

## HyWebMap<sup>1</sup> : un outil d'agrégation, d'organisation et de veille technologique pour le Web

Imad SALEH\*, Fabrice Papy\*\*, Nasserddine BOUHAÏ\*

Université Paris 8, Laboratoire Paragraphe, Département Hypermédia\*, Documentation\*\*  
2 Rue de la liberté, 93526 Saint-Denis Cedex 2, Tél. : 01 49 40 67 58, Fax : 01 49 40 67 83  
Email : isaleh@wanadoo.fr, {papy, nasser}@labart.univ-paris8.fr

### 1. Introduction

Le Web est un espace d'information ouvert, évolutif, hétérogène et non modéré. La navigation dans cet espace est un processus qui nécessite l'activation de connaissances du domaine, de connaissances des rhétoriques qui portent sur l'organisation typique des pages et des connaissances stratégiques qui portent sur la régulation de l'activité de compréhension en fonction des objectifs et des informations rencontrées [ROU 97].

La pertinence d'un service d'information comme le Web dépend de ses capacités à identifier les besoins précis de l'utilisateur dans un contexte donné. Souvent l'utilisateur ne sait pas où chercher l'information, ni comment la rechercher ou encore comment exprimer le but de sa recherche. Les agents intelligents peuvent donc venir à son secours et accomplir plusieurs tâches sur le Web : apprendre un profil d'utilisateur et le guider ensuite dans l'espace du réseau multidimensionnel, retenir les préférences de l'utilisateur et lui proposer des pages intéressantes à visiter, à l'aider à chercher l'information nécessaire, etc. Web est en pleine mutation : envisagé jadis comme "une boîte remplie des liens" il est maintenant considéré de plus en plus comme une "société d'agents" dans laquelle chaque agent est responsable de ses propres tâches et se dirige par sa propre expertise pour mieux satisfaire l'utilisateur concerné.

Nous classons en cinq grandes fonctionnalités les différents types de besoins des utilisateurs :

- **Extraction de l'information** : sur des bases de critères exploitant les résultats des moteurs de recherches existants, l'utilisateur extrait des informations répondant à ses préoccupations, et nous complétons cette extraction par une exploration de catégories arborescentes.
- **Traitement de l'information** : le système fournit à l'utilisateur des outils lui permettant de classer, d'ôter, de répartir, d'afficher, de trier, de filtrer et d'organiser les informations obtenus lors de l'extraction.
- **Veille sur l'information** : l'objectif est de systématiser la recherche d'information et les traitements associés. L'utilisateur introduit une seule fois sa demande, et reçoit ultérieurement de façon automatique des mises à jour régulières sur la base de la demande initiale. Ainsi on peut suivre régulièrement la modification d'un document.
- **Suggestion d'information** : Au moyen de techniques de recoupement automatique d'information plus sophistiquées, le système propose d'une manière automatique ou semi-automatique à l'utilisateur, des informations auxquelles celui-ci n'aurait pas pensé.
- **Personnaliser de l'information** : le système fournit des outils permettant aux utilisateurs de personnaliser l'espace visité en modifiant le parcours de navigation, de renommer les liens entre les documents, d'annoter des documents, d'ajouter des mots-clefs et des descripteurs sémantiques.

Dans cet article, nous présentons le système HyWebMap. Il est développé au laboratoire Paragraphe de l'université Paris 8 avec le langage Java. La version 1.0 de HyWebMap est téléchargeable sur le site : <http://www.labart.univ-paris8.fr/~nasser/hwm.html>

## **1. Problèmes de la désorientation et la 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 demeura 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 de la structure de l'hypertexte, se repérer plus facilement et éviter ainsi de se perdre. Des systèmes tels que NoteCards, Intermedia, StorySpace, Extended Books, et Neptune proposent des "navigateurs graphiques" qui, tels une carte routière, permettent de se situer dans un espace. Un navigateur peut être global ou local selon qu'il représente l'ensemble des nœuds et des liens du réseau hypertexte, ou uniquement les liens rattachés à un nœud particulier. Dans [BRO 88a, BRO 88b, BRO 89] on a cherché à montrer que si la représentation cartographique de la structure du réseau hypertexte pouvait effectivement faciliter la navigation de lecteur et limiter les dangers de la désorientation, elle n'en constituait pas pour autant la seule solution.

En plus des procédés techniques destinés à pallier le problème de désorientation comme le retour en arrière et l'historique de navigation, les chercheurs [BRO 88a, NIL 90] 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. Les milieux professionnels susceptibles d'utiliser un système hypertexte peuvent être aussi variés que les disciplines dont l'hypertexte représente l'information et ils n'ont pas nécessairement l'intuition du fonctionnement des outils hypertextes.

— **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, par exemple comment il a défini la hiérarchie et comment il l'a enrichie par les liens supplémentaires. La perception qu'il peut avoir de la manière dont l'information est organisée peut alors entrer en collision avec le modèle de conception de l'auteur et ainsi expliquer des activations de liens inadéquates.

— **la navigation dans le texte** : le lecteur se perd de la même manière qu'un automobiliste se perd en prenant successivement des routes à droite et à gauche en espérant reprendre son itinéraire initial. 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.

Plus précisément, [FOS 89] a mené une étude empirique visant à déterminer les conditions de désorientation du lecteur — taille du réseau, stratégie d'apprentissage et de gestion d'objectifs, etc.— en dégagant les symptômes observables : nombre de fenêtres ouvertes simultanément, retour systématique sur les mêmes nœuds, fermeture rapide des fenêtres, etc....

Ce constat ne doit pas cependant conduire à considérer la désorientation comme l'inconvénient majeur d'une consultation hypertextuelle. En effet, la désorientation se retrouve tout autant dans le cas d'un livre dont la bibliographie intéresse le lecteur. Si celui-ci consulte les ouvrages relatifs à cette bibliographie et récursivement les bibliographies de ces ouvrages, le phénomène de désorientation par rapport à l'espace de connaissances s'installe de la même manière.

Enfin, au problème de désorientation, vient s'ajouter un phénomène de **surcharge cognitive** (*cognitive overhead*) [AST 93] qui provient de l'incapacité du lecteur à retenir l'essentiel de l'information liée à un nœud parcouru ou de ses difficultés à identifier rapidement la nature des liens qu'il a activé. [ROU 97, CAS96] 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.

### **Personnalisation les pages Web par l'annotation**

L'annotation personnelle de pages Web devient plus en plus nécessaires pour contrôler la prolifération des signets et d'améliorer sa productivité d'utilisation des pages (annoter la partie

importante d'une page permet par exemple de décider de son intérêt sans le charger). Enfin, permet de considérer et échanger les interprétation sur une page ou sur une idées.

L'annotation à reçu une attention particulière dans le cadre de Web ou en document électronique pour :

- Décrire le contenu sémantique des pages,
- Aider à le recherche des informations,
- Résumer le contenu d'une page,
- Personnaliser les informations.

Cette attention a été possible grâce à la convergence de deux facteurs principaux :

A) La faisabilité technique liés au contexte de l'hypertexte qui autorise l'utilisateurs d'une pages à adjoindre à cette page diverses argumentations individualisées, telles que caractérisation de passages, attachement de commentaires, pose de liens entre passage etc. le fait que ces annotations soient enregistrer sur un support numérique en tant que telles signales d'emblée quatre avantages importantes par rapport à la forme traditionnelles de l'activité d'annotative :

- L'ensemble des annotations est exploitable à part, par exemple dans le cadre du travail de rédaction subséquent à la consultation de page. Ce contexte garantit une forme intéressante de continuité lecture-écriture, peu ou non réalisable autrement.
- La relecture, et plus largement les consultation ultérieures de page peuvent être effectivement conduite à travers les annotations. C'est l'un des objectifs à recherchés dans le cadre des annotation classique mais qui est souvent soumis aléas de la mémoire personnelle et aux difficulté de l'accès transversal dans le cadre du livre. On peut ainsi rechercher toutes les annotations où employée telle formule; ou bien collecter et constituer en unité unique tous les commentaires de tel type que j'ai associé à telle ou telles pages.
- La réorientation une relecture particulière en fonction des annotation antérieures, par exemple en ne visualisant que le passage de tels types, et si on le souhaite avec telles propriétés visuelles (en gras, en italique, en rouge, etc.).
- La possibilité exploratoire ou expérimentale de la lecture est instrumentale : par exemple, pour tout passage préalablement défini, il est possible de rechercher s'il existe d'autre passages déjà créés présentent telle ou telles propriétés commune avec ce passage de référence, ou bien de créer automatiquement de tels passages s'ils ne l'ont pas déjà été.

B) La possibilité technique liés au contexte du Web qui représente une vastes du fonds numérisés et la multiplication des profils des utilisateurs potentiels de ces fonds, conduisent à rechercher à privilégier des modalités de gestion personnalisés de ces fonds. L'annotation apparaît comme une de ces modalités de la personnalisation de l'appropriation de page (document) à travers sa lecture, parce qu'elle, dans ses formes externes de manifestation comme dans ses motivations, une technique hautement individualisées.

L'annotation présente du point de vue fonctionnel [VIR 94] comme une technique empirique et individualisée de mémorisation et de capitalisation de résultats de lecture, dès lors utilisable

dans la suite même de la lecture en cours, mais aussi pour la gestion de relecture ou de consultations ultérieurs, et pour facilitation de l'accès transversal ou diagonale du texte. Il s'agit donc d'une écriture intrinsèquement liée à la lecture, sous trois rapports :

- Dans le temps : elle est contemporaine de la lecture et suscitée par celle-ci;
- Dans l'espace : ses traces sont intégrées aux espaces libres de la page ou de livre;
- Dans l'organicité : le texte annotatif n'est pas en général indépendant du texte annoté, mais se rapport au contraire directement à tel passage, formule, ou point du texte lu, comme le sont, chez l'auteur cette fois, certains type de textes greffés (notes bas de page, parenthèses, etc.) l'intelligibilité des commentaires en lecture autonome marquant cette limitation.

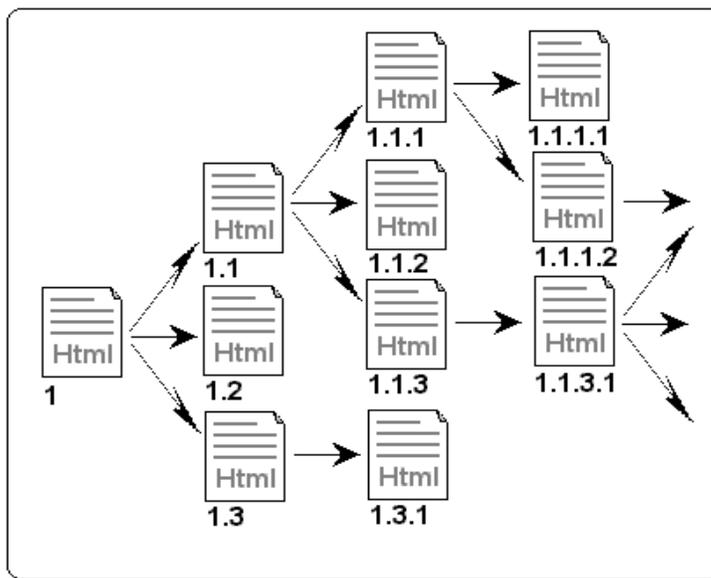
On peut constater que les notions d'indexation et annotation semblent équivalents : on se trouve dans tous les cas avec une indexation conceptuelle du documents qui sera utilisé dans le cadre d'une tâche exploitation. La différence essentielle entre indexer et annoter se situe au niveau de processus d'exploitation, l'indexer, c'est un processus pour décrire un documents pour le retrouver en revanche annoter c'est un processus mentale pour interpréter un document et sauvegarder des traces de lectures. Ces traces peuvent être destinées à soi-même ou bien partager et leurs mise en place peut se faire de façon plus ou moins automatique et plus ou moins collaborative.

## **2. Description du système HyWebMap**

### **2.1. Agent Web du HyWebMap**

Dans le cadre de notre système, nous avons choisi de 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 lecteur, 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 prend en compte les adresses absolues (adresses complètes), il analyse l'ensemble des pages HTML et sauvegarde les liens relatifs, les liens internes, et les liens externes permettant d'accéder à un paragraphe ultérieur sur une même page HTML.



**Figure 1.** Propagation des liens

### Propagation des liens

Phase 1 : A partir d'un document initial, l'agent détecte tous les liens disponibles sur cette page (exemple : les k premiers liens dans la figure 1), tout en stockant dans une base de données les adresses des liens rencontrés.

Phase 2 : il explore tous les liens répertoriés lors de l'analyse de la première page, soit les liens 1.1 à 1.3 dans la figure 1. Les nouvelles adresses des liens trouvées sont sauvegardés dans la base de données.

Phase 3 : l'agent appelle la page suivante dans la liste, donc la page 2 de la figure 1 et fournit le même travail que précédemment jusqu'au lien k. Il construit une table avec l'adresse de la page de départ du lien et l'adresse de la page d'arrivé du lien.

L'agent peut détecter les liens de départ de la page vers d'autres pages du même site (liens internes) ou vers d'autres sites (liens externes). On peut donc constater le caractère exponentiel d'une telle exploration. Supposons que chaque page comporte k liens, l'examen des liens avec une profondeur de 3, renvoie de façon arborescente à l'examen de  $k^3$  pages différentes. Raisonner sur une profondeur de n au lieu d'une profondeur de n-1 prend k fois plus de temps. Pour résoudre ce type de problème, l'agent est paramétrable, et permet à le lecteur de définir le parcours de l'agent en profondeur (verticale) et/ou en largeur (horizontale).

### 2.2. 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 principales caractéristiques de ou des pages explorées se retrouvent ainsi organisées au sein d'un modèle de données comprenant :

**Lien(L),**

**Nœud(N)**

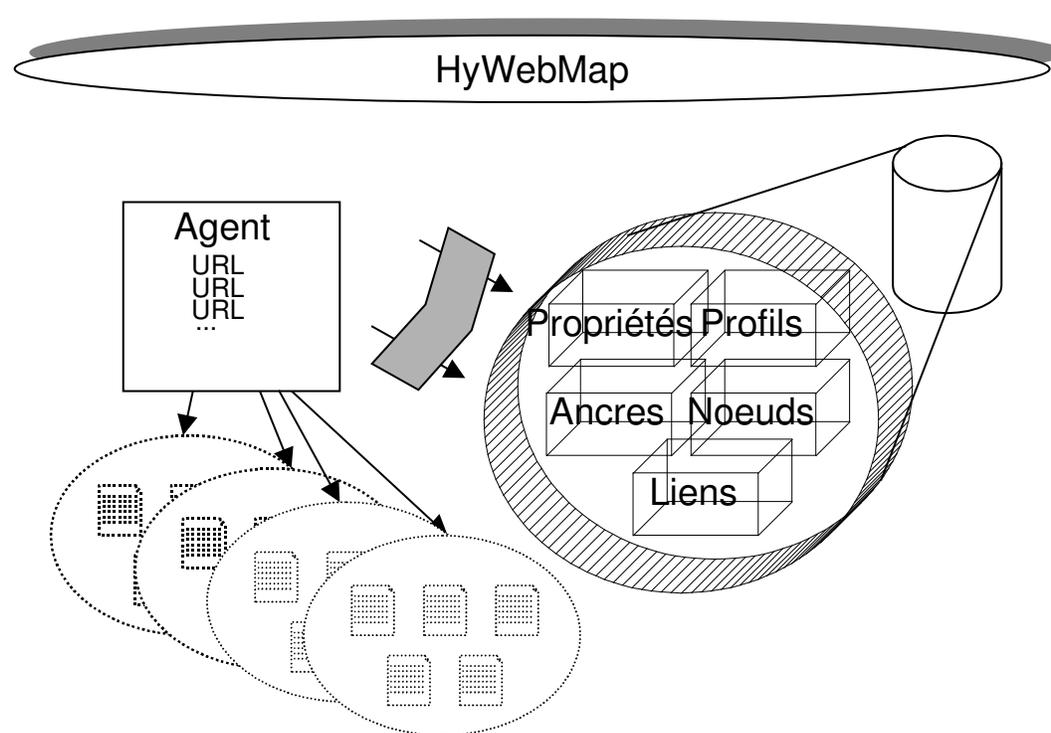
**Ancres(A)**

**Profils**(Nombre-de-caractères, Nombres-d'images, Nombre-de-liens-externes, Nombre-de-liens-internes, Nombre-de-script) 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.

La figure 2 montre des sites Web analysés par l'agent HyWebMap.



**Figure 2. Construction de la base de données après l'analyse des site Web par l'agent HyWebMap.**

### **2.3. La Cartographie sur le Web (Réseau virtuel)**

HyWebMap a été inspiré de l'approche « cartographie », car il est constaté que ces cartes conceptuelles sont d'autant plus pertinentes qu'elles sont construites par les lecteurs eux-mêmes. Dans l'activité de recherche et de sauvegarde des informations pertinentes sur le Web,

le lecteur à constamment besoin d'outils pour l'aider à externaliser et structurer sa propre pensée qui est en interaction avec l'information trouvée. Ces outils doivent permettre à le lecteur d'élaborer à partir d'une cartographie existante, une cartographie personnelle.

La carte HyWebMap dite "réseau virtuel " présente l'espace du site visité (les nœuds et les liens) par l'agent Web. Les lecteurs 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*

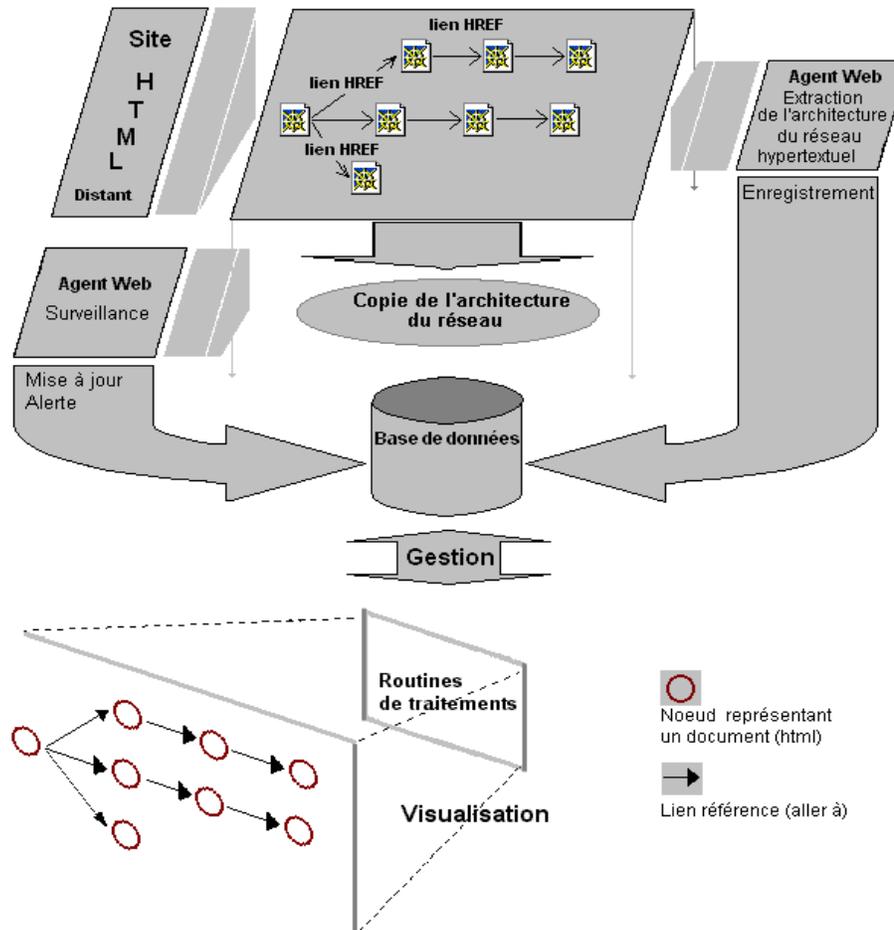
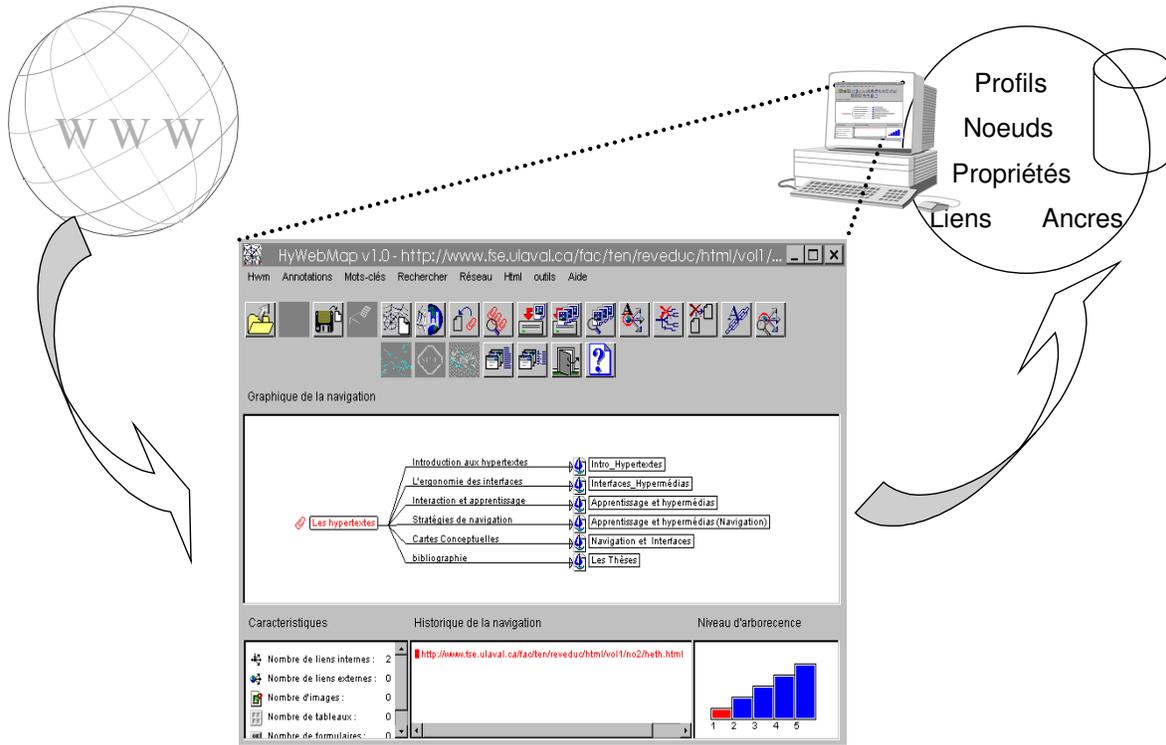


Figure 2. Architecture du système HyWebMap

pages sources, ils sont seulement stockés sur le disque dur de le lecteur et sont actifs lorsque la carte est invoquée. La figure 1 présente l'architecture du système HyWebMap.

**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.



## 2.4. création d'un espace de connaissances personne

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,...).

Ainsi que l'utilisateur peut : 1) naviguer localement dans le réseau virtuel sans établir de connections à internet, ce qui réduit les temps d'accès. Lorsque le lecteur 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.

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, ...).

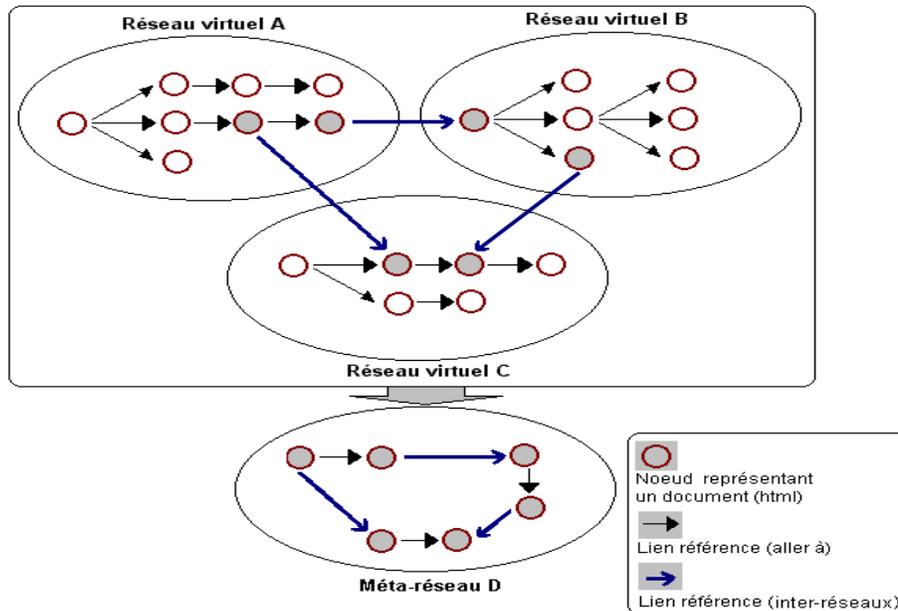


Figure 1 : Un réseau composite HyWebMap

## 2.5. Typologies des nœuds et des liens dans HyWebMap

HyWebMap distingue cinq types des nœuds et sept types des liens pour assurer la cohésion du réseau virtuel ainsi aider le lecteur à se repérer et naviguer dans celui-ci :

### Types de nœuds

**Nœud racine** : c'est un nœud principal défini par l'auteur pour entrer dans son réseau virtuel.

**Nœud parent** : c'est un nœud inférieur du nœud racine, il est lié avec un ou plusieurs nœud/s racine/s, a un ou plusieurs fils.

**Nœud fils** : C'est un nœud du niveau inférieur du nœud parent, il peut être lié avec un ou plusieurs nœud/s parent/s (autrement dit, un fils peut avoir un ou plusieurs pères).

**Nœud orphelin** : c'est un nœud qui n'est lié avec à aucun autre nœud.

**Nœud externe** : c'est un nœud qui n'appartient pas au site d'origine mais on peut y accéder par un lien référence.

Cette typologie des nœuds permet d'identifier la structure de documents et interpréter le sens sémantique entre les nœuds. La fourniture de nœuds est destinée à répondre aux questions suivantes :

*Où suis-je ? Quels sont les types de nœuds suivants ? Pourquoi existe-t-il un lien entre deux types de nœuds ?*

### **Types de liens explicites :**

**Lien ancêtre** : son point de départ un nœud racine et son point d'arrivée un nœud parent.

**Lien parent** : son point de départ un nœud parent et son point d'arrivée un nœud fils.

**Lien fils** : est un lien d'arrivée du lien parent mais il n'a aucun lien vers d'autres nœuds.

**Lien référence** : est un lien libre entre des nœuds (représente le lien "aller à").

**Lien interne** : lien interne intégré dans la même page, c'est-à-dire (région).

**Lien externe** : c'est un lien entre un nœud d'un site avec un nœud d'un autre site. Ce type de liens permet de lier un site à un autre.

**Lien interne rompu** : c'est un lien qui n'existe plus dans le même site (c'est-à-dire un lien orphelin). C'est un point départ mais ce n'est pas un point d'arrivée.

**Lien externe rompu** : c'est un lien qui n'existe plus entre un nœud du site avec un nœud d'un site externe.

Cette typologie des liens explicites aide le lecteur à balayer et interpréter ce qui serait une liste de destinations longues et dénuées de sens. La fourniture de liens sources est destinée pour répondre à la question : *Quoi d'autres pourrais-je avoir obtenu ici ?*

Particulièrement dans le cas de l'interprétation de pages (nœuds) obtenues comme résultat d'une recherche, l'accès de lien entrant peut fournir un accès utile à l'information.

### **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)<sup>1</sup>,
- des descripteurs ou motsclefs,
- 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

---

<sup>1</sup> 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.

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).

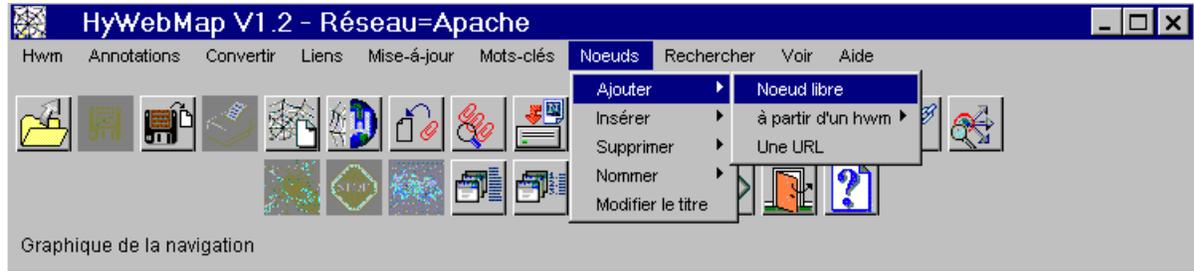


Figure 2a : Menu d'insertion de nœud libre au sein d'un réseau virtuel HyWebMap (Apache.hwm).

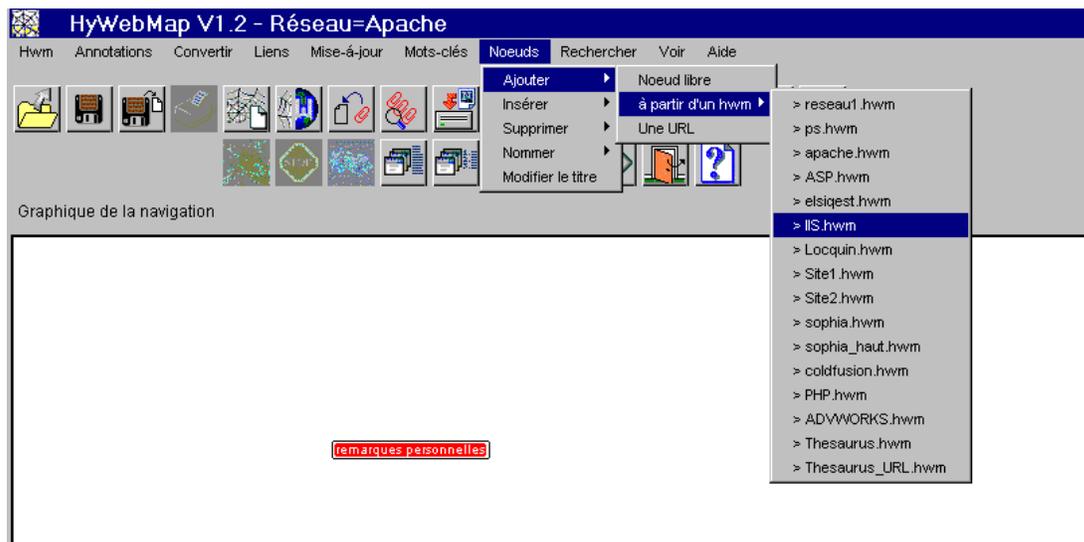
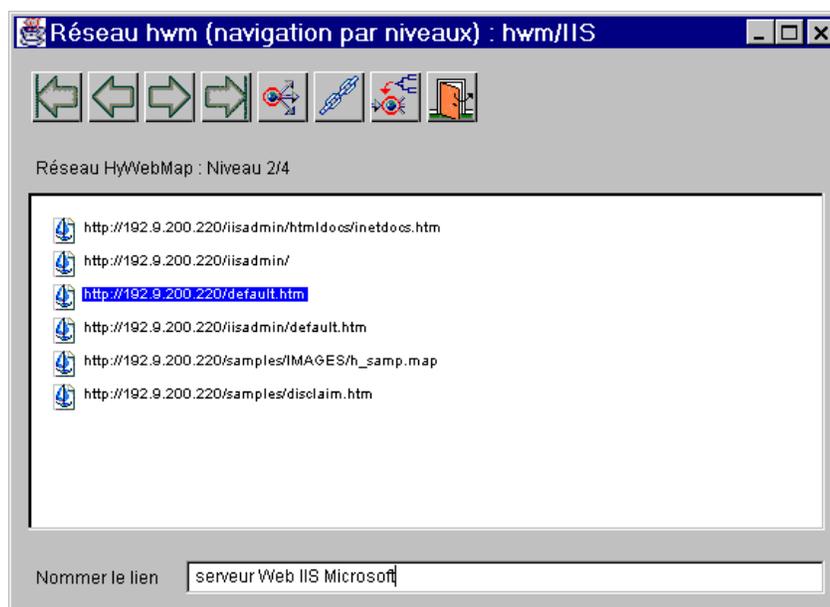
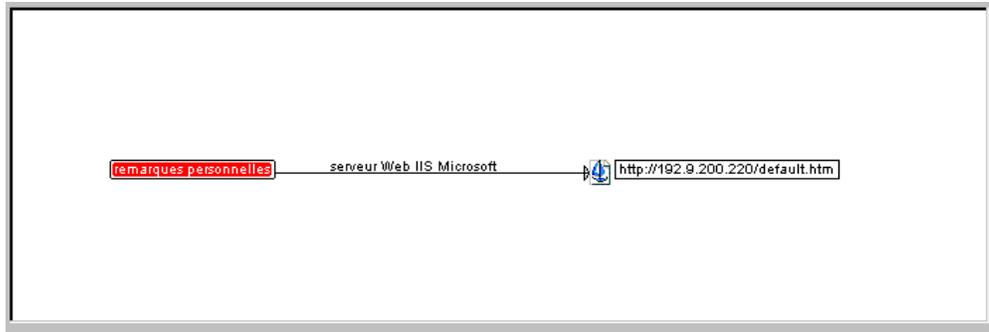


Figure 2b : La sélection d'une ressource distante au travers d'un précédent site HyWebMap



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

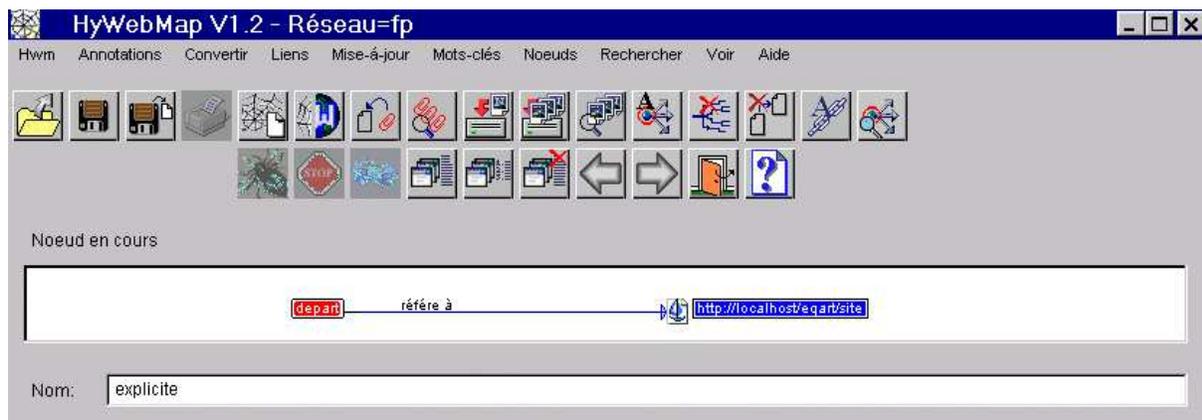


**Figure 2d** : Un nouveau nœud affecté de l'URL <http://192.9.200.220/default.htm> est associé au nœud "remarques personnelles" au moyen du lien "ServeurWeb IIS Microsoft"

## 2.7. 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>.

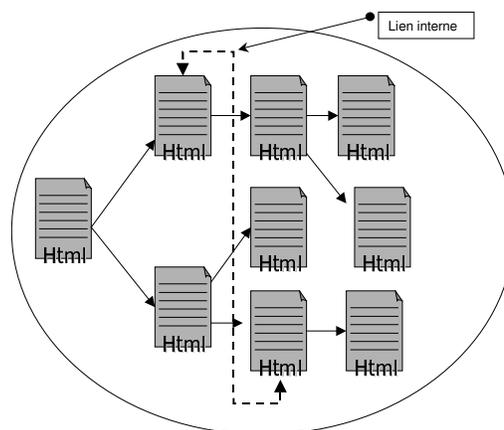


Figure 4 : Lien interne de navigation transversale au sein d'un site virtuel

## 2.6. Agents de surveillance et de mise à jour du réseau virtuel

Une page Web est susceptible d'évoluer au gré des opérations de modification, de suppression et d'insertion des textes et de liens. Cette mise à jour perpétuelle sur le Web pose le problème concret de savoir si le réseau personnel doit seulement garder l'image la plus récente de l'information qu'il représente ou au contraire s'il doit garder en mémoire les modifications successives de cette information.

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, les agents vont détecter son type :

- S'il est du type **nœud fils** alors la suppression sera locale sans réorganisation du réseau virtuel.
- S'il est du type **nœud parent** alors la suppression est plus complexe car celui-ci joue un rôle sémantique dans l'organisation et la compréhension du réseau. HyWebMap propose deux solutions :

La première est manuelle, les agents alertent l'auteur du réseau, en lui demandant de choisir un autre nœud parent sinon ils gardent celui-ci dans le réseau avec ses profils et ses annotations en signalant qu'il est vide (sans contenu).

La deuxième est semi-automatique, les agents cherchent dans la relation profil et dans le champ mots-clés d'annotation, si le nœud parent à supprimer a des mots-clés :

- si oui, ils les prennent, lancent le méta chercheur **copernic** (<http://www.copernic.com/fr>) avec ces mots-clés, puis ils récupèrent le résultat en choisissant les trois premières pages classées par copernic et puis ils les proposent à l'auteur du réseau.
- si non ils appliquent la solution manuelle.
- S'il est du type **nœud racine**, alors les agents alertent l'auteur, en lui demandant de choisir un autre nœud racine.

Ces agents jouent un rôle essentiel pour garantir la cohésion et la mise à jour du réseau virtuel.

Lors de l'expérimentation de HyWebMap, nous avons constaté que la plupart des utilisateurs souhaités avoir sur leur disques locales une trace des contenues des pages Web externe. Pour cela, Nous proposons à l'utilisateur de demander à l'agent de surveillance de rapatrier les ressources externes (c'est-à-dire les contenus des pages de son réseau).

### **Agent d'aide à la création d'un espace personnel**

Lors de nos expérimentations, la plupart des utilisateurs ont souhaité d'avoir un aide pour la création de leurs réseau personnel. Pour cela, le système HyWebMap offre à l'utilisateur une fonctionnalité originale qu'il lui permet de l'aide à la création de son espace personnel. Il s'agit d'un agent Web adaptatif et interactif, il est à la disponibilité de l'utilisateur.

Les principales fonctionnalités :

- *Construction un dictionnaire* : L'utilisateur fournit à l'agent le thème et l'ensemble des mots-clés de son réseau. Par la suite, il évoque le méta-chercheur Copernic avec le mots-clés de l'utilisateur.

- *Construction d'un réseau vide* : L'utilisateur construire un réseau personnel sans ressources externe (c'est-à-dire ou interne en respectant les types des nœuds et de liens, en nommant seulement le nœud et en l'introduisant les mots-clés (métadonnées) de ce nœud. Par la suite, ils évoquent le méta-chercheur Copernic avec les mots-clés du chaque nœud, ils récupèrent les trois premières page concernant les mots-clés de chaque nœud et par la suite l'utilisateur peut choisir une de trois ou sauvegarde les trois.

### **Agent de gestion de profils utilisateurs**

Le système HyWebMap propose un mécanisme de gestion de profils qui consiste à faire de nouvelles propositions à l'utilisateur. Dans un premier temps, l'agent surveille les modifications (actions) apportées par l'utilisateur sur son réseau (de nouveaux mots-clés, annotations, suppression des références (URLs), etc.) pour lui suggérer de nouvelles propositions. Cela, se traduit par de nouvelles versions de réseaux. L'utilisateur aura le choix de les refuser. L'agent suit les différentes actions de l'utilisateurs sur son réseau, pour repérer les modifications (retraits ou enrichissements). Les propositions sont des nouvelles références documentaires sur le Web obtenues par une méta-recherche. La version du réseau en cours est mise à jour sous un nom différent. L'ensemble des nœuds modifier sur le réseau portent une remarque "NP" pour signaler à l'utilisateur qu'il s'agit d'une nouvelle proposition.

### **2.8. Les outils de navigation de HyWebMap**

La navigation dans HyWebMap est locale, il présente à l'écran uniquement le nœud sélectionné avec l'ensemble des liens et nœuds attachés à celui-ci, ainsi ses caractéristiques et son niveau dans le réseau virtuel (figure 4). Chaque nœud visité est représenté par un cercle rouge. Les liens parcourus sont représentés par des flèches rouges. Quand le lecteur revient en arrière pour visiter un nœud et par la suite repart en exploration, HyWebMap affiche en rouge les nœuds et liens déjà visités. Cette représentation facilite l'orientation du lecteur en indiquant les nœuds visités ou non.

HyWebMap est doté de plusieurs outils pour aider le lecteur à naviguer dans son réseau, comme :

- La navigation par historique. C'est un outil classique, car il permet à lecteur de revenir sur un nœud déjà visité.
- La navigation par niveau d'arborescence. Le lecteur peut choisir le niveau souhaité à visiter dans l'arborescence du son réseau virtuel.
- La navigation par date. Le système HyWebMap sauvegarde pendant deux semaines l'historique de navigation du réseau virtuel. Le lecteur 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 qu'une simple liste des nœuds visités dans l'axe du temps.

Les navigateurs Netscape et Explorer de Microsoft proposent des fonctionnalités comme l'historique et les signets qui permettent aux lecteurs de repérer et d'organiser les adresses jugées intéressantes. Ces adresses sont représentées sous forme de listes. Les deux navigateurs permettent la création de catégories et sous-catégories d'adresses. Ces techniques s'avèrent être des aides insuffisantes pour remédier aux problèmes de navigation personnalisée sur le Web et construire un hypertexte personnel à partir des pages Web préférées.

## **2.9. Annotations comme une description sémantique documents.**

Le réseau virtuel personnel présente l'espace Web du lecteur. HyWebMap permet aux lecteurs 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.

### ***Les principales fonctions d'annotation de HyWebMap :***

- ***Ajouter une annotation*** : l'utilisateur peut choisir d'ajouter une annoter à une page donnée ou bien surligne à la souris le texte à annoter et sélectionne l'option "Annoter". HyWebMap propose à l'utilisateur plusieurs champs optionnels : Le titre du document, son URL et le texte surligné sont remplis automatiquement, ainsi un éditeur simple pour que l'utilisateur saisi son annotation. Une annotation peut être de textes, d'images, etc. ou un lien. HyWebMap ajoute automatiquement la date (Heure et Jour) et le numéro d'annotation dans le document.
- ***Voir les annotations du documents courant*** : les annotation sont affiches dans la partie bas de l'écran, leur ancrs sont surlignées en rouge dans le document. D'autres systèmes insèrent une icône avant et après la phrase annoté [GOL 98, MAR 98]. En cliquant sur une annotation dans la fenêtre des annotation, HyWebMap force le navigateur à se positionner sur l'ancre de l'annotation. Cette technique facilite la navigation de la liste des annotations vers le documents.
- ***Changement du document*** : lorsque un changement survient dans le document, le texte surlignés sont parfois introuvable dans le document. Les annotation apparaissent alors en gris dans la liste d'annotation. Ainsi, même lorsque un document n'est plus disponible

alors ses annotations sont sauvegardées. Cette technique améliore l'approche de l'actuel signets [DEN 00].

- **Recherche les annotations** : L'utilisateur peut effectuer de recherche sur ses annotations soit en plein texte, ou bien par leurs attributs : date, titre du document, URL, et le texte surligné. Le résultat est une liste d'annotation regroupées par document où chaque document est représenté par la fusion des textes surlignés qu'il contient. Cette technique est très intéressant car elle permet à l'utilisateur d'avoir une idée sur un document sans le consulter entièrement en quelque sorte un résumé.

### ***Ajouter de metadonnées***

HyWebMap permet à l'utilisateur d'ajouter des métadonnées en plus des texte ou mots surlignés. Il propose à l'utilisateur un champs optionnel mots-clés pour saisie des métadonnées (sans nécessairement faire partie du document). Ce champs est utile à des fins de recherche et personnalisation du document.

### ***Construction d'un dictionnaire personnel***

HyWebMap permet à l'utilisateur de construire son propos dictionnaire où l'utilisateur peut introduire des mots-clés et leur synonyme par rapport ses connaissances. Ce dictionnaire permet d'une part à l'utilisateur d'introduire une seule fois ses mots-clés avec leurs synonymes et d'autres part l'aider à la recherche d'information. Actuellement nous développons cette partie pour que l'utilisateur puisse construire un vrai thésaurus personnel.

Dès qu'un nœud est re-visité, son annotation qui est attachée apparaît dans le navigateur.

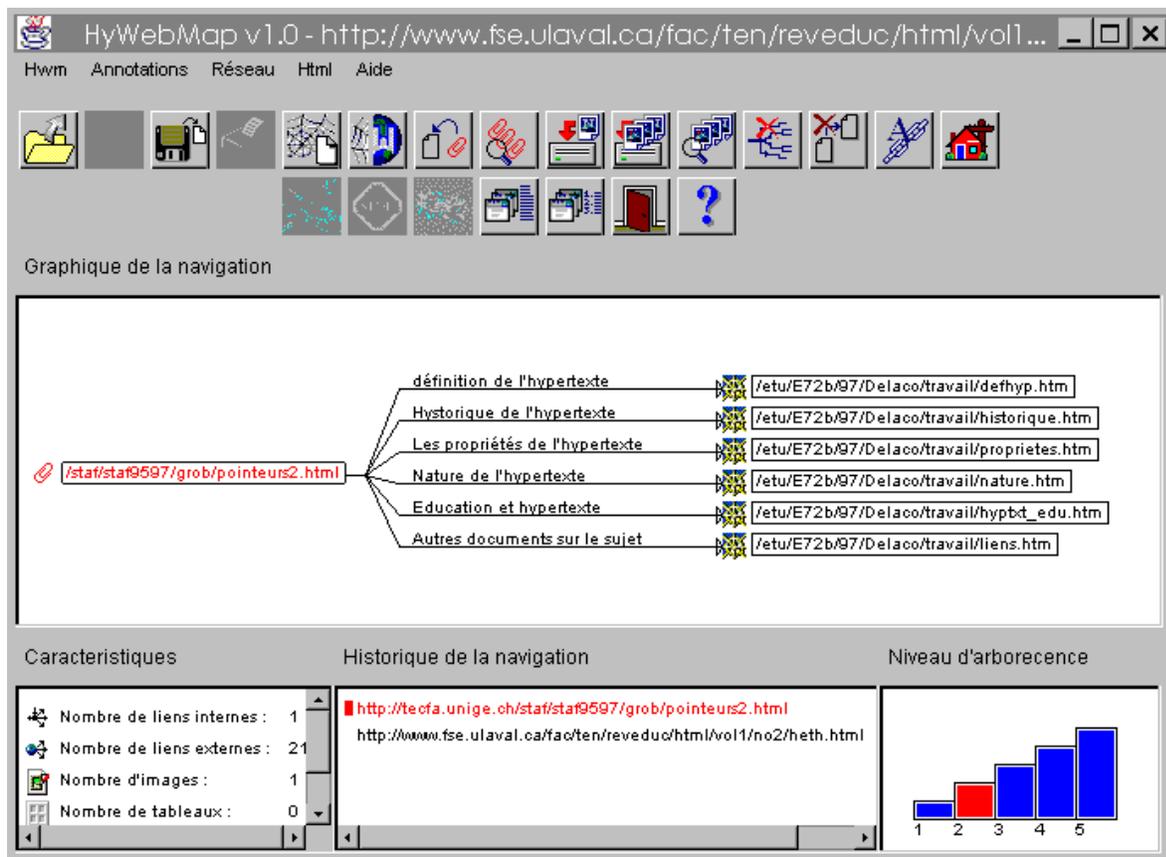


Figure 4. présente un nœud sélectionné avec ses caractéristiques.

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

Nous avons introduit une fonctionnalité pour permettre 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. A gauche se présente la frame de navigation (le nom du lien entre deux nœuds) et à droite la frame présente le profil 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 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 5 présente la génération du réseau virtuel en version html.

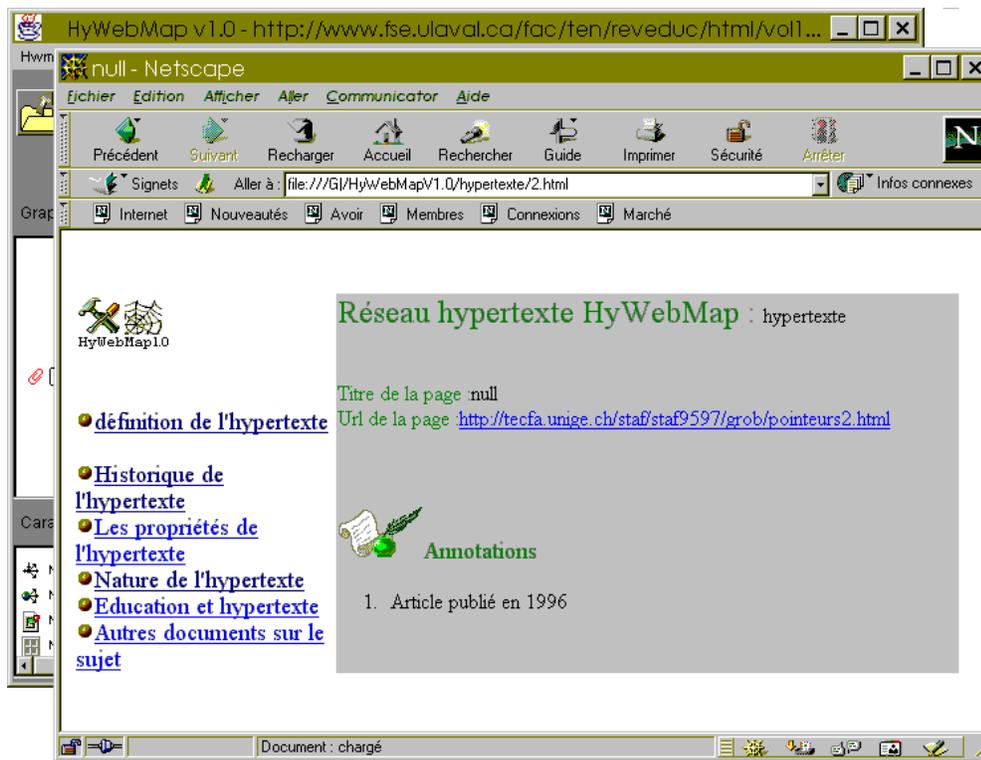


Figure 5. Génération en HTML

Ainsi, HyWebMap permet à l'utilisateur d'avoir une vision global de son site personnel grâce à la génération dynamique de celui-ci en sommaire active et navigable.

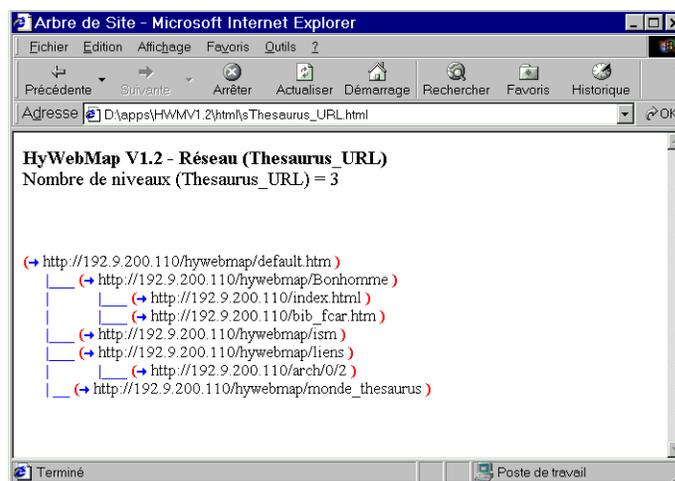


Figure 6 : Sommaire actif d'un site virtuel converti.

### 3. Domaines d'utilisation du système HyWebMap

HyWebMap, un système fonctionnel, est utilisé dans plusieurs domaines :

- L'éducation : les professeurs peuvent réorganiser leur cours par rapport aux thèmes de chaque séance. Ainsi les étudiants peuvent construire leurs réseaux virtuels en fonction de leurs centres d'intérêts. Actuellement, plusieurs professeurs des départements Hypermédia, Documentation, Science de l'éducation de l'université Paris 8 utilisent HyWebMap comme outil de préparation et d'organisation des cours.
- La veille technologique : Plusieurs étudiants utilisent HyWebMap comme un système de recherche documentaire sur le Web. Ils construisent des réseaux virtuels thématiques sur leurs thèmes de recherches, etc.
- La construction de sites portails sur une thématique précise.
- L'enseignement à distance.

#### 4. Discussion

Il existe une quantité importante d'outils permettant la création d'une cartographie d'un site Web [WebAnalyzer (<http://www.WebAnalyzer.com>), CHEOPS (<http://www.crim.ca/ipsi/cheops>), Mapuccino ([www.ibm.com/java/mapuccino](http://www.ibm.com/java/mapuccino)), MAPA (<http://www.DynamicDiagrams.com>), etc.]. L'outil le plus original est MAPA [Kah 95,97 et pour plus amples d'informations, visitez le site <http://www.DynamiqueDiagrams.com>] de Paul Kahn qui permet de construire une cartographie à partir d'un URL en 3D dans une applet Java. Il utilise les couleurs pour séparer les niveaux de hiérarchies des pages HTML affichées. MAPA est un outil de navigation et de gestion d'un site Web. HyWebMap comme MAPA est doté de fonctionnalités permettant la navigation et la gestion du site Web. La différence repose sur le fait que HyWebMap permet au lecteur de réorganiser l'espace visité en créant son réseau personnel par rapport à son point de vue sans mettre en cause les nœuds et les liens de celui-ci, d'ajouter des nœuds et d'annotations sur celui-ci, de renommer les liens et de générer automatiquement un site Web consultable par un navigateur standard.

#### 5. HyWebMap et l'écriture Hypertextuelle

*HyWebMap peut-il être utiliser en tant qu'outil d'écriture hypertextuelle ?* Pour répondre à cette question, nous citons le point de vue de Pierre Levy dans son livre<sup>2</sup>, il écrit « *le navigateur peut se faire auteur de façon plus profonde qu'en parcourant un réseau préétabli : en participation à la structuration de l'hypertexte, en créant de nouveaux liens. Certains systèmes enregistrent les chemins de lecture et renforcent ... ou affaiblissent les liens en fonction de la manière dont ils sont parcourus par la communauté des navigateurs* ». Cette citation a affaibli la distinction entre auteur et lecteur, a condition que le lecteur ait le droit de modifier la navigation hypertextuelle en ajoutant des nœuds et des liens sur l'hypertexte. S'agit-il alors du même hypertexte ? Non, Pour le lecteur suivant qui parcourra l'hypertexte transformé, les significations du document seront, elle aussi modifiées. S'agit-il alors d'écriture ? Incontestablement oui, car la signification des liens et la structure de l'hypertexte sont transformées par rapport au point de vue du lecteur. Le lecteur il devient un auteur de l'hypertexte. Nous concluons que HyWebMap est un outil d'écriture hypertextuelle.

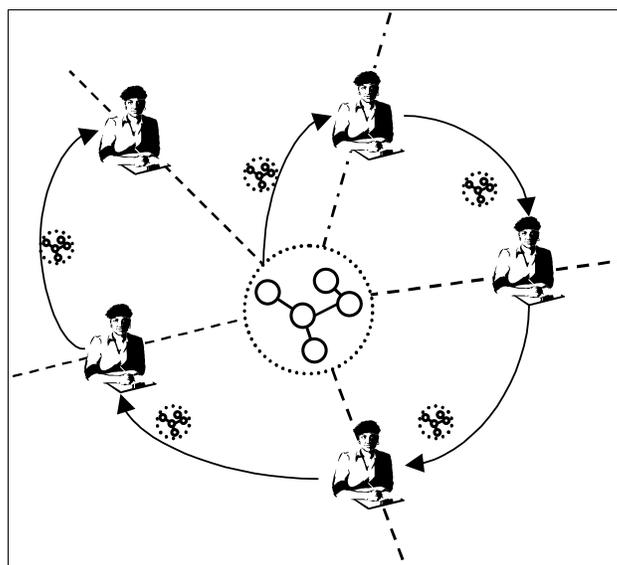
#### Conclusion

---

<sup>2</sup> Qu'est que le virtuel ?, La Découverte, Paris, 1995, P.43.

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 d'HyWebMap (cf. figure 7).



**Figure 7 :** Mise en œuvre collaborative d'un réseau virtuel sous le contrôle des auteurs

## Bibliographie

- [AST 93] ASTOLFI J-P., " Comment les élèves apprennent-ils ?", *Sciences Humaines*, n° 32, octobre 93, pp. 26-29.
- [BAL 96] BALPE J.P., LELU A., SALEH I., PAPY F., « Techniques avancées pour l'hypertexte » », Editions Hermès Paris 1996.
- [BEL 99] BELISLE C., Zeiliger R., Cerratto T., « S'orienter sur le Web en construisant des cartes interactives avec Nestor », Actes H2PTM 99, Editions Hermès 1999.
- [BRO 88a] 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.
- [BRO 88b] BROWN P.J., "Linking and searching within hypertext", *Electronic Publishing*, vol. 1 (1), avril 1988, pp. 45-53.
- [BRO 89] BROWN P.J., "Do we need maps to navigate round hypertext documents?" *Electronic Publishing*, vol. 2 (2), juillet 1989, pp. 91-100.
- [CAS 96] Castelli C., et al. « Getting lost in Hyperspace : Lessons learned and future direction », in ED-MEDIA 96/ED-TELECOM 96, 1996.
- [FOS 89] FOSS C.L., Detecting lost Users: Empirical studies on browsing hypertext, *89H385299*.
- [ROU 97] 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.
- [Kah 95] Kahn P., « 3Global an local hypermedia designe in the encyclopedia african » in Hypermedia Design, Fraïsse, Garzotto, Isakowitz, Nanard an Nanard ed. 1995 Spring, London.
- [Kah 97] Kahn P. et al. « Real space and cyberspace : a comparison of museum maps and electronic publication maps » in ICHIM 97, 1997, Paris.

<sup>1</sup> Il est développé au laboratoire Paragraphe de l'université Paris 8 avec le langage Java. La version 1.0 de HyWebMap est téléchargeable sur le site : <http://www.labart.univ-paris8.fr/~nasser/hwm.html>