

Chercher et réorganiser l'information sur le web

Fabrice Papy, Nasserddine Bouhaï, Imad Saleh

▶ **To cite this version:**

Fabrice Papy, Nasserddine Bouhaï, Imad Saleh. Chercher et réorganiser l'information sur le web. Cinquième colloque Hypermédias et apprentissages, Apr 2001, Grenoble, France. pp.49-58. edutice-00000451

HAL Id: edutice-00000451
https://edutice.hal.science/edutice-00000451

Submitted on 8 Jun 2004

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CHERCHER ET RÉORGANISER L'INFORMATION SUR LE WEB

Fabrice PAPY*, Nasserddine BOUHAÏ et Imad SALEH****

Université Paris VIII, Équipe Hypertexte dynamique,
Départements Documentation* et Hypermédia**,
2, rue de la liberté - 93526 Saint-Denis Cedex 2

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

Résumé : *L'information n'a jamais été aussi facile d'accès. Mais l'augmentation du volume d'informations s'est accompagnée d'un accroissement considérable d'outils destinés à aider ou assister l'utilisateur final à retrouver des informations. Les différentes logiques de fonctionnement de ces outils augmentent les problèmes cognitifs de l'utilisateur : lecteur, étudiant, apprenant... L'utilisateur doit d'une part trouver une cohérence dans les informations pléthoriques qu'il est amené à extraire des réseaux et d'autre part s'adapter aux mécanismes des outils qu'il est amené à utiliser. HyWebMap représente une nouvelle génération d'outils destinée à limiter les problèmes de désorientation et de surcharge cognitive des utilisateurs (étudiants) placés en situation d'apprentissage. L'objectif étant de leur permettre de reconstruire un espace de navigation personnalisé à partir d'éléments importés mais également d'éléments produits par eux.*

Mots-clés : *HyWebMap, écriture, système auteur, système lecteur, formation*

Abstract : *Information has never been as accessible as nowadays. But the increase of the volume of information has been accompanied by a considerable increase of tools that aim at helping and assisting the final user to find information. Different skills used in such tools increase cognitive problems of the user (reader, student, beginner). On the one hand, the user has to find a coherence in plethoric information as he browses through information networks, and on the other hand he has to adapt to the different logics of these tools. HyWebMap represents a new generation of tool designed to reduce disorientation and cognitive overhead of users being in a learning context, the goal being to allow them to organize a customized navigation space built around imported data and information produced by the user himself.*

Keywords : *HyWebMap, authoring, author system, reader system, education*

INTRODUCTION

Comment peut-on souligner le fait que l'information n'a jamais été autant accessible tout en s'accordant à reconnaître que le nombre d'outils destinés à la recherche d'informations – et l'avènement du Web a, du reste, considérablement accentué ce phénomène – ne cesse de proliférer ?

L'explosion de l'offre logicielle est-elle simplement l'expression d'une logique économique de surenchère technique ou au contraire révèle-t-elle une sensibilisation des utilisateurs aux outils les mieux adaptés à leur logique de recherche ?

Peut-on parler véritablement d'interactivité dans ces offres techniques lorsque l'on constate que la capacité d'interagir proposée à l'utilisateur-lecteur se réduit souvent à l'interaction avec les fonctionnalités logicielles des outils au détriment de l'interactivité souhaitable avec l'information qu'il est amené à manipuler ?

Ces singulières contradictions nous ont amené à proposer une nouvelle génération d'outils dont la vocation première n'est pas de se substituer à l'indispensable volonté d'apprendre qui doit animer tout individu placé en situation de recherche d'informations mais de lui proposer un contexte d'épanouissement cognitif où la désorientation et la surcharge cognitive seraient absentes. HyWebMap qui trouve une application flagrante dans le domaine de l'enseignement – et dont nous projetons l'utilisation dans des 1er cycles universitaires en sciences sociales et en sciences humaines – se caractérise par la capacité de produire des espaces de recherche d'informations personnalisés répondant à des exigences d'apprentissage de la recherche documentaire au sein de corpus dédiés.

UNE ORIENTATION DIDACTIQUE

Les travaux de l'équipe de recherche *Hypertexte Dynamique*[1] des départements Hypermédias et Documentation de l'Université Paris 8 ont pour dénominateur commun la conception et l'élaboration de systèmes hypermédias à vocation didactique prise ici dans son acception la plus large.

En effet ces systèmes ne s'adressent pas exclusivement à des pratiques éducatives qui relèvent préférentiellement de la formation initiale ou de la formation continue selon une dichotomie qui veut que les méthodes de l'éducation et de la formation continue aient peu de choses en commun, même si toutes les formes de transmission et d'acquisition des compétences et des savoirs relèvent d'une même souche identitaire (Pelletier, 1996 ; Balpe *et al.*, 1996).

Ces dispositifs élaborés par notre équipe répondent de manière générique aux situations d'apprentissage et d'acquisition de connaissances que l'on retrouve :

- dans des contextes d'accompagnements didactiques tels qu'on peut les rencontrer à l'école, à l'université ou en formation continue. C'est cet aspect que

1 Voir le site h2ptm.hymedia.univ-paris8.fr pour une présentation des membres, réalisations et projets en cours.

nous soulignerons ici au travers de l'utilisation du logiciel HyWebMap auprès d'étudiants de 1er cycle universitaire ;

- en situation d'auto-apprentissage (on parlera plutôt ici d'actualisation de compétences) ;
- dans le cadre de pré-traitement par des professionnels de l'information et de la documentation (documentalistes, ARIST, SVP, CCI...) d'éléments d'information destinés à un usage opérationnel.

ADÉQUATION ENTRE L'OFFRE ET LA DEMANDE D'INFORMATIONS

Les outils que nous développons au sein de notre équipe et dont les prémices s'attachent aux travaux (http://hdyn.hymedia.univ-paris8.fr/neuroweb) conduits dans le cadre du laboratoire Paragraphe de l'Université Paris 8 (Lelu *et al.*, 1999), ne cherchent pas à s'imposer comme des réponses péremptoires à la recherche et à la localisation d'informations à une époque où la numérisation conduit à trouver plus d'un milliard de pages dans les sites Internet. Notre démarche s'inscrit, en amont, dans une réflexion portant sur la pertinence, l'adéquation entre ces différents outils - et les principes moteurs dont ils sont issus – avec une approche ouverte dans l'acquisition de savoirs et de connaissances.

DESCRIPTION DU SYSTÈME HYWEBMAP

Agent Web du HyWebMap

Dans le cadre de notre système, nous avons dû concevoir et développer un outil informatique (dit agent) capable à partir d'une page HTML ou d'une Uniform Resource Locator (URL) quelconque choisie par un utilisateur, de faire l'analyse de cette page, et d'en répertorier ses principales caractéristiques, puis de supprimer les liens internes et externes au site qui sont obsolètes, et enfin de sauvegarder ces informations dans une base de données. L'agent peut détecter quels sont les liens de départ de la page vers d'autres pages du même site (liens internes) ou vers d'autres sites (liens externes). L'agent est paramétrable, et permet à l'utilisateur de définir le parcours de l'agent en profondeur (verticale) et/ou en largeur (horizontale).

Création de la base de données

Le site subit une analyse complète par l'agent Web et sera décrit sous forme de tables et de relations entre des ensembles de nœuds, de liens, d'ancres, de profils, et de propriétés. Nous utilisons la représentation relationnelle pour décrire le contenu des nœuds. Les relations Lien (L), Nœud (N) Ancres (A) et Profils (Nombre-de-caractères, Nombres-d'images, Nombre-de-liens-externes, Nombre-de-liens-internes, etc.) représentent respectivement les liens, les nœuds, les ancres, les profils et les caractéristiques du nœud (propriétés). La relation prédéfinie liaison (Document-A, Document-B, Lien-L) contient tous les nœuds du Document-A qui

sont reliés par un lien-L à un Document-B. La relation Propriété décrit les caractéristiques informationnelles d'un nœud : Titre, Date-Création, Mots-clés, Auteur, etc.

Navigation par réseaux interactifs

La carte HyWebMap dite *réseau virtuel* présente l'espace du site visité (les nœuds et les liens) par l'agent Web. La navigation peut-être globale ou locale, HyWebMap présente l'ensemble des nœuds et des liens de l'espace visité ou uniquement les liens attachés à un nœud particulier.

Les utilisateurs peuvent la modifier et y ajouter des nœuds et des liens personnels. Les liens personnels peuvent être ajoutés soit en utilisant les fonctions copier/coller, soit en reliant graphiquement les nœuds dans la carte par un simple traçage d'une flèche entre deux nœuds. Ainsi est créé un lien hypertextuel. Les liens personnels sont représentés par des flèches de couleur rouge qui les distinguent des liens de la carte visitée. Les liens personnels ne sont pas intégrés dans les pages sources, ils sont seulement stockés sur le disque dur de l'utilisateur et sont actifs lorsque le réseau est invoqué.

Nota : Une page Web contient une séquence de fragments de type MIME (Multipurpose Internet Mail Extension) qui peut être une séquence de texte, de son, de vidéo. HyWebMap considère chaque fragment comme un nœud.

Il est généralement admis qu'un réseau ou une carte est d'autant plus pertinent qu'il est construit par l'utilisateur. Cette construction montre l'engagement actif de l'utilisateur dans les processus d'apprentissage. Dans le processus de recherche d'information pertinente sur le Web, l'utilisateur a besoin d'outils pour l'aider à externaliser et structurer sa pensée qui est en interaction avec l'information trouvée.

La création d'un espace personnel

Le système HyWebMap permet la création d'un méta-réseau virtuel personnel à partir d'autres réseaux virtuels. Ainsi, il permet aux utilisateurs d'enrichir leurs méta-réseaux par des liens et des nœuds personnels. Les nœuds peuvent être des documents publics Web (les sites visités) ou des documents privés (des pages Web personnelles stockées sur le disque de l'utilisateur).

Le méta-réseau est enregistré sur le disque dur de l'utilisateur, que celui-ci a nommé lors de sa création et il prend l'extension .hwm. À ce stade le méta-réseau est considéré à part entière comme un réseau virtuel. L'utilisateur peut : 1) naviguer localement dans le réseau virtuel sans établir de connexions à internet, ce qui économise des temps d'accès. Lorsque l'utilisateur désire visualiser une page, HyWebMap invoque le navigateur installé sur sa machine, et peut basculer à tout moment entre son navigateur et son réseau virtuel, 2) réorganiser la structure d'un site visité par rapport à ses connaissances personnelles sur le domaine, sans mettre en cause la structure du site d'origine, 3) construire des réseaux virtuels thématiques en fonction de ses centres d'intérêts. La figure 1 présente une partie d'un réseau virtuel dans HyWebMap. C'est un réseau thématique concernant l'hypertexte. Il est constitué de cinq niveaux.

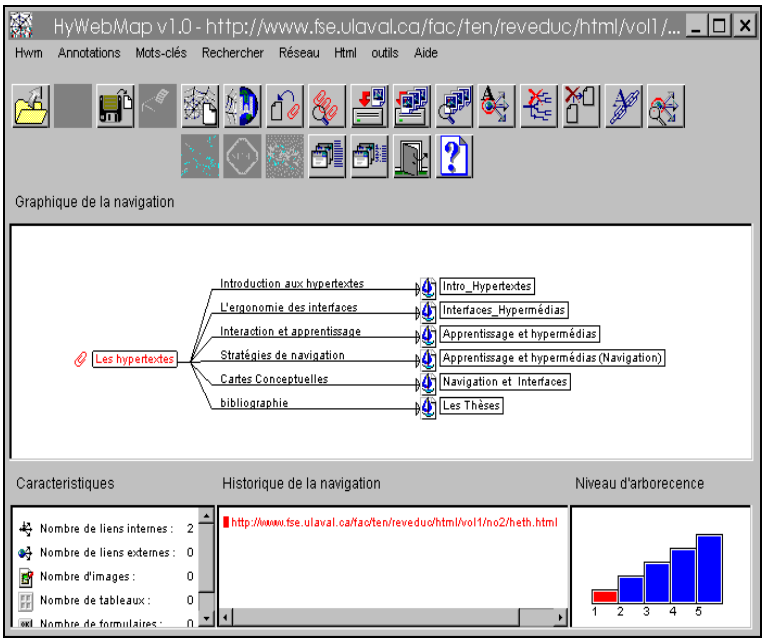


Figure 1. ***: Réseau virtuel dans HyWebMap.***

Agents de surveillance du réseau virtuel

Nous avons développé des agents de surveillance qui contrôlent périodiquement la disponibilité des nœuds du réseau virtuel dans leurs sites d'origine. Dans le cas où un nœud ne serait plus disponible, ils effectuent la modification sur le réseau virtuel et alertent l'utilisateur.

Renommage des nœuds et des liens

Dans le système HyWebMap, nous avons introduit le concept de renommage des nœuds et des liens entre des nœuds pour que l'utilisateur rende son réseau compréhensible. Il peut nommer les liens entre les nœuds en fonction de ses connaissances sur le contenu de ces pages. Ce mode de construction du réseau virtuel favorise chez l'utilisateur une navigation plus réflexive et planifiée. La figure 2 présente les adresses des pages Web avant le renommage des nœuds par l'utilisateur.

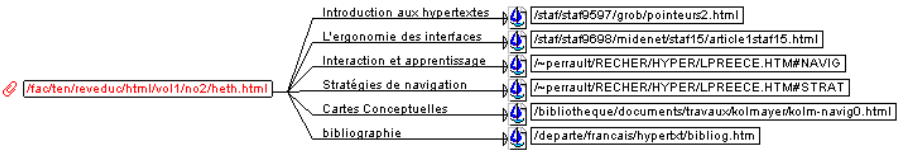


Figure 2. ***Les adresses des pages Web avant le renommage des nœuds.***

Les outils d'aide à la navigation

HyWebMap est doté de plusieurs outils pour aider l'utilisateur à naviguer dans son réseau, comme : La *navigation par historique*, c'est un outil classique, car il permet à l'utilisateur de revenir sur un nœud déjà visité. La *navigation par niveau d'arborescence*, l'utilisateur peut choisir le niveau souhaité à visiter dans l'arborescence du site. La *navigation par date*, le système HyWebMap sauvegarde pendant deux semaines l'historique de navigation du réseau virtuel. Il peut demander à voir les nœuds visités durant un temps défini entre deux dates. HyWebMap présente un réseau virtuel navigable des nœuds déjà visités, ce qui supprime de fait les doublons. Ce qui n'est pas le cas pour les navigateurs standards qui ne présentent seulement qu'une simple liste des nœuds visités dans l'axe du temps.

Création d'annotations, de points de repère et mots-clés

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 avec un réseau de notes (trombone rouge = une ou plusieurs annotation/s sur un nœud).

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.

L'utilisateur a la possibilité de mettre un *point de repère* sur un nœud pour signaler son importance dans le réseau et ainsi se repérer dans celui-ci.

Génération d'un site Web à partir du réseau virtuel de HyWebMap

Une fonctionnalité intéressante fournie par le système HyWebMap est le module permettant la génération transparente du réseau virtuel au format HTML. Elle permet d'enregistrer et d'exporter le réseau virtuel (liens, annotations) au format HTML qui constitue une caractéristique très utile pour la consultation du réseau par un navigateur standard.

Le site généré prend l'extension « .html ». Dès qu'il est chargé, HyWebMap génère deux frames. À gauche se présente la frame de navigation (le nom du lien entre deux nœuds) et à droite la frame présente les profils de la page source : Nom du réseau, Titre de la page, Url de la page source, les mots-clés, nom de l'auteur,

date de la création (Pour les trois derniers, ils sont affichés à condition que la page contienne des méta-balises pour les détecter). HyWebMap affiche les annotations personnelles en bas de la page. La figure 3 présente la génération du réseau virtuel en site html.

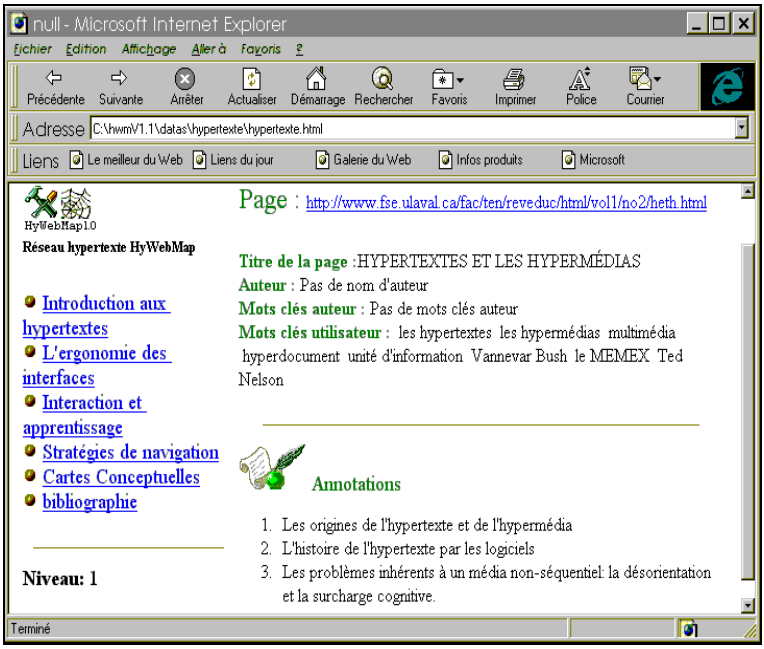


Figure 3. ***La génération du réseau virtuel en site html.***

HYWEBMAP APPLIQUÉ À LA RECHERCHE DOCUMENTAIRE

Soulignons que cette approche de la recherche documentaire au moyen d'un outil comme HyWebMap (Saleh, Bouhaï & Papy, 2000) ne vise pas à remplacer la fréquentation et l'utilisation des bibliothèques universitaires ou les centres de documentation spécialisés. Certes, HyWebMap a été conçu pour gérer des données électroniques véhiculées par le Web, mais il est clair que les lieux ressources disposent de fonds documentaires dont la transposition numérique relève de la pure utopie et qui laissent penser à juste titre que leur fréquentation demeurera un impératif dans l'acquisition du « métier d'étudiant ».

La méthodologie de la recherche documentaire a pour objectif principal de faire acquérir aux étudiants des méthodes de travail et d'étude indispensables à leur réussite universitaire et professionnelle. Cette méthodologie ne concerne pas le travail de recherche documentaire stricto sensu mais introduit également et plus génétiquement les étapes de la formation au travail intellectuel.

Ainsi, apprendre aux étudiants à définir une problématique implique une capacité à :

- délimiter un contexte général : période, zone géographique, domaines disciplinaires, auteurs, thèmes concernés...
- déterminer et sélectionner des documents pertinents de par leur qualité scientifique dans les domaines concernés ;
- cerner le sujet au moyen de rapprochements et d'interrogations ;
- prendre en compte les aspects secondaires (variantes) soulevés par les documents retenus et introduire des hiérarchies ou des nuances.

Leur capacité à organiser les informations qu'ils recueillent en fonction de leur objectif de recherche sera déterminante dans la définition d'une problématique universitaire conforme. C'est ainsi qu'il leur faut apprendre à distinguer :

- les sources : textes et données qui seront sujets à analyse et qui formeront la base de leur travail ;
- les commentaires : études et ouvrages qui proposent des synthèses, des critiques, des analyses, des interprétations de ces textes et données ;
- les opinions et commentaires personnels.

Ces objectifs avoués dans la méthodologie de la recherche documentaire vont directement impliquer l'étudiant de 1er cycle en lecteur/acteur, en lecteur/critique capable d'apprécier la nature de l'information qu'il découvre, d'en apprécier la contextualité, de l'organiser et de la réutiliser afin d'alimenter une problématique qu'il aura été amené à traiter. Ces activités intellectuelles effectives se manifestent alors sous de multiples formes : annotations, résumés, synthèses, fiches de lecture...

L'étudiant se retrouve donc confronté et contraint à organiser sa propre production d'écrits, aussi modeste soit-elle.

HyWebMap répond à cette double activité de lecture et d'écriture mise en œuvre dans l'organisation des éléments d'argumentation destinés à répondre à la problématique définie, en permettant à l'étudiant de se concevoir un réseau constitué des documents (l'intégralité de ceux-ci ou plus succinctement l'ensemble des références qui les caractérisent) qu'il aura pu glaner au cours de ses recherches bibliographiques. Il aura la possibilité d'annoter chacun de ces documents, de créer des liens entre les différents éléments ainsi représentés et de les typer pour faciliter les lectures ultérieures.

L'étudiant pourra ainsi orienter le futur lecteur en préparant des parcours de lecture dans le réseau qu'il aura constitué, parcours de lecture qui représenteront sa ou ses analyses de la problématique à laquelle il aura été confrontée. Ses espaces de connaissances pourront ainsi de manière récurrente faire l'objet de commentaires et d'annotations de la part des lecteurs.

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 environnements 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'opérationnabilité* 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 stéréotypées du précurseur Mosaïc, 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.

Le dispositif logiciel HyWebMap au-delà de ses caractéristiques techniques et de ses modalités de mise en œuvre replace l'utilisateur dans son rôle principal de lecteur/acteur au sein d'un espace de connaissances qu'il construit et qu'il enrichit au fil de ses lectures.

Les différentes situations industrielles et pédagogiques auxquelles nous avons confronté HyWebMap nous encouragent à poursuivre cette expérimentation que nous orienterons vers les métiers de l'information. Les professionnels de l'information sont les premiers à attendre des outils capables de bâtir des systèmes d'informations dédiées avec lesquels leurs usagers seront amenés à interagir

véritablement. Adapter HyWebMap à ce public formé actuellement en 2ème et 3ème cycles de Documentation à l'Université Paris 8 fait partie des orientations effectives de notre équipe de recherche.

Cette adaptation se poursuivra de manière corollaire, d'un point de vue technique par une intégration de fonctionnalités 3D qui offre des avantages évidents en terme de navigation dans les espaces virtuels de données.

BIBLIOGRAPHIE

Balpe J.-P., Lelu A., Saleh I. & Papy F. (1996). *Techniques avancées pour l'hypertexte*, Paris : Hermès.

Lelu A. *et al.* (1999). « Projet Neuroweb : un moteur de recherche mutilingue et cartographique », *in* J.-P. Balpe, A. Lelu, S. Natkin & S. Saleh (dir.), *H2PTM'99*, Hermès, p. 43-55.

Pelletier G. (1996). « Un spécificité pour la formation continue ? », *Sciences humaines*, n° 12, février 1996.

Saleh I., Bouhaï N., Bouyahi S. & Papy F. (2000). « Système HyWebMap : Outil de navigation et de restructuration des connaissances », *Actes du colloque « Technologies de l'Information et de la Communication dans les Enseignements d'Ingénieurs et dans l'Industrie »*, Troyes, 18, 19, 20 octobre 2000.