

INTRODUCTION

Bien que je défende le concept d'une bibliothèque numérique mondiale depuis 1993 [Chen, 1993], ce n'est que vers la fin des années 1990 que j'ai commencé à me consacrer à des travaux de recherche et développement (R&D) sur le développement des bibliothèques numériques. Alors qu'avec le nouveau millénaire nous entrons dans l'ère passionnante du numérique nous commençons à être témoins d'une convergence prometteuse entre les contenus, la technologie et la collaboration mondiale. Il y a des possibilités inédites de développer toutes sortes de bibliothèques numériques mais les défis à relever sont aussi sans précédents. Réalisant l'importance de la relation triangulaire entre le contenu, la technologie et l'utilisateur (Fig. 1), cet article exposera essentiellement les aspects techniques des défis que représentent le développement d'une bibliothèque numérique et d'un portail mondiaux [Chen, 2006a, 2006b].

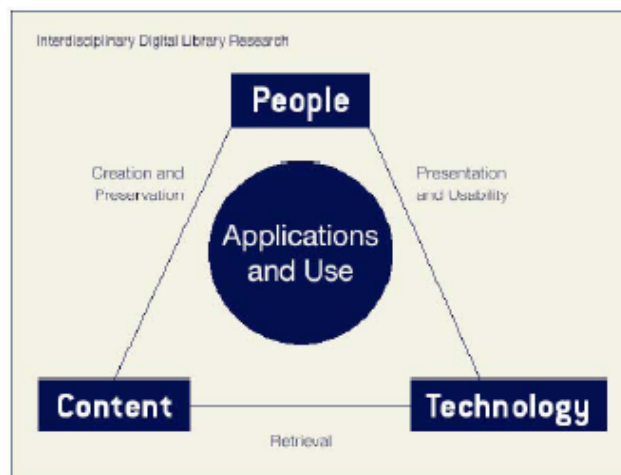


Figure 1. Modèle conceptuel d'une bibliothèque numérique interdisciplinaire.

LES DEFIS DU DEVELOPPEMENT D'UNE BIBLIOTHEQUE NUMERIQUE MONDIALE

Dans le cas de Global Memory Net, qui vise à être à la fois un portail et une bibliothèque mondiale d'images numériques, nous nous sommes fixé, d'un point de

vue conceptuel, certaines conditions au développement d'une bibliothèque numérique. Certaines des conditions retenues sont les suivantes [Chen, 2006a, 2006b] :

- Au lieu d'utiliser le Web pour publier, il faut s'en servir comme d'une plateforme afin de permettre une plus grande participation des utilisateurs ;
- Il faut proposer des services d'information multimédia intégrés et transparents afin d'enrichir le vécu des utilisateurs ;
- Notre système doit fournir des services d'information innovants et non des progiciels pour constituer des bases de données ;
- Les besoins en information et le comportement des utilisateurs ne doivent pas être prédéterminés et les fonctionnalités d'interrogation doivent pouvoir répondre à toute sorte de besoins ;
- Les utilisateurs doivent pouvoir contrôler l'utilisation des données et ils doivent donc pouvoir choisir d'interroger une ou plusieurs collections ;
- Nous devons proposer une base de connaissances et non pas une ou des base(s) de données : s'ils le souhaitent, les utilisateurs doivent pouvoir, après avoir trouvé l'information de base, l'enrichir d'informations multimédia connexes ou d'un mélange d'autres ressources ;
- Il nous faut fournir des ressources web accessibles et interrogeables et pas seulement des liens vers ces ressources ;
- Afin d'améliorer la situation des utilisateurs et leur capacité à consommer les informations obtenues, des fonctionnalités multilingues sont essentielles que ce soit au niveau de la présentation ou de l'interrogation ;
- Il faudra que les utilisateurs puissent partir d'un tout petit bout d'information utile puis l'élargir de façon importante en le reliant aux collections mondiales ainsi qu'aux ressources bibliographique et web du monde entier ;
- Il faudra que les utilisateurs puissent traiter un contenu quelque soit sa granularité ;
- Il faudra fournir un accès géographique aux contenus ;
- La propriété intellectuelle du fournisseur de contenu devra être protégée ;
- Il faudra que les utilisateurs puissent utiliser les données découvertes ou retrouvées afin de développer leurs propres projets ;
- Il faudra que les utilisateurs puissent contribuer activement à la bibliothèque ;

- Etc.

Afin de répondre à ces exigences, nous avons développé dans le cas de Global Memory Net (www.memorynet.org) la dernière version de l'*interactive Multimedia Content Retrieval System* ou système de recherche de contenus multimédia interactifs (provisoirement dénommé *i-M-C-S*) en incorporant de nombreux concepts et fonctions novateurs. Dans cet article, nous illustrerons la manière dont nous avons géré ces conditions. Certains des moyens employés sont les suivants :

- 1) Utilisation du Web comme une plateforme pour intégrer de manière transparente les concepts et les fonctions.
- 2) Possibilité de retrouver des images numériques non seulement de manière traditionnelle en utilisant les champs de métadonnées mais aussi en utilisant des méthodes d'interrogation de pointe fondées sur les contenus (CBIR).
- 3) Une fois localisée(s) la ou les images souhaitée(s), un simple clic de souris permet d'obtenir des bandes sonores, des vidéos numériques, des annotations textuelles les accompagnant ainsi que des ressources web et bibliographiques élargies.
- 4) Les ressources disponibles peuvent être retrouvées en interrogeant une seule collection ou par des recherches croisées multi collections.
- 5) Une présentation ainsi qu'une interrogation multilingue doit être possible.
- 6) La recherche géographique doit être proposée.
- 7) Alors que le but est un accès universel, les utilisateurs doivent aussi avoir la possibilité de contribuer leurs propres ressources.

LE DEVELOPPEMENT DU SYSTEME *i-M-C-S* DE GLOBAL MEMORY NET

L'*interactive Multimedia Content retrieval System* (*i-M-C-S*) a été développé en interne à partir d'un système basé sur Linux/MySQL/PHP avec le schéma de fonctionnalité suivant :

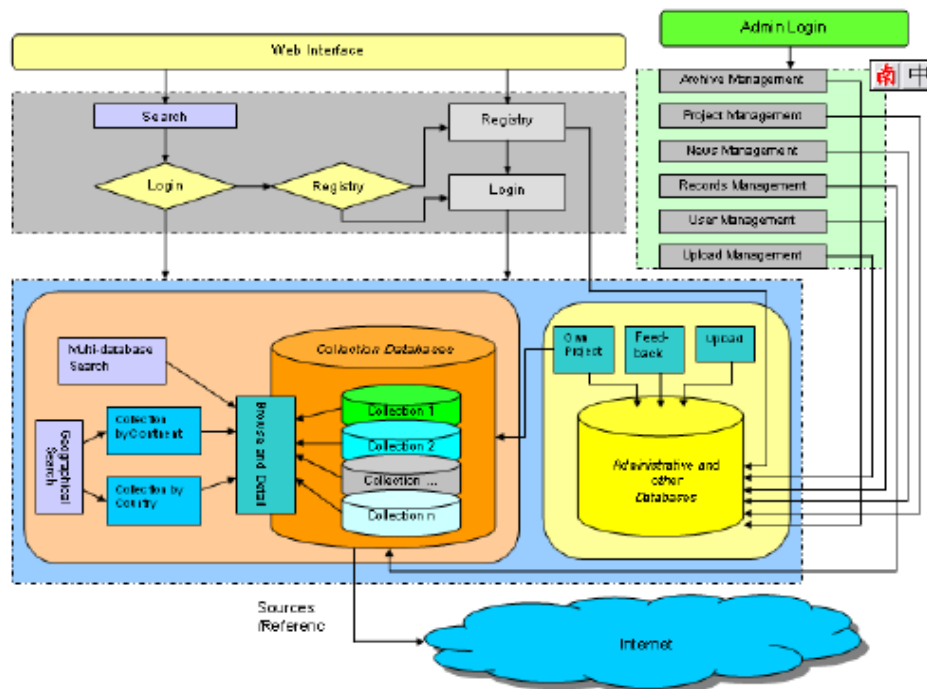


Figure 2. Schéma de fonctionnalité de Global Memory Net [Zhang et Chen, 2005].

Des fonctionnalités ont été ajoutées au fur et à mesure et actuellement la dernière version (Version 4.5) est vraiment capable de répondre à toutes les exigences énumérées ci-dessus. Nous nous pencherons sur ces fonctionnalités dans la section suivante.

De plus, conscients du fait qu'un plus grand nombre de contributeurs de contenus pourraient souhaiter développer leurs propres collections d'images numériques, nous avons développé une version générique de *i-M-C-S* afin de permettre à nos contributeurs de contenus de développer leur propres collections numériques facilement tout simplement en branchant ce qui est demandé une fois créée la base de données d'images élémentaire. Nous l'avons testé avec succès en Croatie lors d'un atelier qui a eu lieu à l'Université de Zadar après la conférence LIDA (Libraries in the Digital Age) 2006 [Chen and Badurina, 2006]. Par ailleurs, cette version générique de *i-M-C-S* est un système à 3 niveaux qui a été développé pour répondre aux besoins des organismes intéressés selon les niveaux et capacités techniques de chacun. Les 3 niveaux sont :

- **Niveau un** : ce système comprend presque toutes les fonctions de système de Global Memory Net actuel. Il demande un serveur et des personnels ayant une connaissance et une expertise de Linux, PHP, MySQL, etc.
- **Niveau deux** : ce système utilise un PC comme serveur et les fonctionnalités du PC en limite les fonctionnalités par rapport au niveau un.
- **Niveau trois** : c'est le niveau le plus bas qui permettra à un organisme ayant beaucoup moins de capacités techniques de commencer à développer sa ou ses base(s) de données de collections numériques, tout en sachant que toutes les applications de bibliothèque numérique nécessitent des bases de données numériques opérationnelles.

SPECIFICITES DE GLOBAL MEMORY NET [Chen, 2006a, 2006b]

Bien qu'il soit impossible ici de couvrir toutes les fonctionnalités, nous allons détailler certaines de celles que nous avons mentionnées plus haut :

- *Le Web est utilisé comme une plateforme* et non comme un support d'édition. La figure 3 montre comment GMNet propose un accès universel via une plateforme web.



Figure 3. La page d'accueil de Global Memory Net

- Un accès instantané à de magnifiques collections d'images* – Avec GMNet on peut accéder à plus de 30 collections d'images contenant plus de 20 000 images ainsi qu'à plus de 2400 collections numériques provenant de plus de 80 pays. Bien que les collections actuelles de GMNet se concentrent plus sur la culture, l'histoire et le patrimoine, la World Digital Collection de GMNet comprend plus de 2400 collections numériques couvrant toutes les thématiques. On y retrouve plus de 100 collections faisant partie du programme Mémoire du monde de l'UNESCO, plus de 290 collections provenant de la Bibliothèque du congrès américain ainsi que des collections provenant d'autres grandes bibliothèques nationales, d'archives, de musées, d'universités, etc. Autrement dit, les ressources inestimables du monde sont accessibles instantanément d'un simple clic de souris. La Figure 3 montre comment on peut accéder à ces collections en sélectionnant à partir du panneau bleu « Collections » à gauche de l'écran. La Figure 4 montre un agrandissement de ce panneau permettant d'en voir les détails. Seules 9

grandes collections s'affichent dans le panneau mais on peut en afficher beaucoup d'autres en cliquant sur « more ».



Figure 4. Panneau de navigation montrant comment accéder aux collections de GMNet.

- *Une interrogation classique facile et souple* – Sélectionnez la collection qui vous intéresse et procédez à une recherche classique en utilisant certains ou tous les champs de métadonnées quand vous savez exactement ce que vous recherchez. Dans ce cas, la syntaxe utilisée est celle de Google.
- *Les utilisateurs peuvent explorer une collection inconnue d'eux et en découvrir la teneur* – Quand les utilisateurs découvrent une collection qu'ils leur est inconnue, ils ne sont souvent pas en mesure d'effectuer des recherches par auteur, titre, sujet ou mot clé comme on le fait d'habitude. Dans ce cas, nous proposons une fonctionnalité permettant de naviguer dans la collection. Plus précisément, la fonction « random » (« au hasard ») permet d'explorer et de se familiariser avec ce que couvre la collection. En quelques secondes on en



Figure 10. Données descriptives (métadonnées) de l'image du poème japonais Waka de la collection Tsurumi disponibles en anglais et en japonais.

- *Filigrane numérique généré dynamiquement pour la protection des droits d'auteur du fournisseur de contenus* – voir Figure 8 ci-dessus. Grâce à cette disposition, les fournisseurs de contenus sont plus enclins à confier au Web les images numériques de leurs trésors.
- *Intégration transparente d'informations multimédia* – Si des informations pertinentes concernant une image sélectionnée sont disponibles sous une forme autre que des textes, les utilisateurs peuvent une fois de plus d'un simple clic de souris retrouver les documents sonores, vidéo, etc. pertinents comme on le voit sur la Figure 11.

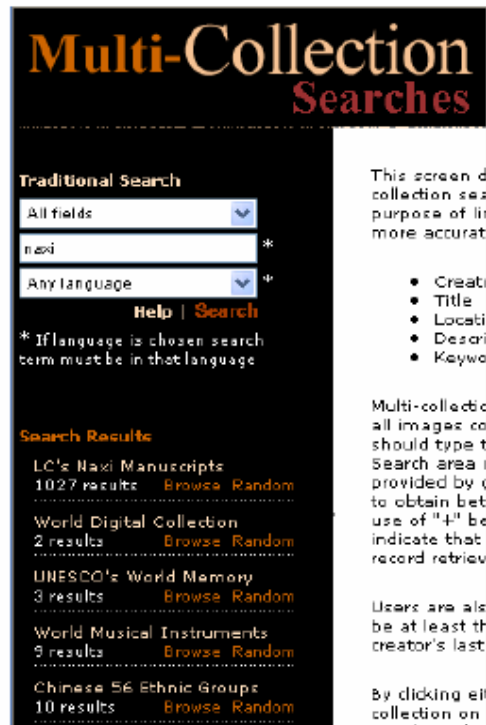


Figure 12. Affichage des résultats d'une recherche dans plusieurs collections

- Les utilisateurs peuvent acquérir des connaissances plus approfondies sur un sujet en interrogeant des ressources bibliographiques et des ressources web – Par exemple, comme le montre la Figure 13, si on souhaite trouver plus d'informations sur une image ayant trait à « Tombouctou » comme des ouvrages, des articles, etc. on peut instantanément trouver des documents pertinents dans le demi-milliard de notices du WorldCat d'OCLC, ou bien dans des ressources web comme Google Scholar, Wikipedia, Internet Archive, Million Books, etc. Si on retrouve les ouvrages que l'on veut dans le WorldCat d'OCLC, on pourra les localiser dans une bibliothèque proche de l'utilisateur. On pourra alors obtenir ces ouvrages par le biais du PEB, etc. Google Scholar et Wikipedia peuvent aussi fournir instantanément de l'information en texte intégral.*



Figure 13. Interroger les ressources bibliographiques et les autres ressources mondiales

- *Affichage multilingue des données descriptives et des résultats* – A ce jour, il y a déjà dans GMNet des contenus en anglais, chinois, croate, français, italien, japonais, espagnol, thaï, vietnamien, etc. comme le montre la Figure 14. Ceci facilite grandement l'utilisation des ressources de GMNet car l'utilisateur peut choisir la ou les langue(s) qui lui sont familières.
- *Les utilisateurs peuvent effectuer des recherches géographiques par continent et/ou par pays* – On peut accéder par continent aux informations de toutes les collections numériques comme le montre la partie supérieure de la Figure 15. De plus, on peut interroger par pays plus de 230 pays comme le montre la partie inférieure de la Figure 15. Malheureusement, à ce jour seulement 80 de ces 230 et quelques pays disposent de collections numériques. De nombreuses bibliothèques nationales dans le monde ne disposent même pas d'une page web. C'est pourquoi GMNet dispose d'une collection séparée sur les bibliothèques nationales du monde comme le montre la Figure 4.



Figure 14. Capacités multilingues

En plus des fonctionnalités mentionnées ci-dessus, il est utile de souligner que GMNet offre aux utilisateurs la possibilité [Chen, 2006b] :

- de participer activement pendant leur processus de recherche ;
- de décider de la manière dont ils veulent utiliser les informations découvertes ou retrouvées ;
- de développer leur(s) propre(s) projet(s) en sélectionnant mélangeant et remélangeant les images retrouvées,
- de contribuer activement à GMNet.

Par exemple, la possibilité de télécharger vers le serveur est fournie au cas où quelqu'un souhaiterait proposer un corpus iconographique important. Une autre point significatif : la possibilité pour les utilisateurs d'utiliser les ressources de GMNet pour leurs propres projets. Celle-ci peut être d'une grande utilité pour les enseignants par exemple car ils peuvent retrouver les images nécessaires sur un sujet donné, aller dans leur classe à l'heure dite, simplement se connecter à GMNet et faire leur cours.

