

LA BIBLIOTHÈQUE NUMÉRIQUE

PRINCIPE ET OUTILS

Université Abdelmalek Essaaidi
Faculté des Sciences et Techniques de Tanger
Pr. Ezziyyani Mostafa
Tel : 00 212 64 43 39 18 – 00 212 39 93 98 84
Fax : 00 212 39 39 39 53
e-Mail : ezziyyani@uae.ma

Dans ce résumé, je présente une synthèse sur les services de base de la Bibliothèque Numérique et les problèmes de la mise en place. Ce papier met également l'accent sur la technologie en matière informatique à déployer pour le développement du contenu avec les normes internationales en vigueur. Dans la partie perspective, je présenterais sommairement l'axe d'intérêts de ma recherche scientifique dans ce domaine d'application, qui fait l'objet du papier que j'ai envoyé précédemment.

I. Fonctionnalités d'une Bibliothèque Numérique.

L'information électronique est devenue une matière première stratégique, dont le coût est certes parfois élevé, mais dont la valorisation est essentielle pour l'entreprise. C'est la numérisation et donc la dématérialisation du document qui a permis la constitution de bibliothèques virtuelles. La bibliothèque numérique est organisée autour de quatre principes :

- Une organisation en fonction d'un public déterminé, on pourrait même dire en fonction d'une « clientèle »,
- Une sélection de documents en fonction d'une identification des besoins et d'un ajustement permanent à ceux-ci,
- Une organisation des connaissances : préparation du catalogue, indexation des documents (en fonction de la clientèle et de ses besoins), etc....
- Une offre de service : aide à la recherche, accès aux documents, etc....

La Bibliothèque Numérique reste basée sur ces principes, mais va élargir ses fonctions et offrir de nouveaux services à un lecteur qui ne se déplace plus et qui trouvera un document toujours disponible du fait de sa dématérialisation. L'un des principes essentiels de la bibliothèque numérique dans l'entreprise est celui de *the right information for the right man*. Mais on ajoutera aussi *in the right time*. Ce dernier point est sans doute l'un des plus importants dans le fonctionnement de la bibliothèque numérique. L'information doit être délivrée au bon moment et doit pouvoir atteindre son destinataire quel que soit le lieu où il se trouve. Pour la mise en place de ces principes, les fonctions de base de la bibliothèque numérique peuvent être résumées en :

Identification des documents : C'est la traditionnelle fonction de consultation de catalogue. Mais sur la bibliothèque virtuelle, l'offre en matière de consultation de catalogues pourra être très riche. Cette richesse devra cependant être raisonnée. Il ne sert à rien de pouvoir accéder à n'importe quel catalogue dont on ne connaît pas l'intérêt et la valeur. La mise à disposition de l'utilisateur, de *bookmarks* ou *signets*, c'est-à-dire une sélection de sites ou catalogues à consulter, sites choisis et évalués en fonction de leur intérêt et des besoins des utilisateurs valorise l'offre disponible.

Cette fonction de consultation de catalogues, de recherche documentaire, est largement facilitée par les logiciels d'interrogation, à partir d'outils banalisés tels que les navigateurs classiques d'Internet. Cette interrogation pourra être faite, en fonction des fichiers et bases de données accessibles sur la bibliothèque virtuelle, soit à partir de mots-clés, soit à partir du texte intégral des documents.

La visualisation du document : **Affichage du contenu** sur l'écran de la station de travail ou du micro-ordinateur portable. Le document peut-être affiché, selon les cas, en mode page restituant le document en fac-similé, véritable *photographie* de l'original, ou en mode texte afin de permettre un accès dans le texte. Certaines applications prévoient d'ailleurs le double affichage.

Manipulation du document : Cette fonction est très intéressante, elle permet l'annotation du document pour l'utilisateur. Celui-ci va pouvoir annoter le texte, le marquer et ajouter ses commentaires. Et ce qui ne pouvait être fait avec le document papier, le document électronique peut, lui, être restructuré selon les pratiques de l'utilisateur.

Analyse bibliométrique : C'est une des voies les plus intéressantes qu'apporte la constitution de fonds numérisés. Sur un corpus, ensemble de documents défini par l'utilisateur, celui-ci, à l'aide d'un logiciel spécifique, va pouvoir faire une analyse sur certains critères permettant de mettre en évidence certaines tendances ou de faire apparaître des phénomènes aberrants.

Analyse lexico-sémantique : Permettant, à partir de l'apparition de termes dans un corpus, de détecter des tendances technologiques ou économiques à travers les signes avant coureurs de ces tendances. Ces tendances peuvent être visualisées à l'aide de logiciels de cartographie des connaissances. Ces cartes de la connaissance permettent de suivre l'évolution d'une technologie dans le temps par comparaisons successives.

Etablissement des liens : Cette fonction hypertexte donne la possibilité à l'utilisateur de créer des liens à l'intérieur d'un document pouvant créer ainsi des cheminements de lecture. Il peut aussi créer des liens, dans un corpus, entre différents documents.

Conservation des documents : La gestion de pérennités documents se pose à trois niveaux. Le premier est celui de la conservation. Que faut-il conserver ? Il est certain qu'une large part des documents qui sont passés sur le Web a désormais disparue entièrement. Par ailleurs, il est hors de question de vouloir conserver *ad vitam aeternam* les millions ou milliards de pages Web. En l'absence aujourd'hui d'un dépôt légal des documents numériques virtuels (à l'exception des documents numériques sur un support physique tel que le cédérom), les pratiques sont encore bien hésitantes en ce domaine. Le second niveau est celui de la durée de vie des nouveaux supports, support magnétique et surtout support optique. En l'absence de toute certitude en la matière, certitude ne pouvant être apportée que par le temps, il convient de recopier régulièrement les documents numériques sur les supports jugés les plus sûrs. Enfin, et cela est sans doute le plus important, il ne faut pas oublier que le document dématérialisé nécessite un système de lecture sans lequel le document ne peut prendre vie.

II. Normalisation de la bibliothèque numérique.

La normalisation est un point important car la bibliothèque numérique doit être un outil de communication et, qui dit communication dit normalisation. Mais c'est sans doute le point sur lequel les difficultés sont résolues ou en voie de résolution. La normalisation est maintenant acquise dans le codage des caractères avec UNICODE qui se substitue peu à peu au codage ASCII. UNICODE permet l'utilisation d'un grand nombre de jeux de caractères et présente ainsi un caractère international avec l'utilisation des langues non latines.

Le balisage et la structuration des documents numériques sont désormais eux aussi normalisés avec l'adoption de la norme XML qui se substitue, elle aussi, à la norme SGML. Une large part de la documentation technique est désormais créée sous XML. Le terme de meta-données (titre, auteur, éditeur, date, etc...) désigne de manière générique, les données créées pour fournir des informations sur les ressources électroniques d'information.

L'utilisation conjointe d'XML et des méta-données normalisées devrait permettre un catalogage automatique des documents dans les bibliothèques numériques. Enfin, PDF (*Portable document format*) est une norme largement reconnue et utilisée pour la diffusion des documents électroniques, de même que l'adoption de la norme Groupe IV du fax pour la compression des documents en noir et blanc et des normes JPEG (*Joint Photographic Export Group – ISO 109181*) et MPEG (*Motion Picture Expert Group – ISO 11172*) pour la compression des images fixes et des images animées.

Dans le domaine de la normalisation du document numérique, il faut réunir, à la fois des professionnels du livre, des éditeurs d'ouvrages, des éditeurs de logiciels, des sociétés développant des livres électroniques et des constructeurs, afin de définir un standard d'échange des documents. Cela nécessite une étude approfondie de la technologie à déployer pour la définition d'une plateforme opérationnelle supportant un contenu variable via des structures riches et extensibles adaptables aux différents éléments qui définissent la base de données de la Bibliologie Numérique et les utilisateurs.

L'informatique et les technologies d'accès permettent également aux personnes handicapées plus particulièrement aux malvoyants ou/et aveugles d'accéder directement aux différentes sources de

données sur le Web. L'accessibilité numérique concerne non seulement les sites Web mais aussi de manière plus générale l'information sous forme numérique tel que la Bibliothèque Numérique quel que soit le mode de consultation. Dans cette optique nous pensons que la réalisation d'un système qui permettra de faciliter l'accès à la Bibliothèque Numérique pour les malvoyants est une nécessité primordiale. Dans cette perspective, la normalisation doit prendre en considération les droits d'accès pour les **malvoyants** et les **aveugles** pour les sites public et/ou semi-public. Par conséquent, Tout lecteur aveugle ou malvoyant, disposant des outils d'accès et de lecture spécifiques, peut s'inscrire et emprunter des livres numériques en suivant une procédure en ligne.

III. Conclusion et perspectives

La mise en œuvre d'une bibliothèque numérique peut se heurter à un certain nombre de difficultés. Mais, au fur et à mesure du développement des technologies, et avec l'expérience acquise en ce domaine, la maîtrise du document numérique est de plus en plus importante. Le document numérique peut être partagé par tous. Il n'y a plus la notion si importante de propriété du document qui freinent tant la circulation de l'information.

L'augmentation et la diversité de ces bibliothèque numérique sur l'Internet ont, en quelques années, profondément modifié l'architecture des systèmes d'information et ont donné de nouvelles réponses aux questions liés aux problèmes de recherche qui s'y rapportent. Cette explosion du nombre de sources digitales accessibles via le Web fait naître le besoin d'outils de médiation entre les utilisateurs et les Bibliothèques Numériques accessibles par Internet. Ces outils doivent dépasser les limites des moteurs de recherche actuels en permettant d'une part aux utilisateurs de poser des requêtes plus sophistiquées que de simples mots-clés, en étant capables d'autre part d'agrèger des éléments de réponses provenant de différentes sources pour construire la réponse globale à la requête de l'utilisateur. Les médiateurs (à l'occurrence le médiateur AXMed¹) ont pour objectif la construction d'une vision virtuelle centralisée et cohérente d'un ensemble de ressources de données distribuées et hétérogènes.

Le problème central de l'intégration de sources digitales est celui de leur hétérogénéité au niveau de leur format, modèle et/ou au niveau sémantique. En effet, l'information numérique est aujourd'hui disponible et distribuée dans une multitude de sources digitales autonomes et fortement hétérogènes (données tabulaires, fichiers structurés, documents XML semi-structurés, Relationnelle, données multimédia telles que images, sons, vidéo, données spatiales). La construction d'un médiateur au dessus de ces sources peut donner l'illusion à un utilisateur qu'il interroge un système d'informations centralisé et homogène.

Dans le cadre d'un projet de recherche, le but est donc de centraliser ces sources digitales hétérogènes via un médiateur (à l'occurrence AXM^{ed}). Parmi les objectifs fondamentaux figure le développement des outils pour l'intégration, la gestion, l'accès, la distribution et l'exploitation de contenus numériques dans ce domaine d'application et les axes qui leur sont attachés. Ce programme s'intéresse également à l'accès multilingue aux produits et services multimédias distribués sur les réseaux numériques, et à leur adaptation aux spécificités des internautes. Cela nécessite, le développement d'une couche logicielle au dessus du médiateur AXMed pour pouvoir communiquer avec l'utilisateur selon sa langue et sa capacité. Et l'utilisation à distance des outils logiciels du traitement numérique prédéfini pour l'exploitation du contenu des sources numériques (lecture, médias, traduction, représentation etc ...), via la technologie Services Web.