

Chapitre 1

De l'hypertexte à l'hypermédia

1.1. Introduction

Depuis le célèbre article de Vannevar Bush en 1945 « As we may think ? » qui fut un des premiers sur les hypertextes décrivant le *Memex (MEmory EXtender)* [BUS 45], l'hypertexte a connu plusieurs évolutions dont la plus importante, celle de Thodor Holm Nelson par son projet *Xanadu* [NEL 65], ainsi que celle de Douglas Engelbart, représentée par son projet *Augment* [ENG 84]. De nombreux outils liés à la technologie hypertextuelle ont vu le jour : l'édition du texte à l'écran, les liens, les fenêtres, la souris, etc. Par ailleurs, cette technologie « hypertextuelle » a marqué une révolution dans la lecture/écriture, dans la communication, dans l'architecture et la consultation de l'information, etc. Nous parlons depuis de la navigation, de l'interactivité, de la non-linéarité, des hyperdocuments, des hypermédias, de l'adaptativité, de la désorientation, de la surcharge cognitive, etc. sans avoir la garantie toujours que ces concepts aient les mêmes connotations chez les uns et les autres que ce qu'ils veulent dire.

Dans ce chapitre, nous essayons de présenter une synthèse sur l'hypertexte, et tout ce qui résulte de cette technologie : des changements au niveau de la lecture et de l'écriture de l'information, ainsi que le passage à l'hypermédia à travers le multimédia. A la fin de ce chapitre nous présentons comme exemple le système hypertexte HyWebMap qui a été développé au laboratoire Paragraphe de l'université Paris 8.

1.2. Hypertexte : éléments de définition

Le terme *Hypertext* est un néologisme anglais créé dans les années soixante [NEL 65], pour décrire un concept inventé déjà vingt ans auparavant [BUS 45]. Cependant, le monde informatique fait la connaissance de ce concept seulement à partir de 1987, quand la société Apple introduit le logiciel HyperCard dans sa gamme de produits. La première conférence internationale consacrée aux hypertextes sous le patronage de l'ACM (*Association of Computer Machinery*) est organisée dans la même année. Depuis, des journaux scientifiques publient des numéros spéciaux et des conférences nationales et internationales se tiennent chaque année sur le sujet.

Ce succès s'explique par le croisement de plusieurs facteurs. En premier lieu, il faut signaler la croissance considérable des espaces disques et l'augmentation importante du rapport qualité/prix des interfaces utilisateurs pendant les dernières années. Cette évolution a conduit à l'apparition de nouvelles applications qui exigent une bonne organisation et une exploitation simple d'une masse gigantesque de documents. L'hypertexte propose des facilités qui relient les informations et les conçoivent sous forme structurée (hiérarchiques, linéaires, etc.) dont l'accès s'opère au moyen de la technique de navigation. En d'autres termes, l'hypertexte gère les flux d'applications informationnelles en constituant un réseau de documents.

En créant le système *Xanadu* (bibliothèque universelle informatisée dont la caractéristique principale est le partage des idées entre les utilisateurs, Ted Nelson a défini l'hypertexte comme étant : *l'association d'un langage naturel avec les possibilités qu'a l'ordinateur d'établir des liaisons interactives et des affichages dynamiques d'un texte non linéaire* [NEL 67].

La plupart des travaux sur les hypertextes font référence à Conklin qui définit un hypertexte comme : *une base de données textuelles, visuelles, graphiques, sonores où chaque îlot d'information est appelé nœud ou cadre..., une interface ou un mode de présentation visuelle permet l'interaction entre l'utilisateur et l'hypermédia* (figure 1.1) [CON 87].

Nous pouvons définir l'hypertexte aussi à travers les actions selon deux niveaux :

- à son niveau le plus simple, c'est un système de gestion de données qui permet d'associer des écrans par des liens ;
- à son niveau le plus sophistiqué, c'est un environnement logiciel pour un travail collaboratif, pour la communication, pour l'acquisition et la représentation des connaissances [CAM 91].

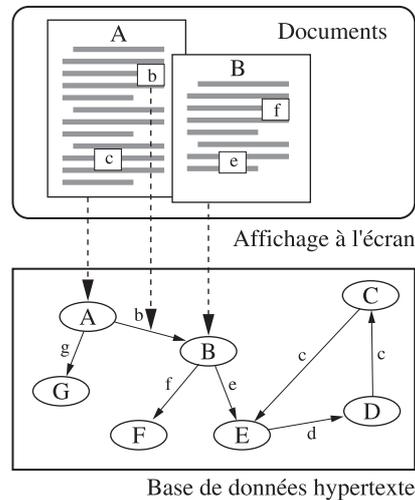


Figure 1.1. L'hypertexte, d'après Conklin [CON 87]

Par ailleurs, [BAL 96] voit qu'on ne peut pas donner facilement une définition précise à l'hypertexte à cause de la diversité des notions qui s'y rapportent. L'hypertexte peut se comprendre à travers ses finalités, ses caractères essentiels et ses qualités propres.

Concepts structurels

En ce qui concerne sa structure, un système hypertexte intègre les concepts suivants :

- *nœuds* : ce sont des unités élémentaires associées à des fragments d'information d'un ou de plusieurs types : texte, graphique, image, son, etc.

- *liens* : les liens constituent le principal moyen pour organiser un document d'une manière non séquentielle. Ils permettent à l'utilisateur de se déplacer d'un endroit à un autre dans un document, ou d'un document à un autre, ou d'un endroit d'un document à un endroit d'un autre document. Un lien peut être uni ou bi-directionnel. Dans le cas d'un lien unidirectionnel on parle de référence pour la source, et de référent pour le but.

- *ancres* : dans le cas d'un texte, l'ancre repère un « endroit » d'un texte, c'est-à-dire une unité sémantique de niveau inférieur à celui du nœud.

Pour un objet : image ou son, une ancre est un objet informatique, en général un *pointeur* (c'est-à-dire l'adresse d'une information) qui contient le point de départ ou d'arrivée d'un lien.

Le domaine ou point d'ancrage pose des problèmes parfois difficiles. En particulier lorsqu'on a affaire à des documents complexes continus. On est alors conduit à définir une *région de lien* sur le document, par exemple sur une carte de géographie. Remarquons que cette ancre peut se trouver enregistrée avec l'objet, voire en faire partie dans le cas des hypertextes géographiques. Lorsqu'un utilisateur a établi entre les objets consultés un ensemble de liens correspondant à sa propre vision ou à un usage déterminé, le tissu (qui peut être très complexe) de ces liens constitue un réseau.

1.3. Une architecture de l'information non linéaire

Le média imprimé est fondé sur la structure figée d'une logique linéaire, de cause à effet, de la thèse et l'antithèse à la synthèse... La particularité de l'organisation hypertextuelle de l'information est, par contre, fondée dans la possible absence de hiérarchie qui a structuré jusqu'à présent notre pensée.

L'hypertexte est fondé sur l'assemblage de fragments textuels indépendants, dont la continuité de lecture n'est pas basée sur l'organisation linéaire et figée de l'information, mais sur ses développements successifs dans le réseau. En conséquence, sa continuité dépend des multiples liaisons faites entre les fragments de textes, images, sons, animations... La forme libre de la technologie de l'information et en particulier de l'hypertexte permet une variation illimitée de formes d'organisations des données et une combinaison illimitée de formes d'expression. A travers les différentes formes d'interconnexion-branchements, chaînes, suites... l'hypertexte génère de multiples séquences de lecture. Il est pratiqué comme un espace non linéaire, ou plutôt multilinéaire ou multiséquentiel, composé de fragments multiples interconnectés et tissés dans un réseau, à l'aide d'hyperliens.

Dans l'espace de l'hypertexte, chaque passage est un point d'intersection dans lequel convergent des parcours menant vers d'autres fragments de texte. La multiplication des séquences de lecture et des entrées en jeu dans le même texte détruit la notion classique du texte ayant un début et une fin et la transpose en une variation indéterminée, configurée en permanence par rapport aux liens. Finalement, la totalité des parcours menant vers les différents passages, les inclut tous, pour former la structure de base, le réseau.

Le sens spécifique du texte dépend du parcours que le lecteur emprunte. Il offre ainsi une nouvelle compréhension de l'espace écrit, non plus dans une organisation prédéterminée et linéaire, mais dans une forme interactive et multilinéaire.

Le lecteur de l'hypertexte est un sujet actif, non seulement dans la déconstruction permanente d'une forme d'organisation figée, mais aussi comme paramètre intervenant dans la structure temporelle du texte. Le déplacement libre du lecteur à travers les différents fragments est une variable qui transforme la structure figée du texte classique en une forme dynamique et multidirectionnelle. Le texte statique s'est transformé en une forme fluide, liquide, dans laquelle l'existence de multiples combinaisons forme une structure dynamique. L'hypertexte est une nouvelle forme discursive remplaçant les systèmes basés sur les idées de centre, de hiérarchie et de linéarité par ceux de nœuds multilinéaires, de liens et de réseau.

1.4. Navigation dans l'hypertexte

A la différence d'une base de données que l'on interroge en lui soumettant des requêtes, un hypertexte est un réseau dont le mode typique de consultation est la navigation de nœud en nœud dans la base [DAC 1990]. Il s'agit des déplacements possibles acheminés par le lecteur d'un nœud à un autre au sein d'un ou de plusieurs documents [SAL 03, SIM 89].

On peut consulter l'information dans un hypertexte à des niveaux de détails plus ou moins grands de trois façons :

- 1/ recherche par mots-clés : elle représente une recherche avec des chaînes de caractères (mots-clés ou valeurs d'attributs) dans les nœuds d'information ;
- 2/ invocation d'un lien : à partir d'un nœud, l'utilisateur peut invoquer librement des liens, choisir son parcours de lecture, et retourner en arrière grâce à un historique consultable sur écran ;
- 3/ utilisation de carte géographique : on peut naviguer dans le document à l'aide d'une carte de structure « browser », où sont représentés des nœuds pertinents et leurs liens [LEC 91, MAR 90].

Il existe cinq possibilités de navigation qui sont [Bal 96] : *Liens typés, références croisées, recherche d'occurrences de chaînes de caractères, tables des matières activables, retour en arrière, consultation du nœud précédant ou du nœud suivant.*

Cependant, cette activité incontournable (navigation) pour circuler dans les ensembles informationnels si prometteurs, était loin d'être aussi facile qu'on ne l'avait imaginée. Deux problèmes sont toujours évoqués, à chaque fois que nous

parlions de la navigation dans des systèmes hypertextes, la *désorientation* et la *surcharge cognitive*. La désorientation est une limite importante, car nous voyons tellement d'informations et d'images (dans un hypertexte) que nous ne savions plus ce qui est relié et nous risquions de ne rien retenir. En outre, nous risquons de perdre le but de recherche à cause de la richesse de représentation non linéaire des textes. En effet, l'utilisateur se trouve parfois perdu dans l'hyperespace puisqu'il ne sait pas comment l'information est organisée, comment trouver l'information qu'il cherche, et si cette information est fiable [MCK 91, SAL 03].

La surcharge cognitive représente, elle aussi, un problème important que rencontre l'utilisateur d'un hypertexte dans le sens où il voit un flux d'information mais il n'arrive pas à en extraire des éléments pertinents et cohérents [EDW 89, NIE 90, SAL 03].

La navigation idéale reviendrait pour le lecteur à déceler, du premier coup, parmi tous les chemins de liens possibles, le plus approprié à ses buts. Pour le dire autrement, il s'agirait pour lui de faire surgir parmi les possibles celui qui correspond à une nécessité qui lui est propre. L'idéal (peut-être inaccessible) serait de concevoir un hypertexte qui allie possibilité (liberté de choisir entre plusieurs chemins différents) et nécessité (détermination d'un seul chemin en fonction d'un contexte donné). En somme, il s'agirait de garantir une multiplicité de choix dont aucun ne serait fait par hasard, c'est-à-dire de rendre le plus clair possible le sens de chaque lien, et le contenu du nœud où il conduit [SAL 01, SAL 03].

Désorientation et surcharge cognitive tiennent donc d'une part à l'inadéquation entre lien sémantique et lien structurel, et d'autre part à un manque de visibilité de l'interface. Mais la résolution de ces deux types de problèmes peut entraîner la disparition du caractère proprement hypertextuel de l'hypertexte : une structuration trop hiérarchisée revient à nier la capacité de créer des associations sémantiques transversales, tandis qu'une interface cartographique où nœuds et liens sont représentés rend inutile l'ancrage des liens au sein de chaque nœud. Autrement dit, quelle solution apporter au problème de la désorientation, sans mettre à mal l'essence hypertextuelle elle-même ?

1.5. De l'Hypertexte minimal hypertexte à modèle utilisateur

1.5.1. Le modèle Dexter [HAL 94]

Créé en octobre 1988 au *Dexter Inn* dans le New Hampshire (Etats-Unis), les concepteurs de Dexter se sont fixés comme objectif de trouver une norme dans la

terminologie et la sémantique des concepts hypertextuels. Ils ont formé un modèle en trois couches :

- la *couche d'exécution (représentation)* qui gère la présentation d'un réseau hypertexte c'est-à-dire les aspects de l'interface utilisateur et les outils de navigation ;
- la *couche de stockage* qui est le noyau du système. Elle modélise le système hypertexte et fournit les fonctionnalités nécessaires à sa gestion ;
- la *couche du contenu informatif (information)* représente et gère le contenu des nœuds hypertextes. Elle encapsule plusieurs éditeurs appropriés à la gestion des éléments multimédias (texte, images, sons, etc.).

L'indépendance est totale entre les couches de stockage et le contenu informatif. Un mécanisme d'ancrage permet l'adressage des objets dans le contenu informatif à partir de la couche de stockage.

Le système Dexter est considéré comme un système *ouvert* c'est-à-dire permettant l'utilisation d'applications externes (traitements et éditeurs de texte, etc.). Ce modèle est réalisé grâce à l'indépendance des couches et sous forme d'une interface fonctionnelle (voir la figure 1.2.).

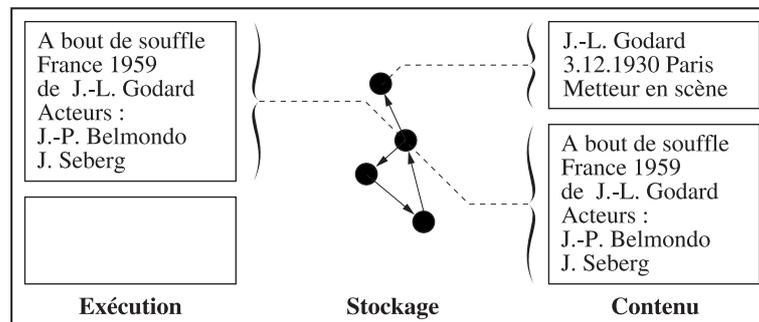


Figure 1.2. Modèle Dexter

Dans ce modèle, l'ensemble des informations et le graphe de représentation sont fixes, l'hypertexte est un ensemble fini de parcours possibles prédéterminés. C'est la solution classique adaptée et appliquée dans la plupart de CD-Rom actuels. Or elle ne peut pas être réalisée correctement que sur des ensemble d'information de volume restreint : une masse importante de données informatives interdisent pratiquement une vue d'ensemble – la métaconnaissance - qui permet seule de les décrire.

Des problèmes apparaissent lorsque l'hypertexte basé sur un modèle ouvert et dynamique c'est-à-dire lorsque l'une des couches des informations ou des représentations n'est pas fixe. Par exemple, lorsque la couche des informations est ouverte, il faut contrôler en permanence le maintien d'une correspondance entre la couche des informations et celle des représentations. L'absence de tel ou tel fragment correspond à une représentation donnée interdisant tout aboutissement d'une demande particulière ; si les fragments ne correspondent à aucune représentation, ils sont donc aveugles à la circulation hypertextuelle, il y a donc modification de la représentation par rapport à de nouveaux éléments.

Le fait que la couche des représentations soit ouverte ou hétérogène oblige à concevoir non plus une couche de représentation unique mais une démultiplication de ce niveau en diverses couches de façon à tenir compte de la diversité de profils possibles des utilisateurs imaginés. C'est la situation des hypertextes « pédagogiques » où les parcours d'information diffèrent suivant les niveaux de connaissances supposés aux utilisateurs.

Une nouvelle notion apparaît, celle de « modèle d'utilisateur », c'est l'éclatement de la couche des représentations en un nombre indéterminé de sous couches dont l'ensemble constitue la métaconnaissance sur les informations (figure 1.3)

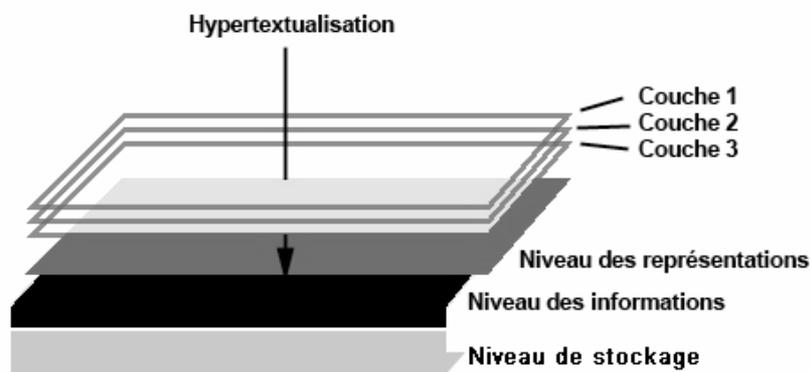


Figure 1.3. Modèle d'utilisateur

1.5.2. L'hypertexte automatique

Cette technologie permet d'aller au-delà d'une représentation fixe d'une métaconnaissance fermée en calculant automatiquement de la représentation. Ces

calculs automatiques font éclater la couche représentation en deux sous-couches distinctes [Bal 97] : une couche de représentation calculé et une couche permettant le calcul de cette représentation. Ils ajoutent ainsi une quatrième couche à la construction des hypertextes : celui de la conceptualisation (figure 1.4).

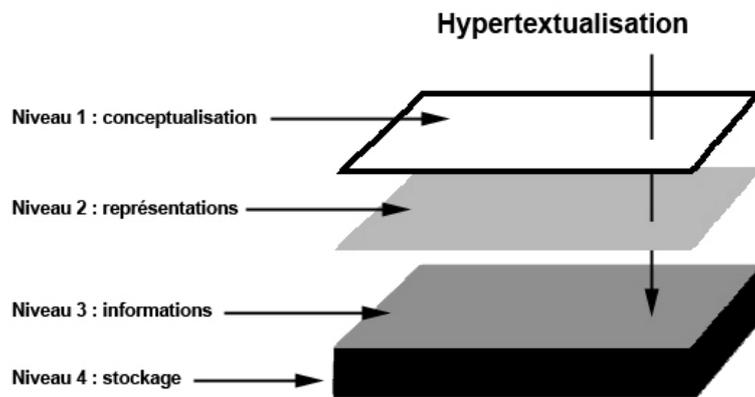


Figure 1.4. Modèle à quatre couches

Cette couche implique une cartographie complète des concepts, Un tel dispositif constitue un générateur de liens. La couche conceptuelle sert de tampon entre la couche des informations et celle de représentations : toute modification de l'état des informations entraîne la modification à la couche des représentations calculées par la couche conceptualisation. L'hypertexte est partiellement auto-adaptatif mais ne s'adapte qu'au contenu puisque le module de calcul – qu'il soit à base linguistique, structurelle (structure logique de textes), statistique – repose sur un modèle préétabli et fixe. Ce dispositif est bien adapté au traitement de volumes d'informations importants mais relativement fixes.

La cartographie pose des problèmes complexes :

- il est difficile de projeter un réseau hypertexte multidimensionnel sur une structure de graphe à deux dimensions ;
- il est impossible de tracer plus d'une vingtaine de nœuds sur un écran sans que le graphe devienne illisible.

Il est donc impossible de représenter de façon bijective un hypertexte dans son entier sur une carte. Il faut donc opérer des choix et choisir des niveaux de représentation.

Une solution consiste à représenter avec précision la structure locale et à « résumer » la structure globale. Une fois de plus, cela suppose d'avoir une théorie sémantique hypertextuelle sur laquelle se baser pour opérer l'agrégation.

Mark Bernstein [BER 92] distingue trois domaines d'automatisation de l'hypertexte :

– la *conversion de documents (hypertextualisation)* : elle consiste à produire un hypertexte à partir de documents de façon automatique. Le statut de cette activité reste à déterminer : s'agit-il d'une activité de transition, ou d'une nécessité économique ou technologique. Ce niveau d'automatisation pose la question du statut du document par rapport à l'hyperdocument ;

– la *structuration automatique de domaines* : elle consiste à traduire la structure d'un domaine en termes d'un modèle hypermédia. Cette question concerne la structuration automatique des hypermédias ;

– la *navigation dans les réseaux hypertextuel* : les structures « virtuelles » peuvent être ajoutées dans un hypertexte au fur et à mesure de la navigation, en fonction des actions de l'utilisateur, ou d'autres événements extérieurs. C'est la question de la génération dynamique.

La conversion de document peut présenter plusieurs scénarii. Il peut s'agir d'un seul document qui est délinéarisé pour donner un hypertexte. Par exemple le cas de la construction d'une encyclopédie hypermédia à partir d'une encyclopédie papier [SAL 93]. Il peut aussi s'agir de réunir un grand ensemble de documents plus petits, par exemple le cas du dossier médical. Le statut du document n'étant pas évident, doit-on préserver son unité ? Dans le cas contraire, comment obtenir des nœuds à partir des documents ?

Les méthodes à employer sont multiples. Toutes reposent, implicitement ou explicitement sur une hypothèse précisant quand il devient intéressant de relier deux nœuds. Les systèmes existants considèrent que c'est le cas lorsque leur sens est proche, ce qui est une hypothèse restrictive. Dans le cas d'un dossier médical par exemple, il peut être intéressant de créer aussi des liens entre des éléments dont le sens est contradictoire, complémentaire, etc. [Salton et Allen 93] construit une structure hypertextuelle adaptée à une requête sur un *corpus* textuel. Les passages sont caractérisés par des vecteurs qui repèrent les mots qui leur sont spécifiques. Une mesure de similarité est appliquée entre les vecteurs pour repérer les passages proches. De tels vecteurs sont construits pour les textes entiers, les sections, les paragraphes et les groupes de phrases.

Le système HYPERMAP du laboratoire Paragraphe de l'université Paris 8 [LEL 98] est basé sur un système d'indexation par lemmatisation, suivi d'une

classification statistique permettant de construire une cartographie sémantique. L'hypertexte obtenu est alors un ensemble de documents fédérés par une carte cliquable. Il s'agit d'une structuration automatique de domaine.

Ce système considère que la proximité de sens justifie l'existence d'un lien. Or ceci pose deux problèmes [BRU 99] :

- la notion même de proximité de sens ne va pas de soi. L'estimation de cette proximité est une activité difficile. Les techniques de l'indexation en texte intégral sont automatiques mais donnent des résultats peu satisfaisants. La lemmatisation est une tâche lourde si l'on souhaite qu'elle soit efficace. Plus fondamentalement le sens des textes repose sur cette technique de la présence de mots, ce qui est pour le moins restrictif dans la notion de sens ;

- il n'y a pas de garantie que l'hypertexte fasse sens, ce type de construction repose sur l'hypothèse de l'existence de passages sémantiquement proches pouvant être liés, sans précision supplémentaire. La sémantique hypertextuelle sous-jacente est grossière et peu effective et suppose, comme nous l'avons vu au-dessus, que la structure fondamentale nœud/liens implique à elle seule une construction de sens. Cela revient un peu à considérer que l'on pourrait construire automatiquement des textes en juxtaposant des phrases, en se basant sur l'observation évidente que les textes sont nécessairement construits à l'aide de phrases.

Pour remédier à ces problèmes, certains systèmes sur le Web s'appuient sur un modèle d'index à deux niveaux : le premier niveau de l'index concerne l'information contenue dans chaque page et le deuxième niveau n'est pas borné à l'information contenue dans une page ; le *voisinage sémantique* de la page est aussi considéré. Ces deux niveaux sont dépendants ; les éléments du deuxième niveau fournissent une notion de contexte sémantique aux éléments du premier niveau, exemple les pages HTML. Autrement dit, le contenu d'une page P_y peut soit compléter le contenu d'une page P_x , soit fournir le contexte du contenu de cette page [DOA 98, FER 00].

REMARQUE.— En ce qui concerne l'état de l'art de la génération automatique des liens et des nœuds, nous distinguons trois approches [BAL 1996] :

- l'*approche structurelle* : elle prend en considération les documents électroniques existants au travers de procédures d'hypertextualisation dont la finalité repose sur un découpage hiérarchique en chapitres, sections, paragraphes, etc. en isolant les références croisées internes et externes. Bien que l'absence de normalisation structurelle des textes ait rendu le processus difficile à mettre en œuvre, l'utilisation généralisée de normes telles que SGML et plus récemment XML a considérablement amélioré la situation ;

– l'*approche linguistique* : elle est issue des logiciels d'interrogation documentaire par requête booléenne sur des mots-clés descriptifs de chaque fragment documentaire, étendus à l'ensemble des mots d'un texte quand des corpus de « texte intégraux » sont disponibles en ligne (bases de dépêche, d'articles, de brevets, etc.). Afin de réduire le « bruit » documentaire, des connaissances linguistiques sont incorporées : normalisation des termes, sélection des seuls mots ou expressions composées pertinentes ;

– l'*approche statistique* : cette approche établit une transition entre les systèmes documentaires et les hypertextes. En effet des recherches ont été développées, depuis les origines de la documentation automatique dans les années 1960, sur une autre façon d'interagir avec les bases documentaires : au lieu d'inviter l'utilisateur à parcourir l'arborescence d'un plan de classement, ou à formuler ses demandes par des combinaisons booléennes de mots-clés, on lui demande seulement de décrire par quelques mots en vrac les sujets du « document idéal » qu'il recherche, ou plus simplement encore de désigner dans une liste de documents issus de son interaction antérieure le document le plus pertinent à ses yeux : le système lui renvoie alors une liste de documents « proches » par ordre de proximité décroissante au document de référence, liste limitée en longueur par un paramètre choisi, généralement quelques dizaines de documents.

Cette procédure d'élargissement de la demande porte généralement le nom d'expansion de requête (*query expansion*) et parfois de bouclage de pertinence (*relevance feed-back*) cette dernière notion englobant tout ce qui peut enrichir une requête initiale à partir du choix de documents ou de termes issus de la réponse à cette requête [Salton 68]. Elle est le plus souvent considérée comme un cas particulier de requête documentaire dans l'optique « question/réponse » mentionnée plus haut. Mais on peut la considérer aussi bien comme un mécanisme de génération automatique de liens hypertextuels : en effet elle fait correspondre à chaque document une liste de documents proches, c'est-à-dire un ensemble de liens (ici : valués, ordonnés par importance décroissante) vers d'autres documents et calculés de façon dynamique, c'est-à-dire qui se trouvent constamment à jour après suppression ou apport de documents nouveaux dans la base documentaire.

Dans le même objectif [CER 04] propose une méthode de la mise en hypertexte s'appuyant sur l'estimation de similarité lexicale entre sources et cibles potentielles d'hyperliens. Cette méthode a été expérimentée sur un document technique.

Pour voir plus d'informations concernant ses approches, consulter [BAL 96, JAC 02, LEL 99, LEL 03, OUE 01, SAL 03].

1.5.3. *L'hypertexte contextuel*

Comme nous avons constaté l'hypertexte devient plus difficile à gérer lorsque aucune couche n'est fixe, c'est le cas d'Internet où les informations en perpétuelle reconfiguration, sont dispersées sur un espace non descriptible et le modèle de parcours réalisé de façon disparate par une infinité mouvante de réalisateurs, est lui-même en constante transformation. Dans ce cas aucun modèle précédent ne peut réellement fonctionner sauf si un modèle unique de conceptualisation peut être imposé à l'ensemble des réalisateurs.

Le choix technologique de l'hypertexte se tourne alors [BAL 97] :

- soit vers une *absence totale de modèle*, c'est la solution d'Internet où chaque concepteur définit les liens qu'il veut vers les objets d'information qui l'intéressent (adresse). Dans ce cas le désordre est complet et elle pose les problèmes des pointeurs périmés, adresses inactives, information peu adéquates, etc. ;

- soit vers un *modèle minimal général sans aucun arrière-plan conceptuel* : celui à base de chaîne de caractère, ce le cas les plupart des moteurs de recherche actuels. Un tel modèle fait rapidement preuve de son inefficacité à réduire le bruit inhérent à ce type d'approche dans les langues naturelles qu'il produit un pourcentage important d'information inutilisable pour des raisons diverses ;

- soit vers un *modèle conceptuel* : celui à base d'ontologie, c'est la solution appliquée dans certaines moteurs de recherche comme NorthOnlight. Un tel modèle est applicable sur un domaine spécifique. Le moteur de recherche Northernlight (northernlight.com) a développé cette approche pour guider l'utilisateur dans sa recherche. L'avantage de cette approche est que les catégories peuvent être plus fines et « parlantes » que des groupes nominaux extraits automatiquement du corpus, mais l'inconvénient majeur est que l'ontologie doit être construite à la main, maintenue à jour en permanence et n'est opérationnelle qu'en une langue donnée [MOT 00] ;

- soit un *modèle adaptatif* mais dont l'adaptabilité ne peut plus dépendre du niveau conceptuel implémenté par le réalisateur de l'hypertexte puisque dans ce cas, il n'y a pas un réalisateur mais une infinité mouvante de réalisateurs. La couche d'analyse définissant l'adaptabilité du modèle doit être, au moins partiellement renseignée par l'utilisateur lui-même. Ce modèle fait appel aux agent dits « intelligents ». Dans ce cas l'utilisateur définit lui-même ce qui doit être le niveau conceptuel qui s'autonomise par rapport à la couche informations. Le module calcul construit des parcours en fonction de ce modèle particulier. Ce modèle est adaptatif à la fois dans les dimensions du temps et de l'espace puisque la description conceptuelle peut toujours être redéfinie.

Quelles que soient les technologies mises en œuvre, cette approche consiste en fait à ajouter une quatrième couche au modèle hypertexte, une couche de définition modifiant le modèle de conceptualisation en fonction de besoins particuliers ou momentanés de l'utilisateur. Cette méthode d'hypertextualisation prenant en compte l'état de l'utilisateur est la seule à introduire une contextualisation puisque, introduisant une relation structurelle à l'extérieur du dispositif hypertextuel proprement dit, elle cherche à prendre en compte les besoins particuliers d'un individu donné, éventuellement à un moment défini, c'est-à-dire en tenant compte d'une dynamique d'évolution.

L'hypertexte idéal étant celui qui, dans un espace informationnel ouvert, répond de façon spécifique à chaque utilisateur particulier, sa technologie doit tendre vers la disparition des graphes préétablis : plus de liens préalables mais une auto-constitution dynamique des liens. Tout ceci se complique alors du fait que l'on se trouve dans des univers d'information à la fois multimodaux et multilingues.

1.6. Typologie des systèmes hypertextes

On peut établir des distinctions entre plusieurs types d'hypertextes selon plusieurs aspects. La typologie des systèmes a été faite en se basant sur la conception et l'utilisation comme critères de distinction.

Selon la structure des systèmes, on peut distinguer quatre types d'hypertextes :

- le microtexte : c'est un texte non linéaire dont les liens explicites renvoient à ses différents fragments ou ses différentes parties ;
- le macrotexte : il s'agit d'un hypertexte avec des liens entre plusieurs documents ou plutôt une collection de documents ;
- l'hypertexte de travail en collaboration : il s'agit d'un texte créé ou consulté par plusieurs personnes en même temps. Il peut être également utilisé dans le cadre de cours, étant donné que les liens entre les groupes de texte permettent d'accéder aux discussions entre les auteurs ;
- l'*expert text* : il s'agit ici, d'une combinaison de la technologie des systèmes experts avec l'hypertexte. En effet, ce type d'hypertexte offre la possibilité de lier les textes, ainsi que de déclencher des réponses différentes suivant que certains liens sont traversés [RAD 91, TEA 95].

Selon leur utilisation, on peut établir une distinction entre quatre types d'applications hypertextuelles [LEC 90] à savoir :

- des bibliothèques universelles (*Macro Literary Systems*) : ces systèmes sont basés sur les liens entre les documents eux mêmes et les liens entre les documents et les commentaires (critiques, annotations...);
- des outils pour la résolution des problèmes (*Problem exploration tools*) : ces outils hypertextes sont particulièrement adaptés à l'écriture collective de logiciels et à l'analyse des situations ;
- des systèmes de feuilletage ou de butinage d'information (*browsing systems*) : ces systèmes sont destinés à être consultés facilement par le public, ils permettent de circuler entre des éléments d'information, de les annoter, d'en extraire des parties ;
- des outils de réalisation d'hypertextes (*General Hypertext Technology*) : ce sont des instruments de gestion de données introduites dans l'hypertexte, d'établissement des liens entre les éléments d'information et de rédaction des informations.

Le type le plus connu des outils proposés aux créateurs d'hypertextes est le logiciel *HYPERCARD* développé par Bill Atkinson pour le Macintosh [CON 87].

Par ailleurs, on peut dresser une typologie des systèmes hypertextes à partir d'une typologie de besoins [DAC 90] :

- pour les scientifiques, le besoin d'établir un environnement qui les assiste dans leurs tâches d'élaboration des savoirs et gestion des connaissances ;
- pour les informaticiens le besoin de gérer les relations qui existent entre les programmes.

En effet, à partir de ces deux types de besoins, on distingue deux types de systèmes :

- des systèmes destinés à un utilisateur qui peut être à la fois lecteur et scripteur, ils se présentent comme étant des outils de prise, de collecte et de synthèse des notes, des outils d'assistance à l'écriture. Bref, des outils du « travailleur intellectuel » ;
- des systèmes qui représentent des environnements permettant de gérer les produits depuis leur conception jusqu'au produit final.

Une autre typologie peut également être faite à partir du contenu de l'hypertexte : hypertexte définitionnel, hypertexte littéraire, hypertexte narratif, hypertexte documentaire et hypertexte éducatif¹.

1. Vous pouvez trouver une synthèse sur cette typologie faite par l'IUFM de l'académie de Grenoble sur le site : <http://www.grenoble.iufm.fr/depart/francais/hypertxt/typolo.htm>.

1.7. Hypertexte et l'interactivité

En prenant la définition de l'interactivité : *inter* et *actif*, tous les deux créant une relation dynamique, un certain pouvoir de décision et de contrôle de l'utilisateur sur la production ou la consultation d'une œuvre sont désormais réalisables. Ces pouvoirs lui permettent d'agir sur l'œuvre en procédant à un nouvel ordonnancement des informations selon ses besoins et ses envies. Il devient auteur parce qu'il réinvente l'œuvre en participant à la structuration de l'hypertexte, en créant de nouveaux liens. Certains systèmes enregistrent les chemins de lecture et renforcent (rendent plus visibles, par exemple) ou affaiblissent les liens en fonction de la manière dont ils sont parcourus par la communauté des utilisateurs. L'interactant en parcourant, sélectionnant, associant des informations reconstruites en permanence, devient en quelque sorte l'interprète d'un hypertexte.

Une partie du travail de l'auteur hypertextuel consiste à évaluer le degré de confiance qu'il doit accorder au lecteur pour pouvoir inventer sa propre lecture, interpréter, comprendre et sentir ce qui n'est pas explicitement exprimé.

L'introduction d'interactivité dans les œuvres permet de rapprocher les concepteurs et utilisateurs, et de faire fusionner de plus en plus l'auteur et le lecteur. Un exemple particulièrement éclairant sur ce rapprochement concernant l'art traditionnel qui présentait sous forme d'objet achevé (exemple : tableau) créé par l'artiste et qu'il donnait à voir, à écouter, rarement à toucher. Contrairement à l'art électronique interactif qui est offert au spectateur, ce n'est plus une œuvre figée mais un système ouvert, un dispositif, qu'il peut manipuler, modifier sur lequel il peut agir, et qui inversement, agira sur lui. C'est donc un système interactif par excellence.

Lecture et écriture hypertextuelle

La lecture de l'hypertexte est la forme interactive entre un environnement et le lecteur qui négocie en permanence la trajectoire de son parcours. Il emprunte des « configurations potentielles », dont certaines ont été planifiées par l'auteur, tandis que d'autres sont le produit de son implication, activant les liens qui lui sont proposés ou qu'il crée suivant les systèmes.

Les différents systèmes fragmentaires sont combinés et liés tout en restant identifiables, ils donnent la possibilité de créer, à travers les possibles parcours, un *scénario* de lecteur dans un hypertexte.

L'hypertexte est un modèle interactif en reconfiguration permanente incluant autant un contexte physique spécifique qu'un environnement global et numérique.

Cette structure ouverte est acquise grâce à l'organisation libre de fragments décontextualisés et dématérialisés offrant au lecteur un parcours fluide et instantané à travers plusieurs possibilités de textes.

L'accès instantané à l'information forme une lecture ininterrompue à travers les fragments délocalisés et renforce le fait d'assister à une expérience en direct. Par conséquent le passage du texte à l'hypertexte est la transformation d'une écriture séquentielle – d'une structure articulée en une écriture événementielle.

La lecture d'hypertexte est devenue un des grands thèmes de controverse de l'hypertexte. En effet, ses détracteurs ont toujours souligné les risques de désorientation qui guettent le lecteur dans l'exploration des espaces d'informations électroniques hypertextuels qui lui sont proposés. A cela ses partisans originels [BRO 88, NIE 90] ont montré que ce phénomène de désorientation engendré par l'hypertexte était un faux problème que l'on résout aisément au moyen d'une méthodologie liée à la construction des réseaux hypertextes et d'écriture dédiée à l'hypertexte. La construction d'un tel réseau va demander de la part de l'auteur des capacités certaines de scripteur mais également d'architecte car va lui échoir l'obligation de définir les unités modulaires mais également celle de créer les liaisons qui existeront. Enfin, la résorption conséquente de la désorientation liée à la lecture d'hypertextes consistera à impliquer l'utilisateur/lecteur. Cette implication que Bernard Stiegler [STI 93] souligne par « lire, c'est écrire » et que l'on retrouve dans le projet de définition du « poste informatique de grands lecteurs » destiné à la Bibliothèque Nationale de France, à l'usage des chercheurs, universitaires et étudiants, recouvre l'accès aux ouvrages numérisés d'une partie du fonds documentaire archivé par la BNF et la possibilité d'associer aux documents consultés des annotations personnelles que le lecteur/auteur (ici les deux profils se retrouvent cumulés) pourrait ensuite modifier, compiler, mettre à la disposition d'autres...

Cette distinction franche entre la fonction de lecteur et celle de scripteur de documents hypertextes s'est sournoisement imposée par l'usage massif du navigateur Internet (que Mosaic, Communicator et Explorer ont largement contribué à populariser), ce fameux feuilleteur (de l'anglais *browser*) dont l'utilisation se limite à un simple « tourne-page » électronique. Le « feuilletage » lié au navigateur Internet cristallise dans une certaine mesure un comportement stéréotypé du lecteur à qui on enlève cette faculté intellectuelle de la lecture au « profit » d'un parcours superficiel des documents auxquels il a accès (on peut noter ainsi le peu de connotation cognitive associée à l'expression « feuilletter un document » où le document dont il est question est souvent un magazine ou un catalogue [JAM 98]. En revanche, « lire en diagonale » n'enlève assurément rien à une forme

d'exploitation cognitive pour les documents sujets à ce type d'exercice. [CHE 92] souligne : *si vous lisez vite, vous comprenez mieux*).

Autrement dit, les activités combinées de lecture et d'écriture de l'utilisateur d'hypertextes se fondent dans une activité dite « lectature » [WEI 99]. Une fois engagé dans une navigation hypertextuelle, le lecteur d'hypertextes devient producteur d'un texte, puisqu'il élabore un programme de lecture à géométrie variable par concaténation de fragments, par ajouts et retraites de nœuds et de liens, par constructions et destructions de réseaux d'informations préexistants. Toutefois, la visée ne concerne pas spécifiquement la navigation. Pour Jean-louis Weissberg [WEI 99] le simple fait de lire un texte est déjà en soi un acte de réécriture : *Lire, c'est réécrire pour soi le texte, et écrire, c'est enchaîner sur des lectures*. Grâce à une infinité de lecture, l'interactivité en plus de vivre, va se réinventer. Pour autant, il est concevable d'imaginer une encyclopédie dont les utilisateurs, au fur et à mesure qu'ils la lisent, créent de nouveaux liens, ajoutant des textes, ainsi de suite. L'hypertexte se prête à ce genre d'opération.

On peut constater que la démarche de lecture sur un support numérique prend une autre dimension car cette démarche n'est plus une démarche de mémorisation et de consommation mais une démarche de réflexion et de construction. Si la lecture d'un document sur un support numérique tel que le cédérom peut différer de la lecture du même document sur le médium livre, ce n'est pas parce que le médium est différent mais parce que les possibilités offertes à la lecture ne sont pas les mêmes.

1.8. Outils indispensables pour l'hypertexte

Créer, maintenir et rendre accessible l'information dans une structure comme l'hypertexte implique des objectifs d'aide à la navigation pour l'utilisateur et d'organisation de l'information pour le gestionnaire.

Pour aider l'utilisateur à naviguer dans un espace hypertextuel, il faut avoir des outils qui permettent de :

- *détecter les profils d'utilisateur* : les utilisateurs dans un espace comme le Web ne se ressemblent pas. Le système idéal devrait être capable de s'adapter de façon interactive aux besoins de chaque utilisateur ou groupe d'utilisateurs, en tirant parti de profils constamment dynamiques ;

- *représenter le contenu sémantique des liens hypertexte* : les liens hypertextuels entre les pages portent des ambiguïtés sur la signification de ces liens. Afin

d'augmenter la qualité de la navigation, il est nécessaire d'étendre le rôle syntaxique de ces liens en leur associant un contenu sémantique.

Pour l'aide à l'organisation de l'information, il faut avoir les outils qui permettent de :

– *régler automatiquement les mises à jour des pages et de liens* : les modifications de contenu de documents et de liens sont source de confusion dans la navigation, et notamment la présence de liens obsolètes. La mise à jour sur un grand ensemble de données étant fastidieuse ou impossible, il convient donc d'automatiser autant que possible cette tâche. L'automatisation ne doit pas mettre en cause la cohésion de la structure de l'hypertexte ;

– *veiller à l'optimisation de la structure hypertexte* : dans sa navigation, l'utilisateur a besoin d'aide ou d'orientation mais aussi de la liberté, de façon à préserver le caractère explorateur de sa démarche. Le chaos – la surabondance des liens – et la rigidité – la pauvreté des possibilités de navigation – font tous les deux obstacle à la recherche d'information. Il est donc précieux de pouvoir définir des procédures automatiques capables de fixer un compromis entre ces deux extrêmes, qu'il s'agisse du nombre de liens ou de la structure sur réseau qu'il forme ;

– *gérer des versions* : un hypertexte est susceptible d'évoluer au gré des opérations de modification, de suppression et d'insertion de nœuds et de liens. Cette mise à jour perpétuelle de l'hypertexte pose le problème concret de savoir si l'hypertexte 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. Le contrôle de version, inspiré des techniques du génie logiciel, est destiné à gérer plusieurs versions du même objet.

On peut constater que ces problèmes sont liés et que leurs solutions respectives devraient s'influencer mutuellement.

1.9. De l'hypertexte à l'hypermédia

Pour ne pas reprendre ici toutes les définitions que nous avons attribuées à l'hypertexte et qui sont valables à l'hypermédia, nous allons plutôt évoquer les différences entre les deux systèmes. Ces différences se situent plutôt du côté du contenu des nœuds. Les nœuds ne contiennent plus seulement du texte, mais peuvent être composés de médias divers, tels que des images, des séquences animées, des vidéos, etc. Toutefois, comme nous avons signalé dans l'introduction, plusieurs auteurs utilisent indifféremment les deux concepts hypertexte et hypermédia pour montrer que le contenu ne représente pas l'intérêt majeur de systèmes, c'est plutôt l'hyperespace et la structure qui intéressent les utilisateurs. L'hypermédia est donc

basé sur la technologie de l'hypertexte et son organisation à base d'hyperliens, et désigne les documents qui résultent d'une combinaison du multimédia et de la structure hypertexte.

Cependant, pour parler du contenu d'un hypermédia, tout en prenant en considération la définition du terme « média » donnée par [LEG 93] : *moyens basés sur la technologie, permettant de diffuser et de conserver des informations ainsi que de développer des attitudes, de susciter une rétroaction et de favoriser l'expression*, nous pouvons évoquer deux types de médias, ceux qui reposent sur les systèmes de conservation d'informations (multimédia) et ceux qui reposent sur les systèmes de communication (hypermédia) (voir la figure 1.5). Nous allons présenter quelques éléments de définition de chacun de ces concepts (médias, multimédia et hypermédia).

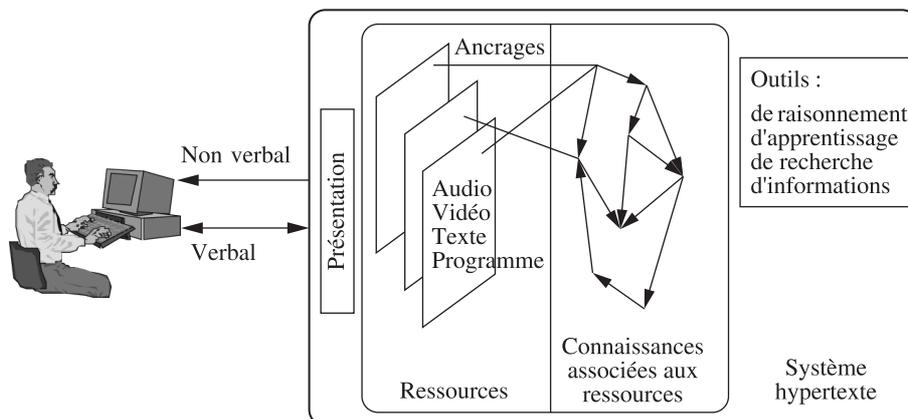


Figure 1.5. Modèle de système hypertexte [NAN 95]

1.9.1. Multimédia

Le terme multimédia, comme son nom l'indique, renvoie à l'intégration des formes multiples de médias (figure 1.6). Ceci inclut le texte, les graphiques, l'audio, la vidéo, etc. Une présentation impliquant audio et clips vidéo par exemple est considérée comme une « présentation de multimédia ». Un logiciel éducatif qui comporte des animations, du son et du texte est un « logiciel multimédia ». CD et DVD sont souvent considérés comme des « formats multimédia » puisqu'ils peuvent stocker beaucoup de données et la plupart des formats multimédia exigent beaucoup d'espace disque.

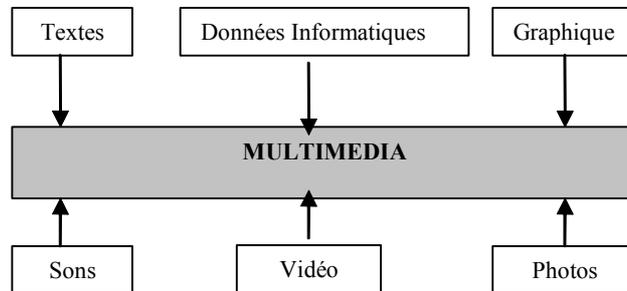


Figure 1.6. Les composants du multimédia

Le multimédia est donc né de l'intégration de différentes techniques. Il fait référence à l'interaction « homme/machine » et résulte de la fusion de l'ordinateur, de la télévision, de la chaîne hi-fi et du téléphone, cumulant les possibilités d'animation de l'une avec les capacités de traitement de l'autre, dans le but de présenter simultanément sur un écran d'ordinateur, avec accompagnement adéquat et sur plusieurs fenêtres si possible, les informations en provenance de diverses sources [CHE 92].

La notion de multimédia implique l'aptitude pour une machine et une application de gérer plusieurs canaux de communication sensorielle avec le son, l'image animée, parfois des informations tactiles, olfactives, éventuellement même des mouvements ou des efforts dans le cas de recours à la réalité virtuelle.

La notion de multimédia n'implique nullement celle d'hyperdocument [NAN 94]. La présence d'images animées et de son n'est pas le garant d'une véritable interactivité entre usager et machine. Il est parfaitement possible de remplacer au sein d'un document linéaire une image par une animation vidéo, ou d'y ajouter un fond sonore. Ceci en fera un document multimédia mais n'en fera en aucun cas un hyperdocument.

Nous définissons un média (électronique) comme un mode particulier d'échange d'informations entre un utilisateur et un gestionnaire de documents, indépendant du modèle physique de la présentation ou du modèle logique du gestionnaire.

Les médias peuvent présenter plusieurs caractéristiques. Un média – un type primitif de données – est *formaté* quand toute la sémantique statique et opérationnelle de sa présentation est reconnue par le système d'informations. En plus, il est nécessaire qu'une partie au moins des opérateurs soit visible au niveau du langage de requêtes et de manipulation de données : notamment certaines opérations de

sélection et de mise à jour. Par exemple, les entiers et les chaînes de caractères sont des médias formatés pour les SGBD. Un média est *non formaté* quand sa sémantique n'est pas reconnue, ou n'est reconnue que partiellement par le système d'informations. Par exemple, pour la plupart des gestionnaires de documents, le type de donnée image est non formaté car la sémantique de sa représentation reste inconnue du système : les images sont de simples blocs de données (*bulk data*) ou des chaînes de bits. En fonction de l'architecture du système, une donnée offre plusieurs visions. Par exemple, pour un gestionnaire d'objets, un texte est non formaté mais la sémantique de cette donnée pourrait être reconnue par le système d'informations.

Il existe des médias actifs et des médias passifs.

– un *média est actif* si, lors de sa présentation, le temps détermine plusieurs de ses propriétés. Les propriétés temporelles des médias actifs nécessitant pour son affichage la représentation au niveau modèle de données, sont de trois types :

- premièrement, les données actives – un son ou une vidéo – nécessitent des intervalles de temps, des durées fixes pour leur présentation. Par exemple, une donnée du type vidéo nécessite un intervalle de temps pour sa restitution,
- deuxièmement, une référence croisée, un lien, d'un média actif, porte sur un intervalle temporel de validité. Par exemple, pour une donnée du type vidéo, une référence croisée pourrait être présentée, à partir de l'apparition d'un personnage quelconque, et jusqu'à sa disparition de la scène,
- troisièmement, lors de l'accès aux médias actifs, la récupération des données pour l'affichage doit se faire sous des contraintes temporelles. Par exemple, la présentation de deux échantillons consécutifs d'une donnée de type son, doit se faire avec une différence de temps inférieure à un certain seuil de manière à ne pas percevoir de distorsions lors de sa restitution sur un haut parleur. Ces médias sont donc « dépendants du retardement » aussi bien pour leur présentation, que pour leur récupération des dispositifs de stockage. Ils nécessitent l'introduction des propriétés temporelles au niveau modèle de données ;

– un *média est passif* si, au niveau modèle de données, sa présentation est indépendante des aspects temporels : entiers, réels, chaînes de caractères, textes ou images. Bien que dans la pratique, les données passives engagent des retardements pour leur présentation, ces propriétés n'ont pas besoin d'être ajoutées au niveau modèle de données. D'un point de vue formel le temps de présentation peut être considéré comme un point dans la dimension temporelle, alors qu'il correspond à point un d'intervalle pour les données actives.

Ainsi, la présentation d'une donnée de type entier ou image peut être réalisée sans introduire des propriétés temporelles au niveau modèle de données. Par ailleurs,

les liens définis sur les données passives portent sur une période de validité indépendante des médias. Par exemple, une référence croisée d'un texte reste normalement active tout au long de la présentation. En outre, lors de la présentation de données passives, le temps de leur présentation n'est pas restreint à des contraintes temporelles. Pourtant, la cohabitation des médias passifs et des médias actifs nécessite l'ajout de propriétés temporelles au niveau modèle.

La taille des données des nouveaux médias exige leur codification pour permettre leur compactage sur deux dimensions différentes. D'une part, les données ne peuvent pas tenir en mémoire centrale. Cette problématique, en ce qui concerne les médias actifs dépasse le cadre des SGBD orientés grande mémoire. Par exemple, dans la plupart des cas la donnée d'un attribut de type vidéo ne pourrait pas tenir en mémoire centrale. En général, les données ne doivent pas être stockées dans leur état final de représentation même dans des dispositifs de stockage de haute densité. Pour rendre plus compacte leur représentation, des données telles que la vidéo ou le son doivent être compactées. Par ailleurs, mêmes sous une représentation compacte, certaines données ne peuvent pas être filtrées efficacement suite à une requête. Cela exige une représentation compactée consultable lors de la résolution d'une requête.

Un média est adressable lorsqu'il est possible de soumettre d'une manière performante des requêtes sur son contenu, par exemple sur les entiers ou les réels. Néanmoins, dans certains gestionnaires pour des données comme les images, la soumission de requêtes sur le contenu n'est pas autorisée : ce média n'est pas adressable vis-à-vis de son gestionnaire.

Un média est considéré comme spécialisé lorsqu'il est spécifique à un domaine d'application quelconque, par exemple une formule mathématique. L'existence des médias spécialisés implique le développement de systèmes ouverts capables de gérer des types comprenant leur représentation et leur sémantique.

Les gestionnaires de documents multimédia étaient, dans les années 1980, la préoccupation essentielle des chercheurs en bases de données. Les systèmes essayaient de considérer la sémantique des nouveaux types de données comme celle des images et quelquefois de certains types spécialisés comme des formules. La notion d'hypertexte fait intervenir une notion de référence et de structure inexistante dans les gestionnaires de documents de ces dernières années. Ces documents introduisent des types de données comme la vidéo, en y ajoutant une dimension temporelle au sein du système d'informations. L'intégration des idées dégagées par les systèmes d'informations multimédia, les nouveaux médias, les techniques hypertextes et l'avancée des technologies de stockage – (disques) – et de restitution – stations de travail – sont à l'origine de la notion de document hypermédia.

1.9.2. Hypermédia

Ce concept fait référence à l'interaction « homme/connaissance ». Il désigne un mode de gestion de l'information où les unités d'information sont connectées entre elles par des liens.

Les hypermédias favorisent un accès facile et direct (non séquentiel) à l'information et permettent le travail collaboratif : les utilisateurs sont plus actifs, plus responsables et plus engagés (figure 1.7). Ils échangent leurs informations, collaborent à la tâche à accomplir. L'utilisateur peut en même temps être l'auteur de ses apprentissages puisque c'est lui qui façonne l'information pour la rendre pertinente.

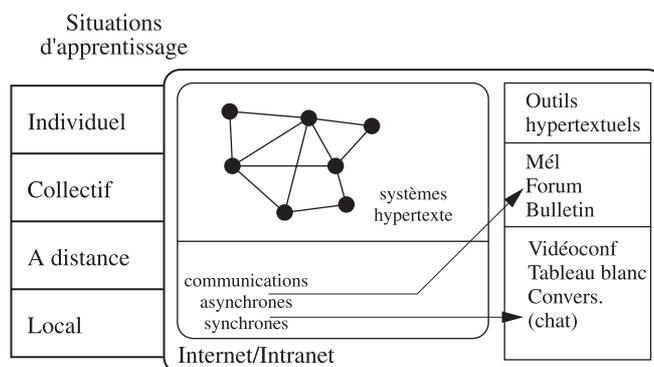


Figure 1.7. Apprentissage au travers d'Internet (ou d'un intranet) vu comme un système hypermédia

Les hypermédias permettent ainsi l'accès à l'information par branchements et un cheminement associatif dans les textes, les graphiques, les vidéos, laissant la porte ouverte à la créativité que suscite ce type de navigation dans l'information. C'est l'interactivité sous-jacente aux possibilités de manipulation et d'appropriation de l'information, qui rend les hypermédias aussi intéressants... L'interactivité rend possible un cheminement individualisé, chaque usager choisissant d'activer tel ou tel lien au gré de son intérêt. [BES 02].

L'hypermédia, qui conjugue hypertexte et multimédia, est à l'heure actuelle le support le plus riche pour l'utilisateur, car il obéit à une structure réseau. Il se rattache aux *médias commutatifs* (type ordinateur, téléphone) par opposition aux *médias irradiants* (type presse, radio, télévision) qui diffusent l'information à partir d'un centre. Il représente un moyen de communication, d'information et de

formation très puissant et très souple qui s'impose de plus en plus. Ajoutons que, dans la pratique, les termes multimédia et hypermédia ont tendance à devenir synonymes.

1.9.2.1. *Hypermédia et Interactivité*

Le réseau Internet et l'arrivée massive des hypertextes et hypermédias ont remis en question nos repères par rapport à la lecture et l'écriture. L'hypermédia implique des grandes réflexions sur les stratégies de lecture et de navigation, mais aussi sur la structure même de la masse sans cesse grandissante d'informations de différents types. Ces réflexions ont fait émerger un nouveau concept, celui de la « lecture interactive » (voir la section 1.7). Par ce concept, nous entendons une lecture d'un document diffusé sur support informatique comprenant des nœuds d'informations reliés entre eux par des liens de natures diverses : textes, images statiques ou animées, séquences sonores, vidéo... Cette lecture appelle à la logique des individus, à leur sens de l'observation, à leur réflexion, à leur imagination, mais aussi à leurs libertés et leurs choix.

L'interactivité est donc la propriété de ce qui permet un échange. Elle s'applique à plusieurs types de communications et de dialogues établis entre un utilisateur et une machine à travers un écran de visualisation. Nous pouvons distinguer dans la littérature plusieurs situations, à savoir l'interactivité de sélection, l'interactivité de contenu, l'interactivité de sélection et de contenu, l'interactivité de création de contenu, etc.

Selon Berners-Lee, la définition de l'interactivité sur le Web ne comprend pas seulement : *l'habilité de choisir mais aussi l'habilité de créer. Nous devons être capables au même temps de trouver n'importe quel type de document sur le Web et d'en créer facilement d'autres. Nous devons être capables de suivre des liens et de les créer – parmi tous les types de médias. Nous devons être capables d'interagir avec d'autres utilisateurs, mais aussi de créer avec les autres. L'interactivité est le processus de créer ou de résoudre des problèmes ensemble.* [BER 99].

1.9.2.2. *Hypermédia adaptatif*

Les hypermédias adaptatifs ont été créés suite à des travaux qui ont essayé de minimiser les aspects négatifs de l'hypermédia en général, ceux qui sont liés à la désorientation et la surcharge cognitive. L'objectif de ces systèmes est d'adapter la présentation des connaissances et d'aider l'utilisateur à se diriger dans l'hyperespace.

Ceci se fait par les possibilités de modifier le contenu des pages ainsi que les liens entre les différentes pages. Plusieurs techniques ont été développées notamment pour l'adaptation des liens, des techniques d'ordonnancement, de

masquage, d'annotations, etc. [BRU 96]. L'architecture de ces systèmes repose sur le modèle du domaine et celui de l'utilisateur. Ces dernières années, et avec le développement de l'enseignement à distance, les recherches se sont orientées plutôt vers les « hypermédias adaptatifs dynamiques » afin d'améliorer la qualité de l'adaptation et de prendre en compte la construction dynamique des pages. Dans ce type de système, nous prenons en compte outre le modèle du domaine et celui de l'utilisateur, la base de données des objets pédagogiques et le générateur de cours (dans un contexte d'enseignement à distance, voir le chapitre 4)

Pour bien illustrer la présentation de ces systèmes, nous signalons ici qu'il faut distinguer entre les hypermédias adaptatifs et les hypermédias adaptables, dont la plus grande différence entre eux est que les premiers s'adaptent automatiquement selon la navigation de l'utilisateur, contrairement aux seconds qui demandent par questionnaire le type du parcours que le navigateur préfère suivre.

Les hypermédias adaptatifs construisent l'interaction en prenant en compte les buts, les préférences et les connaissances de l'individu pour adapter leur utilisation selon les besoins particuliers. Ainsi, il est possible de présenter une version « personnalisée » d'un document hypermédia sans que l'internaute doive programmer cette visualisation. Ces systèmes peuvent avoir en gros trois fonctions :

- enregistrer le parcours et les actions de l'utilisateur dans un document hypermédia ;
- classer les liens en groupes en considérant les connaissances, les intérêts, et les objectifs de l'utilisateur ;
- montrer, cacher, surligner ou souligner quelques fragments de la page lors de sa visualisation [DEB 99].

1.10. HyWebMap (Hypertext Web Mapping) : un système hypertextuel sur le Web

HyWebMap² est un système hypertexte pour la navigation et la structuration des documents sur le Web (figure 1.8) créé dans le cadre de l'élaboration et de la conception des systèmes hypermédias à vocation didactique au laboratoire Paragraphe à l'université Paris 8 [BOU 02, PAP 01]. Il est utilisé entre autres dans le domaine de l'éducation, la veille technologique et l'enseignement à distance et constitué, suite à une étude des besoins des utilisateurs, de plusieurs niveaux.

2. Ce logiciel a été réalisé par Nasreddine Bouhaï dans le cadre de sa thèse [BOU 02].

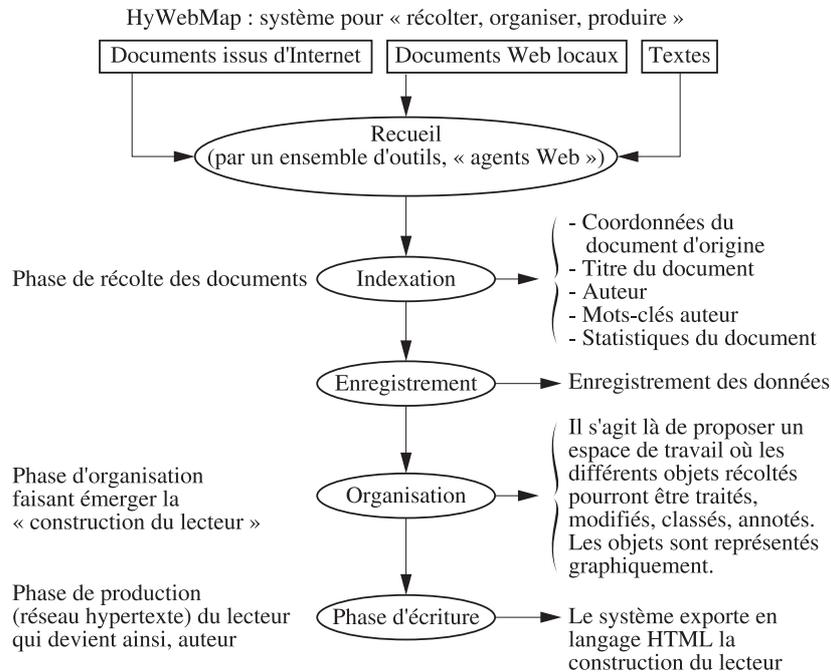


Figure 1.8. Schéma fonctionnel du HyWebMap

Nous reconnaissons 5 tâches essentielles à l'utilisateur [SAL 01] :

– *l'extraction d'information* : sur des bases de critères exploitant les résultats des moteurs de recherche existants, l'utilisateur extrait des informations répondant à ses préoccupations, et nous complétons cette extraction par une exploration de catégories arborescentes ;

– *le 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 obtenues lors de l'extraction ;

– *la « veille »* : 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 ;

– *la « 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 ;

– *la personnalisation 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.

Ces fonctionnalités nous servent de modèle à base de couches logiques pour la conception d'*HyWebMap*. L'originalité d'*HyWebMap* repose d'une part sur ces *cinq fonctionnalités* qui s'appellent les unes les autres, et d'autre part sur la *transversalité des usages* ; l'éducation, la gestion de sites Web, la création de portail, la veille technologique, etc.

1.10.1. Fonctionnalité

1.10.1.1. *Agent Web du HyWebMap*

Le fonctionnement du système *HyWebMap* est basé essentiellement sur l'utilisation des techniques d'agents Web (autrement dit agents logiciels), ces outils qui peuvent avoir une multitude de tâches différentes, se substituent à l'utilisateur pour faire certaines tâches qui s'avèrent souvent lourdes, comme la veille et l'extraction d'information. Il nous était nécessaire de procéder à la conception et au développement de ce type d'outils informatiques. Dans un environnement aussi ouvert que dynamique comme le Web, *HyWebMap* a comme première vocation la réorganisation de l'information issue des espaces d'informations existants (comme les sites Web en ligne).

Cet outil répond à la tâche suivante : 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. Sur le même principe d'un navigateur, l'agent exploite les URL absolus (adresses complètes). Tous les liens relatifs découverts sont convertis en liens absolus pour pouvoir formuler des requêtes de lecture.

1.10.1.2. *La cartographie de site Web (réseau virtuel)*

HyWebMap s'est inspiré de l'approche « cartographique », car on a constaté que les cartes conceptuelles sont plus expressives lorsqu'elles sont construites par les lecteurs eux-mêmes. Dans l'activité de recherche et de sauvegarde des informations pertinentes sur le Web, le lecteur a 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 au lecteur d'élaborer à partir d'une cartographie existante, une cartographie personnelle.

La carte *HyWebMap* dite *réseau virtuel* présente l'espace d'un site visité (les nœuds et les liens) par l'agent Web (figure 8 renvoi, renvoi).

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 pages sources, ils sont seulement stockés sur le disque du lecteur et sont actifs lorsque la carte est invoquée.

1.10.1.3. *Stockage et transformation de données*

Dans un contexte de recherche et de réorganisation d'information issue du Web au sein d'espaces locaux comme le propose le système *HyWebMap* sous forme de cartographie personnelle, il était nécessaire de procéder au stockage et à la transformation des données collectées pour les rendre plus maniables au niveau consultation et présentation.

Dans cette perspective, un espace d'information candidat subit une analyse complète par un agent Web et sera décrit sous forme de tables et de relations entre des ensembles de nœuds, de liens, de pages, de profils, et de propriétés. Nous utilisons la représentation relationnelle pour décrire le contenu des nœuds. Les principales caractéristiques des pages explorées se retrouvent ainsi organisées au sein d'un modèle de données comprenant huit relations.

1.10.1.4. *Les relations de la base de données*

La base de données d'un réseau *HyWebMap* est constituée essentiellement de 8 relations :

- nœuds (N) : les nœuds du réseau *HyWebMap* ;
- pages (P) : les contenus des nœuds, généralement une ressource Web ;
- relations(R) : les liens hypertextes entre les différents nœuds ;
- annotations (A) : contenu annotatif de chaque nœud ;
- profils (P2) : profil statistique du contenu (ressource Web) de chaque nœud ;
- infos (I) : informations (par exemple : nom d'auteur) du contenu (ressource Web) de chaque nœud ;
- mots (M) : cette relation représente les différents mots-clés associés à chaque nœud du réseau ;

- index (I2) : la dernière relation est un index de mots-clés construit manuellement par l'auteur ;

Ainsi, une fois la structure et le contenu de l'espace d'information convertis en cet ensemble de relations, le système *HyWebMap* proposera à l'utilisateur une présentation graphique lui permettant une navigation plus « fluide » et une manipulation « simplifiée » (figure 1.9).

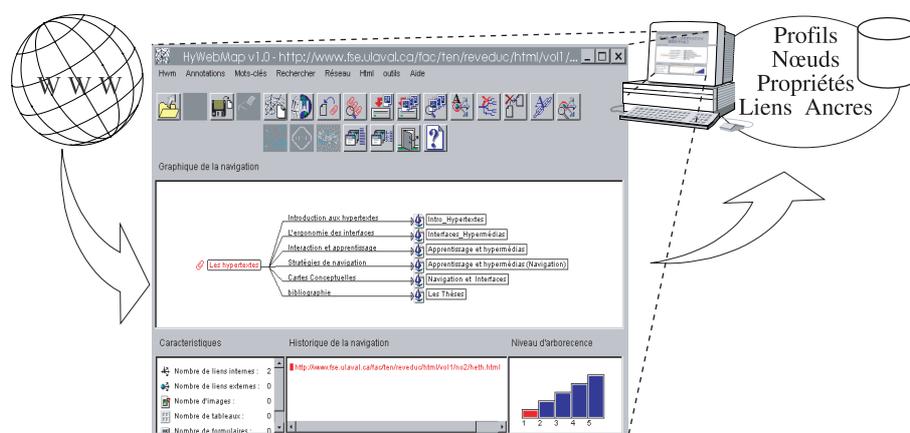


Figure 1.9. Présentation graphique d'un réseau HyWebMap

1.10.1.5. Un système d'agrégation de ressources

Dans sa manipulation la plus élémentaire, *HyWebMap* permet à l'utilisateur/internaute de constituer à partir de ressources documentaires exogènes 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 (figure 9) que nous appelons réseau *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, etc.).

L'utilisateur peut :

– naviguer localement dans le réseau virtuel sans établir de connections à Internet (puisque'il s'agit d'espaces locaux), ce qui améliore 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 ;

- réorganiser la structure d'un espace d'informations cartographié par rapport à ses connaissances personnelles sur le domaine, sans mettre en cause la structure de l'espace d'origine ;

- 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 recherche afin de les mettre à disposition d'utilisateurs spécifiques (agences de communication, bibliothécaires et documentalistes spécialisés, services d'assistance SVP, etc.).

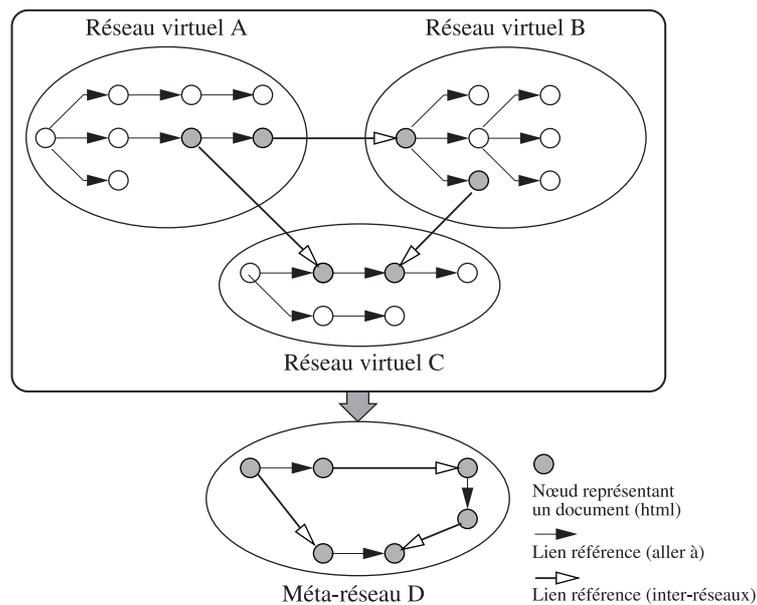


Figure 1.10. Un réseau composite HyWebMap

De manière identique, les 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,

etc.) 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 d'entreprise (sites marchands de comparatifs de prix, sites de fédération offrant à leurs adhérents une information exhaustive, etc.).

1.10.1.6. Typologies des nœuds et des liens dans HyWebMap

HyWebMap distingue cinq types des nœuds et sept types de liens pour assurer la cohésion du réseau virtuel et aider ainsi le lecteur à se repérer et naviguer.

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, à 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 des documents et interpréter le sens sémantique entre les nœuds. La diversité 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 diversité de liens sources est destinée à répondre à la question : *Que pourrais-je obtenir ici ?*

1.10.1.7. Déclaration de nœuds et de liens dans *HyWebMap*

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 doté des attributs suivants (facultatif ou non) :

- un nom (obligatoire) ;
- une ressource documentaire Web (URL de site, de page ou de site *HyWebMap*), (facultatif, en cas de nœud vide) ;
- des descripteurs ou mots-clés (facultatif, en cas de nœud vide) ;
- des annotations (facultatif).

Le système permet 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 » et à y ajouter la ressource Web souhaitée. Ce nœud affecté se retrouve automatiquement intégré au sein du réseau en cours d'élaboration grâce au lien défini.

1.10.1.8. 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'associations arbitraires présents dans *HyWebMap* vont permettre à l'utilisateur/concepteur de préserver les associations *document* ↔ *document* 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.

1.10.1.9. *Agents de surveillance et de mise à jour*

Une page Web est susceptible d'évoluer au gré des opérations de modification, de suppression et d'insertion de 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 comporte des mots-clés :
 - si oui, ils les prennent, lancent le moteur de recherche Google avec ces mots-clés, puis récupèrent le résultat en choisissant les trois premières pages classées par Copernic et 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 souhaitaient disposer sur leur machine d'une trace des contenus des pages Web externes. 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).

1.10.1.10. Agent de gestion des 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 « URL », 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'utilisateur sur son réseau, pour repérer les modifications (retraits ou enrichissements). Les propositions sont des nouvelles références documentaires sur le Web obtenues par une métarecherche. La version du réseau en cours est mise à jour sous un nom différent. L'ensemble des nœuds modifiés sur le réseau porte une remarque « NP » pour signaler à l'utilisateur qu'il s'agit d'une nouvelle proposition (figure 1.11).

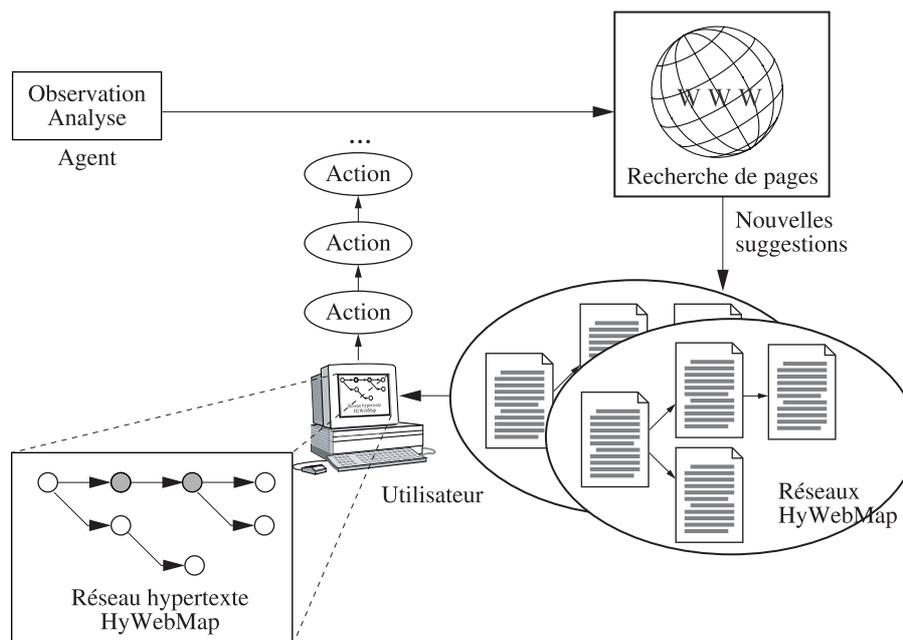


Figure 1.11. Agent de gestion de profils

1.10.1.11. Agent d'aide à la création d'un espace personnel

Lors de l'expérimentation, la plupart des utilisateurs ont souhaité disposer d'une aide pour la création de leurs réseaux personnels. Pour cela, le système *HyWebMap*

offre à l'utilisateur une fonctionnalité originale qui lui permettra d'obtenir 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 d'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 moteur de recherche Google avec les mots-clés de l'utilisateur ;

– *construction d'un réseau vide* : l'utilisateur construit un réseau personnel sans ressources externes (c'est-à-dire interne en respectant les types des nœuds et de liens, en nommant seulement le nœud et en introduisant les mots-clés (métadonnées) de ce nœud. Par la suite, les agents *HyWebMap* évoquent le moteur de recherche Google avec les mots-clés de chaque nœud, ils récupèrent les trois premières pages concernant les mots-clés de chaque nœud et par la suite l'utilisateur peut choisir une des trois ou sauvegarder les trois).

1.10.1.12. *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 que ses caractéristiques et son niveau dans le réseau virtuel. 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 les 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 au 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 de 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 (figure 1.12).

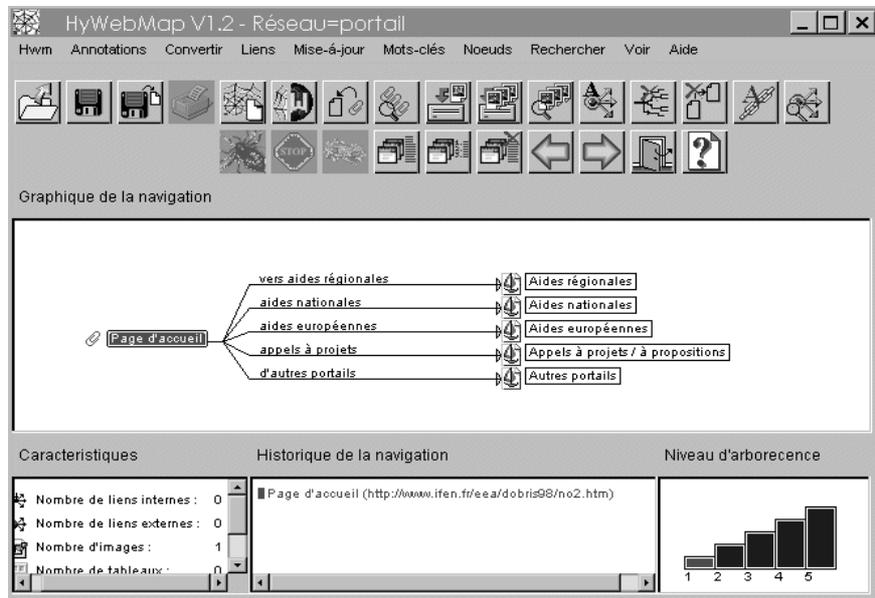


Figure 1.12. Interface Principale de HyWebMap

1.10.1.13. Description sémantique par annotations

HyWebMap permet aux lecteurs d'enrichir leurs réseaux virtuels par la création de leurs annotations personnelles sur un ou plusieurs nœuds. Une 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 annotation 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 saisisse 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 document courant* : les annotations sont affichées dans la partie basse de l'écran, les ancrs sont surlignées en rouge dans le document. D'autres systèmes insèrent une icône avant et après la phrase annotée [GOL 98, MAR 98]. En cliquant sur une annotation dans la fenêtre des annotations,

HyWebMap force le navigateur à se positionner sur l'ancrage de l'annotation. Cette technique facilite la navigation dans la liste des annotations vers le document ;

– *changement du document* : lorsqu'un changement survient dans le document, le texte surligné est parfois introuvable dans le document. Les annotations apparaissent alors en gris dans la liste des annotations. Ainsi, même lorsqu'un document n'est plus disponible alors ses annotations sont sauvegardées. Cette technique améliore l'approche de l'actuel signet [DEN 00] ;

– *rechercher les annotations* : l'utilisateur peut effectuer des recherches sur ses annotations en recherche « plein texte », ou à partir des attributs (date, titre du document, URL) ou bien encore par le texte surligné. Le résultat est une liste d'annotations regroupées dans un document où chaque document est représenté par la fusion des textes surlignés qu'il contient. Cette technique très intéressante, permet à l'utilisateur d'avoir une idée sur un document sans le consulter entièrement (en quelque sorte un résumé).

1.10.1.14. Ajouter des métadonnées

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

Construction d'un dictionnaire personnel

Le système *HyWebMap* permet à l'utilisateur de construire son propre dictionnaire et d'y intégrer si nécessaire des synonymes. Ce dictionnaire permet d'une part à l'utilisateur d'introduire une seule fois ses mots-clés avec leurs synonymes et d'autre part d'améliorer la recherche d'information.

Actuellement nous développons cette partie pour que l'utilisateur puisse construire un véritable *thesaurus* personnel. Dès qu'un nœud est re-visité, l'annotation attachée apparaît dans le navigateur.

1.10.1.15. 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 ou XML. Elle permet d'enregistrer et d'exporter le réseau virtuel (liens, annotations) au format HTML (figure 1.13) ou XML (figures 1.14 et 1.15) qui constitue une caractéristique très utile pour la consultation du réseau par un navigateur standard.

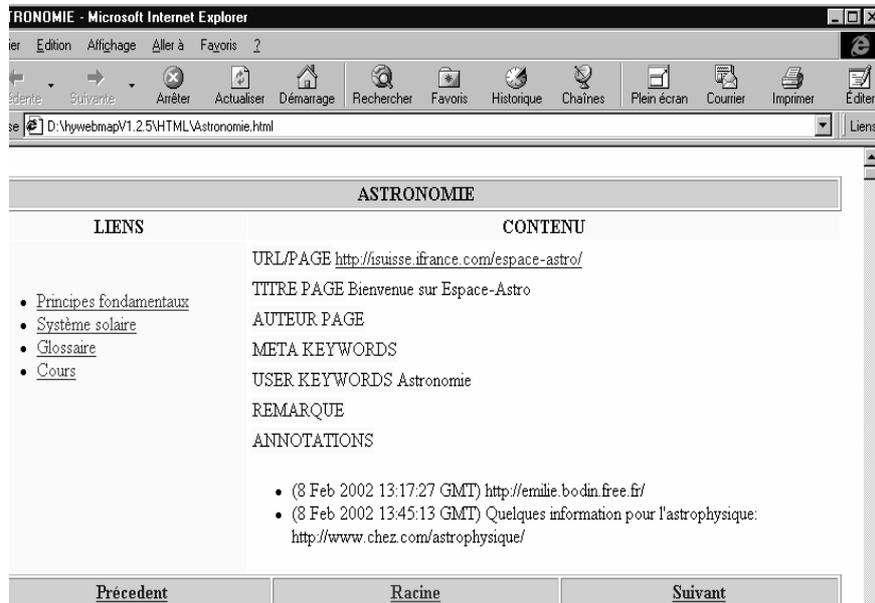


Figure 1.13. Génération du réseau virtuel en version HTML

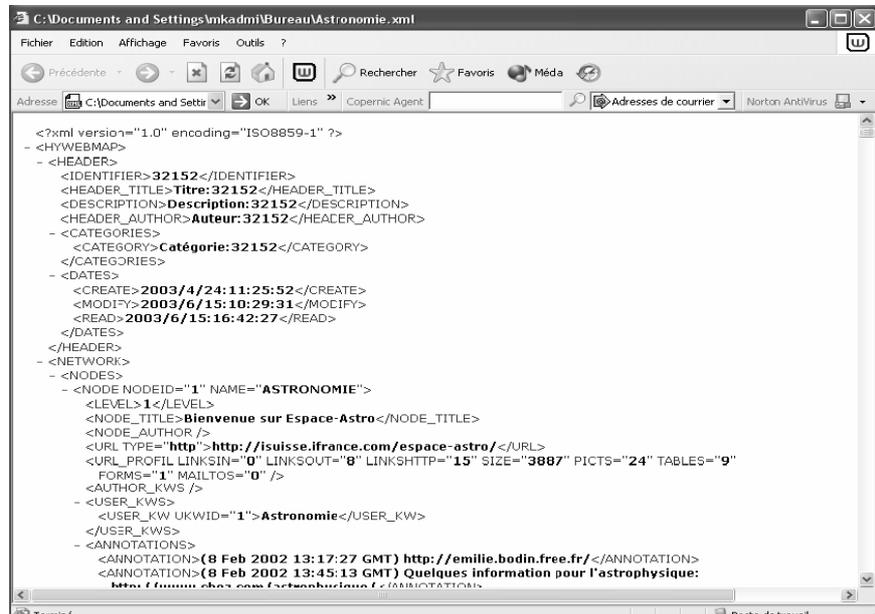


Figure 1.14. Génération du réseau virtuel en version XML

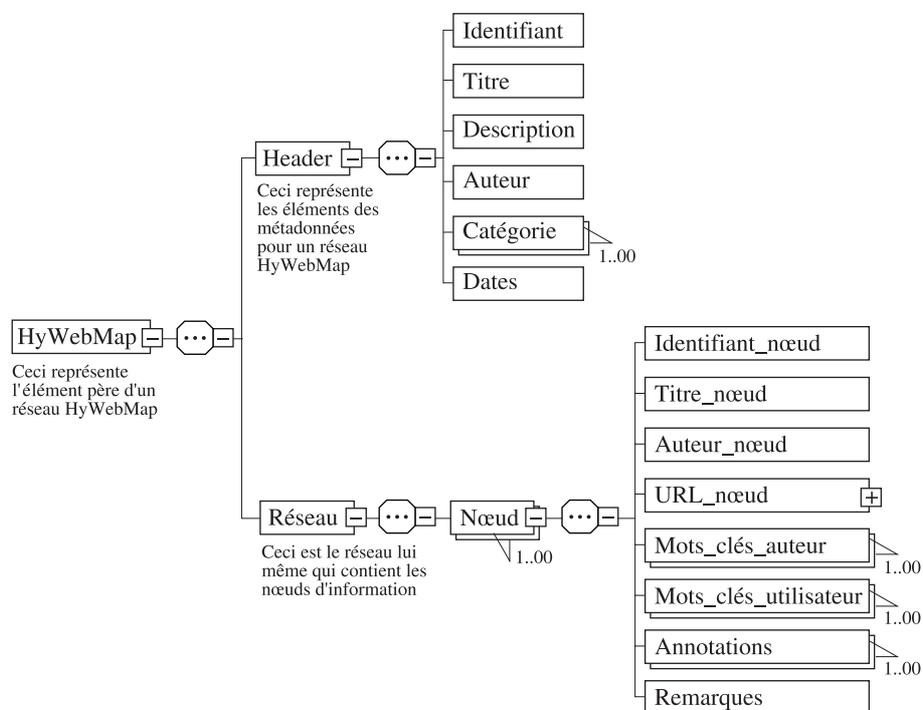


Figure 1.15. Schéma XML des réseaux HyWebMap

Ainsi, *HyWebMap* offre à l'utilisateur une vision globale de son site personnel grâce à la génération dynamique de celui-ci en sommaire.

1.11. Conclusion

Dans ce chapitre, nous nous sommes intéressés à la notion de l'usage de l'hypertexte. Nous avons insisté plus particulièrement sur la typologie des systèmes hypertextes, ainsi que celle des lectures et de lecteurs. En remontant à l'origine de l'hypertexte, celle d'une immense base de données consultée au gré d'un broutage ou d'une navigation par un lecteur, nous nous rendons compte que cette technologie ne cesse d'influencer les modèles de construction de connaissances, et par conséquent les modèles de systèmes de recherche d'informations et des systèmes hypermédias. Beaucoup de travaux théoriques ont été faits dans ce sens dans les années 1990, et depuis la littérature n'en parle pas beaucoup. Ce chapitre se veut une synthèse de rappel de ces différents travaux, mais aussi un travail introductif au sujet des hypermédias collaboratifs : chapitre 2, des hypermédias adaptatifs : chapitre 4.

1.12. Bibliographie

- [BAL 96] BALPE J.P., LELU A., PAPY F., SALEH I., *Techniques avancées pour l'hypertexte*, Paris, Hermès Science Publications, 1996.
- [BAL 97] BALPE J.P., « Hypertexte et interactivité », revue *Hypertextes et hypermédiâs*, Hermès Science Publications, p. 11-22, 1997.
- [BAR 89] BARRETT E., "Introduction: Thought and Language in a Virtual Environment", *The Society of Text: Hypertext, Hypermedia, and the Social Construction of Information*, Edward Barrett (Ed.), Massachusetts, MIT Press, 1989.
- [BER 92] BERNSTEIN M., JOYCE M., LEVINE D., "Contours of Constructive Hypertexts", *ECHT'92 Proceeding of the ACM Conference on Hypertext*, New York, Etats-Unis, 1992.
- [BER 99] BERNERS-LEE, T., *Weaving the web*. Londres, Orion Business Books, 1999.
- [BES 02] BESLILE, C., « Hypermédia et apprentissage. Séminaire DAFCO "L'apprentissage médiatisé : Navigation et hypermédia" », <http://www.ish-lyon.cnrs.fr/labo/LIRE/NavetHyper.htm>, 2002.
- [BOL 91] BOLTER J.D., *Writing Space: The Computer, Hypertext, and the History of Writing*, Hillsdale, New Jersey, Lawrence Erlbaum Associates, Inc., 1991.
- [BOU 02] BOUHAI N., Lire, réécrire et partager le savoir sur le Web : problèmes et solutions, Thèse de Doctorat en Sciences de l'Information et de la Communication, Université Paris 8, 2002.
- [BRO 88a] BROWN P.J., *Hypertext: the way forward*, Editions J.C. Van Vliet, Document manipulation and typography, Proc. of the *International Conference on Electronic Publishing*, Nice, p. 183-191, avril 1988.
- [BRO 88b] BROWN P.J., "Linking and searching within hypertext", *Electronic Publishing*, vol. 1, n° 1, p. 45-53, avril 1988.
- [BRO 89] BROWN P.J., "Do we need maps to navigate round hypertext documents?", *Electronic Publishing*, vol. 2, n° 2, p. 91-100, juillet 1989.
- [BRU 96] BRUSILOVSKY P., "Adaptative hypermedia: attempt and generalize", *Multimedia, Hypermedia and virtual reality*, Berlin, 1996.
- [BRU 99] BRUNIE V., Reconstruction documentaire pour la lecture des hypertexte : problèmes et méthodes, Thèse UTC, 1999.
- [BUS 45] BUSH V., "As we may think", *Atlantic Monthly*, 1945.
- [CAM 91] CAMACHO M., GAUDROY P., KESSACI A., « Hypertextes et hypermédiâs », *L'écho des recherches*, n°146, 4^e trimestre, p. 61-70, 1991.
- [CER 04] CERBAH F. « Exploitation de ressources lexicales pour la mise en hypertexte », *Conférence CIDE.7*, La Rochelle, p. 59-74, 2004.
- [CHE 92] CHENET A., *Eléments pour la conception d'un système multimédia*, Paris, ADBS Editions, 1992.

- [CON 87] CONKLIN, J., "Hypertext: an introduction and survey". *IEEE Computer*, 20 (9), p. 17-41, 1987.
- [CON 92] CONEY M.B., "Technical Readers and Their Rhetorical Roles", *IEEE Transactions on Professional Communication*, vol. 35, n° 2, p. 58-63, juin 1992.
- [DAC 90] DACHELET R., « Hypertexte et hypermédia », *Le document électronique : cours INRIA*, p. 135-161, juin 1990.
- [DEB 99] DE BRA, P., BRUSILOVSKY, P., HOUBEN, G., "Adaptive hypermedia: from systems to framework", *ACM Computing Surveys* 31(4), <http://www.acm.org/surveys/Formatting.html>, décembre 1999.
- [DEN 00] DENOUE L., VIGNOLLET L., « L'importance des annotations », *Document numérique*, vol. 4, n° 1-2, p. 37-57, 2000.
- [DOA 98] DOAN B.L. *et al.*, « Utiliser les méta-informations pour décrire les documents www : application aux échanges d'informations » *Proc. Of NTICF*, novembre 1998.
- [EDW 89] EDWARDS D.M., HARDMAN L., "Lost in hyperspace: cognitive mapping and navigation in a hypertext environment", *Hypertext: theory in to practise*, Norwood NJ : Ablex, p. 105-125, 1989.
- [ENG 84] ENGLEBART D., "Autoships provisions in Augment", *Proc. Of IEEE COMPCON*, Los Angeles, Etats-Unis, p. 51-58, 1984.
- [FER 00] FERNANDO A., BEIGBEDER M., « Des moteurs de recherche efficaces pour des systèmes hypertextes grâce au contexte des nœuds », *Conf. TICE'2000*, Lyon, p. 31-42, 2000.
- [GLU 89] GLUSHKO R.J., "Design Issues for Multi-Document Hypertexts", *Hypertext '89 Proceedings*, p. 51-60, novembre 1989.
- [HOR 89] HORN R.E., *Mapping Hypertext*, Lexington, MA, The Lexington Institute, 1989.
- [JAC 02] JACKIEWICZ A., « Repérage et délimitation des cadres organisationnels pour la segmentation automatique des textes », In *actes CIFT'02*, Hammamt, Tunisie, p. 95-105, 2002.
- [JAM 98] JAMET E., « Comment lisons-nous ? », *Dossier la lecture, Sciences Humaines*, n° 82, p. 20-25, avril 1998.
- [LAN 91] LANDOW G.P., DELANY P., "Hypertext, Hypermedia and Literary Studies: the State of the Art", *Hypermedia and Literary Studies*, George P. Landow and Paul Delany (Eds.), Massachusetts, MIT Press, 1991.
- [LAN 97] LANDOW P., *Hypertext 2.0*, Baltimore and London, The Johns Hopkins University Press, p. 185-186, 1997.
- [LEC 90] LE CROSNIER H., *Systèmes d'accès à des ressources documentaires : vers des antéserveurs intelligents*, Thèse de doctorat, Marseille : Université de droit, d'économie et des sciences d'Aix-Marseille, 1990.
- [LEC 91] LE CROSNIER H., « Introduction à l'hypertexte », *Bulletin des Bibliothèques de France*, tome 36, n° 4, 1991.

- [LEG 93] LEGENDRE R., *Dictionnaire actuel de l'éducation*, 2^e édition, Montréal : Guérin, Paris : ESKA, 1993.
- [LEL 98] LELU A., RHISSASSI H., « Projet Hypermap : pour un environnement complet de génération automatique d'hypertexte », *Actes de EP'98*, J. André ed., Saint-Malo, avril 1998.
- [LEL 99] LELU A., « Représentations cartographiques de corpus textuels : codage, algorithmes, ergonomie », HDR, Université Paris 8, 1999.
- [LEL 03] LELU A. "Inscription sociales des outils d'observation massive des mots et des liens : quelques pistes", *Conf. JFT'2003*, Tours, p. 81-90, 2003.
- [MAR 90] MARCHIONI G., XIA L., DWINGINGS S., "Effects of seeking in a hypertext environment", *Proceeding of the 53rd annual meeting of ASIS*, Toronto, 1990.
- [MAR 98] MARSHALL C.C., "Towards an ecology of hypertext annotation", Actes de 9th ACM Hypertext and Hypermedia Conference, Pittsburgh, Etats-Unis, 1998.
- [GOL 98] GOLDFARD C.F. *et al.*, *SGML buyer's guide*, Edition Prentice Hall PTR, Upper Saddle River, NJ, Etats-Unis, 1998.
- [MIC 93] MICHALAK S., CONEY M., "Hypertext and the author/reader dialogue", *Proceedings of the fifth ACM conference on Hypertext*, décembre 1993.
- [MOT 00] MOTTA E. *et al.*, "Ontology-Driven document enrichment: principles, tools and applications", in *International Journal of Human Computer Studies*, vol. 52, n° 5, p. 1 071-1 109, 2000.
- [NAN 94] NANARD M., « L'apport des travaux de recherche dans les hypertextes aux techniques éducatives », Actes du séminaire *Hypermédiat, éducation et formation*, E. Bruillard, B. de la Passardière, G.L. Baron (dir.), 1994.
- [NAN 95] NANARD J., NANARD M., "Hypertext design environments and the hypertext design process", *Communication of the ACM*, Vol. 38, N° 8, août 1995.
- [NEL 65] NELSON T.H., "à file structure for the complex, the changing and the indeterminate", *Proceedings of the 20th ACM National Conference*, New York, NY : ACM Press, 1965.
- [NEL 67] NELSON T.H., "Getting it out of our system", *Information retrieval: a critical review*, p. 191-210, 1967
- [NIE 90] NIELSON J., *Navigating through large information spaces, hypertext and Hypermedia*, Paris, Academic Press, 1990.
- [OUE 01] OUEFELLI T., La segmentation des documents techniques composites dans une perspective d'indexation : vers la définition d'un modèle dans une optique d'automatisation, Thèse, Université de Grenoble 3, 2001.
- [PAP 01] PAPY F., SALEH I., BOUHAÏ N., « Chercher et réorganiser l'information sur le Web », actes du 5^e colloque *Hypermédiat et Apprentissages*, Grenoble , p. 49-58, avril 2001.
- [RAD 91] RADA R., *Hypertext: from text to hypertext*, McGraw Hill Publishers, 1991.

- [SAL 01] SALEH I., « Web sémantique, Dynamique et Coopératif », HDR, Université de Paris 8, 23 novembre 2001.
- [SAL 03] SALEH I., PAPY F., « Navigation et recherche dans un système hypertexte », in *Méthodes avancées pour les systèmes de recherche d'information*, Madjid Ihadjadene (dir.), Hermès Science Publications, p. 19-33, 2003.
- [SAL 68] SALTON G., "*Automatic Information Organization and Retrieval*", Mac Graw Hill, New York, Etats-Unis, 1968.
- [SAL 93] SALTON G., ALLEN J., "Selective text utilization and text traversal, ACM Hypertext '93", *Proceedings of the fifth ACM conference on Hypertext '93*, Seattle, Etats-Unis, novembre 1993.
- [SIM 89] SIMPSON A., "Navigation in hypertext: design issue", *Proceedings of the 13th online information*, Londres, 1989.
- [SLA 91] SLATIN J., "Reading Hypertext: Order and Coherence in a New Medium", in *Hypermedia and Literary Studies*, George P. Landow and Paul Delany (Eds.), Massachusetts, MIT Press, 1991.
- [STI 93] STIEGLER B., « Les temps de la lecture », *Science de l'information et de la communication*, Larousse, 1993.
- [TAN 98] TANNENBAUM R.S., *Theoretical Foundations of Multimedia*, New York, Computer Science Press, 1998.
- [TEA 95] TEASDALE G., « L'Hypertexte : historique et applications en bibliothéconomie », *Cursus*, vol. 1, n° 1, octobre 1995.
- [WEI 99] WEISSBERG J.L., *Présences à distance, déplacement virtuel et réseaux numériques. Pourquoi nous ne croyons plus à la télévision*, l'Harmattan, 1999.

- [THE 02] THEVENIN D., COUTAZ J., « Adaptation des IHM : taxonomies et architecture logicielle », *14^e conférence francophone sur l'interaction homme-machine IHM'2002*, Poitiers, 2002.
- [VAN 03] VANDERDONCKT J., LIMBOURG Q., FLORINS M., “Deriving the Navigational Structure of a User Interface”, *Proceedings of conference Interact 2003*, 2003.
- [VER 00] VERNIER F., NIGAY L., “à Framework for the Combination and Characterization of Output Modalities”, *Proceedings of DSV-IS200*, Springer Verlag, p. 32-48, 2000.
- [ZAK 03] ZAKARIA M.R., MOORE A., STEWART C.D., BRAILSFORD T.J., “Pluggable user models for adaptive hypermedia in education”, *Proceedings of Hypertext'2003*, ACM Press, p. 170-171, 2003.