International Conference on Document Image Analysis for Libraries, DIAL '06*, 2006, Lyon France, p. 145-155.

Ravindranathan U., Shen R., Gonçalves M.A., Fan W., Fox E.A., Flanagan J.W., «ETANA-DL: a digital library for integrated handling of heterogeneous archaeological data», *International Conference on Digital Libraries, ACM*, Tuscon, AZ, USA, 2004, p. 76-77.

Savoy J., «Indexation manuelle et automatique: une évaluation comparative basée sur un corpus en langue française», *CORIA 2005*, 2005, p. 9-24.

Tamine L., Boughanem M., Zemirli N., «Inferring the user interests using the search history», *LWA 2006*, 2006, p. 108-110.