

HyWebMap : structurer l'information du Web

Fabrice Papy^{*}, Imad Saleh^{**}, Nasserddine Bouhai^{**}

Université Paris 8, Equipe Hypertexte dynamique,
Départements Documentation* & Hypermédia**
2 rue de la liberté, 93526 Saint-Denis Cedex 2
Tél. : 01 49 40 67 58, Fax : 01 49 40 67 83

{papy, saleh, nasser}@labart.univ-paris8.fr

Résumé :

Les travaux réalisés par T. Nelson, J. Nielsen, J.-P. Balpe, P.J. Brown, L. Conklin, et P.K. Garg, relatifs à l'Hypertexte, l'hypertextualisation, les hyperdocuments, les systèmes lecteur/auteur, les parcours de lecture, l'exploration documentaire, ...ont apporté le contexte théorique dans lequel nous avons élaboré HyWebMap, notre système d'organisation de connaissances pour le Web.

Depuis quelques années, la distribution de l'information électronique repose sur les technologies de l'Information et la Communication, que l'Internet et sa déclinaison intranet ont très largement contribué à généraliser et à développer. Les nombreux serveurs Web qui participent à l'architecture du réseau des réseaux mettent ainsi à la disposition de tous, des sources d'informations pérennes, pertinentes et conséquentes d'une même communauté sémantique. Alors qu'il est impossible de contester la facilité avec laquelle les navigateurs Internet rendent possible l'accès aux documents électroniques disponibles, toutes les réserves peuvent être émises en ce qui concerne leur capacité à proposer une exploitation effective de sources d'informations réparties que l'internaute aura su reconnaître et identifier préférentiellement. La dissémination des informations dans les différents berceaux technologiques devient un inconvénient majeur à une exploitation conjointe des informations recensées.

HyWebMap au travers des dispositifs techniques qu'il met en œuvre permet de définir une structure virtuelle composite où sera agrégée la totalité ou une partie des informations préalablement localisées par l'internaute. Cette particularité de construction d'une structure virtuelle dynamique s'accompagne d'un ensemble de fonctionnalités qui confère à notre système la double dimension lecteur et auteur.

- La dimension lecteur est assurée par les agents web qui sont lancés vers les URL distantes à la demande de l'utilisateur/lecteur et qui seront rassemblés au sein d'un espace d'information personnel que l'utilisateur aura tout loisir de nommer et d'enregistrer.

- La dimension auteur est très largement développée dans HyWebMap au travers des opérations qu'il est possible d'appliquer sur les éléments collectés antérieurement : intégration au site virtuel de nouveaux nœuds, pose de liens, nommage explicite des nœuds et des liens, enrichissement des nœuds par mots-clefs, annotations...

Ces caractéristiques auteur/lecteur sont associés à des mécanismes sophistiqués d'actualisation du site virtuel ainsi déclaré. Ponctuellement le créateur du site virtuel pourra s'assurer grâce aux agents Web de la cohérence de son réseau composite en vérifiant l'état des éléments exogènes (nœuds distants modifiés ou supprimés). Des sollicitations automatiques à des moteurs de recherche ont été prévues afin de proposer de nouveaux candidats au remplacement de nœuds distants détruits ou définitivement indisponibles.

HyWebMap se propose de mettre en avant les concepts fondateurs de l'hypertexte en les appliquant au monde de l'Internet en mettant en évidence la qualité sémantique des nœuds et des liens d'un réseau virtuel. Au travers d'HyWebMap nous retrouvons l'expertise que représente les parcours de navigation d'un lecteur averti au sein d'un espace d'informations hypertextuelles. Nous avons étendu cette conception originale des livres électroniques à la construction de sites Web virtuels, thématiques et répartis qui s'appuient sur l'agrégation de ressources distantes afin de former une communauté de documents sémantiquement homogènes.

Mots-clefs : HyWebMap, hypermédia, écriture, système auteur, système lecteur, apprentissage, formation

Problèmes de désorientation et de surcharge cognitive dans un hypertexte

La désorientation dans l'espace hypertexte est présentée comme étant une conséquence de la navigation au sein d'un hypertexte. Toutefois certains auteurs relativisent cette désorientation parce qu'elle n'a lieu que dans les hypertextes de taille importante et qu'elle dépend de l'expérience de l'auteur en matière de réalisation de réseaux hypertextes.

La prolifération des moteurs de recherche et autres méta-chercheurs qui recensent des millions de sites Internet et qui ramènent à l'issue d'une interrogation plusieurs milliers d'adresses estimées pertinentes, illustre parfaitement ce phénomène de désorientation qui guette l'internaute-chercheur. Le parcours rigoureux ou plus aléatoire des adresses proposées le menant d'un type de site à un autre introduira fatalement une distorsion dans le fil conducteur de sa recherche et conduira inéluctablement vers cette forme de désorientation que la pratique des premiers réseaux hypertextes ont mis en évidence.

Il est vrai que les sites portails proposés aujourd'hui forment une alternative qualitative alléchante par rapport aux listes démesurées d'URL proposées par les moteurs de recherche à l'issue d'une requête exprimée par l'internaute. Certes plus pertinents, ces sites portails n'en sont pas moins de forts vecteurs de désorientation pour

le visiteur puisque l'information proposée subitement agrégée autour de quelques thèmes génériques prend une "épaisseur" insoupçonnée de sorte que l'internaute - peut-être tout près du but de sa recherche - en demeure néanmoins éloigné.

Une stratégie habituellement utilisée dans les systèmes hypertextes consiste à proposer à le lecteur une carte du réseau afin qu'il puisse, à partir de la vision globale ou locale de la structure de l'hypertexte, se repérer plus facilement et éviter ainsi de se perdre. les chercheurs [Brown 88, Balpe 96] se sont penchés sur le comportement des lecteurs pour expliquer le phénomène de désorientation. Ils ont mis en évidence trois niveaux de problèmes liés a :

— **la navigation dans l'espace logiciel** : le lecteur ne comprend pas les facilités de navigation que lui offre le système hypertexte.

— **la navigation dans l'espace conceptuel** : le lecteur ne comprend pas comment l'auteur a organisé l'information pour la représenter sous une forme hypertextuelle.

— **la navigation dans le texte** : Le lecteur d'hypertexte, sans cesse sollicité par des liens à activer, se laisse tenter par des digressions, qui lui font perdre le fil de sa navigation de départ. Les multiples décrochements qu'il a effectués finissent par le laisser à un point du réseau qui ne correspond plus à sa recherche première.

Enfin, au problème de désorientation, vient s'ajouter un phénomène de **surcharge cognitive** (*cognitive overhead*) [Astolfi 93] qui provient de l'incapacité du lecteur à retenir l'essentiel de l'information liée à un noeud parcouru ou de ses difficultés à identifier rapidement la nature des liens qu'il a activé. [Rouet 97, Castelli 96] ont identifié six variables cognitives clés dans la navigation en hyperespace : *efficacité intellectuelle, capacité analytique, flexibilité analytique, capacité de synthèse, raisonnement abstrait, indépendance du champs cognitif*. La navigation s'avère un processus complexe.

Dans l'espace Web, nous ajoutons aux problèmes de la désorientation et de la surcharge cognitive des **incidents de navigation**, qui peuvent survenir en raison de :

- l'augmentation considérable du trafic d'informations hétérogènes,
- l'existence de nombreux liens obsolètes soient internes ou externes,
- l'organisation des pages non adaptées à le lecteur.

On constate que le lecteur doit se repérer, s'orienter et naviguer dans un espace d'information en perpétuelle évolution et cette évolution ne semble guère se ralentir.

Ces différents constats nous ont amené à proposer HyWebMap.

HyWebMap : Un système agrégatif de ressources exogènes

Dans sa manipulation la plus élémentaire, HyWebMap permet à l'utilisateur/internaute de constituer à partir de ressources documentaires exogènes (cf figure 1) puisées sur Internet et/ou au sein d'Intranets – l'exploitation de l'un et des autres étant bien sûr possible – et également à partir de ressources personnelles (documents HTML créés spécifiquement par l'utilisateur), un *book* électronique composite que nous appelons *site virtuel ou espace de connaissances personnel*. HyWebMap offre un ensemble de fonctionnalités permettant de mettre en œuvre ces réseaux virtuels au travers d'une part des mécanismes de gestion de nœuds et de liens et d'autre part des dispositifs d'administration de sites (agents d'actualisation, interrogation de moteurs de recherche, recherche intégrée, génération dynamique de sites,...)

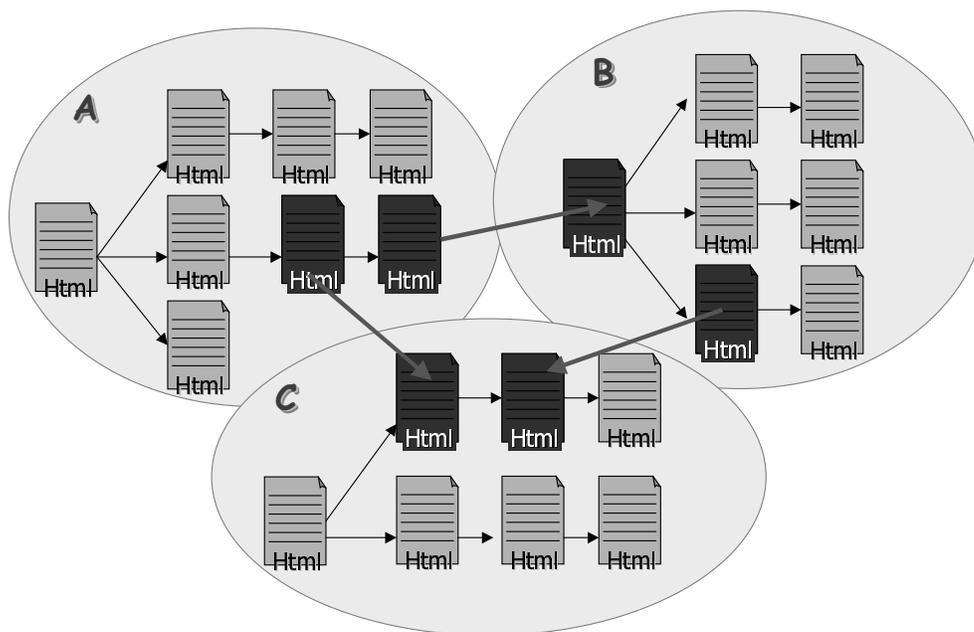


Figure 1 : Un réseau composite HyWebMap

Cette démarche agrégative est assez caractéristique de ce que l'on peut observer chez les veilleurs (cyberlimier, cyberdocumentalistes), dont la principale activité consiste à repérer (à rapatrier et à actualiser) les informations les plus pertinentes liées à un domaine, un sujet ou un thème de recherches afin de les mettre à disposition d'usagers spécifiques (agences de communication, bibliothécaires et documentalistes spécialisés, services d'assistance SVP,...).

De manière identique, l'animateurs de réseaux ; dont le travail est à la croisée des métiers de webmaster, documentaliste et chargé de communication a pour tâche la recherche d'informations précises et exhaustives (par exemple dans les domaines d'activités bancaires concernant les produits financiers et les mouvements boursiers,

...) mais endosse également la charge d'éditorialiste puisque lui échoit la fonction de synthétiser l'information collectée et d'en préparer une consultation structurée. Ces démarches sophistiquées de recherche et de réorganisation des données se retrouvent fréquemment dans les sites portails ou *vortails* d'entreprise (site marchands de comparatifs de prix, sites de fédération offrant à leurs adhérents une information exhaustive, ...).

Améliorer les outils rudimentaires d'organisation des navigateurs

L'élaboration d'arborescence utilisant les signets et favoris que proposent les navigateurs courants répondent peu ou mal à la nécessaire réorganisation des informations ainsi collectées. Une information repérée par l'internaute/lecteur ne prendra sa véritable dimension qu'une fois rapprochée de cette autre information collectée à l'occasion d'une précédente recherche. Pouvoir les rassembler au sein d'un même dossier ou rubrique portant un nom évocateur et fédérateur offre certes l'assurance de ne pas perdre l'adresse URL précise de la ressource localisée et d'introduire une forme primaire d'organisation mais n'est d'aucune utilité pour exprimer les relations plus étroites qui peuvent lier un document à un autre document et qui pourtant sont à l'origine du rapprochement de ces informations.

Le nom du dossier ou de la rubrique ainsi créé jouera le rôle unificateur de classe sémantique qui justifiera la co-présence de ressources documentaires distinctes. Ce subterfuge proposé par les navigateurs est un pis-aller qui à le défaut majeur d'externaliser les relations entre les documents et de transformer des associations de type *document* ↔ *document* à des associations *document* ↔ *classe* ↔ *document*. Outre la disparition de cette association privilégiée entre des sources documentaires distinctes, chaque élément assujetti à la classe sémantique du dossier se retrouve doté d'associations n-aires avec tous les autres éléments contenus dans le dossier. Des systèmes tel qu'IronWeb [Dussaux 00], ont amélioré l'usage des favoris – qui demeure extrêmement individualisé autant dans le choix des ressources documentaires retenues que dans l'organisation de regroupements aménagés - en proposant des mécanismes de partage des favoris personnels. Les utilisateurs choisissant d'adopter IronWeb mettent à disposition des autres utilisateurs l'ensemble de leurs favoris. Ce mode de stockage centralisé offre l'avantage de mettre à la disposition de l'ensemble de la communauté l'intégralité des favoris identifiés par chacun de ses membres tout en préservant l'appartenance et l'organisation personnelle de chacun. L'avantage évident d'IronWeb réside dans ces collections de favoris « expertisés » qui forment alors des « autorités » dont l'usage en terme de recherche pourra s'avérer plus pertinent que les résultats produits par les moteurs de recherche généralistes.

Déclaration de nœuds et de liens dans HyWebMap

Les nœuds : support de ressources documentaires

L'architecture d'HyWebMap repose sur la notion de « nœud libre » non typé dont la vocation première est de servir de support à l'élément d'information endogène ou exogène que l'utilisateur lui affectera. Ce nœud libre peut se voir doter des attributs suivants :

- un nom
- une ressource documentaire Web (URL de site, de page ou de site HyWebMap)¹,
- des descripteurs ou mots-clés,
- une annotation.

Les figures 2b, 2c et 2d illustrent l'affectation d'un nœud libre par l'URL d'une page HTML provenant d'un autre site virtuel HyWebMap. La démarche consiste à déclarer un « nœud libre » (figure 2a) et à y ajouter la ressource Web souhaitée (figures 2b et 2c). Ce nœud affecté se retrouve automatiquement intégré au sein du réseau en cours d'élaboration grâce au lien défini (figure 2d).

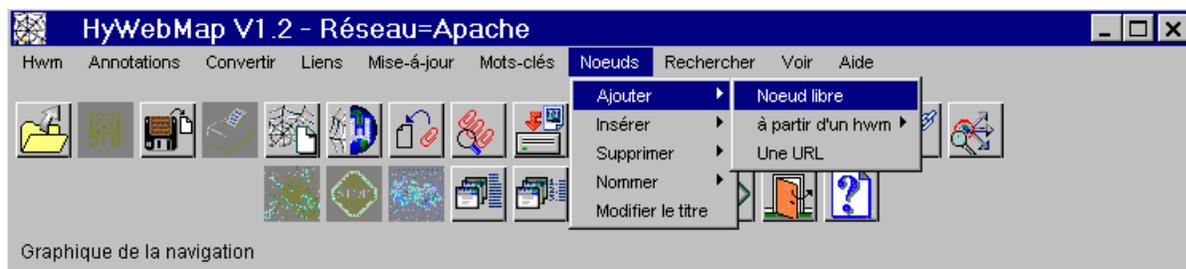


Figure 2a : Menu d'insertion de nœud libre au sein d'un réseau virtuel HyWebMap (Apache.hwm).

¹ Les ressources externes associées aux nœuds peuvent provenir d'un site virtuel HyWebMap, d'une page distante (par exemple <http://www-bu.univ-paris8.fr/Ref/index.html>) ou encore d'une page locale (les serveurs web personnels; *Personal web server*, *Apache*, *WebStar*,... équipent en standard la plupart des systèmes d'exploitation même ceux destinés à un usage de poste de travail (Windows 9x, Mac OS x, BeOs,...)). Ils permettent aux rédacteurs de simuler en environnement de travail local un fonctionnement « en ligne » où toutes les composantes techniques liées au protocole HTTP et FTP sont utilisables. Du point de vue technique, ces « pages locales » sont manipulées et traitées de façon totalement identique aux pages distantes. Cette possibilité de création de pages locales appelle une activité rédactionnelle de la part du concepteur du site virtuel dont le rôle n'est pas réduit à la seule collecte d'informations.

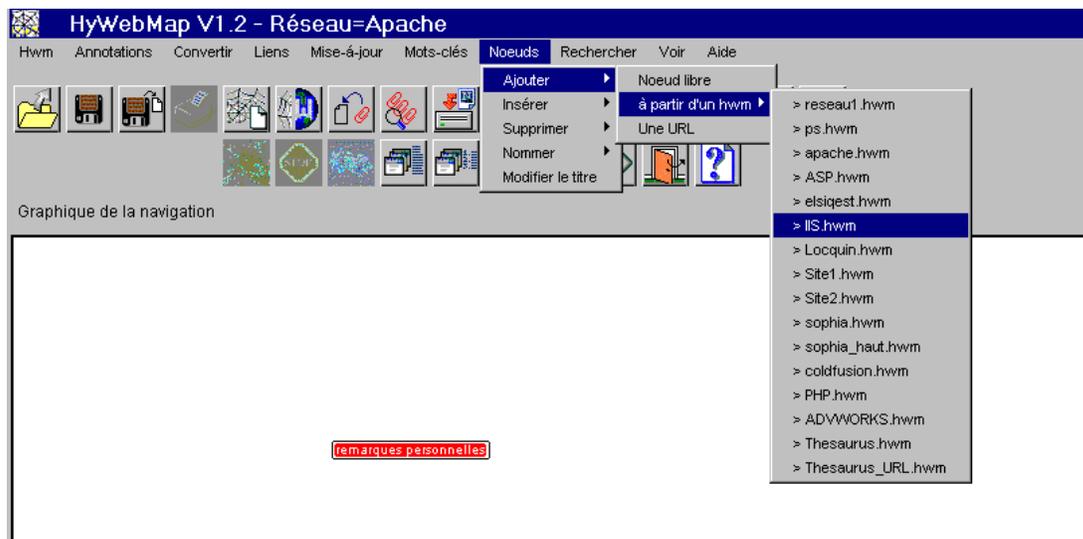


Figure 2b : La sélection d'une ressource distante au travers d'un précédent site HyWebMap

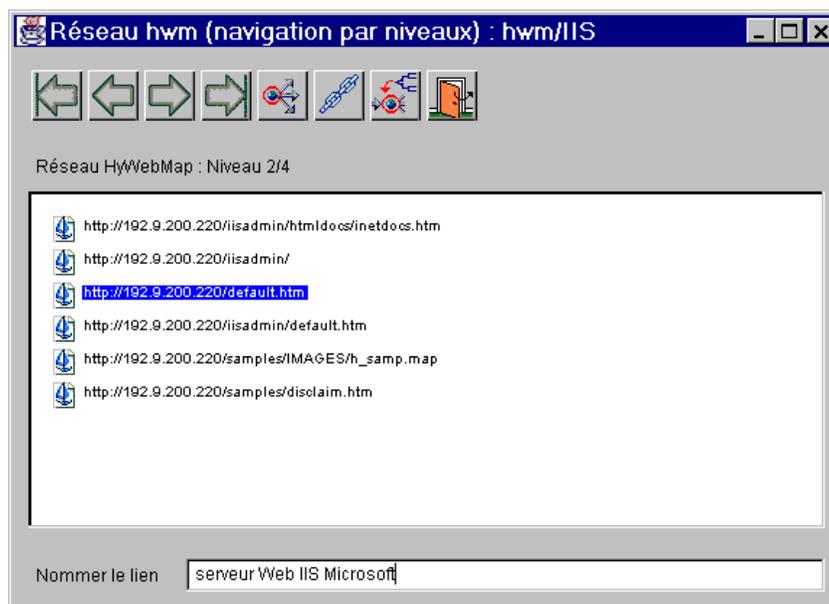


Figure 2c : Sélection d'un nœud et identification du lien

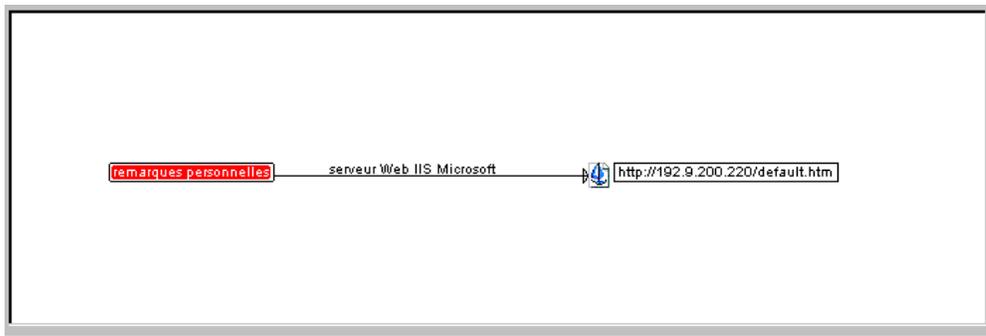


Figure 2d : Un nouveau nœud affecté de l'URL <http://192.9.200.220/default.htm> est associé au nœud "remarques personnelles" au moyen du lien "ServeurWeb IIS Microsoft"

Définir des associations entre ressources distinctes

L'agrégation significative de sources d'informations exogènes repose sur la capacité à exprimer la nature des liaisons pouvant exister entre plusieurs éléments. Les fonctionnalités de nommage de liens et d'association arbitraire présents dans HyWebMap vont permettre à l'utilisateur/concepteur de préserver les associations *document ↔ document* (cf. figure 3) que les systèmes de signets et favoris des navigateurs tendent à faire disparaître au profit d'une simple organisation arborescente.

Les liens univoques posés au sein d'un réseau virtuel HyWebMap permettent grâce aux liens internes de définir des relations transversales qui aboutiront à la création d'une véritable architecture maillée (cf. figure 4).

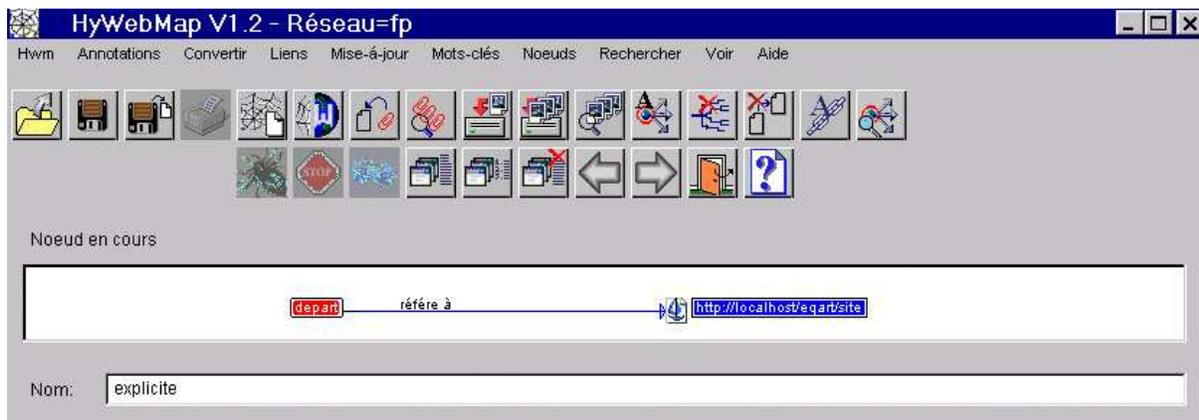


Figure 3 : remplacement du lien « réfère à » par le lien « explicite » entre les nœuds « depart » et <http://localhost/eqart.site>.

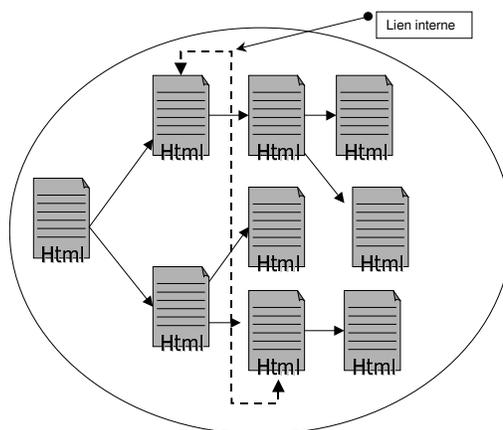


Figure 4 : Lien interne de navigation transversale au sein d'un site virtuel

Mode auteur : Annotations et génération de sites

Annotations sur les nœuds

Le réseau virtuel personnel présente l'espace Web de l'utilisateur. HyWebMap permet aux utilisateurs d'enrichir leurs réseaux virtuels par la création de leurs annotations personnelles sur un ou plusieurs nœuds. Cette annotation peut être un texte, des liens ou un texte avec des liens. Le réseau virtuel se trouve annoté alors avec un réseau de notes.

Une annotation comprend trois champs distincts :

- un champ **commentaire** pour les commentaires libres (fond) de textes, d'images, etc. ou un lien.
- un champ **mots-clés** sans nécessairement faire partie du nœud (les mots-clés peuvent être créés par l'utilisateur et peuvent être utilisés lors des interrogations sur le réseau.
- un champs pour les **variantes** (synonyme ou correction de syntaxe) (forme).

Pour un nœud donné, chacun de ces champs peut être rempli ou non. Ils sont complètement indépendants. Cette distinction entre les annotations est utile à des fins de recherche. Les annotations sont au format HTML. Dès qu'un nœud est re-visité, son annotation qui est attachée apparaît dans le navigateur.

Génération automatique de sites Web

Au delà de la construction du site virtuel répondant à ses propres exigences structurels, HyWebMap peut sur demande de l'utilisateur instancier le schéma du site virtuel sous la forme d'un site Web autonome (cf.figure 5) qui reprendra les propriétés de chacun des nœuds et des liens (nom, descripteurs auteur, descripteurs lecteur, annotations,..) et dont l'accès sera facilité par un sommaire actif (cf. figure 6).

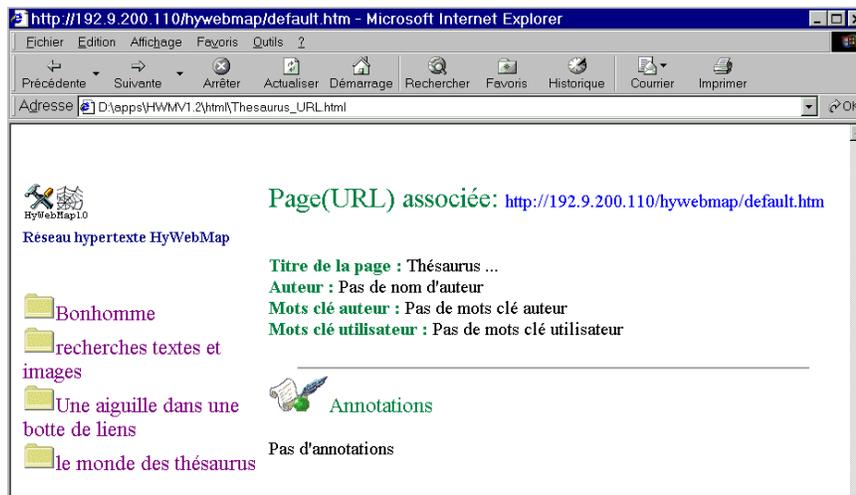


Figure 5 : Conversion HTML d'un nœud du réseau virtuel avec tous les liens associés.

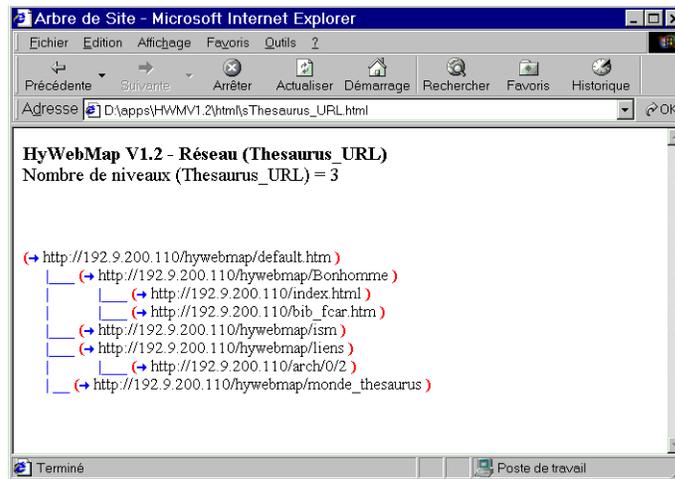


Figure 6 : Sommaire actif d'un site virtuel converti.

Conclusion

L'évolution des Technologies de l'information et de la Communication, libérées des contingences techniques coercitives de l'informatique a ouvert des perspectives extraordinaires en matière d'accessibilité à l'information numérique (accessibilité liée de façon univoque à la consultation mais où l'aisance voire même la facilité se retrouve également dans la production de l'information). La fin des années 80 a vu apparaître pléthore de dispositifs logiciels (KMS, InterMedia, MetaCard, HyperCard, InterSect, NoteCards, HAM, Guide, ...) empruntant les fonctionnalités de l'hypertexte, exploitant les récentes avancées en matière de technologies de réseaux et destinées à faciliter le transfert et la consultation d'informations multimédia.

Néanmoins, les débuts de la décennie 90 n'ont pas vu de véritables solutions logicielles s'imposer, freinées par les difficultés de migration des fonds documentaires électroniques existants (hypertextualisation) et par les problèmes méthodologiques résultant de la conception de systèmes d'informations hypertextes.

Cette difficile période de mutation technologique de l'information a été extrêmement riche en réflexions et interrogations quant aux profils des utilisateurs et à leurs objectifs de recherche. Une géométrie de type de systèmes allant de ceux spécifiquement "auteur" à ceux résolument "lecteur" a ainsi témoigné des diverses approches favorisant la production de documents multimédias, le point de vue de l'auteur, du lecteur, l'accessibilité de l'information, l'organisation, les environnement collaboratifs, etc.

Le déploiement tentaculaire du World Wide Web a définitivement contrarié ces différentes approches en amorçant une ère de production industrielle de données numériques. L'explosion démographique de sites Web atteste de "*l'opérationnalité*" des systèmes logiciels de production d'information et des efforts de développement faits en ce sens. Du point de vue du lecteur, les outils de recherche n'ont guère évolué, on retrouve dans les versions "modernes" des navigateurs Internet les fonctionnalités stéréotypés du précurseur Mosaic, fruit des travaux du NCSA. On ne peut que constater que l'accroissement du volume d'informations accessibles en ligne ne s'est nullement accompagné d'outils qualitatifs susceptibles d'assister efficacement l'utilisateur dans sa recherche d'informations. Les problèmes de désorientation dans les espaces de données électroniques et de surcharge cognitive qui perdurent jusqu'à présent montrent bien que le modèle de l'utilisateur considéré comme un simple lecteur "presse-boutons" est définitivement obsolète.

Les fonctionnalités du logiciel HyWebMap qui ont été présentées introduisent des variations dans la manière d'appréhender les systèmes auteur. Sans pour autant se distinguer comme un véritable outil de publication, HyWebMap dépasse largement le contexte d'utilisation des systèmes essentiellement lecteur. La capacité à pouvoir organiser des sources d'informations distinctes et distantes, de les compléter par des éléments

d'informations personnels (pages personnelles, annotations) font d'HyWebMap un produit hybride entre les 2 tendances de systèmes. Cependant, du seul point de vue de l'opération d'agrégation de ressources documentaires qui conduit à produire un nouvel espace de données cohérent et sémantiquement homogène au sein d'un même réseau maillé, il est incontestable que le logiciel relève bien d'un comportement auteur. Les possibilités de génération de sites web inédits accentuent encore cette tendance.

Les réseaux virtuels élaborés par les auteurs, instanciés ou non en sites Web, forment des "parcours de navigation experts" qui pourront être mis à la disposition de nouveaux utilisateurs, qui eux-mêmes s'en inspireront éventuellement pour concevoir à leur tour de nouveaux réseaux et cela dans une dynamique de "consultation/production" sans fin. Ce mode de fonctionnement introduit les prémisses d'une utilisation collaborative d'HyWebMap dans la création et la consultation d'espaces de données. Dans la situation d'implémentation actuelle d'HyWebMap, il est clair que cette forme d'activité collaborative reste extrêmement dépendante de la rigueur de tous les auteurs participant activement aux évolutions d'un même espace virtuel. C'est cette rigueur qui sera la garantie de préserver la cohérence et l'intégrité des informations apportées au fil des opérations d'actualisation, sans laquelle la désorganisation et l'incohérence guetteront le réseau virtuel collectif en l'absence de contrôle de synchronisation et de version dans l'actuelle version

Bibliographie

- ASTOLFI J-P., *Comment les élèves apprennent-ils ?*, Sciences Humaines, n° 32, octobre 93, pp. 26-29.
- BALPE J.P., LELU A., SALEH I., PAPY F., *Techniques avancées pour l'hypertexte*, Editions Hermès Paris 1996.
- BELISLE C., Zeiliger R., Cerratto T., *S'orienter sur le Web en construisant des cartes interactives avec Nestor*, Actes H2PTM 99, Editions Hermès 1999.
- BROWN P.J., *Hypertext : the way forward*, éditions J.C. Van Vliet, *Document manipulation and typography, Proc. of the International Conference on Electronic Publishing*, Nice, avril 1988, pp.183-191.
- CASTELLI C., et al. *Getting lost in Hyperspace : Lessons learned and future direction*, in ED-MEDIA 96/ED-TELECOM 96, 1996.
- DUSSAUX G., PECUCHET J-P., *IronWeb : une architecture distribuée pour la création de connaissances pédagogiques sur le Web*, in Technologies de l'Information et de la Communication dans les Enseignements d'ingénieurs et dans l'Industrie, 18-20 octobre 2000, Troyes, France
- ROUET J.F., « Le lecteur face à l'hypertexte », in apprendre avec le multimédia Où en est-on ? Sous la direction de GRIMONT A., Paris Editions Retz, 1997.

