



HAL
open science

De l'hypertexte au document numérique personnalisable : accessibilité, usages, interfaces

Fabrice Papy

► **To cite this version:**

Fabrice Papy. De l'hypertexte au document numérique personnalisable : accessibilité, usages, interfaces. Sciences de l'information et de la communication. Université Paris 8, 2004. tel-02511668

HAL Id: tel-02511668

<https://hal.science/tel-02511668>

Submitted on 20 Mar 2020

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

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

Université Paris 8

Habilitation à Diriger les Recherches
en
Sciences de l'Information et de la Communication

De l'Hypertexte au Document Numérique Personnalisable.
Accessibilité. Usages. Interfaces

par

Fabrice PAPY

soutenue le 30 novembre 2004 devant un jury composé de :

Mme le Prof. Sylvie Leleu-Merviel, Université de Valenciennes (Rapporteur)

M. le Prof. Michel Mathien, Université Strasbourg III (Rapporteur)

M. le Prof. Gérard-Michel Cochard, Université de Picardie (Rapporteur)

M. le Prof. Jean-Pierre Balpe, Université Paris VIII (Examineur)

M. le Prof. Claude Baltz, Université Paris VIII (Président)

M. le Prof. Imad Saleh, Université Paris VIII (Directeur de recherches)

TABLE DES MATIERES

Introduction	1
Informations et Technologies : la position centrale de l'individu.....	4
Document électronique, Document structuré, Document numérique	7
Usage et Usagers de l'information à l'ère numérique	11
Hypertextualisation, Format de données, Méta-réseau et Navigation par la structure	33
La question de l'universalité des formats d'échange des documents électroniques hypertextualisés	36
De l'identification d'une structure physique à un étiquetage logique du contenu de documents électroniques linéaires.....	39
Méta-réseau, méta-description, actualisation et navigation.....	44
Conclusion	46
Pour l'élaboration d'Interfaces Homme / Connaissances	51
Accès aux données et profil utilisateur.....	56
NeuroWeb : vers une interface Homme-Connaissances	59
Conclusion	67
De HyWebMap à K-Web Organizer. Vers la généralisation du document numérique personnalisable. Le cas du Document Réticulaire Structuré	71
Espaces virtuels personnels, espaces virtuels partagés, espaces sémantiques	78
HyWebMap.....	79
K-Web Organizer.....	82
Un document numérique personnalisable : Le Document Réticulaire Structuré	84
Conclusion	94
Conclusion.....	99
Annexe	112
Bibliographie.....	125

CHAPITRE 1

INTRODUCTION

Depuis plus de quinze ans, le laboratoire Paragraphe¹ est reconnu pour sa démarche distincte et innovante relevant des notions complexes et souvent ambiguës de l'Hypertexte. Les approches théoriques et les approches pratiques qui s'enrichissent mutuellement donnent son identité au Laboratoire. Les éditions régulières de la conférence "Hypertexte / Hypermédia, Réalisations, Outils et Méthodes"² [H2ptm 03, 01, 99, 97, 95] témoignent de cet équilibre entre exigences théoriques et approches pratiques. Par-delà cette complémentarité, on constate que les travaux initiés au sein du Laboratoire accompagnent ou précèdent fréquemment les évolutions des TIC et leur impact sur notre société.

Ce mémoire d'habilitation décrit mes activités de recherche depuis ma thèse, plus particulièrement celles que j'effectue au sein du Laboratoire Paragraphe de l'Université Paris 8 et qui concernent :

- les dispositifs d'hypertextualisation automatique permettant la production automatisée de réseaux hypertextes à partir de documents électroniques textuels [Smolczewska 04] [Le Quellec 98] [Papy 95] [Papy 93].
- la recherche automatisée d'informations électroniques, qui implique la mise en œuvre de moteurs de recherches exploitant les technologies du Web dynamique, de l'indexation automatique et du calcul statistique de pertinence et de voisinage (termes et documents) [Papy 03a] [Lelu 01] [Lelu 99] [Balpe 96] [Lelu 93].
- la structuration normative des documents numériques issus des technologies du Web et largement représentés par les dialectes XML (XSLT, XSL-FO,

¹ <http://paragraphe.univ-paris8.fr>

² "Hypertext, Hypermedia, Products, Tools, Methods" : <http://h2ptm.univ-paris8.fr/>

CSS, SVG, SMIL, DOM,...) [Papy 04b] [Bui Thi 03] [Michard 02] [Roisin 99] [Hudrisier 99] [XML 98],

- les usages individuels et collectifs qui découlent de ce déploiement massif de documents numériques en matière d'organisation et de coopération numérique (mutualisation de contenus) [Papy 04d] [Papy 03c] [Papy 02] [Pincemin 99],
- les extensions vers le travail collaboratif et l'organisation numérique que les Technologies de l'Information et de la Communication appliquées à l'Enseignement (TICE) concourent à déployer aussi bien dans les environnements numériques de travail que dans les plates-formes d'enseignement à distance [Papy 03c] [Papy 04] [Papy 04] [Campos 03] [Samain 01] [Harvey 95].
- l'utilisation des principes conjugués de l'hypertexte, des interfaces et des documents numériques structurés dans des environnements fortement structurés du point de vue de la connaissance proposés par les bibliothèques universitaires (bases de données, thésaurus, normes, classification,...) au moyen du catalogue en ligne [Papy 04c] [Roy 02] [Christoffel 02] [Harvey 01] [Ihadjadene 01] [Lecolinet 01] [Yee 99] [Cubaud 98] [Maignien 95].

Il est important de remarquer que ce mémoire d'habilitation place mes différents travaux dans le cadre plus général de l'interaction des technologies de l'information avec les pratiques sociales et les usages [Dubey 01] [Jeanneret 00] [Wolton 00] [Baltz 98] [Le Coadic 97] [Castells 96]. Ce sont principalement les usages induits par les documents électroniques [Roisin 99] [Nanard 98] [Quint 92] [Furuta 89] revisités par la vague numérique elle-même impulsée par les technologies protéiformes issues du Web qui donnent le fil directeur de mes travaux engagés depuis dix ans. Car le document numérique et les technologies *ad hoc* ont indéniablement exacerbé la relation duale "Homme / Information" [Baltz 03] [Bernat 98] de sorte que les questions d'accessibilité aux ressources numériques, de désorientation, de surcharge cognitive, d'organisation, de partage et de coopération numériques, largement évoquées avec l'Hypertexte [Kobayashi 00] [Belisle 99] [Nielsen 90] [Nielsen 89] [Brown 89a] se trouvent à nouveau convoquées et de façon plus insistante dans une

reconfiguration de notre société en société de l'Information qui attribue une part de taille à la dimension technologique de l'information [Projet 01] [Mayère 97] [Castells 96] [Miléo 96] [Richard 96].

Mes travaux de recherches ont largement subi l'influence des techniques d'information, de l'informatique et de l'Internet qui, depuis une quinzaine d'années, bousculent considérablement le champ des Sciences de l'Information et de la Communication. La théorie se construit fréquemment après et à partir des expérimentations. Ce constat s'applique à mes travaux qui portent d'une part sur le positionnement central de l'individu par rapport aux Technologies et à l'information numérique et d'autre part sur le document numérique et les technologies afférentes.

Mes travaux concernent tant le contenu de l'information que sa distribution, ouvrant aujourd'hui des perspectives théoriques conséquentes aussi bien en matière de circulation, d'échange et de partage des connaissances, qu'en ce qui concerne leurs conditions d'appropriation et d'adaptation des contenus numériques.

Mes contributions techniques s'insèrent dans une démarche volontaire de ne pas laisser les Sciences de l'Information et de la Communication dans une attitude essentiellement critique à l'égard des technologies, et à l'extérieur des dispositifs dont les usages relèvent précisément des SIC (systèmes de production et d'accès à l'information, étude des acteurs individuels et collectifs,...) [Fondin 02] [Fondin 01] [Bougnoux 01] [Jeanneret 00] [Wolton 00] [Breton 95]. Cette position active dans le prototypage de dispositifs techniques (et que je manifeste encore dans la plus récente réalisation du dispositif de métaphore graphique des résultats de recherche d'un OPAC universitaire, cf. chapitre 5), n'a eu pour motivation que de donner forme aux exigences des SIC en matière d'usages, exigences largement maltraitées par les concepteurs d'artefacts TIC³.

³ Sur ce point, on peut remarquer que ce sont les psychologues-ergonomes [Rabardel 98] qui mesurent effectivement le décalage qui existent entre d'une part les idées des concepteurs et les dispositifs qu'ils réalisent et d'autre part les usages réels que l'on peut observer chez les "utilisateurs". Ainsi le travail de recherche de [Cerratto 03] a clairement montré que le logiciel d'écriture collaborative qu'elle avait étudié permettait de tout faire ...sauf de l'écriture collaborative : "The results obtained in the study showed important failures in the computer-mediated collaboration of remote groups, but it is important to point out that the specific problems that were found in the activities of the groups might be due to the specific collaborative writing program used in the study rather than being intrinsic problems of the computer mediated collaboration."

Informations et Technologies : la position centrale de l'individu

L'imposant et prolifique appareillage technologique qui cherche à rendre accessible et exploitable **automatiquement** cette information numérique toujours plus pléthorique, montre paradoxalement ses limites, précisément par cette tendance à la production surnuméraire de "solutions" techniques. L'invention de nouveaux usages, les détournements de la technologie que l'on peut observer (très récemment les Weblogs, les journaux intimes et les wikis en sont d'excellentes illustrations) rendent très rapidement obsolètes les solutions techniques innovantes élaborées pour répondre à des usages inédits qui mettent en échec les technologies précédentes ; la demande impérieuse d'informations actualisées a amené le Web dynamique [Saleh 99] à supplanter le "Web statique", les méta-moteurs puis les portails ont cherché à répondre aux exigences thématiques et spécialisées que les moteurs de recherches généralistes et les annuaires résolvaient avec difficultés, la *push-technology* a tenté d'apporter une solution à l'actualisation effrénée des sites d'information (conséquence du Web dynamique), le Peer-to-Peer (Kazaa, Gnutella, Groove, ...), les agents intelligents (Cybion, Copernic, HyWebMap,...), les systèmes cartographiques (Kartoo, MapStan, CartoWeb,...) les dispositifs de collaboration et plus récemment les interfaces adaptatives peinent à raccrocher les wagons de l'ingéniosité avec laquelle les usagers du Web inventent leurs propres usages. Jusqu'à l'hypothétique moment où les machines seront en mesure de le faire, l'être humain demeure seul capable de donner sens à l'information [H2ptm 03] et d'opérer la synthèse cognitive de ce qui a été collecté [Cerratto 03] [Saleh 01] [Saleh 01b] [Rabardel 95].

Ainsi, le contexte socio-cognitif de l'utilisateur fait aujourd'hui l'objet d'une attention particulière avec le déploiement des Technologies de la Communication et de l'Information appliquées à l'Enseignement (TICE) et invite à une indispensable réflexion sur les nouvelles médiations à envisager pour faire passer les connaissances produites pour un cadre donné et objectivé à cette fin, vers les contextes plus subjectifs des utilisateurs [Arnaud 02] [Senteni 01] [Belisle 99] [Tazi 99] [Rabardel 95]. Des travaux de recherche anciens, antérieurs à l'industrialisation de la production d'information électronique apparue avec l'avènement du Web, avaient

déjà mis l'accent sur le contexte cognitif du lecteur/utilisateur, mais en l'appréhendant sous l'angle étroit des inconvénients majeurs habituellement constatés dans la consultation des fonds documentaires hypertextuels : la désorientation et la surcharge cognitive [Saleh 04] [Balpe 96] [Nielsen 90] [Brown 89a]. D'une certaine façon, c'est tout l'intérêt de l'accessibilité de l'information proposée par l'Hypertexte qui était remise en question par ces inconvénients. Il est difficile de nier qu'aucune véritable solution n'a été trouvée pour pallier ces problèmes épineux. Certes des solutions ponctuelles à base de cartes, d'historique de navigation, de parcours de lectures pré-enregistrés, etc. ont été avancées mais sans apporter de réponses satisfaisantes à ces invariants problématiques de la lecture hypertextuelle. On peut avancer aujourd'hui que c'est l'inexistence de réseaux hypertextes de grande taille qui ont permis de relativiser l'importance de ces inconvénients. En effet les quelques réalisations de réseaux hypertextes "de production" ont été conçues par des publics de spécialistes et d'experts à l'intention de publics de même qualité, sensibilisés à la nature des informations disponibles. La connaissance de la thématique du réseau, des informations qu'ils étaient susceptibles d'y trouver, la capacité à analyser et à définir leurs propres besoins correspondent à une activité d'exploration (*browsing*) et de recherche active qui a maintenu à distance ces questions embarrassantes de la désorientation et de la surcharge cognitive.

Or, le déploiement électronique du World Wide Web, l'explosion documentaire, la multiplicité des profils utilisateurs et la reconfiguration progressive de nos sociétés en sociétés de l'Information ont exercé, en les complexifiant, un rappel immédiat de ces problèmes irrésolus.

L'Internet des premières années, émanation des travaux de recherche dans les domaines de l'informatique et des télécommunications a rapidement muté dans les esprits et dans les usages en un système mondialisé de production, d'échange et de publication d'informations qui s'immisce toujours plus profondément dans les sphères d'activité sociale de l'Homme. Les loisirs, l'économie, l'administration, l'enseignement, etc. se trouvent impliqués par la déferlante des Technologies de l'Information et de la Communication et exigent d'un nombre grandissant d'individus des capacités inédites de recherche, de collecte et de synthèse d'informations. La

présence de l'Internet est passée d'un état "diffus"⁴ à un état "contraint" où l'alternative n'existe plus vraiment, où les modalités fonctionnelles induites par le principe fédérateur de l'Internet ne sont plus contournables. La mutation vers ce mode contraint est d'ores et déjà amorcée, elle est observable de façon symptomatique avec le vaste projet Copernic du Ministère de l'Economie et des Finances qui réalise une dématérialisation progressive de l'administration⁵. Ce projet prévu pour se déployer sur 7 ans préfigure non seulement l'e-administration de demain mais assurément l'e-gouvernement ou gouvernement en ligne que les Canadiens expérimentent [Hudon 03] et qui est mis en place progressivement dans l'Union Européenne. Une grande partie du projet Copernic a montré son efficacité fonctionnelle en 2002 avec la télé-déclaration de l'impôt sur le revenu qui a concerné plus de 500 000 contribuables, devenus pour l'occasion, électroniques⁶.

Or, les Technologies de l'Information et de la Communication, seules, ont chaque jour, de plus en plus de mal à répondre aux besoins de recherche, de représentation et d'organisation des informations et connaissances dont les volumes mis à disposition ne cessent de s'accroître sur l'ensemble des réseaux. Face à cette quadrature du cercle, la problématique majeure de la société de l'Information porte sur la capacité d'autonomie des individus dans un environnement informationnel [Baltz 03] [Acquier 02] [Chaudiron 02] [Dufour 01] [Fayet 99] pour lequel on entrevoit mal une organisation rationnelle permettant à tout un chacun – et c'est bien là, la gageure - qu'elles que soient ses demandes, sa formation, ses acquis, ... de s'y mouvoir avec aisance, acuité et efficacité⁷. Devant l'impasse technologique, c'est à l'individu de se

⁴ De nombreux services sont depuis de nombreuses années accessibles à partir des sites Web : se tenir informé des nouvelles nationales et internationales, consulter son compte bancaire, suivre les remboursements de sa mutuelle, de son dossier d'assurance, connaître l'état du trafic en temps réel dans les transports en commun, obtenir le calcul d'itinéraire en fonction d'un mode de transport lors d'un voyage annoncé, etc. sont autant d'interactions que nous qualifions de diffuses dans le sens où elles existent sans introduire d'impacts sociaux-économiques significatifs. Ces espaces électroniques interactifs dont on ne peut nier l'utilité sont souvent à la recherche d'un modèle économique qui à défaut d'être trouvé conduit à la disparition du service proposé (par exemple : www.pariscope.fr).

⁵ <http://www.adae.gouv.fr/> Agence pour le Développement de l'Administration Electronique (circulaire du 12 septembre 2003 relative au développement de l'administration électronique, parue au J.O n° 216 du 18 septembre 2003 page 16006)

⁶ Le succès de l'opération très largement repris par les organes d'information nationaux a masqué la manifeste "inégalité numérique" qui s'est instaurée. En effet, les télé-déclarants ont bénéficié d'un délai supplémentaire pour remplir électroniquement leurs déclarations de revenus, délai dont ne bénéficiaient pas les contribuables utilisant les moyens traditionnels.

⁷ "...Les Quinze estiment que les entreprises et les citoyens doivent avoir accès à une infrastructure de communication peu coûteuse que chaque citoyen doit être doté des compétences nécessaires pour

doter de moyens lui permettant de se confronter à la complexité naissante de la société de l'Information [Bretelle 99] [Coulon 99] [Coulon 97]. De cette confrontation, il doit pouvoir tirer une ligne de conduite, c'est-à-dire la capacité d'auto-définir des stratégies - toujours mouvantes - d'acquisition. Ses connaissances de base devront être la capacité à définir des objectifs, dominer des stratégies d'acquisition de connaissances provisoires, à reconstruire en permanence une pensée au moins autant analogique qu'analytique et ceci dans un regard toujours critique [Balpe 97] [Bouhai 02] [Souchier 03].

Document électronique, Document structuré, Document numérique

La question du document numérique peu présente à l'époque antérieure au World Wide Web (historiquement et informatiquement liée à celle des "Documents Structurés" [Quint 92] [Newcomb 91] [Furuta 89]) se pose avec vigueur depuis que la numérisation a investi tous les médias et que les Technologies du Web ont imposé des formats moins propriétaires et adaptés aux capacités de transmission des réseaux de télécommunications. De multiples formats ont été conçus pour coder texte, image (fixe et animée), son et vidéo en permettant une étonnante interopérabilité applicative.

Ainsi, les nouveaux modes de production et de mise à disposition de l'information numérique remettent fondamentalement en cause la notion même de document [Stockinger 01] [Ballay 99]. Un document numérique n'est plus un objet stable s'exprimant dans une entité physique (le fichier) aux formes connues, et déterminées par les processus d'édition, comme a pu l'être un document traditionnel papier.

Les évolutions en terme de formats de document (XML, Xlink, Xpointer, Xpath, CSS, XSL, RDF, WOL, SVG, SMIL, MPEG, etc.) portent toujours plus loin les contenus électroniques vectorisés par les documents numériques mais également leurs conditions d'acquisition, de diffusion et de consultation. De multimédia au multimodal, le document numérique prend donc de multiples formes en des formats

vivre et travailler dans cette nouvelle société de l'information. Ils veulent notamment que toutes les écoles de l'Union disposent d'un accès et de ressources multimédias d'ici la fin 2001, qu'un nombre suffisant d'enseignants soient à même de les utiliser d'ici la fin 2002. Un accès généralisé par voie électronique devra être assuré à tous les services publics de base d'ici 2003. " (Le Monde, 26-27 mars 2000)

divers mais toujours ouverts. Ils permettent de combiner à l'infini les contenus au point où la notion même de document disparaît devant celle de contenus riches [Papy 03d] [Gandon 02] [Bourda 01] et de gestion de contenus. Le document numérique doté de ces orthèses langagières peut désormais assumer les multiples fonctions textuelles, métatextuelles, paratextuelles et hypertextuelles [Stockinger 01].

La capacité désormais banalisée de pouvoir créer, modifier, assembler et partager des séquences numériques relevant de médias distincts a bousculé les seules problématiques techniques pour investir celles des usages [Leleu-Merviel 04] [Papy 03d] [Chartron 02] [Salaun 00] [Ballay 99]. Dans la majorité des cybercommunautés, le succès, la productivité ou la cohésion sont souvent liés à l'intégration très libre d'outils de communication dans des environnements souples qui sont en mesure de satisfaire les exigences de communication de la communauté virtuelle. Les groupware et autres intranets [Le Bœuf 01] [Courbon 97] [Levan 95] véritable catalyseurs informatiques des premières communautés virtuelles viennent à perdre leur hégémonie et s'effacent devant le document numérique, objet complexe, polymorphe et adaptable qui devient la clef de la cohésion et le point de focale de toute collaboration. Il devient un objet engendré en partie par des auteurs et parfois reconstruit par celui qui cherche à l'exploiter. Cependant, une reconstruction performante, au sens où elle serait correctement adaptée à l'usage que son utilisateur (auteur, lecteur, auteur/lecteur, lecteur/auteur) prévoit d'en faire, ne pourra avoir lieu que si les documents introduits par leurs producteurs dans un système d'information possèdent des caractéristiques qui leur permettent de se prêter à cette reconstruction [Roisin 99].

Le "document numérique" tend à prendre la dimension conceptuelle, technique et sociale dont était chargé l'Hypertexte des précurseurs. Aujourd'hui, il ouvre des perspectives de reconfiguration dynamique, d'utilisation personnalisée, de collaboration, d'organisation, etc. qui dépassent largement le cadre informatique *stricto sensu* où il était originellement maintenu [Gaudin 00] [Furuta 89]. La virtualisation du document [Garlatti 99], l'adaptabilité des contenus [Papy 03c] [Gagneux 02] aux dispositifs de restitution et aux profils des utilisateurs / lecteurs et les modes d'organisation / coopération qu'il autorise sont autant de facteurs qui font du document numérique un objet d'étude légitime des Sciences de l'Information et de la Communication.

En effet, la nature des informations retrouvées est une chose, sa réalisation sous la forme de documents exploitables en est une autre, quant à sa transformation en connaissances, le processus est loin d'être évident. Un constat s'impose ; la prolifération des dispositifs (modèles, méthodes, systèmes, outils, etc.) destinés à la recherche d'informations électroniques nous donne un aperçu de l'étendue des problématiques sous-jacentes. On peut considérer que les étendues conceptuelles relevant distinctement du document (production, archivage, nature, statut,...) et de la connaissance (construction, médiation, organisation, partage, gestion, transmission,...) sont autant voire peut-être même plus conséquentes encore.

L'explosion de l'Internet ces dernières années, de cette explosion industrielle qui a laissée loin derrière les considérations universitaires du "Réseau des réseaux", a précipité des réflexions plus info-communicationnelles concernant les approches inévitablement liées au document numérique ; formats d'échange, accessibilité (non plus technique mais cognitive), indexation des documents complexes, distribution des traitements, archivage, fouille de textes, transmission des connaissances, édition, veille (stratégique, technologique, économique, concurrentielle,), etc.

Documenter le savoir-faire, faire circuler l'information, partager et transmettre des connaissances sont autant d'activités quotidiennes qui génèrent autant (de nouveaux types) de documents et qui relèvent bien de contextes socio-culturels distincts.

Le statut du document numérique est effectivement au centre des préoccupations informationnelles scientifiques. Les ouvrages, revues, colloques et conférences qui

lui sont dédiés, à ses multiples avatars et métamorphoses auxquelles les normes donnent le "fond" et la "forme" attestent qu'il existe désormais pour un contenu autant de formes qu'il y a de profils de lecteurs, de conditions cognitives et de circonstances d'utilisation [Isko 01]. Assurément une certaine confusion règne aujourd'hui autour du document numérique, comme elle a régné naguère autour de l'Hypertexte [Balpe 96] [Nielsen 89] [Giffard 97] [Verroust 97].

On ne peut ignorer, dès lors que le document numérique n'est plus directement lié à des problématiques de stockage ou d'archivage (GED, GEIDE), qu'il participe véritablement à des procédures dynamiques de production, de transmission, d'organisation et de partage de connaissances qui ne sauraient être séparées des contextes culturels, sociaux, cognitifs, intellectuels et organisationnels qui font partie intégrante des individus concernés par les objets numériques mis à leur disposition. Impossible donc d'éliminer ce que Pierre Bourdieu nommait l'habitus, ces dispositions sociales intériorisées par l'individu au cours de sa vie, de ses apprentissages, de ces inclinations, souvent inconscientes, à percevoir, faire et penser d'une certaine manière issues de nos relations sociales, impossible d'ignorer le triangle sémiotique de Peirce qui, schématisant, a mis en évidence l'importance de l'interprétation de la connaissance (objet) par rapport à la place de l'interprétant et au signe qui le représente. Le cycle sémiotique démontre que toute représentation ne peut exister sans le partage d'un espace commun entre émetteur et interprétation. Les expériences ont montré qu'il existe presque autant de logiques d'accès à l'information que de contenus publiés.

Aujourd'hui le document numérique et l'imposant appareillage technologique qui l'accompagne rend possible la déclaration de méta-données de toutes sortes, qui confère au document numérique une dimension agrégative inédite. Des situations d'usage telles que le travail collaboratif, l'enseignement à distance et la cartographie d'informations illustrent l'apport de ces technologies à l'élaboration de véritables documents hypermédias adaptables, ressources numériques agrégatives s'il en est.

Usage et Usagers de l'information à l'ère numérique

Mes travaux de recherche, amorcés avec ma soutenance de thèse en février 1995, ont confirmé la pertinence d'une démarche associative dans l'accès à l'information n'extrapolant pas le profil social, culturel et cognitif du sujet, de l'utilisateur (terme infocommunicationnellement plus correct que celui d'utilisateur). L'Hypertexte considéré comme technique, méthode, concept ou genre littéraire a bousculé une approche essentiellement mathématique et technique du traitement de l'information [Parrochia 93]⁸. La science informatique s'est imposée comme le "bras armé" des systèmes d'information (pris ici dans un sens très large cf. [Miege 03]) en proposant des heuristiques toujours plus sophistiquées, instrumentalisant l'acteur humain au point de le transformer en "utilisateur" [Chaudiron 02] [Rabardel 98]. Enfermé dans les modèles établis à partir des pré-supposés des concepteurs [Pavé 89][Creis 84], d'un imaginaire technique [Flichy 02] qui fait abstraction de sa dimension sociale, cognitive et culturelle [Habernas 73], le sujet n'a d'autres choix que se fondre dans les fonctionnalités des dispositifs techniques idéalement conçus pour un modèle, un *pattern* (pour reprendre l'expression informatique consacrée) dont rien ne permet d'apprécier la qualité d'appariement. L'Hypertexte en replaçant l'individu (tantôt lecteur, tantôt auteur, cumulant parfois les deux profils [Weissberg 99] [Giffard 97] [Verroust 97] [Balpe 96] [Clement 95] [Brown 88] [Conklin 87]) au centre du dispositif, en rétablissant la prépondérance du facteur humain, a imposé une approche assurément moins artificielle, certainement moins calculée du traitement de l'information et *ipso facto* naturellement anthropomorphique.

Plus encore l'Hypertexte invite à une approche qui met en exergue l'exploration et la navigation au sein d'espaces de données, réduisant du même coup le traitement informatique à une portion congrue : le calcul élémentaire d'une trajectoire à parcourir vers un objectif connu et informatiquement référencé. C'est au lecteur qu'échoient les stratégies de lecture, c'est à lui que revient de droit la responsabilité des liens qu'il active, au gré des associations que sa lecture le conduit à opérer,

⁸ "Le but de la théorie de Shannon qui se présente explicitement comme une théorie mathématique de la communication est de calculer à quelles conditions une information maximum peut être transmise d'une source S à un destinataire D au moyen d'un certain canal, compte tenu des contraintes qui affectent aussi bien le codage/décodage de cette information que sa transmission proprement dit (bruit)".

déterminées par ses (mé)connaissances initiales, ses carences, ses faiblesses, ses certitudes, son contexte social et/ou professionnel... bref autant de paramètres, complexes, imprévisibles que les systèmes de recherche d'information évacuent au nom de l'efficacité des heuristiques implémentées.

La connaissance et la maîtrise technique des Technologies de l'Information de la Communication qui façonnent notre société vouée à l'Information (Langages de développement pour le Web, bases de données, architecture client/serveur,...) me permettent de donner forme à ces dispositifs qui préservent non seulement la position centrale de l'individu, du sujet, de l'utilisateur mais l'encouragent également à l'assumer. Les expériences que j'ai pu mener auprès de différentes communautés relevant de différents contextes sociaux, professionnels et culturels (professionnels de l'information, industriels, grand public, étudiants,...) m'ont amené à élaborer des systèmes de recherche et des modalités d'organisation d'informations qui ne nient pas ces environnements distincts.

Au contraire, les propositions que j'émetts aujourd'hui s'orientent vers l'adaptabilité des dispositifs de recherche (et de collecte) d'informations numériques d'une part aux usages de plus en plus inédits (pratiques humaines et sociales de l'information) et d'autre part aux usagers de plus en plus pluri-communautaires (communautés scientifique, professionnelle, culturelle, nationale, internationale, entreprise, syndicat, université,...) au sein de communautés fréquemment en recouvrement et en recoupement (journaliste, enseignant, étudiant, salarié, spécialiste d'un domaine, professionnels de l'information, ...). Les Technologies de l'Information et de la Communication dans ces différents lieux qu'elles ont investis et virtualisés (entreprise, administration, université,...), ont certainement eu un effet amplificateur dans l'innovation des usages et l'ambiguïté des statuts des usagers [Le Coadic 97]. Epiphénomènes ? Etats provisoires d'un processus technologique encore en devenir ? Ces statuts émergents confirment et rendent visible la pluralité et la diversité des pratiques sociales, culturelles et cognitives de l'information qui ne peuvent plus être assujetties et de facto réduites à un quelconque "diktat technique".

A. Moles a montré dans [Moles 72] que la théorie de l'information de Shannon et Weaver porte essentiellement sur une information exprimée comme une quantité

mesurable et qui caractérise selon ses auteurs, le processus de communication . Dans son ouvrage, A. Moles a mis en évidence que la communication transporte avant tout de la complexité et qu'elle relève principalement de deux notions : la signification et la valeur de l'information [Milon 99]. Ces notions de signification et de valeur trop fluctuantes et trop anthropomorphiques pour être directement utilisables, ont été minimisées par le détour de la technique afin de rendre ces questions solubles.

Dès que l'on s'intéresse au savoir et à l'information, on ne peut ignorer l'humain et le social [Jeanneret 00] [Mayère 97]. Quelles sont les attentes en matière d'accès au savoir ? notait Jean-Louis Metzger [Metzger 04] en rappelant que ces attentes augmentent avec l'augmentation du volume de savoir disponible. L'avènement de l'Internet et du World Wide Web a engagé une démocratisation mondialisée (du moins dans les pays qui l'autorise !) de la production, de la diffusion et de l'accès à l'information numérique.

Or la question de la recherche d'information et de l'évaluation des systèmes *ad hoc* est principalement appréhendée sous l'angle techniciste de l'amélioration des performances (TREC, Amaryllis) avec une évidente amnésie en ce qui concerne l'adéquation des systèmes avec les pratiques sociales. Les moteurs de recherche sur le Web n'échappent pas à cette règle puisque les paradigmes techniques des systèmes de recherche documentaire traditionnels se retrouvent dans ces incontournables de la recherche d'informations sur le Web [Favier 02]. Cependant les usagers et leurs profils sont autrement plus larges que celui des spécialistes et professionnels de la recherche d'informations. Les concepteurs doivent innover pour s'adapter à ce contexte particulier qui a un impact important sur le fonctionnement des outils et leur véritable performance [Ferret 01].

La nature heuristique des approches qui sous-entendent de tels systèmes est inadéquate compte tenu nouveaux environnements éducatifs, d'enseignement, d'apprentissage, de documentation dédiés à l'amélioration des connaissances humaines. Il devient impérieux de développer de nouveaux modèles de création du sens dans la gestion de l'information et du savoir pour faire face aux nouveaux usages et aux nouveaux comportements des personnes et des communautés d'apprentissage dans le contexte de la convergence [Harvey 01]. La socialisation

(apprendre n'est pas qu'une activité individuelle, c'est aussi une activité d'échanges et d'interaction), l'externalisation (rendre les connaissances tacites, explicites et conscientes), la combinaison (intégrer diverses sortes de connaissances pour en produire de nouvelles), et l'internalisation (convertir les savoirs explicites en savoirs tacites) sont des phénomènes qui ne peuvent plus être abstraits des dispositifs techniques.

De ce point de vue, le matériel documentaire n'est pas une information et l'information n'est pas une donnée pré-construite qu'il s'agirait d'offrir ou de rendre accessible. Elle est plutôt un processus par lequel une entité de nature humaine et sociale est transformée (et c'est cette transformation que l'on peut demander aux TIC d'accompagner avec les facilités techniques d'accessibilité à l'information qu'elle propose dès lors que les projections qui sont envisagées ne vont pas dans le sens des heuristiques techniques qui sont élaborés aujourd'hui). Car il ne faut pas se leurrer, la connaissance des individus est loin d'être connue des sciences humaines et sociales, on peut s'interroger sur la capacité des dispositifs techniques - et de leurs pré-supposés de conception - à pouvoir répondre aux attentes des individus tournées vers cette soif de connaître [Dinet 02]. Cependant, la perspective de recherche ne peut être que constructiviste et non cognitiviste car l'activité est construite par les acteurs en situation compte tenu de leur état et de leur culture et en fonction de leur activité antérieure dans cette situation ou dans d'autres y compris hors travail [Favier 02]. A cela vient s'ajouter le fait que l'activité de recherche d'information est motivée par un besoin initial qui s'inscrit dans un contexte social où l'information est recherchée pour une utilisation précise. Ce contexte influence la pertinence de l'information retrouvée [Chevallet 02].

Cette question de la pertinence dépasse très largement les propriétés objectives d'attribution de la valeur intrinsèque des documents concernés. La pertinence, qualité attribuée subjectivement par l'utilisateur relève davantage de l'adéquation ternaire entre nature de l'information, individu et contexte. Replacée dans le cadre de la formation, cette adéquation s'exprime sous la forme de connaissances de référence destinées à un public d'apprenants en situation d'acquisition et de construction de connaissances. Ces connaissances référentes se trouveront être différemment décrites en fonction du "profil" des apprenants et des conditions grâce auxquelles s'opère cette construction.

Les volcans, les éruptions volcaniques et les coulées de lave par exemple, sont des phénomènes étudiés aussi bien par les élèves de collège, les étudiants en 1er cycle universitaire "Sciences de la Terre" et les chercheurs en géodésie et en géophysique. Confrontée à ces profils concernés par des approches distinctes des mêmes phénomènes, il est évident qu'une même information bien que scientifiquement fiable ne pourra être accueillie avec le même rapport de qualité et de pertinence par les acteurs de ces profils différents.

En d'autres termes, l'information est pertinente d'une part si elle est utile à l'utilisateur pour atteindre le but initialement fixé en apportant par exemple de nouvelles informations et d'autre part si son traitement ne nécessite pas un effort trop important [Rouet 01] [Harvey 01].

L'Internet est devenu en quelques années un lieu de publication d'informations stratégiques, fiables, pertinentes et pérennes [Dufour 01] [Ben Hamadou 01] [Dkaki 99] [Timimi 99] où les pratiques de recherche d'informations ne sont pas toujours tournées vers la recherche de «LA» réponse [Ferret 01], mais plus fréquemment vers la constitution de références documentaires convergentes que les dispositifs de veille ou les portails ont très largement mis en évidence. Les systèmes d'aspirations de site aujourd'hui intégrés dans les navigateurs, les gestionnaires de signets bien que fonctionnellement limités ont depuis longtemps offert aux internautes le loisir de se constituer de véritables sites thématiques au sein d'espaces locaux d'informations ré-organisées, véritables copies off-line des sites en ligne.

A la croisée des chemins

L'intérêt que je porte depuis 2002 au *e-learning* et aux plates-formes d'enseignement à distance [Papy 04a] [Papy 03c] [Papy 03b] n'est pas directement lié à des problématiques éducatives ou au développement d'environnements propices à l'EIAH mais initialement alimenté par une occasion de convoquer les différentes approches que j'ai énoncé précédemment : technologique (Technologies informatiques du Web qui intègrent le document numérique), sociale (pratiques personnelles et collectives des usagers), informationnelle (données, informations recherchées, collectées, agrégées, etc., pertinence, interprétation, ...) et communicationnelle (outils de communications, nature des communications, communautés, collaboration inter-

individuelles, collective, communication contrainte, libre,...). Plus qu'un lieu de convergences, le cadre général du E-learning représente en fait un véritable "pôle magnétique" exerçant une attraction très forte sur l'ensemble de ces paramètres. C'est le lieu que j'ai choisi pour tenter de rassembler l'ensemble des dispositifs ou des objets que j'ai élaboré, - ou que j'ai contribué à élaborer - au sein du Laboratoire Paragraphe (Document Réticulaire Structuré, HyWebMap, K-Web Organizer, NeuroWeb, ...) [Papy 04b] [Papy 04c]. L'expérience menée de 2002 à 2004 a nécessité l'adaptation d'un certain nombre de fonctionnalités techniques provenant de la plate-forme INES de l'Université Picardie Jules Verne pour remédier non seulement aux limites des logiques des outils mais également à l'omniprésence des dimensions culturelles et sociales au sein des environnements communicants [Papy 04d]. Néanmoins, la capacité à pouvoir construire, déconstruire, partager, améliorer, ..., en fonction des profils cognitifs des individus, des sources d'informations qui pourront s'avérer source de connaissances voire même sources de savoir est un élément déterminant que le document numérique apporte sans conteste.

C'est dans ce sens que s'orientent mes travaux ; intégrer dans la conception même des dispositifs de recherche d'information et de collaboration, les dimensions irréductibles de l'information et de la communication, pour éviter la difficile expérience que relate Dominique Boullier lorsqu'il a été amené à "faire plier la logique de l'outil" [Boullier 00].

Le partenariat que j'ai initié avec la Bibliothèque Universitaire de Paris 8 afin de proposer un dispositif de consultation du catalogue en ligne ne venant pas contrarier les pratiques des usagers de la bibliothèque (public et professionnel) s'insère précisément dans cette démarche.

De mon parcours, je peux ainsi identifier cinq étapes majeures marquées par:

1 – Définir des modalités opérationnelles de production automatique de réseaux hypertextes en s'appuyant sur la structure physique et logique de documents électroniques faiblement structurés. Ces modalités opérationnelles prennent en compte d'une part la création automatique de réseaux et d'autre part la nécessité d'une méta-description des réseaux automatiquement générés afin d'une part de rendre possible l'actualisation (meta-réseau) et d'autre part d'améliorer les parcours de

lecture. Ces travaux de recherche doctorale réalisés dans le cadre d'un CIFRE avec la Direction Etudes et Recherches d'Electricité de France, ont montré l'application réelle, en milieu industriel des principes de l'hypertexte en matière d'exploration et de navigation dans des fonds documentaires électroniques volumineux. Ils ont été intégré dans d'autres travaux : thèse de doctorat de Didier Bourigault, thèse de doctorat de Bertrand Le Quellec, mémoire d'ingénieur CNAM de Véronique Jolly.

2- Avec l'avènement du Web, proposer de nouvelles modalités de recherche sur des documents électroniques dont les structures physique et logique sont peu significatives (HTML). L'interface de recherche et d'interrogation que j'ai mis en place dans le moteur de recherches NeuroWeb a confirmé l'intérêt de combiner les modes d'exploration textuel et graphique dans un fonds documentaire peu structuré au fil des rétroactions autorisées par le moteur et l'interface à partir des sélections de l'utilisateur (*relevance feedback*).

3- Exploiter les formats de documents dans la définition de document numériques hétérogènes susceptible de suivre l'évolution cognitive des usagers placés en situation individuelle et collective de productions et de construction de connaissances (document numérique XML pour HyWebMap et K-WebOrganizer).

4- Envisager les modes d'organisation des communautés de pratiques et des communautés d'intérêt autour du document numérique ; modes d'organisation qui s'affirment aujourd'hui avec les Environnements Numériques de Travail, les Formations A Distance et plus généralement les Campus et Universités Numériques⁹. De façon plus générale, les questions relatives à la veille (stratégique, technologique, économique et plus directement informationnelle), à l'émergence des communautés d'intérêt et de pratique (de tous ordres), à la prolifération des forums de discussions, aux systèmes de publication partagée, aux "Wiki" et autres "Weblogs" montrent une tendance avérée de l'exploitation agrégative - quelle soit à vocation personnelle, collective ou professionnelle -, de l'information recherchée et collectée. Cette approche de l'agrégation et de la syndication de contenus numériques, personnalisables et partageables a inspiré un moteur de recherches

⁹ <http://www.educnet.education.fr/superieur/campus.htm>

agrégatif (Web Feedback¹⁰) que j'ai prototypé à partir des acquis de K-Web Organizer.

5 – Transposer les acquis relevant des problématiques larges de l'hypertexte dans des environnements de connaissances fortement organisés. Depuis octobre 2003, j'ai initié le projet "VisualCatalog" qui implique une équipe pluri-disciplinaire d'une douzaine de membres (conservateurs des bibliothèques, enseignants-chercheurs) et qui introduit une convergence entre principes de l'hypertexte, documents numériques, cartographie dynamique, exploration, navigation dans les OPAC des bibliothèques. C'est le catalogue en ligne (250 000 notices) de la bibliothèque universitaire de l'Université Paris 8 qui fait actuellement l'objet de cette transposition. Ce projet d'envergure est inséré dans le contrat quadriennal 2004-2007 de l'Université Paris 8.

Croiser les multiples approches relatives aux interfaces, aux documents numériques structurés, à l'analyse sémantique et statistique de fonds documentaires numériques, expérimenter des techniques de réorganisation et de visualisation de documents en exploitant les formats de documents aujourd'hui disponibles (XML et dialectes associés) forment ainsi le fil conducteur de mes travaux de recherches qui se sont articulés en quatre grandes périodes :

¹⁰ <http://feedback.univ-paris8.fr/index.php>

1) "Hypertextualisation automatique" (1991-1996)

- Introduire dans des documents électronique faiblement structurés des modalités de lecture inspirées de l'hypertexte moins contraignantes que les techniques de consultation de données disponibles à cette époque et largement influencées par les approches informatiques : algorithmique, bases de données et systèmes experts.
- Produire automatiquement des hyperdocuments en s'appuyant sur les idées sous-jacentes associées aux formats des documents structurés (principalement SGML et ODA) permettant d'affranchir ces "*documents hypertextes*" de solutions de description propriétaires éphémères (RTF).
- "Dé-linéariser" des textes électroniques à partir de la structure physique des documents. Une structure logique établie en tenant compte des repères physiques disposés dans les documents initiaux forme, grâce à une grammaire adéquate, une grille de définition des unités logiques composant les réseaux hypertextes produits et servant de base à une navigation structurelle (structure logique).
- "Re-linéariser" par des parcours de lecture multiples (1 document auteur | 1 hyperdocument généré \Leftrightarrow N document(s) lecteur(s)).
- Introduire une méta-description des réseaux automatiquement générés afin d'en permettre l'actualisation (couplage de l'hypertextualisation avec une organisation relationnelle des données et un mode d'interrogation normalisé de type SQL).

2) Technologies du Web (1996-2000)

L'avènement du Web et la prolifération de pages HTML ont introduit la nécessité de disposer de mécanismes de recherche permettant de localiser les documents statiques existants (URL) à partir d'une recherche sur les contenus de ces documents.

Intégrer la gestion des pages HTML (contenu et informations para-textuelles) aux systèmes de gestion de bases de données relationnelles s'est progressivement imposée pour constituer le "Web dynamique".

Plus que la dimension technique de gestion de données, le rapprochement des technologies du Web vers celles des bases de données relationnelles s'explique par la nécessité d'une structuration logique des données. Le modèle relationnel impose par définition cette forme de description logique afin d'assurer un fonctionnement optimal des systèmes associés en respectant (a) l'indispensable minimisation de la redondance de données, (b) l'indépendance logique/physique et, (c) la non-procéduralité des interrogations permettant à un utilisateur non informaticien d'exploiter ses propres données au travers d'un langage de définition et de manipulation de type SQL.

Ce rapprochement a contribué à imposer la démarche structurante (d'un point de vue logique) des documents que l'on retrouve dans les extensions actuelles du document numérique au travers des technologies du Web (XML et XSLT notamment).

3) "Document numérique et applications" (depuis 1998)

On constate que la présence de documents de toutes natures (texte, images, animations, sons, vidéo) dans l'espace électronique mondialisé du Web met en échec annuaires et moteurs de recherches. La structuration confuse (mélange de description logique, physique, logique/physique) et insuffisante des documents HTML inhibe les utilisations innovantes sous-jacentes (méta-écriture, collaboration, partage) que le Peer-to-Peer a amorcé mais suivant une approche essentiellement technique.

Ces utilisations nouvelles qui font une large part à la personnalisation de contenus, au partage et à la collaboration ont mis en évidence l'adaptabilité des contenus en fonction des profils cognitifs des utilisateurs et des conditions d'acquisition/réception de ces contenus (réseaux hauts-débits, câble, xDSL, téléphones mobiles, PDA, TabletPC, portables).

J'ai intégré au dispositif HyWebMap, conçu à l'origine pour répondre aux problématiques de la désorientation et de la surcharge cognitive révélées lors de l'exploration de fonds documentaires hypertextuels issus du Web, un formalisme XML permettant d'appréhender les réseaux virtuels produits par le système HyWebMap comme un véritable document-réseau extensible. Ce

document-réseau que j'ai nommé "Document réticulaire Structuré" devient l'entité de base partageable au sein d'un dispositif collaboratif K-Web Organizer.

Le moteur de recherches multilingue NeuroWeb¹¹ développé au sein du Laboratoire Paragraphe, pour lequel j'ai défini l'interface de navigation et agrégé les différents composants fonctionnels, synthétise l'apport des technologies du Web, de l'ingénierie linguistique et du traitement statistique.

Les différentes réalisations que j'ai menées en milieu industriel sous la forme de collaboration Recherche/Industrie (l'intranet de capitalisation de connaissances pour la métallurgie chez le constructeur automobile Renault, les sites Internet destinés à proposer aux internautes l'intégralité des ressources documentaires liées aux émissions diffusées par Arte-TV ; l'une de géopolitique "Le Dessous Des Cartes", l'autre de vulgarisation scientifique "Archimède") ont confirmé la nécessité d'interfaces "inframince" Homme / Information.

4) "Structuration des connaissances, Navigation, Interface" (depuis 2003)

Face à l'expansion de l'information numérique, la structuration logique de l'information s'est imposée comme la seule alternative pour éviter que le Web ne se transforme en un capharnaüm électronique. Les travaux amorcés pour la sémantisation du Web (Web Sémantique via Dublin Core Metadata Initiative, Resource Description Framework et Web Ontology Language) confirment l'indispensable nécessité de la méta-description des documents numériques. Le monde de la documentation et des bibliothèques, principalement celui afférent aux bibliothèques universitaires, de par leurs proximités avec le monde de l'enseignement supérieur et de la recherche, se présente comme un terrain de recherches idéal pour l'organisation d'un système d'exploration de ressources électroniques dans un environnement fortement sémantisé et structuré (normes UNIMARC, Z39.50, Thésaurus, répertoire d'autorités-matières,...).

Introduire les principes de l'hypertexte (navigation, feuilletage) au sein du "catalogue en ligne" en exploitant les méta-données des notices

bibliographiques (Classification Décimale Universelle, descripteurs RAMEAU) et en y adjoignant deux approches ; l'une relevant de l'ergonomie (sujet, artefact, instrument) et l'autre de la "visualisation métaphorique" constitue la démarche que j'ai amorcé dans le projet VisualCatalog¹² depuis septembre 2003 et qui se propose de faire converger ces approches distinctes.

La plupart des activités décrites ici sont depuis ma recherche doctorale, le fruit d'un travail d'équipe avec les différentes personnes qui font partie de "l'Equipe Hypertexte Dynamique"¹³. Les signatures multiples des publications auxquelles j'ai contribué reflètent ce mode de travail. La première personne du pluriel employé fréquemment dans ce mémoire correspond donc effectivement à des participations de plusieurs personnes dans les choix, les spécifications et les réalisations décrites. Cependant, je m'efforce dans la mesure du possible d'identifier ma part personnelle dans ces travaux.

Pour rendre compte de mes activités de recherches, j'ai choisi d'organiser mon mémoire en quatre chapitres qui proposent un compromis entre la chronologie de mes recherches et les éléments déterminants de cette recherche :

- **"Hypertextualisation, Métaréseau et Navigation par la structure"** (chapitre 1). Ce chapitre qui illustre les principes de la granularisation des documents à partir de grammaires physico-logiques contenues dans la signature des formats de documents à hypertextualiser (SGML, RTF) préfigure les conditions de construction dynamique de contenus qui sont aujourd'hui proposées avec les technologies issues du Web. En effet, les langages de définition et de déclaration de structure établis autour du XML permettent aujourd'hui de définir des séquences d'informations et de leur adjoindre des règles de décomposition/recomposition/assemblage, notamment avec Xlink, Xpointer, Xpath et XSLT. Par ce travail de thèse, j'ai pu aborder de façon extrêmement pratique et opérationnelle les opérations applicables sur les documents structurés que les documents numériques continuent à être aujourd'hui. Les normes

¹¹ <http://neuroweb.univ-paris8.fr>

¹² <http://visualcatalog.univ-paris8.fr>

¹³ <http://h2ptm.univ-paris8.fr/ehd/>

émanant du Consortium W3 ont assurément assaini un milieu qui, jusqu'à un passé récent était laissé aux initiatives multiples des sociétés informatiques qui voyait là un terrain de batailles pour instaurer une hégémonie commerciale.

- **"Pour l'élaboration d'Interfaces Homme / Connaissances"** (chapitre 2). Le processus complexe de recherche d'information et l'appréciation de la pertinence des informations rapatriées par l'offre pléthorique de dispositifs ad hoc, passent inmanquablement par les interfaces des logiciels. Cette question d'interface relève tout autant des modalités fonctionnelles d'interaction intégrés par les concepteurs au sein des systèmes que des subterfuges techniques destinés à attirer l'attention de l'utilisateur sur les informations (estimées) pertinentes (listes de termes, de documents, résumés automatiques, cartographie, estimation statistique,...). L'interface qui rapproche le monde humain du monde numérique de l'information alimente de façon conséquente le désarroi de l'utilisateur qui pourra rester étranger à la logique du système opératoire de proximité (IHM) et/ou à la logique de sélection des informations collectées (moteurs de recherche, SRI). Différentes situations d'expérimentations à partir d'environnements techniques que j'ai développés en grande partie, m'ont apporté des éléments de réponse quant aux paramètres à intégrer dans les interfaces des dispositifs de recherche d'informations, et qui gravitent essentiellement autour de l'axe central du contexte informationnel de l'utilisateur (besoins, connaissances préalables). Ces interfaces étaient destinées à des publics variés : spécialistes de la métallurgie automobile (capitalisation de connaissances en incidentologie de fonderie, Fonderies du Poitou et FunFrap/Renault), professionnels de l'information (NeuroWeb/ANRT)¹⁴ et du grand public intéressé par la géopolitique (Le Dessous Des Cartes/ARTE). Ces trois expérimentations ; l'une de recherche et les deux autres industrielles, ont confirmé que la problématique de l'interface ne pouvait plus être considérée comme le seul apanage de la science informatique [Aristote 02] [Coutaz 90] principalement depuis que les questions d'interfaces initialement appréhendées comme un cadre de communication avec des machines physiques (calculateur, périphériques) s'attachaient aujourd'hui à des interactions avec une matière abstraite : l'Information. Les travaux relatifs à la

¹⁴ C'est ce dispositif que je décrirai principalement dans le chapitre 2

"Communication Personne-Système (d'information)" [Miège 02] et à la "Communication Médiatisée par Ordinateur" [Cmo 01] ont légitimé une position info-communicationnelle des interfaces. L'idée sous-jacente que j'introduis dans ce deuxième chapitre est que les "Interfaces Homme-Connaissances" représentent une voie de recherches totalement distincte des approches strictement informatiques consacrées aux IHM et aux IHE, où l'Interface, l'Usager et l'Information occupent respectivement les fonctions du message, de l'émetteur et du destinataire. En effet, le développement de l'information numérique sous les multiples formes que nous lui connaissons aujourd'hui tend l'utilisateur (l'émetteur), à exprimer des requêtes qui ne sont plus uniquement des sollicitations à des équipements techniques variés ("informatique diffuse") mais des propositions d'interaction avec un fonds informationnel (le destinataire). Ces allers-retours ont pour objet d'extraire d'une part des informations répondant à un ensemble de questions ou de besoins relevant de positions socio-cognitives, et d'autre part d'obtenir une représentation du fonds informationnel interrogé (Web, Cyberspace). Dans un contexte où l'on sait que la théorie de l'information ne peut pas être réduite à une expression mathématique de la quantité d'informations transmises, l'interface (le vecteur du message) ne peut plus être considéré comme une expression informatique aussi sophistiquée soit-elle, répondant à des propositions heuristiques liées à des traitements pré-supposés. Il s'agit au contraire de permettre à l'utilisateur d'investir totalement le vecteur de communication, en restreignant les traitements spécifiques à l'interface, pour favoriser un feedback constructif entre ses sollicitations et l'interprétation des résultats de réponses adressées par le fonds informationnel. Car ne l'oublions pas, l'utilisateur de ressources électroniques est aujourd'hui de plus en plus confronté aux modalités opératoires de l'Internet (Intranet et extranet compris) qui lui demande de savoir chercher l'information, d'avoir une représentation claire de sa situation cognitive, de définir ses propres stratégies d'acquisition de connaissances, de mesurer le contexte dans lequel s'insère cette démarche de recherche et d'exploiter à la convergence de ces différents facteurs les bons artefacts informatiques. Face à cette complexité, mes travaux actuels - que j'applique dans le projet VisualCatalog -, tendent à affirmer l'Interface comme un vecteur de la communication médiatisée par

l'ordinateur, affecté de propriétés les plus neutres possibles, afin d'élaborer une interface "inframince" [Buci-Glucksmann 02].

- **"De HyWebMap à K-Web Organizer. Vers la généralisation du document numérique personnalisable. Le cas du Document Réticulaire Structuré"** (chapitre 3),

Les dispositifs HyWebMap et K-Web Organizer ont été élaborés au sein de l'équipe Hypertexte Dynamique du Laboratoire Paragraphe dans l'idée de répondre aux phénomènes de désorientation et de surcharge cognitive largement observés dans la consultation de réseaux hypertextes. Le dispositif HyWebMap est un système d'organisation des documents lus et collectés sur le Web par l'utilisateur. Totalement différent d'un aspirateur de sites et plus adapté à la recherche d'informations qu'un gestionnaire de signets, HyWebMap propose des fonctionnalités d'annotations (mots-clés, commentaires) de description de nœuds et de liens. Outre cette vocation à retarder les situations de désorientation et de surcharge cognitive, HyWebMap a été conçu dans une optique d'organisation technique de données (et non pas de document) empruntant beaucoup à la modélisation relationnelle telle que je l'avais adoptée dans mon travail de thèse pour l'actualisation des réseaux hypertextualisés [Papy 95] [Papy 94] [Vuldy 93]. Ma participation effective à l'élaboration de ce dispositif a consisté à reconsidérer la position "donnée" pour privilégier la position "document" en exploitant les formats de type XML. La "Document Type Definition" que j'ai arrêtée et qui a débouché sur un "Document Réticulaire Structuré" a apporté une très grande interopérabilité avec d'autres SRI et a permis d'entrevoir la construction d'espaces de connaissances comme un activité collaborative. Le système K-Web Organizer, né de cette proposition collaborative, consiste en un moteur de recherches sur des documents réticulaires structurés partagés au sein de communautés d'intérêts.

Ces travaux sur HyWebMap et K-Web Organizer ont alimenté de façon déterminante mes réflexions sur l'utilisation des formats de documents de type XML afin de promouvoir l'accessibilité individuelle et collective de réseaux de connaissances élaborés par les usagers des systèmes d'informations (SRI, moteurs de recherches). La capacité à pouvoir envisager, grâce aux Technologies du Web inhérentes aux

formats de documents [XML 98], un document numérique *fragmentaire, évolutif, virtuellement et physiquement agrégatif* me paraît correspondre aux exigences sociales, culturelles et cognitives de la recherche d'information. La grande adaptabilité et le fort potentiel d'intégration intrinsèque de ces "technologies de documents" offrent enfin une véritable assistance technique à la dimension stratégique du traitement de l'information pour tous les acteurs impliqués.

C'est cette adaptativité du document numérique établie à partir des formats de document qui le rend propice à une composition / fragmentation / décomposition illimitée. Le document numérique est ainsi en mesure d'offrir un fond et une forme circonstanciellement appariés à tout contexte informationnel inédit (cognitif, pédagogique, organisationnel, social,...).

- "Conclusion et axes de recherches : Environnements Numériques de Travail. Navigation hypertextuelle et Visualisation graphique dynamique et métaphorique des recherches d'information sur un OPAC Web" (chapitre 4).

Mes travaux sur la conversion de documents électroniques en réseaux hypertextes ont clairement montré les avantages de la navigation hypertextuelle au sein de fonds documentaires techniques homogènes (documentations techniques relatives à l'ingénierie électrique) destinés à un public du domaine constitué de spécialistes ou de novices en formation¹⁵. Il est important de noter que les réseaux hypertextes automatiquement générés restaient assujettis aux mêmes exigences de gestion documentaire (contrôle qualité, plan de classement, comité de lecture, diffusion, confidentialité, archivage, publications,...) que les documents produits de façon plus traditionnelle au moyen de logiciels de traitement de textes. Les réseaux hypertextes mis à la disposition de ce public ont constitué finalement une banque de réseaux hypertextes dont les modalités de consultation ont été accueillies positivement par la plupart des utilisateurs experts puisque d'autres projets associant documents structurés et hypertextes ont suivi [Le Quellec 98] [Jolly 94] [Vuldy 93].

¹⁵ Ces novices issus du monde de l'électricité se trouvent placés, comme souvent en milieu industriel, dans un processus de formation visant à les informer des spécificités techniques et industrielles propres à l'entreprise, pour lesquelles leurs études initiales, fréquemment en sciences de l'ingénieur, les ont préparés en théorie.

L'exploitation de la structure logique des documents électroniques linéaires (révélées par les marqueurs physiques) pour déterminer les unités logiques (nœuds) des réseaux hypertextes automatiquement générés, signés (*tag*) de sorte qu'ils puissent être intégrés de façon univoque au sein d'environnements hypertextes différents, se retrouvent dans les formats des documents du Web sémantique (XML, DCMI, WOL, RDF,...). Ce sont ces puissantes et incontournables technologies que j'ai intégrées dans les travaux que j'ai menés par la suite (NeuroWeb, HyWebMap, K-Web Organizer, ...), et qui ont été alimentés par les problématiques de navigation dans les documents hypertextuels du Web (les pages HTML).

A l'occasion de ces travaux, les quelques pistes que j'ai eu l'occasion d'explorer avec mes collègues chercheurs du Laboratoire Paragraphe sont demeurés, à mon sens anecdotique en raison d'une part de la diversité des pratiques de "Recherche d'Information"¹⁶ et d'autre part des présupposés informatiques concernant les modèles informationnels des usagers [Carbonell 04] [Chaudiron 02] et les habiletés documentaires sous-jacentes [Baltz 03] [Dinet 02] [Curbières 02] [Dufour 01] [Dkaki 99] [Coulon 99] [Chaumier 77].

Il apparaît que le Web propose une multitude de documents, de cette multitude qui s'exprime en milliard [Andrieu 02] [Foenix-Riou 01], et qui d'un point de vue de la pertinence et de la qualité des documents, est d'une extrême hétérogénéité. La préoccupation de pouvoir retrouver une information ne représente en fait qu'un des nombreux aspects des problématiques majeures de la "Recherche d'Information". L'adéquation des réponses avec le profil du demandeur, la réorganisation des informations collectées en fonction du contexte des demandes, la familiarité avec les artefacts de recherche, etc. ne sont pas de moindre importance.

Il est essentiel de ne pas omettre que c'est en regard au contexte très spécialisé dans lequel l'Hypertexte a introduit des démonstrateurs [ACM 88] [Nielsen 90] [Conklin 87] où le public de "lecteurs" est familiarisé avec le fonds documentaire, qu'il est possible de le considérer comme un adjuvant véritable à la "Recherche d'Information" [Saleh 04] [Simonnot 02] [Dufour 01] [Hudon 01] [Holzem 99] [Kobayashi 00] [Lefèvre 00] [Marchionini 88]. Le principe d'accessibilité à

l'information introduit par l'hypertexte, amplifiant l'exploration associative, apparaît comme faisant véritablement sens dans une démarche de "Recherche d'Information" où les usagers et les interactions de ces derniers avec les fonds d'informations, sont a priori identifiés¹⁷. Les projets que j'ai menés chez EDF et Renault ont corroboré ce postulat, comme ils ont corroboré la nécessité d'une homogénéité intellectuelle des contenus.

C'est parce que l'Hypertexte a offert à un public d'experts une plus grande accessibilité¹⁸ à des informations dont ils avaient la connaissance, mais pour lesquels il était nécessaire de connaître les modalités fonctionnelles des systèmes informatiques complexes qui en avaient la charge de gestion [Papy 95], que j'ai envisagé de transposer cet acquis dans l'environnement documentaire d'une bibliothèque universitaire. On retrouve en ce lieu consacré à la pratique documentaire une évidente dimension conceptuelle de l'organisation des connaissances basée sur des règles explicites de classification du savoir (Classification décimale de Dewey, Classification décimale universelle, Library of Congress Classification, Classification de Bliss, National Library of Medicine,, Colon classification, Unimarc,...). Cette dimension conceptuelle de la structuration des connaissances est de façon contradictoire, partagée par les usagers-experts (conservateurs, bibliothécaires, assistants, catalogueurs, acquéreurs,...) qui en assurent la pérennité et ignorée par les usagers-visiteurs (étudiants, enseignants) qui constituent pourtant bien le public privilégié de ces lieux-ressources.

Bien que le rapprochement de l'hypertexte et des environnements documentaires proposés par les bibliothèques ne soit pas récent [Lefèvre 00] [Crosnier 95], j'ai pu constater qu'il relevait essentiellement de modalités techniques (réseau, protocoles, interfaces banalisées, liens hypertextes) destinées à améliorer les conditions d'accès

¹⁶ pris ici dans l'acception des professions de l'information. La RI résume l'ensemble des pratiques documentaires de recherche, de collecte, de traitement, de stockage et de restitution d'informations.

¹⁷ les usagers et les usages "identifiés" ne signifiant pas nécessairement "connus". Cette distinction qui peut paraître subtile révèle la richesse des pratiques documentaires pour des profils d'usagers attendus. La bibliothèque universitaire, les collections et les usagers (personnels de la BU, étudiants, enseignants) sont effectivement identifiés. Clairement pour les ouvrages disponibles dans l'édifice, beaucoup moins pour les usagers. Nous verrons dans le chapitre 4 que les profils de ces différents usagers sont non seulement diversifiés mais qu'au sein même de chacune de ces grandes catégories (personnels, étudiants, enseignants) des variations apparaissent. Le constat est très similaire pour les usages.

¹⁸ accès technique facilité, accès intuitif privilégié.

aux collections dans le cadre des programmes d'informatisation des bibliothèques [Jacquesson 94] au moyen d'un Système Intégré de Gestion des Bibliothèques (SIGB)¹⁹.

Comme le fait remarquer Richard Roy [Roy 03], les catalogues informatisés des bibliothèques par leur mise en ligne (OPAC Web) ont vu leur usage se démultiplier et s'accompagner dans le même temps d'une dégradation qualitative du fait d'une conception réductrice et malheureusement dominante de ces interfaces "Web". Le libreaccès technique porté par les SIGB ne signifie pas pour autant accessibilité des informations afférentes aux ressources électroniques rendus disponibles [Rouet 01] [Polit 04] [Yee 99]. Hors, depuis [Corson 82], on sait que la dimension psychologique liée à l'utilisation d'environnements informatiques (en l'occurrence les bases de données) est importante. Ces travaux ont souligné la nécessité d'apporter aux utilisateurs (usagers) la compréhension des principes d'organisation des systèmes d'informatisation auxquels ils se trouvent confrontés et ce, afin qu'ils puissent bénéficier pleinement des potentialités des systèmes informatisés de recherche d'informations²⁰.

L'orientation que je donne à ce projet de recherches conséquent²¹ pour lequel j'ai constitué une équipe pluridisciplinaire de 13 personnes, (conservateurs,

¹⁹ "Le système d'information documentaire est un dispositif informatique global d'accès à la documentation, aux ressources pédagogiques et à l'information multimédia, depuis un poste de travail ou de consultation banalisé, grâce à une interface unique rendant à l'utilisateur transparents les différents langages et normes des systèmes agrégés constitutifs et lui permettant l'interrogation de bases de données et de différentes ressources, ainsi que l'accès à des services, tels que la commande, la réservation de documents ou d'autres prestations, un album personnel ; ainsi le catalogue Opac du SIGB, le réseau de CD-Rom, les documents internes numérisés (thèses, cours), les abonnements électroniques, les bases de données bibliographiques, textuelles, factuelles, les campus numériques, etc., sont disponibles, selon les habilitations des usagers (gérées par un annuaire central) sur l'Intranet de l'établissement d'enseignement supérieur ou sur Internet, localement ou à distance. Le SID est constitutif du système d'information de l'établissement (ou portail) ; souvent, il en est la pierre angulaire. Il résulte d'une organisation et d'une architecture réfléchies au niveau de la direction de l'établissement et de la mise en commun des savoir-faire de la bibliothèque, du centre de ressources informatiques (CRI) et des enseignants-chercheurs pilotes dans les TICE." (Enseignement supérieur La sous-direction des bibliothèques. <http://www.sup.adc.education.fr/bib/Acti/Infor/infor.htm>)

²⁰ "Les connaissances actuelles en ce domaine laissent supposer que les difficultés éventuelles d'utilisation du système seraient en grande partie dues à une discordance trop importante entre les représentations des informations proposées et l'organisation de ces informations telle que la conçoit l'opérateur" (p. 53)

²¹ <http://visualcatalog.univ-paris8.fr>

bibliothécaires, psycho-ergonomes, géographe-cartographe,...) ne porte pas sur une n-ième édition d'une transposition technique de l'Hypertexte revisitée par l'attrait particulier de technologies informatiques récentes (XML, ingénierie linguistique) mais sur l'intérêt d'utiliser ce principe d'accessibilité pour favoriser la préhension mutuelle des environnements intellectuels des usagers-experts ("auteurs" de modalités pratiques d'organisation et d'adaptation des schémas génériques d'organisation de la bibliothèque²²) et des usagers-visiteurs.

Il ne s'agit pas ici d'imposer un modèle techniciste quelconque aux uns et aux autres, fréquemment inadapté parce qu'en rupture avec les situations sociales, culturelles et cognitives des acteurs, mais de rendre effectivement accessibles les univers intellectuels des uns et les exigences organisationnelles des autres, en rendant visible les attentes des premiers, les pré-supposés des seconds...au moyen d'un élément commun susceptible d'assurer la fonction d'un système de communication permettant aux uns d'appréhender les usages des autres. On peut observer que les approches informatiques en ingénierie des connaissances [Bui Thui 03] [Laublet 02] [Gandon 02] qui tout en laissant transparaître ces multiplicités de point de vue [Poissenot 02] cherchent invariablement à les homogénéiser, à les contraindre dans un modèle qui - le paradoxe est bien là – éloigne les acteurs susceptibles de proposer des situations réelles de confrontation, de vérification et donc de validation. Le degré d'abstraction des modèles produits est tel, que ces acteurs, ne percevant plus le rapport de parenté entre cette abstraction et leur réalité [Sperber 01], évitent ces confrontations enrichissantes qui permettraient d'élaborer des modèles plus proches des usages. Si les usagers sont contraints de passer sous les fourches Caudines d'un tel modèle, ils proposeront une manifestation tronquée de leur réalité pour pouvoir bénéficier malgré tout du processus de validation associé au modèle. A moindre échelle, on retrouve cette problématique du modèle et des instances dans les DTD

²² les bibliothèques universitaires fonctionnent sur des principes identiques en ce qui concerne les services proposés aux usagers (accès libre aux ouvrages, catalogue, PEB, accès aux ressources électroniques, bouquets de presse,...). En revanche, l'identité scientifique de la bibliothèque est directement influencée par l'identité de l'université qui l'héberge (sciences exactes, médecine, sciences humaines et sociales). Au sein de chacun de ces établissements, et par voie de conséquence de chacune de ces bibliothèques, se développent alors des spécialités, des pôles d'excellence qui se constituent au fil des évolutions scientifiques des champs disciplinaires concernés. La bibliothèque universitaire accompagne cette évolution et archive les témoignages de cette évolution. Ces témoignages ont des répercussions directes sur la classification utilisée dans la bibliothèque. En effet, la Classification Universelle Décimale se trouvera adaptée et aménagée dans l'établissement impliqué.

SGML [Role 01] [Bradley 92] [GoldFarb 91] [Herwijnen 90] et dans les variations de documents susceptibles de s'y conformer. En prenant pour exemple l'initiative CALS²³ (*Computer-aided Acquisition and Logistics Support* ou *Continuous Acquisition and Life-Cycle Support*), on peut trouver exceptionnel, à la lumière de ses objectifs initiaux, la capacité d'intégration de ce méta-modèle de documents SGML mais on peut également regretter les contraintes qu'un tel modèle a imposé aux entreprises, organismes, correspondants,... arrivés après l'élaboration de ce modèle et de la nature des concessions qu'elles ont du apporter dans la redéfinition de leurs documents afin de se conformer aux exigences du meta-modèle CALS. Qui pourrait raisonnablement soutenir l'idée que ce "lissage" rendu nécessaire par le méta-modèle n'a pas tué dans l'œuf une originalité documentaire reflet de l'originalité de l'entreprise ou de l'institution concernée ?

Or en matière de Recherche d'Information, la question de la modélisation reste en suspens car les usages sont complexes, reflets d'univers documentaires dont la cohérence réside dans la façon dont les usagers, subjectivement, se les approprient [Ghitalla 02]. Ils demeureront le fait de phénomènes de centration, de démarches anthropocentrées dont il est difficile de nier ou de confirmer la pertinence. A ce titre, le catalogue en ligne se présente comme le lieu de rencontre, vecteur d'échanges et support de communication entre différents usagers porteurs d'usages extrêmement différents liées aux fonctions qu'ils occupent, aux tâches qui leurs incombent et aux travaux qu'ils ont à assumer.

Les travaux que je mène depuis septembre 2002, qui sont insérés dans le contrat quadriennal 2005-2008 de l'université Paris 8, trouvent une exploitation judicieuse des technologies du Web (bases de données, langages de développement, XML, ...) et visent à donner fond et forme à de véritables interfaces de sens permettant à des communautés disjointes mais extrêmement proches, d'appréhender les univers informationnels respectifs dans lesquels ils sont impliqués. Ceci, selon moi, déterminera l'amélioration de la performance documentaire de tous les acteurs.

²³ La meta-DTD CALS élaborée à l'initiative du DoD (Department of Defense américain) est un programme de rationalisation du processus d'acquisition de matériel militaire qui vise, en ce qui a trait à la documentation technique, à réduire les coûts reliés à la production, au stockage, à la mise à jour et à la consultation de l'ensemble des manuels techniques d'équipement militaire. (<http://navycals.dt.navy.mil/cals.html>)

Cette recherche appliquée fait converger les résultats de mes recherches antérieures significatives décrites dans les chapitres de ce mémoire d'habilitation.

L'organisation, l'ingénierie et la communication des connaissances, les particularités de la navigation hypertexte en ce qu'elle favorise la lecture associative ou qu'elle permette par effet de sérendipité de rencontrer une information qui trouve un écho dans les recherches menées, constituent une occasion exceptionnelle de combiner des approches interdisciplinaires pour rendre les usages plus visibles (Information, Documentation, Psychologie-Ergonomie, Art, géographie-Cartographie).

C'est une perspective de recherches alléchante que de pouvoir (ré)introduire la dimension métaphorique [Chauvin 04] [Papy 03] comme alternative pertinente à des réalisations techniques (CartoWeb, Kartoo, Influo, Mapstan, Miner3D, Perspectives Wall, ...) [Lecolinet 01] [Lelu 01] [Card 99] [Cubaud 98] [Lelu 98] [Brown 89a] [Keyes 85] largement inspirées de la Visualisation Scientifique pour des résultats dont on ignore aujourd'hui s'ils répondent de fait aux besoins informationnels des usagers [Hascoët 04] [Bihanic 03] [Chaudiron 02] [Ghitalla 02] [Jeanneret 99]. A ce propos, le regard de la psychologie-ergonomie qui est largement mis à contribution dans ce projet [Papy 04] [Rabardel 01] [Rabardel 95], nous apportera des éléments de réponse quant aux pistes que mon intuition m'aura amené à explorer.

CHAPITRE 2

HYPertextUALISATION, FORMAT DE DONNEES, META-RESEAU ET NAVIGATION PAR LA STRUCTURE

Dès la fin des années 80, de nombreuses réalisations techniques ont permis de vérifier les perspectives innovantes de l'Hypertexte [Nielsen 89] [ACM 88] en matière d'accessibilité des fonds documentaires électroniques. A cette époque, les systèmes experts issus de l'intelligence artificielle, les systèmes de gestion de bases de données et plus généralement le génie logiciel exigeaient une organisation spécifique des systèmes d'information déterminant chez les utilisateurs non seulement une connaissance précise des dispositifs techniques assurant la médiation avec les données stockées, mais également une connaissance approfondie de l'organisation sous-jacente des données (bases de fait, bases de règles, bases de connaissances, logiques prédicatives, chaînages avant et arrière pour les systèmes experts, langages de définition et de manipulation dédiés pour les SGBD relationnelles couplés à une modélisation des données imposée en amont).

Les systèmes hypertextes développés en ce début des années 90 (HyperCard, Guide, KMS, etc...) ont été à l'origine de démonstrations et de prototypes qui ont mis en évidence l'intérêt de l'accès intuitif à l'information sans pour autant réussir à proposer une "industrialisation" de l'utilisation des réseaux hypertextes.

En dépit de cette "frilosité" industrielle, la difficulté à en définir précisément les contours et à proposer une formalisation théorique efficace [Balpe 96] [Clement 95] [Giffard 97] , l'Hypertexte s'est imposé malgré tout comme une alternative originale dans la façon d'appréhender l'information en permettant de réaliser des associations qui ne s'expliquent pas nécessairement par des raisons déductives, de cause à effet ou de tout autre nature logique mais pour des raisons souvent intuitives. Le lecteur grâce à des interfaces logicielles simples qui le dispensent d'acquérir une technicité particulière, se déplace en activant les liens disponibles dans un espace de connaissances en fonction de ses thèmes de recherche, de sa curiosité ou plus

simplement d'une forme de fortuité. Dans ce but, l'hypertexte exploite au mieux les possibilités technologiques des calculateurs (processeurs rapides, mémoires centrales volumineuses, mémoires de masse à accès direct, écrans graphiques, multifenêtrage, dispositifs de pointage et de sélection) en les mettant au service des associations entre informations peu structurées conduites par l'intelligence ou l'intuition des lecteurs [Nielsen 90] [Sandvad 89] [ACM 88].

Mes travaux de recherche réalisés dans le cadre d'une Convention Industrielle de Formation par la Recherche se sont déroulés à la Direction des Etudes et Recherches d'Electricité de France au sein du service "Informatique et Mathématiques Appliquées" et répondaient aux problèmes majeurs rencontrés par les instances scientifiques de la Direction des Etudes et Recherches (prestataire de services pour les autres directions opérationnelles d'EDF) en matière de disponibilité et d'exploitation (actualisation des documents, formations des techniciens et ingénieurs, "mémoire industrielle",...) d'importantes documentations techniques liées aux ouvrages électriques (Isolement des lignes à haute tension, incidentologie des transformateurs électriques, Systèmes Elémentaires des Centrales Nucléaires,...) et aux applications informatiques placées sous contrôle qualité (code de mécanique ASTER²⁴, logiciel PDMS utilisé pour la modélisation des centrales nucléaires...).

Dans cet environnement, mon travail de thèse a largement contribué à ce que l'hypertexte soit envisagé comme une alternative sérieuse aux dispositifs largement utilisés à l'époque que sont les systèmes documentaires, les bases de données et les systèmes experts mais dont le déploiement et les conditions d'utilisation restaient assujettis à de lourdes obligations de formalisation du système d'information, de formation des utilisateurs et d'objectif de conception. L'hypertextualisation a montré les possibilités d'industrialisation de la transformation de documents électroniques faiblement structurés (au format Rich Text Format), essentiellement techniques (documents de type sciences de l'ingénieur) afin de généraliser la production de réseaux hypertextes mieux adaptés à des navigations exploratoires de la part des utilisateurs et utilisant efficacement les interfaces graphiques de type WIMP (*Window Icon Menu Pointer*).

²⁴ Ce code de mécanique représentait plus de 400 000 lignes de programmes FORTRAN exécutables sur super-calculateur CRAY.

L'ensemble applicatif, car il s'agissait bien d'un ensemble de briques logicielles²⁵ que j'ai développé, a permis de pérenniser l'activité de recherche sur les hypertextes au sein de la Direction d'Etudes et Recherches d'EDF. En effet, d'autres travaux de recherches ont suivis sur la question de l'hypertextualisation :

- une thèse orientée vers la "Construction Automatique de documents hypermédias à caractère technique dans un cadre industriel "[Le Quellec 98],
- un mémoire d'ingénieur CNAM portant sur la "Mise en place d'une chaîne d'hypertextualisation automatique de documents techniques" en mettant en œuvre la norme SGML [Jolly 95] [Vuldy 93].

Il faut également noter l'usage de la chaîne d'hypertextualisation dans la réalisation sous HyperCard de la première interface de navigation des résultats d'extraction du logiciel LEXTER²⁶ élaboré dans le cadre d'une thèse en Traitement Automatique des Langues [Gros 94] [Bourigault 92] [Bourigault 94] (cf. figure 1).

²⁵ 4 modules composaient la chaîne d'hypertextualisation APTE : un logiciel d'écriture et d'analyse de patrons syntaxiques permettant de repérer les références croisées dans les documents techniques traités (formalisme BNF), un logiciel de granularisation fonctionnant de concert avec la grammaire BNF rédigé, un module de portage des unités physiques composant le réseau hypertexte vers les systèmes hypertextes dédiés (Toolbook, HyperCard, Multicard) et enfin le dispositif de supervision du méta-réseau destiné à l'actualisation des réseaux automatiquement générés.

²⁶ LEXTER (Logiciel d'EX de TERminologie) est un analyseur syntaxique robuste dédié à l'extraction de syntagmes nominaux et adjectivaux à partir de corpus spécialisés rédigés en français, dans une perspective d'acquisition terminologique.

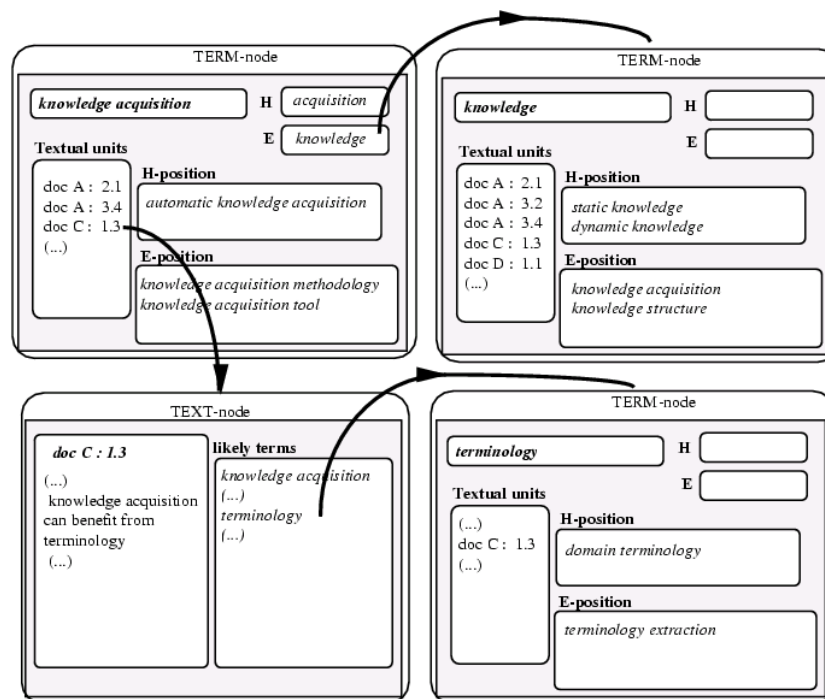


Figure 2.1. Liens hypertextes de l'hypertexte terminologique [Gros 94]

La question de l'universalité des formats d'échange des documents électroniques hypertextualisés

L'hypertextualisation, plus qu'un processus fonctionnel représente un "lieu de traitement" qui transforme le document électronique linéaire initial en un autre objet finalisé sous la forme d'un réseau de nœuds et de liens appelé de façon usuelle "réseau hypertexte". Mais ce lieu de traitement n'aboutit de fait qu'à la conversion d'un document électronique en un autre document électronique²⁷. La dimension hypertextuelle des objets électroniques résultant de l'hypertextualisation est proposée de manière essentiellement factice par les interfaces de consultation des systèmes hypertextes. Réseaux, piles, livres, cartes, pages, fiches, liens, ancres (de départ et d'arrivée) sont autant de métaphores destinées à suggérer des nœuds et des liens au sein de réseaux hypertextes qui n'ont pas d'existence électronique effective.

²⁷ à cette époque, les logiciels et autres systèmes hypertextes et hypermédias tels qu'HyperCard, Toolbook, MetaCard ou Dynatext exploitaient bel et bien des fichiers informatiques aussi monolithiques d'un point de vue technique que les fichiers de traitement de textes soumis au processus d'hypertextualisation

L'hypertextualisation est ce lieu du transitoire et du temporaire qui produit de nouvelles informations (ne répondant plus à l'organisation conceptuelle et matérielle des informations présentes dans le document initial) qui ne seront pourtant pas celles proposées à la lecture hypertextuelle ; la matérialisation et la représentation finale de l'information hypertextualisée demeurant dépendantes des dispositifs techniques adoptés.

En effet, à cette époque, l'absence de normes (hyper)documentaires exploitables²⁸ ne permettait pas d'envisager la production de documents hypertextes en dehors de formats de données imposés par les systèmes propriétaires existant. L'accessibilité de l'information apportée par l'hypertexte se trouvait remise en question par l'assujettissement fonctionnel des réseaux hypertextes aux spécificités des systèmes employés. Dans ce contexte, l'hypertextualisation a déterminé des formats d'échanges de données numériques le plus ouvert possible d'un point de vue du codage des caractères et de l'organisation des éléments logiques du document. Il s'agissait ici de préserver les données (indépendamment de leur mise en forme) et les informations de description permettant de préparer la production des documents hypertextes en s'appuyant sur la structure logique des documents de départ.

Les environnements matériels et logiciels hétérogènes auxquels j'ai été confronté lors de la mise en place de la chaîne d'hypertextualisation a conduit à adopter la norme ASCII (American Standard Code for Information Interchange) pour coder les caractères des fichiers de données. Tous les caractères étendus (lettres minuscules et majuscules associées à des signes diacritiques) qui ne s'accordaient pas sur une interprétation univoque d'un système d'exploitation à un autre ont ainsi fait l'objet de table de correspondances (cf. Tableau 1) entre les différentes plate-formes logicielles du moment (Microsoft, Novell, Apple, Unix, IBM).

L'organisation des objets documentaires temporaires (pré-réseau hypertexte) devait d'une part être aisément importable au sein des systèmes hypertextes et d'autre part offrir une grande souplesse dans la déclaration des structures logiques des documents hypertextualisés. Un formalisme de balises inspiré du SGML a été retenu (cf. tableau 2)

[CharacterSet]
\mac=MAC
\pc=ASCII
\pca=ASCII
\ansi=ANSI
[CharacterTranscription]
MAC,ASCII,ANSI
181,230,181 # μ
136,133,224 # à
137,131,226 # â
138,132,228 # ä
141,135,231 # ç
142,130,233 # é
143,138,232 # è
144,136,234 # ê
145,137,235 # ë
148,140,238 # î
149,139,239 # ï
153,147,244 # ô
154,148,246 # ö
157,151,249 # ù
158,150,251 # û
159,129,252 # ü
161,248,176 # °
165,45,45 # (puce)
201,32,32 # ...
213,39,146 # '
216,152,255 # ÿ
Tableau 1 : correspondance numérique des "caractères étendus" pour les plate-formes logicielles les plus répandues

```

<Version>1.0</Version>
<LigProjet1>Coordination de l'Isolement Version 1.0</LigProjet1>
<Version>1.0</Version>
<LigProjet1>Coordination de l'Isolement Version 1.0</LigProjet1>
<Titre>l'Enclenchement et au Réenclenchement de ligne à vide</Titre>
<StrDate>11.02.93</StrDate>
<Auteurs>A. SABOT</Auteurs>
<Code_Manuel>C1</Code_Manuel>
<Part_Num>30</Part_Num>
<Doc_Num>01</Doc_Num>
<Document>C1.30.01</Document>
<NbPage>32</NbPage>
<Departement>Service ERMEL/LGE</Departement>
<Indice>C</Indice>
<Diffusion>Participants PPRD</Diffusion>
===== _NOD00000
<Section>1</Section>
<TitreSection>Description du phénomène</TitreSection>
<Ext1_Manuel>C1</Ext1_Manuel>
<Ext1_Part>01</Ext1_Part>
<Ext1_Doc>01</Ext1_Doc>

```

²⁸ les formats tels que HyTime, MHEG, ODA étaient effectivement définis voire normalisés mais aucune boîte à outils logicielle n'était disponible !

<pre> <Ext1_Manuel>C1</Ext1_Manuel> <Ext1_Part>01</Ext1_Part> <Ext1_Doc>01</Ext1_Doc> <Ext2_Manuel>C1</Ext2_Manuel> <Ext2_Part>01</Ext2_Part> <Ext2_Doc>01</Ext2_Doc> <Ext2_Manuel>C1</Ext2_Manuel> <Ext2_Part>01</Ext2_Part> <Ext2_Doc>01</Ext2_Doc> <LigReference>[C1.01.01]</LigReference> ===== _NOD00001 <Section>1.1</Section> <TitreSection>Le phénomène : approche simplifiée</TitreSection> <Ext1_Manuel>C1</Ext1_Manuel> <Ext1_Part>35</Ext1_Part> <Ext1_Doc>01</Ext1_Doc> <Ext1_Manuel>C1</Ext1_Manuel> <Ext1_Part>35</Ext1_Part> <Ext1_Doc>01</Ext1_Doc> <Ext2_Manuel>C1</Ext2_Manuel> <Ext2_Part>35</Ext2_Part> <Ext2_Doc>01</Ext2_Doc> <Ext2_Manuel>C1</Ext2_Manuel> <Ext2_Part>35</Ext2_Part> <Ext2_Doc>01</Ext2_Doc> <LigReference>[C1.35.01]</LigReference> </pre>
<p>Tableau 2 : Un fichier temporaire structuré "SGML" candidat à la production de réseaux hypertextes.</p>

Ces choix de codage et d'organisation se sont avérés dans la réalité parfaitement judicieux puisqu'ils ont permis de s'adapter aux environnements techniques et logiciels distincts à partir desquels des réseaux hypertextes ont été automatiquement générés (HyperCard, ToolBook, DynaText, MetaCard,...).

De l'identification d'une structure physique à un étiquetage logique du contenu de documents électroniques linéaires

Les différentes entités génériques de type chapitre, section ou paragraphe, participant à la structure logique des documents, sont souvent mis en évidence par des attributs de présentation, qui confèrent aux documents une structure physique. La combinaison et la diversité de ces entités logiques peuvent évidemment varier d'un document à l'autre. Une lettre, un journal, un magazine ou un roman fera appel à certains éléments logiques alors que le manuel de référence d'un logiciel en utilisera d'autres.

La structure physique d'un document associe à chacune des entités logiques une représentation graphique et précise leur emplacement physique sur une page. Ceci est rendu possible grâce aux options qu'offrent les logiciels de formatage de texte (et les périphériques d'impression) en terme de polices de caractères, de tailles, de marges, d'interlignage, de cadrage...

Dès l'instant où la structure physique est homogène (une même représentation physique pour un même type d'entité logique), il devient possible de manipuler les entités logiques au travers de leur représentation physique. Ainsi l'utilisation d'un formatage particulier pour un même type d'entité logique permet son identification voire même sa caractérisation

La granularisation logique des documents soumis à l'hypertextualisation, est ainsi déterminée par des "règles de découpage physique".

Le Rich Text Format [Microsoft 90], est un format de document proposé par Microsoft qui permet de représenter dans un même document (au format ASCII) les informations et les caractéristiques typographiques de ces informations (l'information demeure en clair, alors que les attributs de formatage sont décrits par des marqueurs ou balises).

En matière de documents techniques, la structure physique caractérise fortement les entités logiques et il est donc possible de manipuler ces entités logiques seulement par l'image de leur représentation physique. Les documents techniques structurés que j'ai traité correspondent schématiquement à une grammaire "logique" (cf. tableau 1) et vérifient également une grammaire "physique" (cf. tableau 2), conformément à leur modèle de document.

C'est à partir de la structure physique et de la représentation des styles en R.T.F. que le programme détermine les unités d'information. Une correspondance directe existe entre le type de l'entité logique, ses caractéristiques typographiques et son identification R.T.F (cf. tableau 3).

A chaque document est associé un fichier contenant les "règles" de découpage qui sont de fait la description des identifiants R.T.F. permettant de distinguer les différentes entités logiques.

Le découpage en utilisant ces règles permet d'une part de s'adapter à tout type de document structuré à partir d'un modèle (le travail initial consistera à identifier les balises R.T.F. assurant le formatage physique) et d'autre part d'autoriser une certaine flexibilité quant à la détermination du grain d'information (cf. figure 2.2).

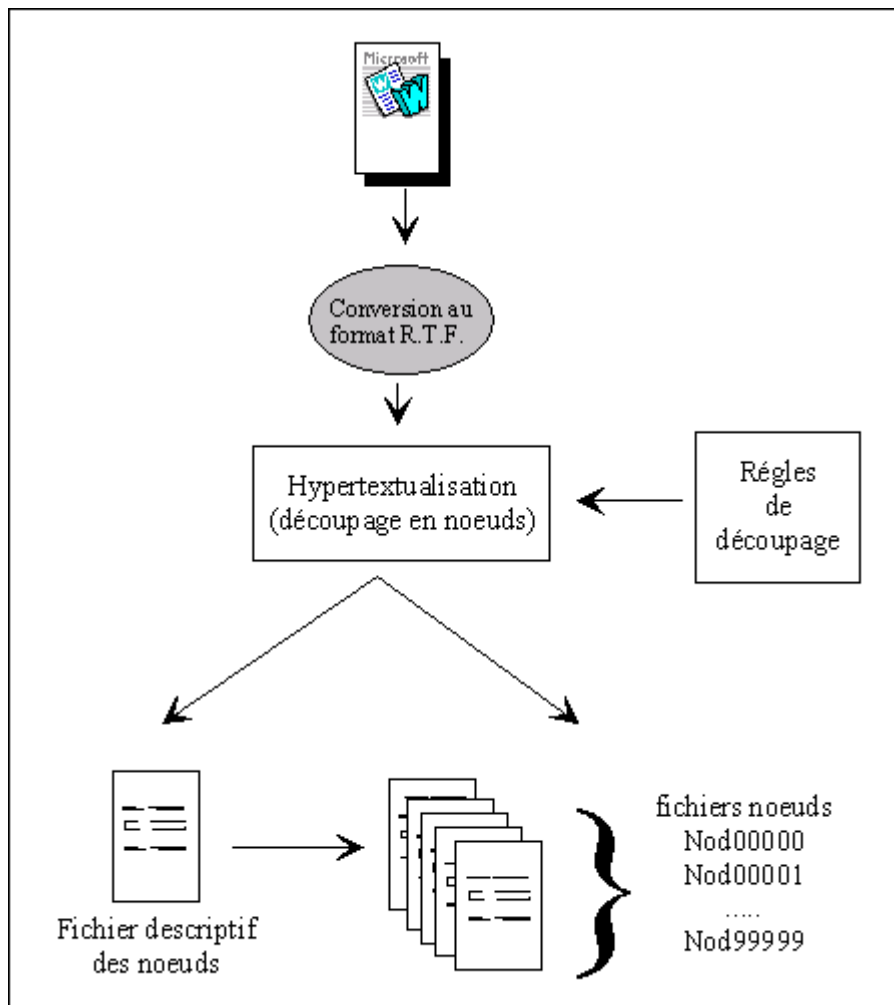


Figure 2.2 : *Le processus d'hypertextualisation*

Ce programme qui utilise en entrée le document au format R.T.F. délivre en fin de processus un ensemble de fichiers (les noeuds) et un descriptif destiné à garder la séquentialité de lecture du document initial (approche auteur).

TechnicalDocument ::= DocTitle Text Sect1
 Sect1 ::= Title Text Sect1 | Title Text Sect2
 Sect2 ::= Title Text Sect1 | Title Text Sect2 | Title Text Sect3
 Sect3 ::= Title Text Sect1 | Title Text Sect2 | Title Text Sect3 | Title Text Sect4
 Sect4 ::= Title Text Sect1 | Title Text Sect2 | Title Text Sect3 | Title Text Sect4
 Title ::= *character string describing the section title*
 Text ::= *character string*
 DocTitle ::= *character string describing the document title*

Tableau 1

TechnicalDocument ::= DocTitle Text Sect1
 Sect1 ::= Sect1Title Text Sect1 | Sect1Title Text Sect2
 Sect2 ::= Sect2Title Text Sect1 | Sect2Title Text Sect2 | Sect2Title Text Sect3
 Sect3 ::= Sect3Title Text Sect1 | Sect3Title Text Sect2 | Sect3Title Text Sect3 Sect3Title Text Sect4
 Sect4 ::= Sect4Title Text Sect1 | Sect4Title Text Sect2 | Sect4Title Text Sect3 Sect4Title Text Sect4
 DocTitle ::= <CharacterSize 18 point><bold><left indent 1.4cm><left aligned><double bottom border>
 Sect1Title ::= <CharacterSize 14pt> <bold> <left indent 1.4cm><first indent -1.4cm> <left aligned><single bottom border>
 Sect2Title ::= <CharacterSize 12pt> <bold> <left indent 1.4cm><first indent -1.4cm> <left aligned>
 Sect3Title ::= <CharacterSize 10pt> <bold> <left indent 1.4cm><first indent -1.4cm> <left aligned>
 Sect4Title ::= <CharacterSize 10pt> <bold> <left indent 2.1cm>< first indent -2.1cm> <left aligned>
 Text ::= <first indent 0.7cm> <justified> <CharacterSize 10pt>

Tableau 2

Element logique	Attributs typographiques	Identification RTF
DocTitle	<CharacterSize 18pt> <bold> <left indent 1.4cm> <left aligned><double bottom border>	\s4
Sect1Title	<CharacterSize 14pt><bold><left indent 1.4cm> <first indent -1.4cm><single bottom border>	\s251
Sect2Title	<CharacterSize 12pt> <bold> <left indent 1.4cm> <first indent -1.4cm> <left aligned>	\s252
Sect3Title	<CharacterSize 10pt> <bold> <left indent 1.4cm> <first indent -1.4cm> <left aligned>	\s253
Sect4Title	<CharacterSize 10pt><bold> <left indent 2.1cm> < first indent -2.1cm><left aligned>	\s254
Text	<CharacterSize 10pt> <first indent 0.7cm><justified>	\s2

Tableau 3

Une fois les nœuds créés, il s'agit de repérer les liens explicites (ou structurels) qui existent entre eux. Ces liens explicites sont les références croisées intégrées dans le texte par l'auteur afin d'orienter les lecteurs vers une autre partie du document en cours de consultation ou vers un autre document susceptible de leur fournir de plus amples informations sur un point particulier. Automatiser l'extraction des références croisées sous-entendait s'adapter à une multiplicité de syntaxes de références croisées proportionnelle à l'ensemble, aux sur-ensembles et aux sous-ensembles attachés au document cible. Le programme d'analyse permettant cette extraction s'appuie sur une description générique de la grammaire des références croisées. Cette description est faite en utilisant la convention Backus Naur Form (BNF) qui permet de décrire facilement les constructions autorisées par les grammaires (en ce qui concerne ces grammaires logiques, on parle aujourd'hui de DTD) dont voici un extrait :

```
Entree[10]= LigReference
LigReference*= CrochetO Blancs ExtDocSect Blancs CrochetF |
                CrochetO Blancs ExtDocSect CrochetF |
                CrochetO ExtDocSect Blancs CrochetF |
                CrochetO ExtDocSect CrochetF |
                CrochetO Blancs ExtDoc Blancs CrochetF |
                CrochetO Blancs ExtDoc CrochetF |
                CrochetO ExtDoc Blancs CrochetF |
                CrochetO ExtDoc CrochetF |
                CrochetO Blancs IntDoc Blancs CrochetF |
                CrochetO Blancs IntDoc CrochetF |
                CrochetO IntDoc Blancs CrochetF |
                CrochetO IntDoc CrochetF
IntDoc= SepDocSection IntSection |
        IntSection
IntSection*= Niveau1 Point Niveau2 Point Niveau3 Point Niveau4 Point |
            Niveau1 Point Niveau2 Point Niveau3 Point Niveau4 |
```

```

Niveau1 Point Niveau2 Point Niveau3 Point |
Niveau1 Point Niveau2 Point Niveau3 |
Niveau1 Point Niveau2 Point |
Niveau1 Point Niveau2 |
Niveau1 Point
ExtDocSect= Document1 SepDocSection Ext1_Section
Document1= Ext1_Manuel Point Ext1_Part Point Ext1_Doc
Ext1_Manuel*= LetMaj Chiffre
Ext1_Part*= Chiffre Chiffre | Chiffre
Ext1_Doc*= Chiffre Chiffre | Chiffre
Ext1_Section*= Niveau1 Point Niveau2 Point Niveau3 Point Niveau4 Point |
Niveau1 Point Niveau2 Point Niveau3 Point Niveau4 |
Niveau1 Point Niveau2 Point Niveau3 Point |
Niveau1
Niveau1= Nombre
Niveau2= Nombre
Niveau3= Nombre
Niveau4= Nombre
ExtDoc= Document2
Document2= Ext2_Manuel Point Ext2_Part Point Ext2_Doc
Ext2_Manuel*= LetMaj Chiffre
Ext2_Part*= Chiffre Chiffre | Chiffre
Ext2_Doc*= Chiffre Chiffre | Chiffre
Annexe= LetMaj Chiffre Chiffre Point |
LetMaj Chiffre Chiffre |
LetMaj Chiffre Point |
LetMaj Chiffre
Blancs= Espace Blancs
Nombre= Chiffre Nombre
Chiffre= "0" | "1" | "2" | "3" | "4" | "5" | "6" | "7" | "8" | "9"
LetMin= "a" | "b" | "c" | "d" | "e" | "f" | "g" | "h" | "i" | "j" | ...
LetMaj= "A" | "B" | "C" | "D" | "E" | "F" | "G" | "H" | "I" | "J" | ...
Lettres= LetMin | LetMaj
Ponctuation= "!" | "" | "&" | "!" | "(" | ")" | "¿" | ...
Romain= "I" | "V" | "X"
Espace= " "
SeqEspace= TAB SeqEspace | Espace SeqEspace
FinLigne= LF FinLigne
Point= "."
Tiret= "-"
CrochetO= "["
CrochetF= "]"
Parag= "§"
SepDocSection= Blancs Tiret Blancs |
Blancs Parag Blancs |
Parag Blancs |
Blancs
CharSpeciaux= "~" | "_" | "ß" | "_ " | "Œ" | "_ " | "_ " | "µ" | "_ " | "_ " | ...

```

Méta-réseau, méta-description, actualisation et navigation

La production automatique de réseaux hypertextualisés a rapidement posé la question de la gestion des mises à jour des réseaux livrés à la consultation des utilisateurs.

En effet, l'hypertextualisation automatique conduit rapidement à une importante production de réseaux hypertextes inter-reliés par la structure et par les références croisées fréquemment présentes dans les documents techniques. Afin de prévenir les incohérences dans le référencement des documents, résultant des modifications apportées par les auteurs, la nécessité de disposer d'une vision fédératrice de l'ensemble des documents hypertextualisés et des liens s'y rapportant, est apparue.

C'est cette vision fédératrice que j'appelle le "méta-réseau" et qui sera transposée dans un formalisme relationnel afin d'être géré par un SGBD relationnelles. La méta-description de ce méta-réseau est composé de l'ensemble des attributs para-textuels déduits du processus de découpage de l'hypertextualisation et mis en évidence par les différentes grammaires (découpage, identification des références croisées).

La réalisation du couplage hypertexte / bases de données a permis de montrer le consensus fonctionnel entre les différents outils de recherche d'informations. On ne peut nier l'existence des différences fonctionnelles entre hypertextes et bases de données : les premiers favorisant un mode de navigation dans l'information basé sur l'association alors que les seconds davantage orientés interrogation, considèrent l'utilisateur comme familiarisé avec un langage de requête portant sur une structuration de l'information qu'il doit nécessairement connaître. Dans les systèmes hypertextes, la plus grande part fonctionnelle est dédiée à la consultation de l'information par le lecteur. Dans les bases de données, la gestion physique des données et de leur intégrité revêt une importance aussi grande que l'aspect interrogation. Ici, l'objectif était de coupler la consultation associative de l'information dans les hypertextes à la gestion de la structure de l'information proposée par les bases de données. Cette gestion de la structure pouvant être mise à la disposition des lecteurs non pas dans un but d'administration, mais pour qu'ils prennent connaissance des liens structurels qui peuvent exister entre les informations.

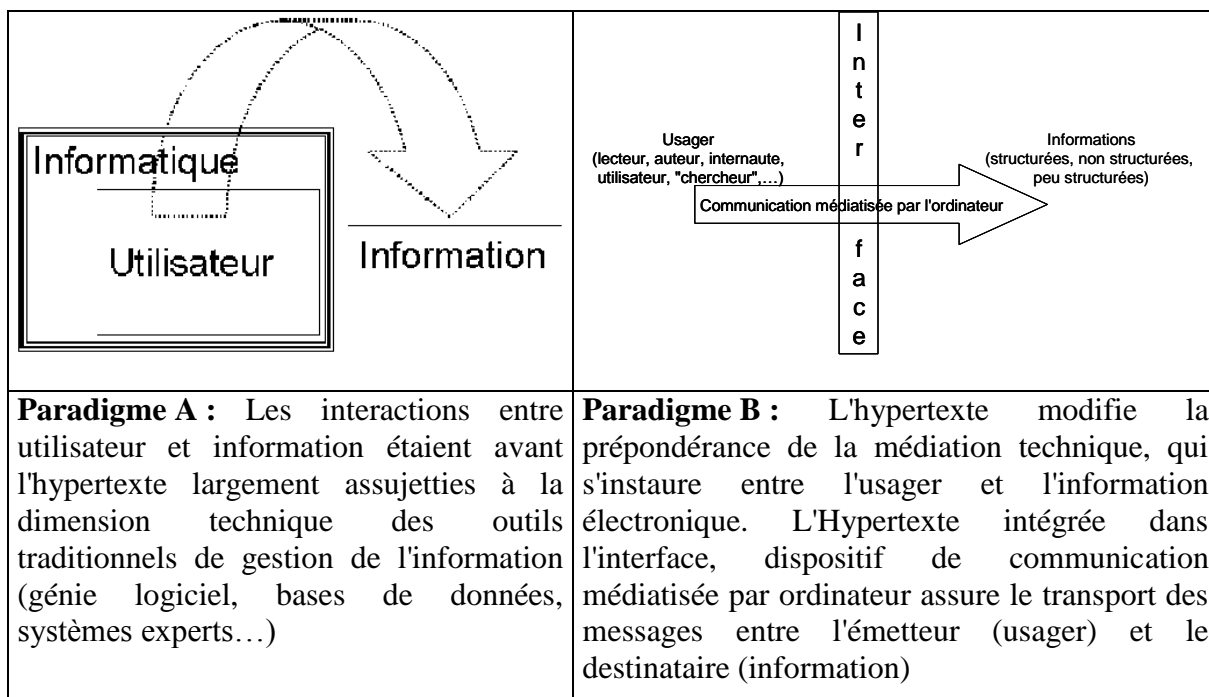
Conclusion

L'approche structurelle que j'ai abordée lors de ma recherche doctorale a alimenté une partie de l'ouvrage collectif de 288 pages publié en 1996 par Jean-Pierre Balpe, Alain Lelu, Imad Saleh et moi-même sur "Les techniques avancées de l'hypertexte". Cet ouvrage se présentait comme un état de l'art sur les questions pratiques de l'hypertexte et de l'hypertextualisation. Ce sont ainsi trois approches qui ont été rassemblées sur ces questions : l'une sur la structure des textes, l'autre inspirée par la linguistique appliquée à la littérature et la dernière par les statistiques et l'analyse de données [Balpe 96].

Les avancées et les orientations technologiques qui ont déterminé en grande partie le développement mondial du Web que nous connaissons aujourd'hui s'expliquent par de nombreux facteurs : technologiques (télécommunications, informatique), économiques, politiques et sociaux. Il est certainement difficile d'apprécier la part effective de chacun d'entre eux ou de mesurer les parts d'influence qu'ils ont pu entretenir mais il est certain que la mutation de nos sociétés nationales en une société de l'information largement interconnectée et mondialisée synthétise ces apports distincts.

Sans être en mesure d'en définir la proportion effective, il est difficile de nier que la part technologique a joué un rôle significatif dans la "libéralisation" de l'accès à l'information. Les travaux significatifs sur l'hypertexte engagés dès les années 80, avaient d'une certaine façon anticipé en pondérant autrement les rôles des différents "acteurs" des systèmes d'informations automatisés et en revisitant l'association ternaire : utilisateur, information et système informatiques.

Avec l'hypertexte, c'est plus qu'une transformation des interactions "homme/machine/information" qui est proposée mais un véritable redéploiement des prépondérances associées aux trois parties majeures. La machine et plus précisément la machine informatique a incontestablement modifié le rapport à l'information et à son traitement. Force est de constater que d'un paradigme A, nous sommes passés à un paradigme B, comme l'illustre le tableau suivant :



La question complexe de l'accessibilité à l'information qu'aborde de front l'Hypertexte à trouvé avec l'avènement du Web une expression technologique qui se caractérise essentiellement par une distinction de plus en plus marquée entre d'une part, les données et les documents et d'autre part, le traitement (de ces données) et la présentation (de ces documents).

Il ne s'agit pas ici de soutenir l'idée que cette distinction n'existait pas dans les systèmes experts et les systèmes de gestion de bases de données ; puisque dans l'un et l'autre cas la formalisation des données (bases de faits et de règles pour les uns, schéma relationnel pour les autres) était guidée par des principes d'indépendance logique / physique. Mais une indépendance logique / physique acquise est vérifiée uniquement au sein même de l'environnement de gestion de données défini. Les données normalisées pour un système de gestion de bases de données relationnelles peuvent ainsi être transférées sans véritable problème vers un autre système de gestion de bases de données relationnelles. Mais ce transfert, de part la constitution même de ces données, directement liée aux contraintes de formalisation n'est plus envisageable dès lors que l'on désire exploiter ces données au sein d'un système expert par exemple.

L'indépendance logique / physique observée intrinsèquement dans ces différents systèmes ne préserve pas pour autant l'indépendance de l'utilisateur qui, confronté à l'un ou à l'autre de ces systèmes de natures différentes, se devra d'en maîtriser les principes fonctionnels pour pouvoir interagir et comprendre les résultats des interrogations qu'il leur adresse. Les exigences structurelles d'une base de données ne sont que la reproduction de celles d'une communauté savante qui produit et utilise ces données, à une période spécifique.

La proposition du "document numérique"²⁹ qui accompagne le déploiement mondial de l'Internet pousse à l'extrême cette notion d'indépendance logique / physique en s'attachant à donner une cohérence "idéale" aux données le composant. Par idéal, je signifie que l'agrégation de données au sein d'un document (au sens où l'entendent les recommandations du W3C au travers des spécifications de la famille XML) n'est pas dirigée vers une quelconque considération applicative ou sémantique des données.

Le document numérique exacerbe les considérations de structuration et de déstructuration de l'information au moyen de données circonstanciellement agrégées au sein de fichiers numériques indépendants des plates-formes informatiques matérielles et logicielles.

On observe aujourd'hui que les technologies du Web ont confirmé l'indépendance du document numérique par rapport à toutes les formes de traitement dont sont porteuses les données constituantes. Les traitements appliqués à ces données peuvent ainsi se recomposer en un ou plusieurs documents susceptibles d'être à nouveau transformés et ce dans un cycle infini.

Ainsi, les procédures de granularisation que j'ai mis en place et qui étaient destinées à distinguer au sein des documents techniques traités les unités de sens révélées par la "Mise en Forme Matérielle" [Virbel 95] se retrouvent aujourd'hui largement exploitées dans les formats de documents de type XML. La granularisation des documents XML, on parlera davantage dans ce cas de fragmentation virtuelle, est apportée par les mécanismes de balisages et d'espaces de nom³⁰ (*namespaces*) inhérents à la structuration XML. Grâce à ces procédés, il est concevable de déclarer

²⁹ Le document numérique dont il est question en regard à l'approche structurelle n'a plus guère de parenté - si ce n'est occasionnel - avec l'étymologie du terme document [Stockinger 2001]

³⁰ <http://www.w3.org/TR/1999/REC-xml-names-19990114/>

explicitement au sein de documents XML, des séquences sémantiquement distinctes³¹, regroupées pour des questions de gestion (mais qui ne remettent pas en cause d'autres conditions d'utilisation).

Séparer le fond de la forme, définir des grammaires de structures logique et physique, distinguer traitement et présentation, appliquer sélectivement des traitements différents déterminés par des conditions d'exploitation spécifiques (transformations, "parsing syntaxique", bases de données) est en l'occurrence la démarche d'une certaine façon "anticipatrice" que j'ai appliquée dans mon travail de thèse et que l'on retrouve aujourd'hui dans XML (DTD et parser), le Cascading StyleSheet (CSS), les Transformations XSL (XSLT) et le "Web dynamique" [Papy 93] [Papy 95] [Saleh 99].

Cette phase très technique dans mon parcours de recherches en Sciences de l'Information et de la Communication s'est imposée pour que l'Hypertexte ne soit pas considérée de façon caricaturalement binaire : d'une part un système technique que l'informatique aurait su matérialiser et d'autre part un principe littéraire qui offre de reconsidérer les positions discursives des auteurs et des lecteurs. Les procédures d'industrialisation que j'ai écrites n'ont jamais été fictives, dans le sens où elles n'ont pas été élaborées en laboratoire, en rupture avec les conditions habituelles d'utilisation des documentations techniques par les acteurs concernés au sein d'EDF (ingénieurs-chercheurs, ingénieurs de production, techniciens,...). Les questions sensibles liées à la capitalisation de connaissances et à la mémoire d'entreprise [Ballay 99] [Jolly 94] étaient constamment présentes voire même déterminantes lorsque l'on connaît aujourd'hui les enjeux du démantèlement des centrales nucléaires arrivées en fin d'activité.

Rendre des documentations techniques accessibles au moyen de l'Hypertexte ne relevait pas d'une simple gageure technique ou d'une problématique d'ingénierie seule mais d'une accessibilité stratégique d'informations dont la valeur intrinsèque dépassait le cadre confidentiel dans lequel se sont développées les techniques

³¹ Dans un formalisme Entité-Association, le prêt d'ouvrages par les usagers d'une bibliothèque est représenté par les entités "Usager" et "Ouvrage" reliées par l'association binaire "Prêt". La transposition en Modèle Physique de Données propres aux SGBD relationnelles donne isomorphiquement trois tables distinctes dotées d'attributs associées à des domaines différents mis en

d'hypertextualisation. Les Dossiers des Systèmes Élémentaires relatifs aux cahiers des charges de construction des centrales nucléaires, les conditions d'isolement des lignes électriques à haute tension et les incidents d'exploitation sur les transformateurs de moyenne et haute tension forment des informations particulièrement sensibles qui, jusqu'à un passé récent, étaient incarcérées au sein de dispositifs informatiques spécifiques (bases de données, systèmes propriétaires, systèmes experts). Les destinataires privilégiés de l'information maîtrisant mal les fonctionnalités internes de manipulation de ces systèmes dédiés (langages de requêtes, procédures, logiques d'implémentation) se trouvaient alors mis à l'écart d'informations qui devaient leur permettre d'élaborer des solutions satisfaisantes en réponse à des situations critiques qui leur étaient soumises.

Dans ce contexte, mes travaux sur l'hypertextualisation ont confirmé la nécessité d'une maîtrise technologique (bases de données, langages de développement du Web, technologies des documents numériques) pour éviter que les Technologies de l'Information et de la Communication ne s'approprient une position dominante qui revient *de jure* à l'information et à ses usagers.

Toutes mes recherches, jusqu'aux plus récentes, se sont attachées à préserver cet équilibre.

relation grâce aux clefs primaires et étrangères déclarées explicitement dans chacune des tables. En XML, il est possible de juxtaposer et de distinguer ces mêmes éléments dans le même fichier.

CHAPITRE 3

POUR L'ELABORATION D'INTERFACES HOMME / CONNAISSANCES

L'Hypertexte confirme bien la possibilité d'offrir un accès associatif à une information électronique en restreignant les manipulations de nature essentiellement technique susceptible de parasiter la lecture. La lecture résultant de l'activation successive de différents nœuds d'informations distribués dans un réseau construit le "parcours de lecture" du lecteur/utilisateur [Balpe 96] [Nielsen 90]. L'individualisation des parcours apparaît dès lors que la structure modulaire d'une œuvre, d'un réseau, d'un fonds documentaire, etc. va permettre à chacun d'aborder les contenus proposés dans l'ordre qui lui convient, indépendamment de celui défini par l'auteur ou le concepteur.

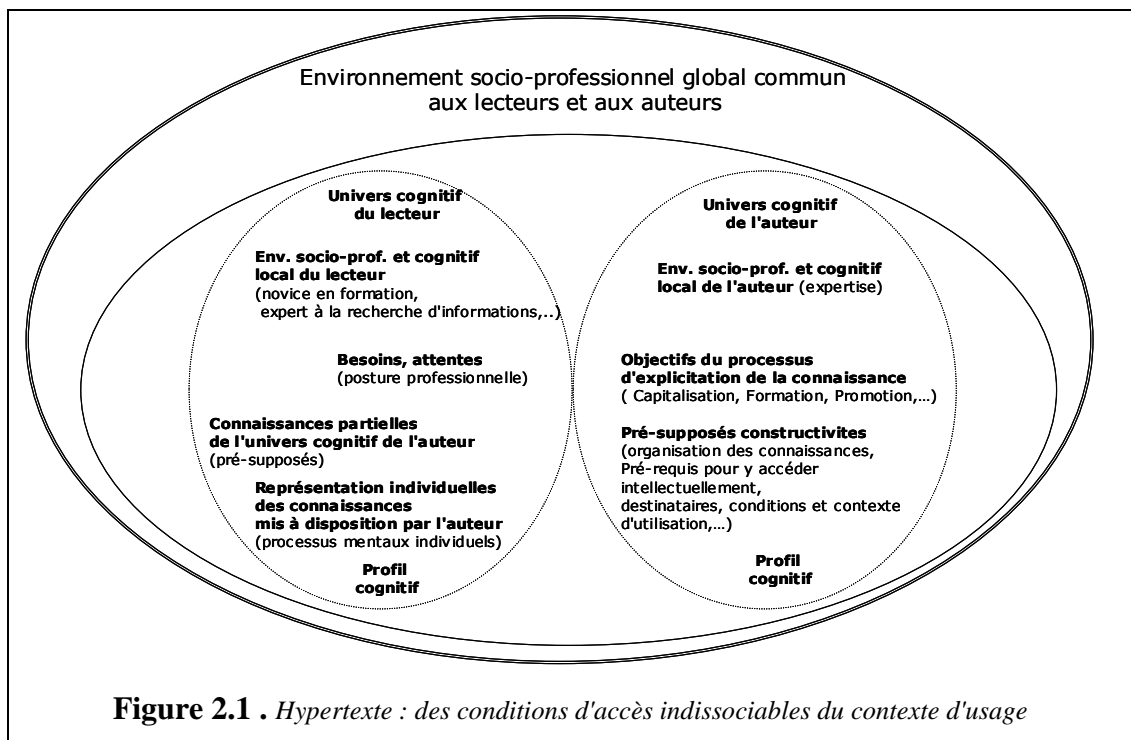
Ce particularisme de la lecture hypertextuelle répond à ce que *"l'homme futur est appelé à s'orienter de façon individuelle dans une information de plus en plus diversifiée, il semble que ce soit dans cette direction que se joue l'avenir non seulement Internet mais aussi de toutes les grandes bases de données que vont devenir les réseaux de télévision câblée"* [Balpe 97].

Mes travaux sur l'hypertextualisation se sont appuyés sur les préliminaires d'organisation des connaissances (acteurs, conception, objectif), qui ont rendu possible l'automatisation de méthodes d'accès de type hypertextuelles en exploitant les indicateurs explicites (références croisées, mise en forme) de documents techniques conçus linéairement (fonctionnement intrinsèque des logiciels de traitement de texte).

Le succès de ces travaux expérimentaux – ils ont bien été suivis par une industrialisation de certaines documentations techniques [Jolly 94][Vuldy 93] – a confirmé d'une part la nécessité d'une organisation intellectuelle préalable des connaissances proposées à la diversité des lectures hypertextuelles et d'autre part

l'omniprésence de la dimension sociale qui circonscrit le cadre des interactions auteur-lecteur.

Il est important de ne pas sous-évaluer les conditions pratiques d'élaboration de réseaux hypertextes en regard à la production d'informations et de connaissances déterminée par des questions d'investissements intellectuel, organisationnel et financier. La production de réseaux hypertextes, manuelle et/ou automatique, n'a jamais conduit à une réalisation "désincarnée" des connaissances recomposées. Nous sommes bien ici dans une modalité d'accès certes différente de celles que l'informatique du moment proposait, mais dans des contextes socio-professionnels précis, où les producteurs de l'information et des connaissances sont identifiés, comme le sont les usagers potentiels (cf. figure 2.1). Le contexte d'usage qui déterminera la mise à la disposition de ces connaissances hypertextualisées est également identifié, fréquemment lié à la capitalisation de connaissances, non pas dans une visée exclusive d'archivage mais plutôt d'exploitation effective (formation des novices, augmentation du champs de connaissances des experts).



En cette période antérieure au World Wide Web, la question des interfaces est largement absente des problématiques liées à la navigation et à l'exploration des réseaux hypertextes³². En effet, les réseaux hypertextes conçus directement par les auteurs des contenus, ou circonstanciellement par des concepteurs travaillant de concert avec les auteurs, étaient avant tout une invitation à une multitude de lectures individualisées d'une œuvre collective élaborée si ce n'est de façon cohérente, a minima concordante. La structure initiale du réseau (qui sous-entend également un point d'entrée particulier) se présente comme la simple transposition de la logique de l'auteur³³ et peut intégrer éventuellement, des suggestions de lecture transversale. Les réseaux hypertextes automatiquement générés ne dérogeaient pas à ce principe. La logique de l'auteur dans le document initial était préservée et pouvait être reprise à tout moment par le lecteur occasionnel. En revanche la table des matières et les références croisées réactives, la recherche en texte intégral permettaient à tout moment d'initier des lectures originales.

Les interfaces des premiers réseaux hypertextes étaient ainsi réduites à leur plus simple expression : visualisation des nœuds, identification des zones sensibles (*hotwords*), retour arrière et historique (cf. figure 2.2, figure 2.3).

L'interaction technique n'était qu'une suite logique aux résultats des processus intellectuels mis en œuvre par le lecteur pour appréhender la logique de l'auteur et pour l'adopter, le cas échéant, comme proposition de lecture. Les situations de désorientation observées dans la lecture hypertextuelle relèvent davantage d'une incompréhension de la logique auteur par le lecteur ; incompréhension qui résulte manifestement de l'inadéquation du profil cognitif du lecteur ou d'une absence de familiarité avec les contenus du réseau consulté. En effet, la lecture hypertextuelle peut s'apparenter à une forme de Recherche d'Information [Saleh 04] [Lefèvre 00] [Papy 95] [Lucarella 90] [Brown 88b] qui dispense le lecteur de l'apprentissage d'un langage d'interrogation spécialisé. Mais le filtre de l'interrogation n'existant pas, le

³² On peut noter par exemple que le système Guide d'OWL [Brown 89a] proposait un dispositif extrêmement simple de "réseau hypertexte-livre électronique" qui se déploie en fonction des opérations d'approfondissement de lecture, à l'image de la carte routière qui se déploie, se replie, se consulte sélectivement ou globalement en fonction du parcours routier qui est envisagé. Le retour arrière dans Guide consistant simplement par le fait de "replier" la partie dépliée.

³³ qui était systématiquement un collectif d'auteurs pour les projets documentaires associés à mes travaux au sein d'EDF.

lecteur inexpérimenté subit les affres de la désorientation au gré des liens qu'il active de façon intempestive.

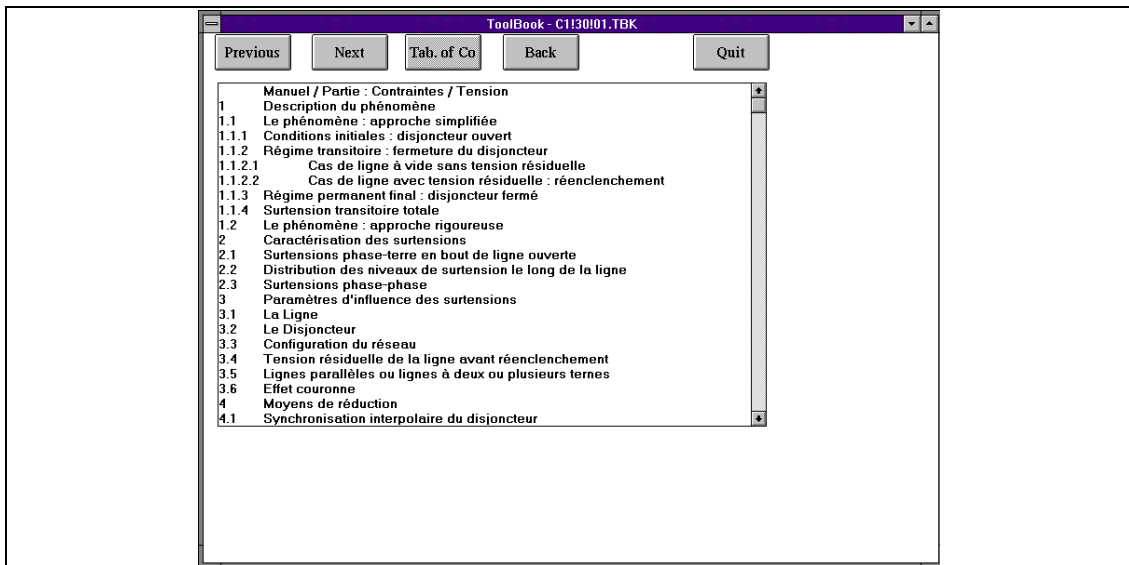


Figure 2.2 : Interface simplifiée pour la navigation au sein d'un document hypertextualisé (les "boutons" situés en haut de fenêtre constituent la navigation globale disponible sur toutes les fenêtres. Le bouton "Next" permet de suivre le cheminement de lecture tel que l'auteur l'a déterminé en élaborant le document initial. Le bouton "Back" est un historique pas à pas qui permet de parcourir à rebours les nœuds consultés. "Tab of Co" représente un raccourci vers la page d'accueil, en l'occurrence la table des matières, où chaque élément qui la constitue est une zone sensible donnant un accès direct au nœud concerné).

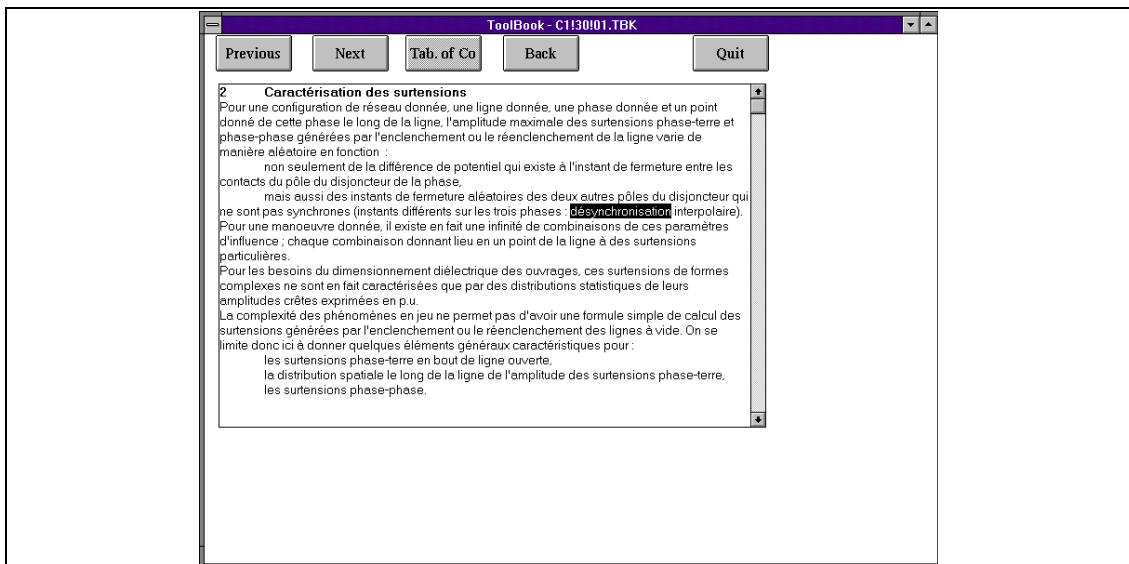


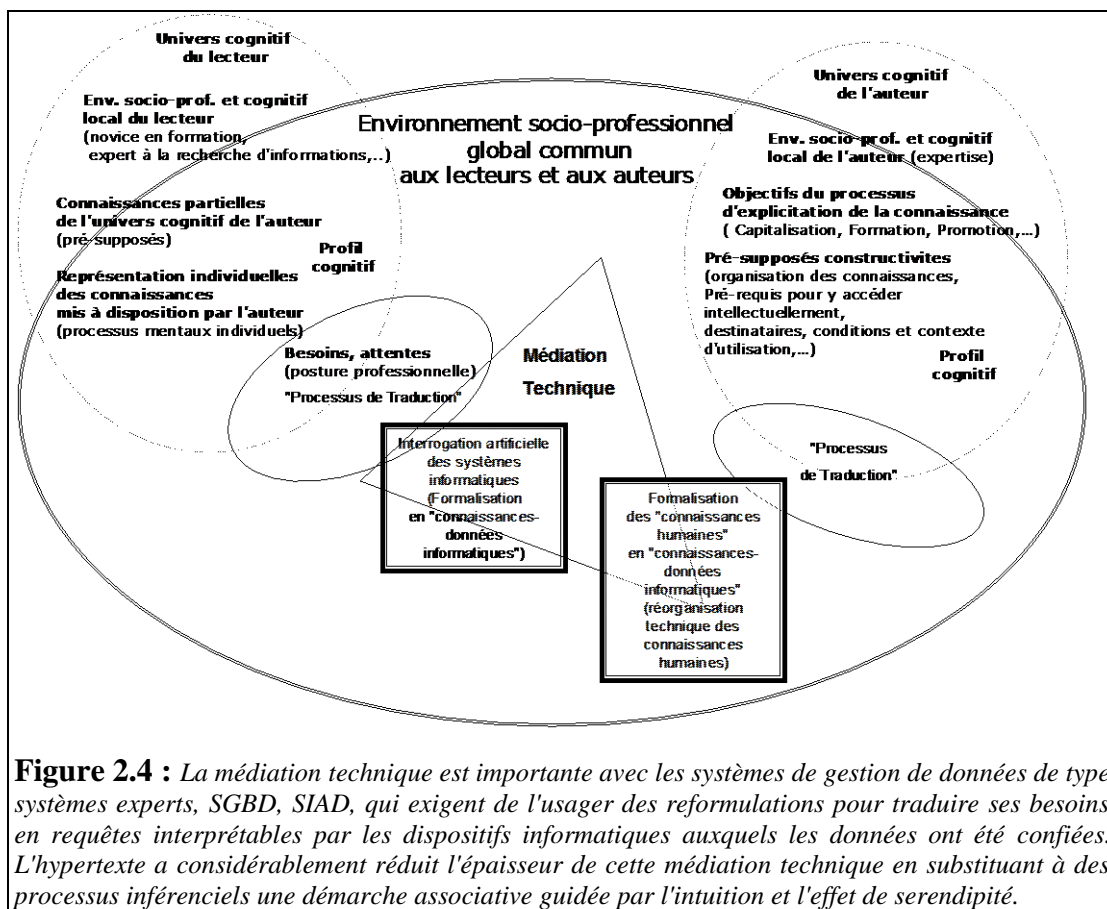
Figure 2.3 : accès à un nœud du réseau et navigation transversale par une recherche en texte intégral (ici le terme désynchronisation)

La finalité des réseaux hypertextes construits automatiquement, relais de l'intérêt manifeste pour un accès associatif aux fragments de textes, ne remet pas en cause pour autant les dispositifs traditionnels d'accès aux données (systèmes d'informations spécifiques, GED, bases de données, systèmes experts, SIAD, ...) qui répondent à des modélisations destinées à des usages connus. De par leur conception, ces dispositifs font subir aux données une reconfiguration structurelle importante³⁴ afin de trouver un optimum dans une performance technique (informatique) qui garantira aux données : stockage pérenne, disponibilité concurrente, confidentialité, sécurité, etc (cf. figure 2.4).

Pour illustration, j'ai consigné en annexe (cf. annexe A) de ce mémoire, un exercice de modélisation³⁵ que je demande aux étudiants inscrits à mon cours "Ingénierie des Connaissances" de la formation de DESS "Conception et Gestion des Réseaux d'Information et de Documentation". Ce cours de 50 heures que je dispense depuis 5 ans, associe cours magistraux et ateliers et propose aux étudiants (praticiens de l'information, futurs responsables de centres de ressources documentaires ou de systèmes d'informations documentaires) une réflexion sur l'ingénierie informatique appliquée à la documentation et des exigences idiosyncrasiques qui y sont liées. Les adaptations des situations informationnelles sont doubles ; elles concernent d'une part les contraintes propres à la modélisation (Entité-Association, UML, KADS,...) et d'autre part les modalités spécifiques de fonctionnement des dispositifs techniques au sein d'une même famille de produits (les SGBD relationnelles, bien qu'ils soient unifiés sous le vocable "relationnelles" montrent des différences techniques conséquentes qui affectent jusqu'au langage d'interrogation SQL).

³⁴ cela est particulièrement vrai avec les systèmes experts, les bases de données relationnelles et objet.

³⁵ Les étudiants sont nécessairement en binôme. Le premier propose la situation à modéliser, le second cherche à en établir une abstraction par l'application d'une démarche de modélisation de type Entité-Association. L'objectif étant de transposer la situation réelle dans un système de gestion de bases de données relationnelles (Access, Oracle, MySQL, SQL Server,...). Il est important que l'étudiant faisant la proposition de la situation à modéliser soit familier de celle-ci. Il ne s'agit pas de témoigner d'un cas d'école mais d'une réalité informationnelle. De la même manière, il est important que le second étudiant soit étranger à la réalité informationnelle d'étude. Le cours leur a donné les principes méthodologiques pour avancer la conceptualisation de l'univers informationnel qu'ils sont invités à modéliser. A l'issue de la modélisation et de la transposition technique dans le SGBDr de leur choix, le premier étudiant élabore cinq questions en rapport avec la réalité de l'univers informationnel de référence que le second étudiant doit formuler dans le langage d'interrogation du SGBD retenu. Il s'agit pour eux d'analyser les résultats des requêtes formulées et d'en vérifier la cohérence.



Accès aux données et profil utilisateur

Dans ces environnements informatiques dédiés, c'est généralement l'accès aux données qui pâtit des conséquences des modélisations incontournables. Ils induisent de véritables traductions des connaissances de l'auteur, qui vont progressivement transformer sa pensée, la "déshumaniser", en modifier les angles de vue, mettre en avant-plan des éléments d'arrière-plan, créer des données inexistantes (clefs primaires des bases de données,...), imposer des procédures d'accès aux données et des transitions artificielles,...

Ces adjuvants techniques vont introduire une médiation technique prépondérante et modifier les interactions privilégiées entre l'auteur et les lecteurs basées sur l'environnement social, culturel et cognitif dans lequel ils évoluent. Dans les réseaux hypertextes directement conçus par l'auteur ou élaborés automatiquement à partir de

documents rédigés par lui, cet environnement partagé va déterminer trois types d'interactions ; l'une à l'origine de l'auteur, les deux autres du lecteur :

- **L'interaction latente.** L'auteur dans l'élaboration des sources qui constitueront les éléments d'informations du réseau hypertexte plaque son propre modèle mental modifié par le contexte socio-professionnel dans lequel il évolue. La configuration du réseau hypertexte, réalisée manuellement ou automatiquement, témoignera de ce processus mental externalisé (cf. figure 1).
- **L'interaction passive.** Pour le lecteur, cette interaction passive constitue "l'effet-miroir" de l'interaction latente introduite par l'auteur. Le lecteur suit la logique auteur pas à pas. Cette interaction éloigne les interfaces de consultation de toute fioriture technique et les transforme en simple "tourne-page" électronique.
- **L'interaction active.** Cette interaction relève principalement de situation liée à la Recherche d'Informations. Nous sommes ici à la convergence des besoins de l'utilisateur, de son profil socio-cognitif, du champs des connaissances du domaine, de la représentation du champs de connaissance de l'auteur (traduit par l'interaction latente), de ses conjectures sur la possibilité de trouver les réponses adéquates, ...

Le Web et sa masse documentaire en expansion constante a introduit une gageure de taille à la proposition d'accessibilité de l'Hypertexte. L'explosion des champs de connaissances, la multiplicité des logiques auteurs, la diversité des contextes socio-culturels, l'indétermination de la qualité et de la pertinence des informations disponibles ont rendu improbable la recherche d'informations menée uniquement de façon associative. L'artefactualisation logicielle s'est imposée (analyse sémantique, statistique des fonds collectés, filtrage, résumés [Isko 01] [Pincemin 99]) et par voie de conséquence "l'épaisseur" de la médiation technique. L'univers socio-cognitif global des usagers et des producteurs d'information, dans ce qu'il a de commun, devient de plus en plus transparent alors que les univers socio-cognitifs locaux se retrouvent séparés par les algorithmes complexes des logiciels de traitement de l'information. Au nom de "qui peut le plus, peut le moins", les dispositifs techniques

de traitement et de recherche d'informations [Leloup 98] utilisent le territoire numérique du Web [Foenix-Riou 01] [Andrieu 01] [Andrieu 00] comme "vitrine technologique" pour investir des secteurs où le traitement de l'information reste avant tout une question de contexte et de circonstances [Simonnot 02] [Rouet 01] [Bourdoncle 00] [Mattelard 99] [Ballay 99].

Avec cette sur-production de dispositifs techniques de traitement et de recherches d'information, on comprend que le champs de l'informatique relatif aux IHM [Coutaz 90] et aux IHE [Aristote 02] continuent à maintenir une position dominante dans une situation qui concerne pourtant davantage la dimension sociale, culturelle et cognitive d'une communication entre un usager et des fonds d'informations. L'idéologie du ratissage exhaustif de toutes informations numériques par l'omnipotence technologique des moteurs de recherches et des arpenteurs place aujourd'hui l'usager en interaction avec un fonds d'information dont il a de plus en plus de difficulté à appréhender les proportions³⁶, le degré d'organisation, de mise à jour, de pertinence et d'adéquation *a priori* avec des besoins qu'il n'est pas toujours en mesure d'exprimer avec précision et avec la rigueur que les SRI exigent par ailleurs³⁷.

Les travaux plus récents que je conduis au sein de la bibliothèque universitaire de Paris 8 (cf. chapitre 4) confirment non seulement la nécessité de préserver le rapprochement entre les univers sociaux, culturels et informationnels des usagers et des producteurs d'informations mais plus encore de les rendre visibles, l'un pour l'autre. R. Barthes soulignait dans son introduction à l'analyse structurale des récits qu'en ce qui concerne les niveaux du sens "*...une phrase, on le sait, peut-être décrite linguistiquement, à plusieurs niveaux (phonétique, phonologique, grammatical, contextuel); ces niveaux sont dans un rapport hiérarchique, car, si chacun a ses propres unités et ses propres corrélations, obligeant pour chacun d'eux à une description indépendante, aucun niveau ne peut à lui seul produire du sens : toute*

³⁶ l'unité de mesure (octet, Ko, Mo, Go) permettant d'exprimer les tailles des fichiers numériques est peu significative. En effet, les images en fonction de leur format d'enregistrement peuvent varier du simple au centuple (JPEG, TIFF), les textes publiés sur le Web sont largement alourdis par les balises HTML qui gonflent artificiellement la taille de la page. Les fichiers vidéo et audio en fonction des types de compression et de format ne renseignent pas sur les durées effectives et la qualité des enregistrements.

³⁷ demande mal formulée = bruit ou silence !

unité qui appartient à un certain niveau ne prend de sens que si elle peut s'intégrer dans un niveau supérieur..." [Barthes 66].

L'élément supérieur en Recherche d'Informations concerne les conditions de création, d'organisation et de gestion des fonds documentaires qu'il convient de rendre, a minima accessible et au mieux visible, aux usagers afin que leurs recherches prennent sens.

C'est cette démarche que j'ai initiée dans l'élaboration du système d'interface du moteur de recherches NeuroWeb développé par l'équipe Hypertexte Dynamique du Laboratoire Paragraphe [Lelu 01] [Lelu 00] [Lelu 99]. Outre le développement de l'interface, j'ai réalisé l'assemblage des différents modules logiciels du moteur, qui sont assujettis par l'interface.

NeuroWeb : vers une interface Homme-Connaissances

NeuroWeb est avant tout un moteur de recherches³⁸ développé à la suite d'un appel d'offres lancé par le Ministère de l'Éducation Nationale et l'Agence Nationale pour la Recherche Technologique, qui intègre au moyen de requêtes fines et des cartographies à la demande, deux approches complémentaires ; la première exploitant les méthodes d'exploration de corpus à partir de cartographies textuelles utilisant les N-grammes, la seconde utilisant la lemmatisation de textes et les réseaux sémantiques.

Le moteur de recherche alimenté par un robot de téléchargement de sites développé spécifiquement [Papy 03a], est doté d'une série de modules activés circonstanciellement en fonction des opérations demandées par l'utilisateur à partir du navigateur : texte intégral, approximation lexicale, interrogation sur lemme, expansion (à un lemme correspondent N lemmes, à un document correspondent N documents), N-grammes dotés de fonctions de proximité lexicale, cartographie dynamique.

NeuroWeb procède à partir de 2 ensembles de données interdépendants : les termes et les documents. Un terme peut être présent dans plusieurs documents (et au moins

³⁸ <http://neuroweb.univ-paris8.fr>

dans un document), et chaque document contient au moins un terme. Les termes dont il est question ici sont issus du traitement linguistique du logiciel Nomino³⁹. Ce logiciel de TALN permet d'extraire des "Unités Complexes Nominales", expressions terminologiques candidates lemmatisées (éventuellement constituées de plusieurs syntagmes) considérées comme significatives par le logiciel.

Tous les traitements statistiques de NeuroWeb portent sur les calculs de fréquence (*TfIdf* aménagé [Lelu 98] [Lelu 93]) des "Unités Complexes Nominales" issues des documents retenus dans le fonds documentaire.

Le principe interactif de NeuroWeb repose sur un système de panier-requête que l'utilisateur peut remplir et vider à sa convenance en fonction des résultats obtenus. Le panier-requête peut être rempli avec des termes (expression terminologique) et des documents. En fonction des choix de l'utilisateur ("Vers docs", "Vers Mots"), l'activation de la recherche va déterminer une liste triée des expressions candidates les plus proches du contenu du panier-requête.

Au-delà de l'originalité de ce moteur de recherches dont le bilan est consultable en [Lelu 01], c'est le mode d'accès et de recherche d'information que ce moteur met particulièrement en évidence et que les spécialistes de la recherche d'informations appellent le *relevance feedback* [Lefevre 01] [Cattenat 92].

Feedback sur le profil cognitif de l'utilisateur

Avec NeuroWeb, les interventions rétroactives de l'utilisateur demeurent indispensables ; d'une part parce qu'elles permettent à celui-ci, d'appréhender le vocabulaire d'indexation sur lequel aboutissent des requêtes formulées avec un vocabulaire inadapté, d'autre part, elles lui permettent de construire une requête qui fait appel à davantage de termes voire même à un ensemble de termes qu'il n'est pas en mesure d'exprimer (parce que trop loin de ses connaissances initiales) mais dont il a l'intuition et que lui suggère la thématique d'un document par exemple.

Par l'intermédiaire du "panier-requête", c'est cette rétroaction qui est mise en œuvre et en évidence dans le prototype NeuroWeb. Ce "panier-requête", un "caddie de

³⁹ <http://www.ling.uqam.ca/nomino/synopsis.htm>

recherches" cumule les expressions dont l'utilisateur a la certitude (et qui d'une certaine mesure atteste de son niveau de connaissances) sur les expressions qui émanent des index construits dynamiquement par le système (cf. figures 2.5, 2.6, 2.7, 2.8, 2.9 et 2.10).

Ce mode de rétroaction qui s'appuie sur la performance statistique du système apporte un éclairage original sur la façon dont se construisent de nouvelles connaissances dès que l'utilisateur aborde une recherche d'information sur des ensembles documentaires pour lesquels il ne dispose pas des acquis lui permettant de les appréhender. Car la *relevance feedback* va effectivement permettre à l'utilisateur de se familiariser avec le vocabulaire d'indexation des documents interrogés et d'approcher progressivement l'univers cognitif sous-jacent au fonds documentaire.

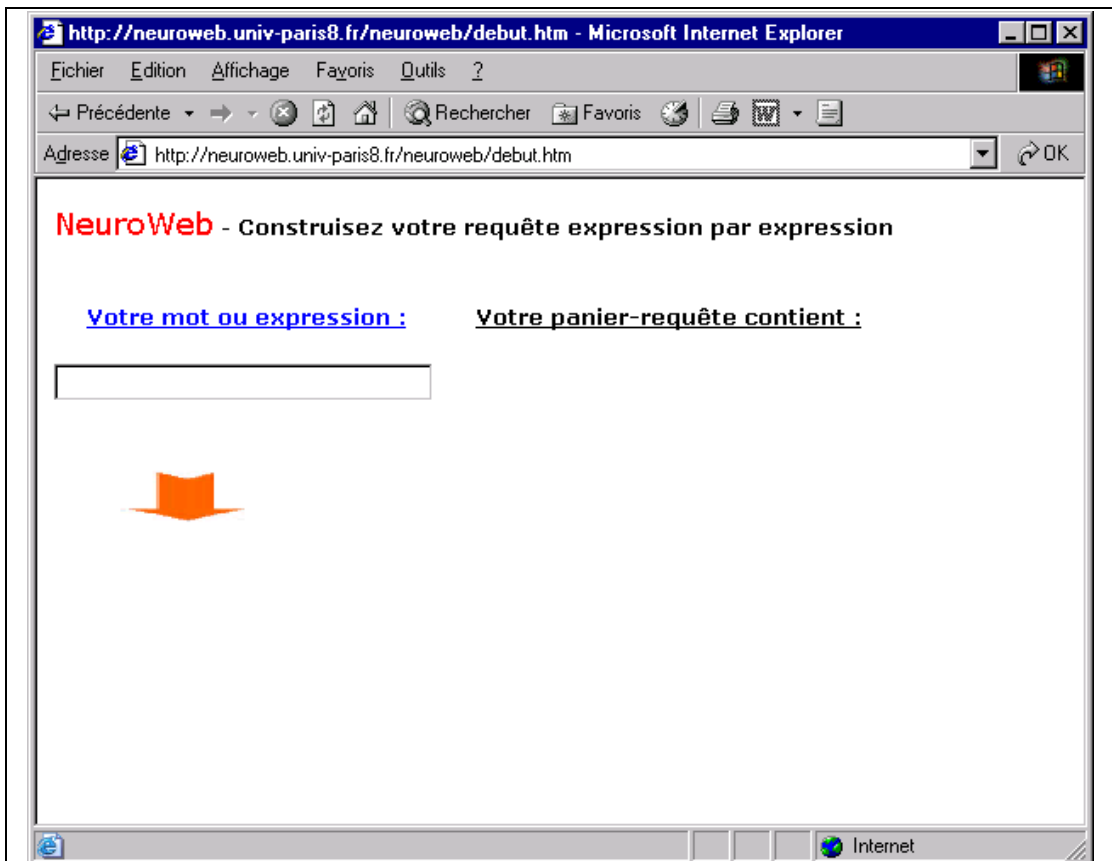


Figure 2.5 . Interface NeuroWeb: écran initial. L'utilisateur est invité à saisir un terme pour démarrer le processus de remplissage du panier-requête (invisible ici). L'interface va se modifier dynamiquement à l'issue de cette première requête pour faire apparaître le panier-requête. La zone de résultats va apparaître sous la forme d'une liste de termes ou de documents triés sur des critères statistiques. Les termes ou les documents estimés les plus en corrélation avec le contenu du panier-requête apparaissent en début de liste.

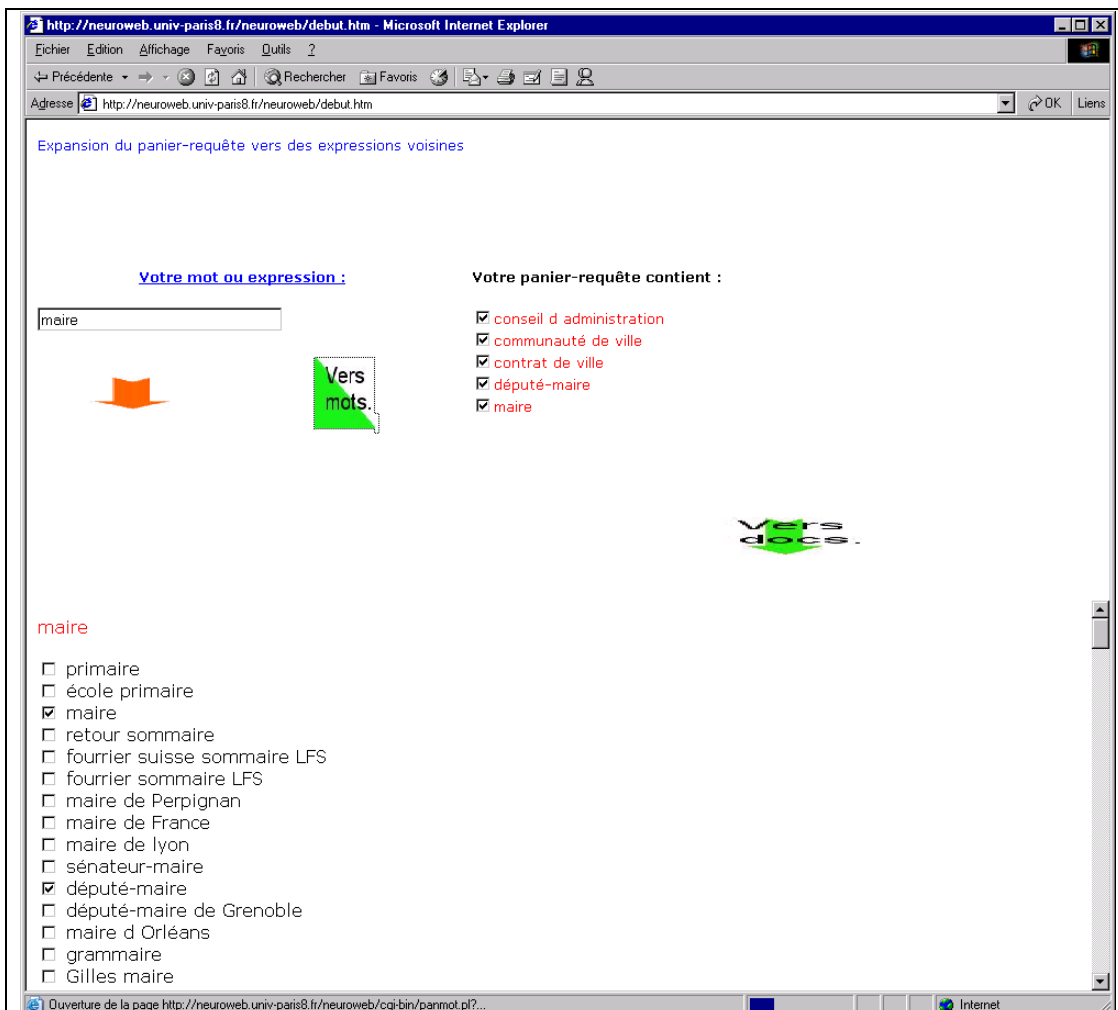


Figure 2.6 - NeuroWeb : recherche des termes associés au mot "maire" (terme saisi par l'utilisateur). A l'issue de la liste proposée par NeuroWeb, les termes "maire" et "député-maire" sont ajoutés au panier-requête qui contenait déjà les termes "conseil d'administration" (obtenu à partir de la saisie du mot "administration"), "communauté de ville" et "contrat de ville" (obtenus par la saisie du mot "ville"). L'utilisateur peut générer une nouvelle liste de termes en sélectionnant le bouton "Vers Mots" ou une nouvelle liste de documents en sélectionnant le bouton "Vers Docs".

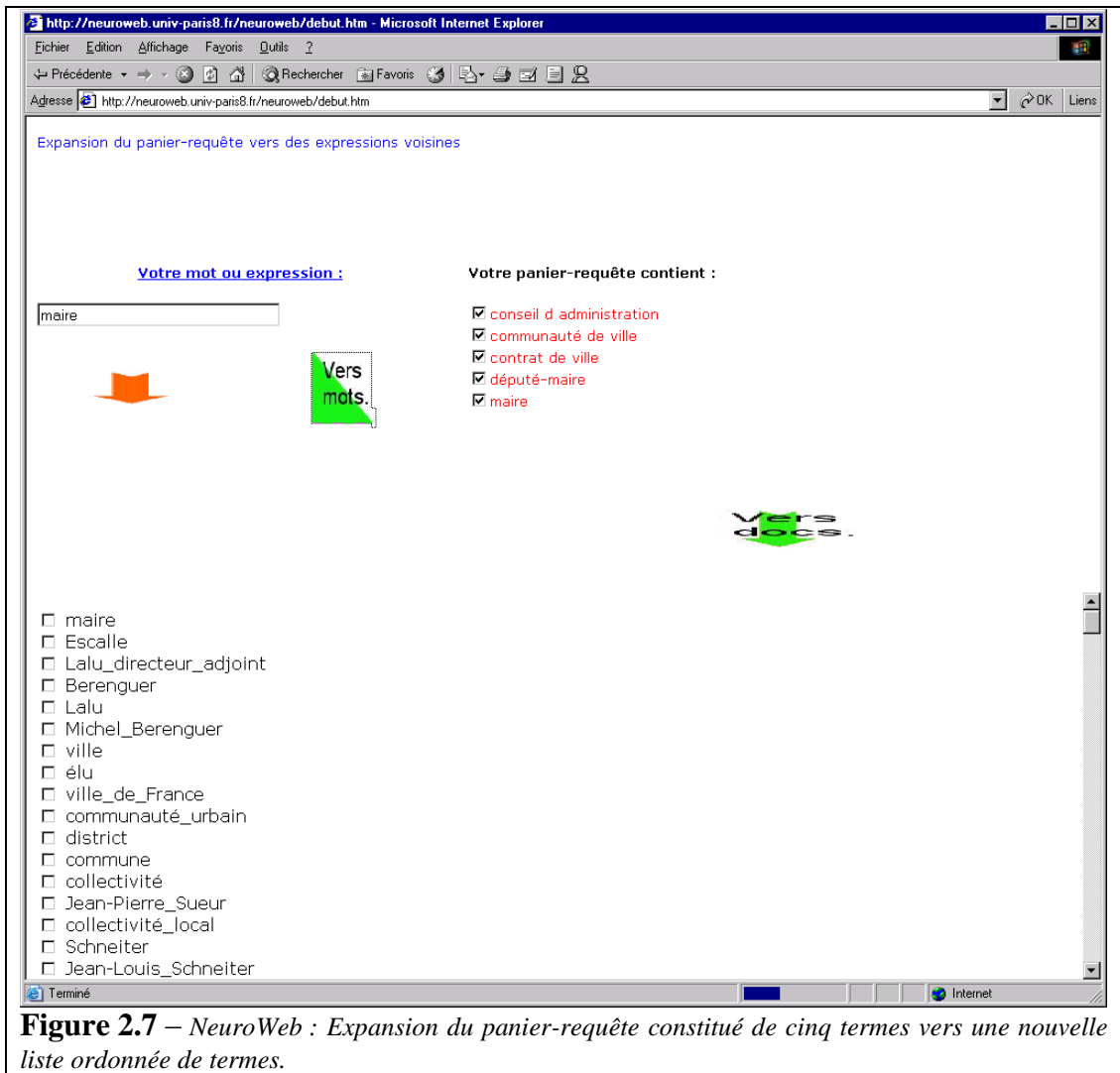




Figure 2.7 – NeuroWeb : Expansion du panier-requête constitué de cinq termes vers une nouvelle liste ordonnée de termes.

Expansion du panier-requête vers les pages Web


Votre mot ou expression :

maire

  **Vers mots.**

Votre panier-requête contient :

- conseil d administration
- communauté de ville
- contrat de ville
- député-maire
- maire

 **Vers docs.**

- [1007 - Numéro 346 - 36251](#) **Texte**
- [1049 - Grandes Villes Hebdo 265 - numéro 265](#) **Texte**
- [1036 - Grandes Villes Hebdo 278 - numéro 278](#) **Texte**
- [1010 - Numéro spécial 343 - 36249](#) **Texte**
- [1005 - Numéro 349 - 36277](#) **Texte**
- [1063 - Grandes Villes Hebdo - numéro 251](#) **Texte**
- [1038 - Grandes Villes Hebdo 276 - numéro 276](#) **Texte**
- [1013 - Numéro 341 - 36235](#) **Texte**
- [1011 - Numéro 343 - 36249](#) **Texte**
- [4403 - Qui fait quoi ? Le Conseil d'administration - Le](#) **Texte**
- [4402 - Les statuts de l'association - But et](#) **Texte**
- [1031 - Grandes Villes Hebdo 283 - numéro 283](#) **Texte**
- [1033 - Grandes Villes Hebdo 281 - numéro 281](#) **Texte**
- [1040 - Grandes Villes Hebdo 274 - numéro 274](#) **Texte**
- [1021 - Numéro 334 - 36184](#) **Texte**
- [1032 - Grandes Villes Hebdo 282 - numéro 282](#) **Texte**
- [1015 - Numéro 339 - 36221](#) **Texte**


Terminé

Figure 2.8 - NeuroWeb : Expansion du panier-requête constitué de termes (même panier-requête que figure 2.7) vers une liste ordonnée de documents

Expansion du panier-requête vers les pages Web


Votre mot ou expression :

maire



Votre panier-requête contient :

- conseil d administration
- communauté de ville
- contrat de ville
- député-maire
- maire
- 1063 - Grandes Villes Hebdo - numéro 251
- 4403 - Qui fait quoi ? Le Conseil d'administration - Le



- 1010 - Numéro spécial 343 - 36249 [Texte](#)
- 1005 - Numéro 349 - 36277 [Texte](#)
- 1063 - Grandes Villes Hebdo - numéro 251 [Texte](#)
- 1038 - Grandes Villes Hebdo 276 - numéro 276 [Texte](#)
- 1013 - Numéro 341 - 36235 [Texte](#)
- 1011 - Numéro 343 - 36249 [Texte](#)
- 4403 - Qui fait quoi ? Le Conseil d'administration - Le [Texte](#)
- 4402 - Les statuts de l'association - But et [Texte](#)
- 1031 - Grandes Villes Hebdo 283 - numéro 283 [Texte](#)
- 1033 - Grandes Villes Hebdo 281 - numéro 281 [Texte](#)
- 1040 - Grandes Villes Hebdo 274 - numéro 274 [Texte](#)
- 1021 - Numéro 334 - 36184 [Texte](#)
- 1032 - Grandes Villes Hebdo 282 - numéro 282 [Texte](#)
- 1015 - Numéro 339 - 36221 [Texte](#)
- 1041 - Grandes Villes Hebdo 273 - numéro 273 [Texte](#)
- 1023 - Numéro 332 - 36172 [Texte](#)
- 1071 - Numéro 322 - 36095 [Texte](#)
- 4407 - Qui fait quoi ? Le Bureau - Le [Texte](#)

Terminé

Figure 2.9 - NeuroWeb : A partir de la liste des documents proposés par le moteur, les références de deux documents sont ajoutées au panier requête constitué jusqu'à présent uniquement de termes. (la sélection au moyen de cases à cocher dans la liste des documents, ici en bas à gauche et en bleue, fait apparaître les références sélectionnées dans le panier-requête)

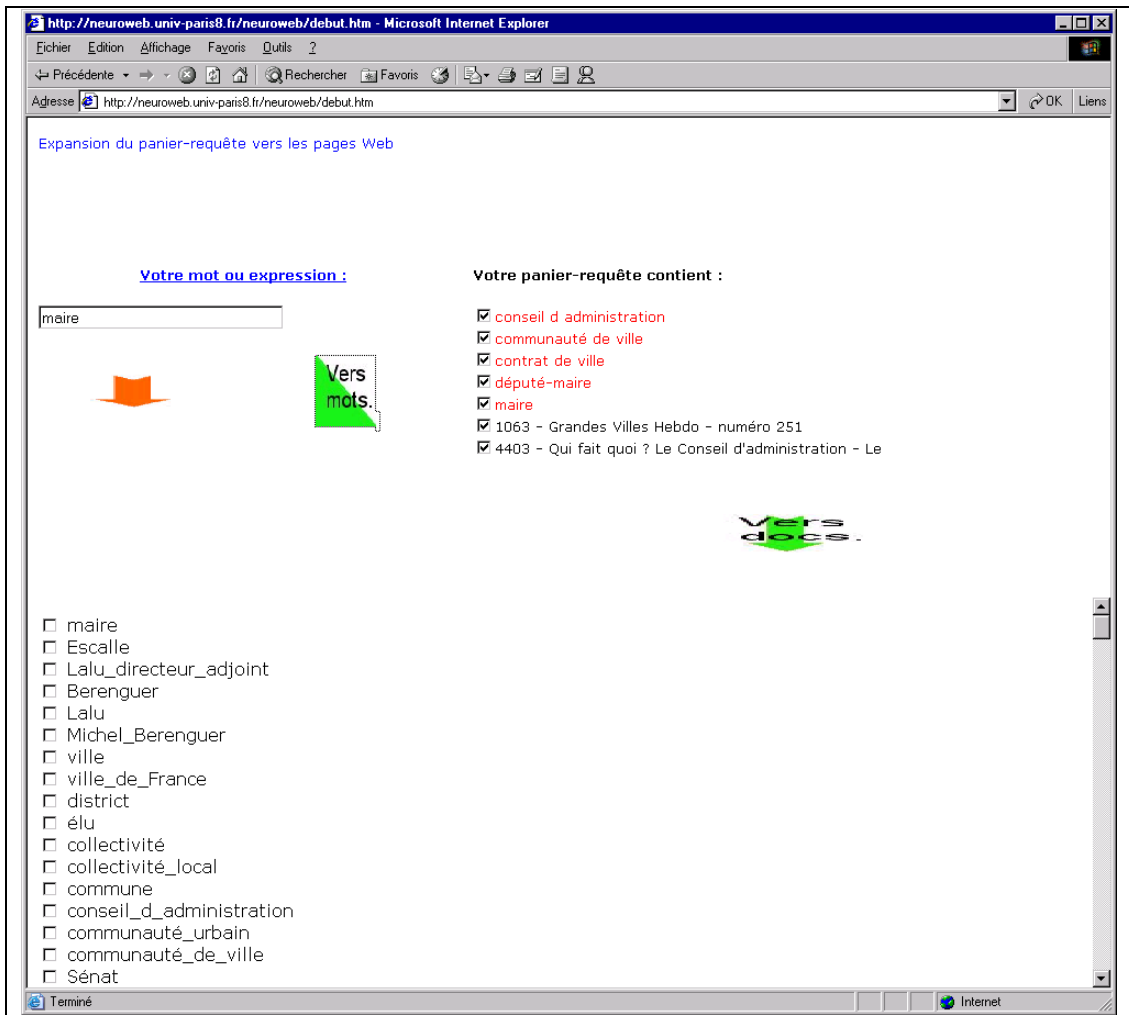


Figure 2.10 - - NeuroWeb : Expansion du panier-requête constitué de termes et de documents, vers une liste ordonnée de termes. Par rapport aux résultats de la figure 2.7 on voit que la liste est modifiée à partir du 8^{ème} terme (ici "ville_de_France", dans la figure 2.7, c'est le terme "élu" qui apparaît).

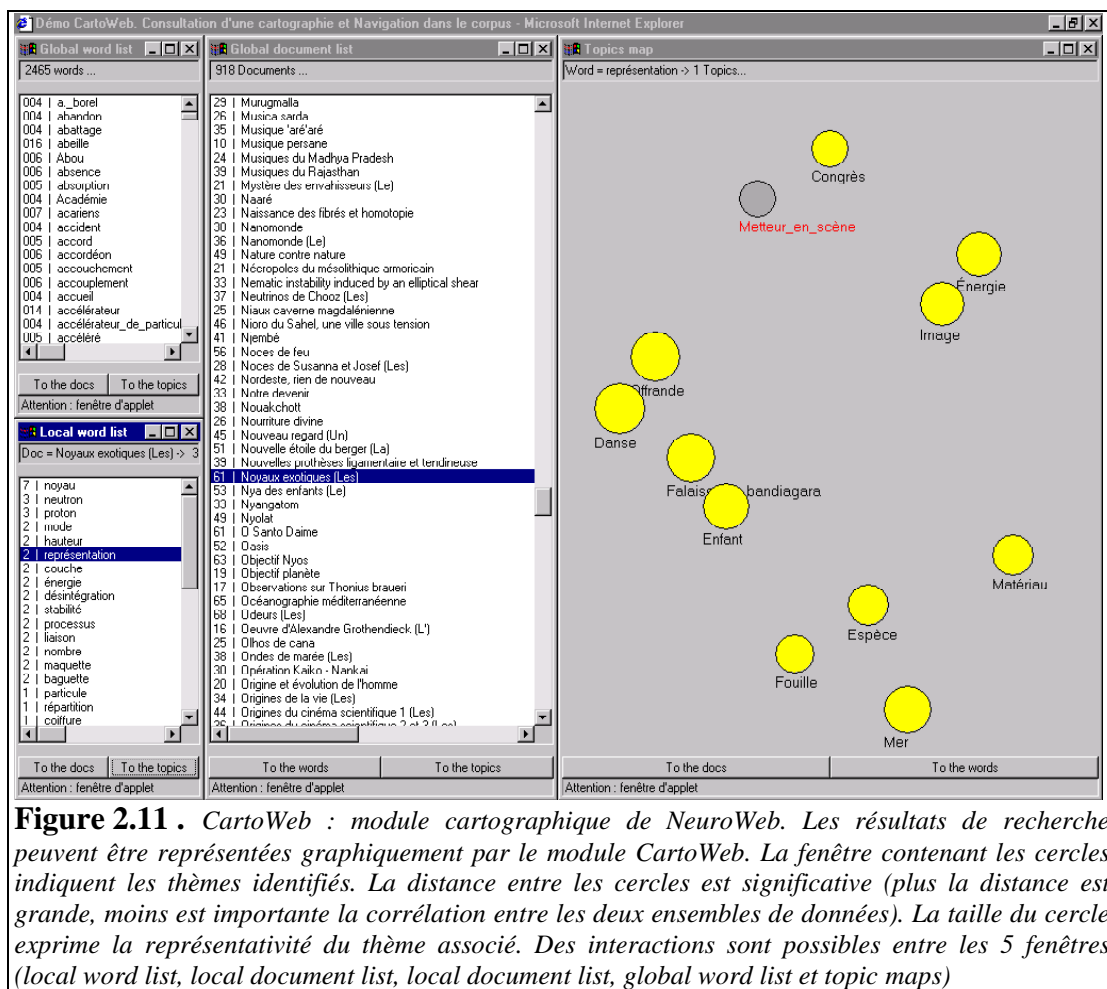


Figure 2.11 . CartoWeb : module cartographique de NeuroWeb. Les résultats de recherche peuvent être représentés graphiquement par le module CartoWeb. La fenêtre contenant les cercles indiquent les thèmes identifiés. La distance entre les cercles est significative (plus la distance est grande, moins est importante la corrélation entre les deux ensembles de données). La taille du cercle exprime la représentativité du thème associé. Des interactions sont possibles entre les 5 fenêtres (local word list, local document list, local document list, global word list et topic maps)

Conclusion

Neuroweb testé à l'Université Paris VIII pendant deux stages intensifs de 8 jours par des étudiants de trois diplômes de 3^{ème} cycle en Sciences de l'Information et de la Communication (DESS hypermédia, DEA Enjeux Sociaux et Technologies de la Communication, DESS Conception et Gestion des réseaux d'Information et de Documentation) [Lelu 01] a confirmé l'intérêt du dispositif d'expansion de requêtes vers les mots et vers les documents.

Lors de ces tests, NeuroWeb a fonctionné sur deux corpus de textes ; l'un provenant d'une extraction de 8 000 pages issues du Web et pré-sélectionnées au moyen d'un dispositif de classification automatique de documents [Papy 03a] [Lelu 98], et l'autre

composé des 70 articles de la revue électronique Solaris⁴⁰ (numéros 1 à 7). Le second corpus certes moins volumineux que le premier présentait l'avantage d'être thématiquement plus homogène (thèmes relatifs à l'Information, la Documentation, la Communication). C'est ce second corpus qui a été préférentiellement utilisé par les étudiants des différents 3^{ème} cycle en raison d'une part de leur affinité disciplinaire et d'autre part d'une sensibilité, liée certainement à leurs études en cours, à certaines thématiques développées dans les numéros de la revue.

Au delà de la validation des principes interactifs de l'interface, cette étude a mis en évidence l'autonomie technique extrêmement rapide des usagers. Il est certain que la simplicité fonctionnelle de l'interface (adaptation dynamique, absence de booléen, de syntaxe d'inclusion ou d'exclusion, remplissage/déremplissage du panier-requête) et l'explicitation des modalités d'indexation des documents (UCN, Nomino) ont contribué à la prise en main rapide du dispositif de recherche d'information.

Le module d'expansion sémantique d'un ou de plusieurs mots vers les mots du vocabulaire a été reconnu comme riche, pertinent, et parfois amusant ou poétique, du fait des rapprochements opérés. Ce module "donnait des idées" d'expressions lexicales pour poursuivre la recherche, et l'activation en chaîne de cette fonction d'expansion de requête représentait à elle seule un plaisir en soi, celui de dériver au hasard des associations de termes sémantiquement proches. Pourtant, la difficulté de la reformulation en raison d'une méconnaissance du vocabulaire d'indexation est le point d'achoppement fréquemment constaté dans l'interrogation des SRI⁴¹.

Ce test en particulier confirme l'intérêt de l'hypothèse (à défaut de l'avoir vérifiée) que j'é mets concernant l'univers socio-cognitif de l'auteur qu'il convient de rendre visible à l'utilisateur. La cohésion du fonds documentaire (la revue électronique Solaris) a sans nul doute joué un rôle important dans l'attraction ludique que représentait le module d'expansion sémantique de mots vers les mots.

⁴⁰ La revue Solaris qui vise un certain nombre d'interrogations dans le domaine de la production-circulation des informations, des savoirs et ce dans le contexte général du processus de numérisation du signe, du développement des réseaux électroniques.
<http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/index.html>

⁴¹ "...il ne va pas facilement de soi de faire varier les mots, pour exprimer de toutes les façons possibles les concepts que l'on évoque..." [Cattenat 92]

La cohésion de ce fonds documentaire a permis aux usagers de rapprocher leur propre profil socio-cognitif de celui de l'auteur (en l'occurrence du collectif d'auteurs tourné vers des problématiques convergentes). Et sur ce point, rien ne permet de contredire que les interrogations successives opérées par l'utilisateur n'avaient pour autre objectif que d'appréhender le contexte global de la revue, d'en circonscrire l'étendue conceptuelle afin de déterminer plus spécifiquement des stratégies d'interrogation personnalisées.

Il est utile de préciser que l'effet parasite de l'augmentation artificielle du niveau de l'interrogation (au fil des interrogations, les mots appelant d'autres mots, NeuroWeb finit par proposer les expressions lexicales du domaine, du moins celles utilisées par les auteurs des articles) a eu un impact positif sur les usagers. En effet, en pouvant rapidement identifier les termes – dont ils n'appréhendaient pas nécessairement la portée sémantique – leur permettant d'aboutir aux documents les plus pertinents, les usagers ne sont pas tombés dans la situation absurde de se contenter de documents hors de leur niveau de connaissances du moment.

Cet impact positif s'est manifesté concrètement par les demandes répétées d'une sauvegarde des réponses successives de NeuroWeb aux sollicitations d'interrogation.

Les recherches effectuées par ces usagers-étudiants englobaient un processus complet allant des interrogations multiples à la consultation d'un ou de plusieurs documents. Le parcours de recherches, inédit pour chacun d'entre eux, devenait alors une construction intellectuelle où les différentes étapes apparaissaient aussi importante que l'issue de la recherche (les documents consultés).

La possibilité de préserver ce parcours, ne consiste pas en un simple marquage des étapes de la recherche (comme le font les historiques) mais revêt bien la capacité de construire un document qui résume le cheminement en lui associant les éléments de la recherche (mots utilisés, panier-requête, réponses du système).

Aujourd'hui, les formats de document liés au XML offrent une adaptabilité des contenus qui rend possible la conception de documents numériques permettant de sauvegarder des informations résultant de processus inédits et individuels. La recherche d'informations relève de tels processus, la transformation des profils cognitifs tout autant !

L'expérience NeuroWeb a confirmé la nécessité d'impliquer les usagers dans des univers informationnels cohérents (de cette cohérence que j'ai pu appréhender avec les documents que j'ai traité lors de mes travaux sur l'hypertextualisation), susceptibles de fournir les réponses aux besoins déterminés par l'environnement social et culturel local à l'utilisateur.

Mais la nécessité de donner aux usagers la perception de leur propre univers informationnel grâce aux interactions qu'ils conduisent avec l'univers informationnel de référence me semble aussi impérieux que de leur donner la perception de cet univers de référence.

Les documents numériques et leurs avatars technologiques (structure logique adaptée, meta-description contextuelle, transformation des contenus, formats d'affichage), en s'insérant dans un fonctionnement technique devenu indépendant des applications informatiques, sont plus que jamais en mesure de relever ce défi informationnel. Les "Interfaces Hommes Connaissances" sont cette tentative de rendre appréhendable par l'utilisateur, l'univers informationnel de l'auteur, que mes travaux sur le document numérique, la recherche d'informations et les espaces de connaissances individuels et partagés, cherchent à rendre effective.

CHAPITRE 4

DE HYWEBMAP A K-WEB ORGANIZER : VERS LA GENERALISATION DU DOCUMENT NUMERIQUE PERSONNALISABLE. LE CAS DU DOCUMENT RETICULAIRE STRUCTURE

La généralisation des principes techniques de communication du World Wide Web (le service *http* du protocole TCP/IP notamment) a entraîné une production industrielle d'informations numériques en constante croissance. Des mécanismes automatiques conçus pour réguler et gérer cette information numérique pléthorique, hérités des dispositifs informatiques tels que les moteurs de recherches, les méta-moteurs, les systèmes de gestion collaboratifs de signets, les portails, les agents intelligents, etc. se multiplient et mesurent même leurs performances de recherches à l'occasion de concours organisés à cette intention (TREC⁴², Amaryllis, CLEF⁴³, INEX⁴⁴, NTCIR⁴⁵, ...).

Cependant l'information numérique, au-delà de l'effcience de ces intermédiaires techniques de corrélation et de rapprochement, demeure comme n'importe quelle autre information, matière à appropriation par des acteurs-consommateurs fréquemment différents des acteurs-producteurs. Avec toutes les précautions qu'une telle comparaison exige, on peut observer que les systèmes automatisés de recherche d'informations numériques reproduisent en définitive l'environnement opératoire des connaissances et savoirs proposés par les bibliothèques⁴⁶. Les connaissances conservées en ces lieux, avec ce qu'elles suggèrent de diversités de "fond" (les champs de la connaissance) et de "forme" (les supports), sont mis à la disposition

⁴² Text REtrieval Conference. <http://trec.nist.gov/>

⁴³ Cross Language Evaluation Forum. <http://clef.iei.pi.cnr.it/>

⁴⁴ Initiative for the Evaluation of XML retrieval. [http://www.is.informatik.uni-
duisburg.de/projects/inex03/](http://www.is.informatik.uni-
duisburg.de/projects/inex03/)

⁴⁵ NII-NACSIS Test Collection for IR Systems

⁴⁶ principalement en ce qui concerne le caractère institutionnel de la conservation des savoirs et des connaissances, et l'organisation préalable des connaissances conservées.

d'utilisateurs porteurs de problématiques de recherches inédites et de questionnements originaux [Courbières 02] [Bernat 98] [Coulon 97]. La conservation des œuvres intellectuelles - et les modes d'organisation qui en découlent (les classifications) - doit pouvoir, en s'affranchissant de toutes les démarches individuelles de recherche d'informations, permettre à n'importe quel usager, peu ou prou averti, d'exploiter à son profit les fonds documentaires conservés [Fayet 99].

De la même manière, l'information numérique, comme le souligne Claude Baltz, *"...n'est pas un simple "élément immatériel" dans l'entreprise ou l'institution, qu'elle induit sur celle-ci des effets structurants, qu'elle ne se communique pas sans un minimum de précautions humaines, juridiques et stratégiques, que sa diffusion et sa compréhension posent des problèmes de culture et de langue propre à chaque milieu, interne ou externe, etc.."* [Baltz 03]. L'information numérique, sous tous ses formats, peut être largement traitée par les nombreux systèmes informatiques *ad hoc* qui, en répondant à des fonctions similaires (**LES** moteurs de recherches, **LES** annuaires) relèvent néanmoins de logiques algorithmiques spécifiques (Exalead, Inluo, Google, ...). Au final, c'est à l'utilisateur que revient la mission de valider les corrélations, de commenter, d'adapter, d'ajuster les informations filtrées par ces systèmes généralistes de recherches d'informations⁴⁷ [Isko 01] [Ben Hamadou 01].

Le système HyWebMap représente le fer de lance des travaux sur la constitution et la ré-organisation de fonds documentaires thématiques à partir de sources électroniques collectées sur le Web. La raison d'être de ce dispositif et les hypothèses formulées par l'équipe "Hypertexte Dynamique" ont fait l'objet de nombreuses publications qui ont validé l'originalité de notre approche [Papy 03b] [Papy 03a] [Papy 02] [Papy 99] [Saleh 04...99].

Je ne présenterai ici qu'un rappel des principes⁴⁸ d'HyWebMap et de K-Web Organizer (principe communautaire d'HyWebMap), pour ne m'intéresser qu'à ce qui constitue ma participation à ces deux dispositifs : la définition d'une DTD au format XML d'un document numérique évolutif, adaptatif et partageable ; le

⁴⁷ "Le rôle de médiateur entre l'utilisateur et les sources sur Internet n'a jamais été aussi important puisque les systèmes d'aide à la recherche disponibles sur Internet sont largement déficients et fortement irrationnels" [Rostaing 01]

⁴⁸ La plupart des articles sont disponibles en ligne et au format pdf à l'adresse <http://h2ptm.univ-paris8.fr/fab/fab.htm> (rubrique Publications)

"Document Réticulaire Structuré" (DRS) [Papy 04b] [Papy 03c] [Papy 02], vecteur de communication entre les deux environnements de gestion de connaissances.

Ce sont d'une part mes travaux antérieurs sur la nécessité de limiter la part instrumentale (indépendance instrument/objet documentaire) dans l'élaboration d'un objet informationnel synthétique, réponse partielle aux recherches d'informations initiées par l'utilisateur spécifiquement dans son cadre socio-informationnel, et d'autre part les analyses sur les comportements de recherche à l'issue de cours que je dispense depuis 1998 à des étudiants de 1^{er}, 2^{ème} et 3^{ème} cycle⁴⁹, qui m'ont amené à élaborer cette proposition.

En effet, HyWebMap est un prototype élaboré autour d'une tentative de résolution des phénomènes de désorientation et de surcharge cognitive observés lors de l'exploration de réseaux hypertextes. Le postulat initial selon lequel, le Web, réseau hypertexte de taille gigantesque, composé de documents ne bénéficiant pas de la cohérence des réseaux hypertextes antérieurs au Web, ne pouvait qu'amplifier ces phénomènes parasites d'errance, a alimenté les travaux qui ont abouti à la réalisation d'HyWebMap.

Cependant, cette hypothèse de départ m'a paru insatisfaisante car, sans pour autant nier le contexte informationnel, social, culturel, et cognitif de l'utilisateur, elle a eu tendance à le minorer. En quelque sorte, HyWebMap s'est transformé en dispositif destiné à éviter la désorientation et la surcharge cognitive d'un utilisateur désincarné dont l'environnement informationnel est supputé, mais jamais connu.

Or le cours de "Méthodologie de la Recherche et du Traitement de l'Information"⁵⁰ [Fayet 99] [Coulon 99] [Panigel 98] [Coulon 97] que je dispense à des étudiants de 1^{er} cycle d'un cursus de Médiation culturelle et communication spécialisation Information et Communication, montre que l'apprentissage pour un étudiant-novice de la méthodologie à la Recherche d'informations relève bien d'un processus distinct

⁴⁹ respectivement DEUG de Communication, Maîtrise en Sciences de l'Information Documentation et DESS orienté métier de la Documentation "Conception et Gestion des Réseaux d'Information et de Documentation"

⁵⁰ L'objectif de l'enseignement est de donner aux étudiants, en complément des apprentissages propres à leur spécialité, des méthodes de travail et d'étude indispensables à leur réussite universitaire et professionnelle en leur faisant prendre conscience de la dimension économique et stratégique de l'information, en les initiant à la recherche d'informations sur support manuel et multimédia, en les formant à l'exploitation de l'information (résumé, synthèse, analyse, ...)

s'acquérant avant toute véritable démarche de recherche opérationnelle (DEA, Doctorat) et permettant de développer les réflexes de "l'étudiant affilié"⁵¹ [Coulon 97]. Cette démarche méthodique de la recherche d'informations correspond donc à des habiletés que l'étudiant se devra d'intégrer au plus vite puisque non seulement elle conditionnera l'approche intellectuelle généralement observée à l'Université^{52, 53} mais de surcroît elle augmentera ses probabilités de réussite [Fressard 02] [Acquier 02] [Bretelle-Desmazières 99].

Ainsi un "étudiant affilié" est un usager de systèmes d'informations moins sujet à la désorientation et à la surcharge cognitive puisqu'il sera à même d'établir un plan de route conceptuel d'une recherche d'informations orienté vers un objectif dont le contexte est reconnu. C'est là que la médiation technique prend toute sa mesure puisqu'elle devient bien l'accessoire d'un processus catalyseur d'Information et de Communication où l'individu se trouve placé en position centrale.

La technique ne résout pas les déficiences méthodologiques individuelles de la recherche d'informations. L'utilisateur potentiel d'HyWebMap, usager de systèmes d'informations [Le Coadic 97] est nécessairement aguerris aux exigences et aux difficultés de la recherche d'informations et sous-entend des stratégies de recherches qui incluent une utilisation adéquate des technologies *ad hoc*. Dans le cas contraire, l'environnement logiciel complexe sous-jacent à HyWebMap (agents intelligents, restructuration cartographique, assistant de mise à jour, générateur de site,...) ne

leur apprendre à valoriser leur travaux et à les communiquer oralement ou par écrit.

⁵¹ *"Il permet en effet d'accéder à une maîtrise instrumentale de l'information et, en particulier, des ressources de la bibliothèque. Plus encore, il permet d'accéder à la maîtrise symbolique des règles du monde universitaire et à l'intériorisation du « métier d'étudiant"*

⁵² accéder aux principales ressources imprimées et électroniques (outils informatiques), développer une démarche méthodique à travers ces différentes ressources, se familiariser avec la lecture et la consultation de l'écrit d'écran, maîtriser les notions logico-linguistiques élémentaires nécessaires à l'interrogation des banques de données (opérateurs, booléens et autres), développer l'usage critique d'Internet...

⁵³ *"L'apprentissage de la connaissance doit devenir actif : il faut écouter, discuter, remettre en ordre, disséquer, s'interroger sur la validité de l'information présentée, confronter son savoir avec d'autres, réfléchir plutôt qu'emmagasiner. Les nouvelles technologies devraient nous permettre de repenser l'acte d'apprendre, c'est du moins ce que l'on nous promet. Les énormes masses d'information doivent conduire à observer les faits et les comparer selon différentes sources, puis se faire sa propre opinion. De plus le multimédia doit pouvoir s'adapter à l'apprenant, pour lui offrir des conditions d'apprentissages optimales. L'éducation, mais aussi la formation professionnelle et personnelle, doivent passer non plus par l'apprentissage seul des contenus, mais aussi par l'apprentissage de méthodes de travail, d'acquisition et de manipulation des connaissances : ces méthodes doivent privilégier la réflexion face à l'information, la confrontation de savoirs et des idées."* [Richard 96].

contribuera qu'à l'accentuation des phénomènes duaux ; désorientation et surcharge cognitive.

Les autres cours que j'assure au département Documentation de l'Université Paris 8 concernent respectivement les "Technologies de l'Information" (Maîtrise Sciences de l'Information-Documentation) et l'"Ingénierie des Connaissances" (DESS). Le premier cours s'adresse à un public d'étudiant(e)s, futurs professionnel(le)s de l'information, et le second à des étudiant(e)s généralement professionnel(le)s de l'information confirmés.

J'ai pu observer chez ces (futurs) praticiens de l'Information-Documentation, au gré des enseignements qui leurs sont dispensés et des travaux qu'ils rendent (bibliographie, Thésaurus, RID, relations avec les usagers,..) la position intermédiaire qu'ils finissent par occuper (d'infomédiaire pour citer l'expression de Joël de Rosnay⁵⁴) et qui se situent à la confluence des attentes informationnelles (émises fréquemment par la structure où ils travaillent) et des systèmes d'informations (où ils puisent pour répondre aux besoins).

Ces étudiant(e)s mettent en place de véritables stratégies s'adressant aussi bien aux usagers-clients pour qui ils effectuent les recherches qu'aux environnements pourvoyeurs d'informations dont ils font usage (lieux ressources, bases de données spécialisées, Internet,...). Les stratégies destinées aux usagers-clients consistent fréquemment en une explicitation de l'environnement de ce dernier pour déterminer le contexte informationnel où il est impliqué et mieux cerner par voie de conséquence ses attentes et ses besoins. Les stratégies destinées aux systèmes d'informations relèvent : a) de l'analyse critique liées aux informations collectées au moyen des moteurs de recherches, d'agents intelligents, b) du choix des différentes ressources sollicités, c) de mécanismes de vérification et de validation des informations collectées (disponibilité, propriété),...

Chez ces étudiant(e)s, la méthodologie de la recherche documentaire tournée ou non vers des exigences de veille, est omniprésente (information, technologie, économie, concurrence,...). Le chemin qui les conduit d'une attente ou d'un besoin exprimé par un client, à la synthèse des informations trouvées, est ainsi parfaitement balisé.

⁵⁴ 01 Informatique, 23/03/2003 consultable à l'adresse : <http://www.01net.com/article/204290.html>

Ce parcours de recherche balisé est d'une importance capitale dès lors qu'un fonctionnement communautaire s'impose. En effet, comme j'ai pu le constater, les professionnels de l'information-documentation (documentalistes, bibliothécaires spécialisés) travaillent fréquemment en réseau afin de mener à bien leurs recherches⁵⁵ [Baltz 03] [Campos 03] [Chartron 02] [Dufour 01] [Bernat 98] [Courbon 97].

C'est cette réalité de la recherche documentaire relayée par des professionnels de l'information qui rassemble expertise technique (les outils de recherche), expertise informationnelle (ressources documentaires, méthodologie) et dimension sociale qu'il me paraissait indispensable d'intégrer dans le dispositif HyWebMap.

Cependant, ces caractéristiques du fonctionnement des professionnels de l'information, qui demeurent génériques, laissent entrevoir des modalités pratiques d'organisation, directement liées au savoir-faire personnel, qu'il est difficile de pré-établir. Les transpositions techniques de l'informatique ont l'inconvénient majeur, parce qu'elle sont portées par des modélisations (cf. Chapitre 2), de ne plus laisser la place à des expressions symptomatiques qui font sens pour celui ou ceux qui les initient. Les démarches de recherches sont d'une extrême diversité, et au delà de l'exercice de la méthode, les initiatives originales, les stratégies personnelles et collectives [Balpe 97] [Richard 96] [Breton 95] vont déterminer la bonne avancée d'une recherche d'informations.

Un professionnel de l'information travaille très circonstancié et plus généralement au sein de différentes communautés de pratique (documentation, archive, veille, Knowledge Management,...) qui dépassent le cadre exclusif de leur structure d'accueil [Campos 03] [Simon 03] [Flake 02] [Favier 02] [Crosnier 98] [Miléo 96] [Harvey 95] [Michel 93]. Quelques mémoires de DESS dont j'ai assuré la direction témoignent tout particulièrement de la réalité opérationnelle de ce mode de fonctionnement [Simon 03] [Devaud 03] [Coquin 02] [Lavaivre 01] [Lacroix 01].

Les formats de document numérique inspirés du XML, de Xlink, Xpath, Xpointer et XSLT [Michard 02] [XML 98] proposent une adaptabilité des contenus qui encouragent les transformations de structures initialement adoptées en de nouvelles

⁵⁵ l'activité des listes de diffusion de l'adbs (www.adbs.fr) et de Biblio-fr en témoigne.

structures répondant mieux aux besoins circonstanciels d'une recherche d'informations. La DTD du "Document Réticulaire Structuré" que j'ai proposée répond à l'exigence d'une structure de base, un "document-réseau" dont on trouve les prémices de description dans [Chartron 02] [Salaun 02] [Tazi 99] [Tilley 93], sur lequel peut venir s'agréger des informations (informations elles-mêmes et informations du processus d'acquisition) rendues nécessaires par le caractère original de la recherche ou par des modes de fonctionnement (communauté de pratique, communauté d'intérêt). A ce stade, le DRS, est plus proche d'un "produit d'information" [Leleu-Merviel 97] individualisé que d'un document destiné à l'enseignement [Stockinger 01].

En désenclavant l'approche de type "donnée" (très inspirée d'une démarche de modélisation Entité-Association cf. figure 4.1) au profit d'une approche de type "document" [Papy 03d] [Papy 03c], le Document Réticulaire Structuré a permis de mettre en place une première version du système communautaire K-Web Organizer, assujetti à des pratiques réelles de recherches et d'agrégation d'informations [Papy 03e] [Papy 03a] [Papy 02] [Bouhaï 02] [Saleh 01c].

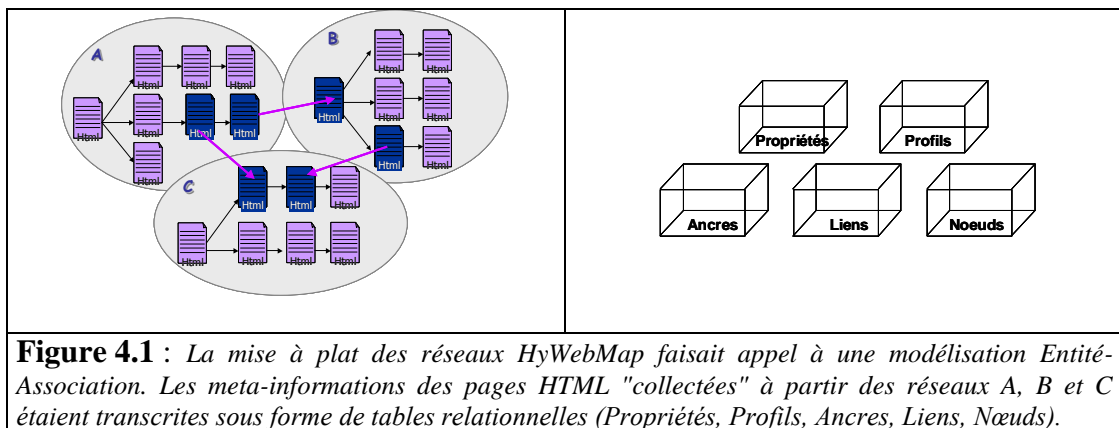


Figure 4.1 : La mise à plat des réseaux HyWebMap faisait appel à une modélisation Entité-Association. Les meta-informations des pages HTML "collectées" à partir des réseaux A, B et C étaient transcrites sous forme de tables relationnelles (Propriétés, Profils, Ancres, Liens, Noeuds).

Espaces virtuels personnels, espaces virtuels partagés, espaces sémantiques

On constate que les pratiques de recherche d'informations sur le Web ne sont pas toujours tournées vers la recherche de « LA » réponse⁵⁶ [Ferret 01] mais fréquemment vers la constitution de références documentaires convergentes que les dispositifs de veille ou les portails ont très largement mis en évidence. Les systèmes d'aspirations de site intégrés dans les navigateurs et dans les mécanismes de favoris/signets ont depuis longtemps offert aux internautes le loisir de se constituer de véritables sites thématiques au sein d'espaces locaux d'informations ré-organisées (de façon primitive puisque essentiellement hiérarchique), véritables copies off-line des sites en ligne. L'utilisation désormais presque systématique de SGBD dans la publication de sites d'informations (Web dynamique) rend cette pratique caduque tout en exacerbant la nécessité de disposer de moyen d'exprimer les associations repérées entre les différentes sources de publications.

HyWebMap et K-Web Organizer représentent des systèmes dont les intérêts majeurs résident dans la collecte, l'assemblage et la réorganisation de connaissances au sein de réseaux virtuels à partir d'informations collectées sur le Web ou produites spécifiquement par les auteurs des réseaux. HyWebMap est ainsi un dispositif lecteur/auteur d'agrégation de connaissances ; cette agrégation étant indissociable d'une fréquente voire indispensable (ré)organisation. K-Web Organizer est le principe communautaire d'HyWebMap, il permet de constituer de vastes bibliothèques de réseaux virtuels agrégatifs mises à la disposition de communautés virtuelles explicitement déclarées ou se découvrant au fil des interactions sur les réseaux HyWebMap. Les membres de ces communautés peuvent participer à la constitution de nouveaux réseaux virtuels, totalement inédits ou constitués de parties des réseaux mises à la disposition des communautés repérées.

⁵⁶ Par exemple : quelle est la hauteur de la Tour Eiffel ? du Mont-Blanc ? Quand a eu lieu le premier vol spatial habité ?

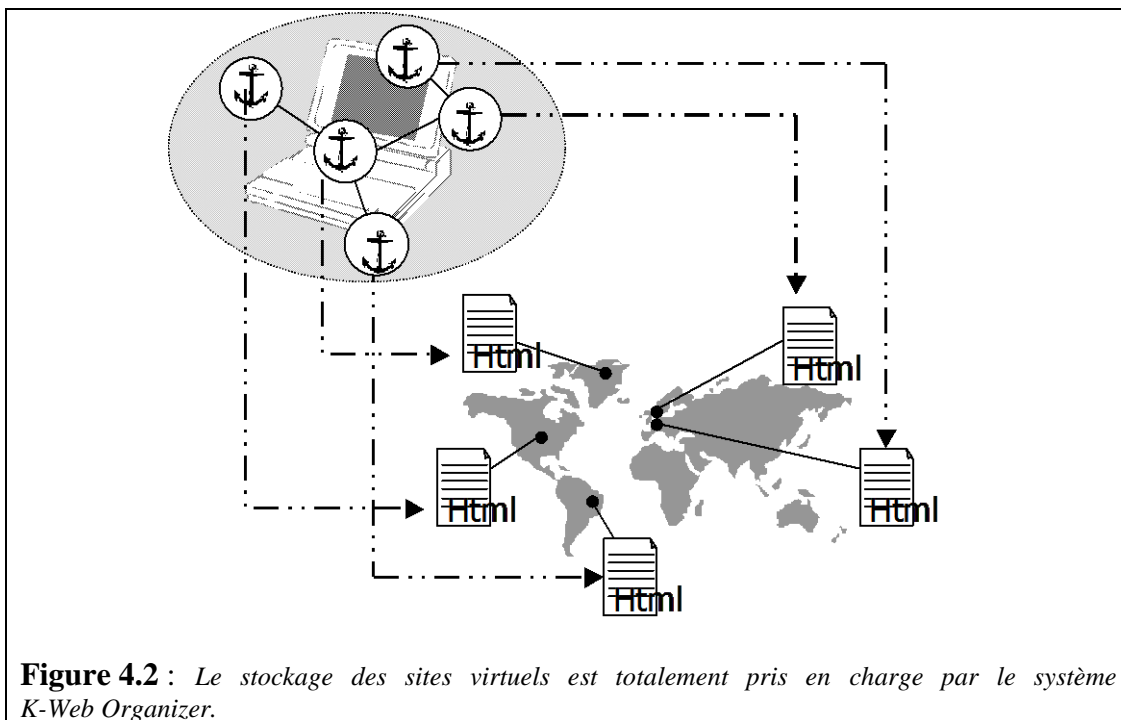
HyWebMap

Directement associé aux ressources numériques du Web, HyWebMap grâce à ses fonctionnalités, permet de définir une structure virtuelle composite où sera agrégée la totalité ou une partie des informations préalablement localisées par l'internaute. Cette particularité de construction d'une structure virtuelle dynamique s'accompagne d'un ensemble de fonctionnalités qui confère à ce système la double dimension lecteur et auteur.

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

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

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



Au travers d'HyWebMap c'est l'expertise que représentent les parcours de navigation d'un lecteur averti au sein d'un espace d'informations hypertextuelles, qui est recherchée. Cette conception originale des livres électroniques s'est enrichie d'une construction de sites Web virtuels, thématiques et répartis qui s'appuient sur l'agrégation de ressources distantes afin de former un ensemble de documents sémantiquement homogènes.

Agent Web du HyWebMap

HyWebMap inclut un outil informatique (dit agent) capable à partir d'une page HTML ou d'une Uniform Resource Locator (URL) quelconque choisie par un utilisateur, de faire l'analyse de cette page [Papy 03a] [Lelu 93], et d'en répertorier ses principales caractéristiques (page carrefour interne, page carrefour externe, page informative textuelle,...).

Navigation par réseaux interactifs personnels

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

nœuds et des liens personnels. Les liens personnels peuvent être ajoutés soit en utilisant les fonctions copier/coller, soit en reliant graphiquement les nœuds dans la carte par un simple traçage d'une flèche entre deux nœuds. Ainsi est créé un lien hypertextuel. Les liens personnels sont représentés par des flèches de couleur rouge qui les distinguent des liens de la carte visitée.

Agents de surveillance du réseau virtuel

Un agent de surveillance contrôle périodiquement la disponibilité des nœuds du réseau virtuel dans leurs sites d'origine. Dans le cas où un nœud ne serait plus disponible, il effectue la modification sur le réseau virtuel, et alerte l'utilisateur.

Etiquette des nœuds et des liens

Dans le système HyWebMap, les nœuds et les liens sont étiquetés afin de rendre le réseau plus explicite. L'utilisateur est encouragé à nommer les liens entre les nœuds afin de faciliter les recherches postérieures.

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

Le réseau virtuel personnel présente l'espace Web de l'utilisateur. HyWebMap autorise la création d'annotations sur les nœuds des réseaux virtuels. Cette annotation peut être un texte, des liens ou un texte avec des liens. Le réseau virtuel peut donc se trouver annoté par un réseau de notes. L'utilisateur a la possibilité de mettre un point de repère sur un nœud pour signaler son importance dans le réseau.

Aides à la navigation

En plus des mécanismes traditionnels de recherche (sur les mots-clefs associés aux nœuds, sur les labels des liens et des nœuds, dans les annotations, dans les nœuds directement), HyWebMap est doté de plusieurs fonctionnalités d'assistance à la navigation : la navigation par historique (outil classique qui offre à l'utilisateur le retour sur un nœud déjà visité), la navigation par niveau d'arborescence (l'utilisateur sélectionne directement le niveau de l'arborescence), la navigation par date (HyWebMap sauvegarde pendant deux semaines l'historique de navigation du réseau virtuel, l'utilisateur peut demander à voir les nœuds visités durant une période précise).

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

HyWebMap offre la possibilité de générer automatiquement un site Web à partir des réseaux virtuels. Cette fonctionnalité offre l'avantage de rendre les réseaux virtuels visibles de tout navigateur banalisé.

K-Web Organizer (http://h2ptm.univ-paris8.fr/k_web)

Le système collaboratif, d'échange et de partage de connaissances baptisé K_Web Organizer, fonctionne à partir de réseaux HyWebMap, entités agrégatives complexes réunissant ressources électroniques "hors ligne et/ou en ligne", annotations et descripteurs auteurs/lecteurs.

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

Le système K-Web Organizer rend possible la création d'une communauté d'utilisateurs fédérée autour de ressources documentaires thématiques produites directement par ses membres ou de contributions réalisées à partir de compilations de documents pour lesquelles ils imaginent une organisation particulière. Ces productions originales ou composites mais toujours individuelles sont mises à la disposition de tous et participent à un seul et même fonds documentaire dans lequel les interactions entre auteurs et documents seront signalées en temps réel à toute la communauté.

La communauté dont nous traitons ici est une conséquence de l'association des réseaux virtuels HyWebMap engendrés par les utilisateurs-auteurs ; associations opérables en pratique par le système à partir d'objets documentaires (les réseaux virtuels personnels) dont les contenus peuvent être associés.

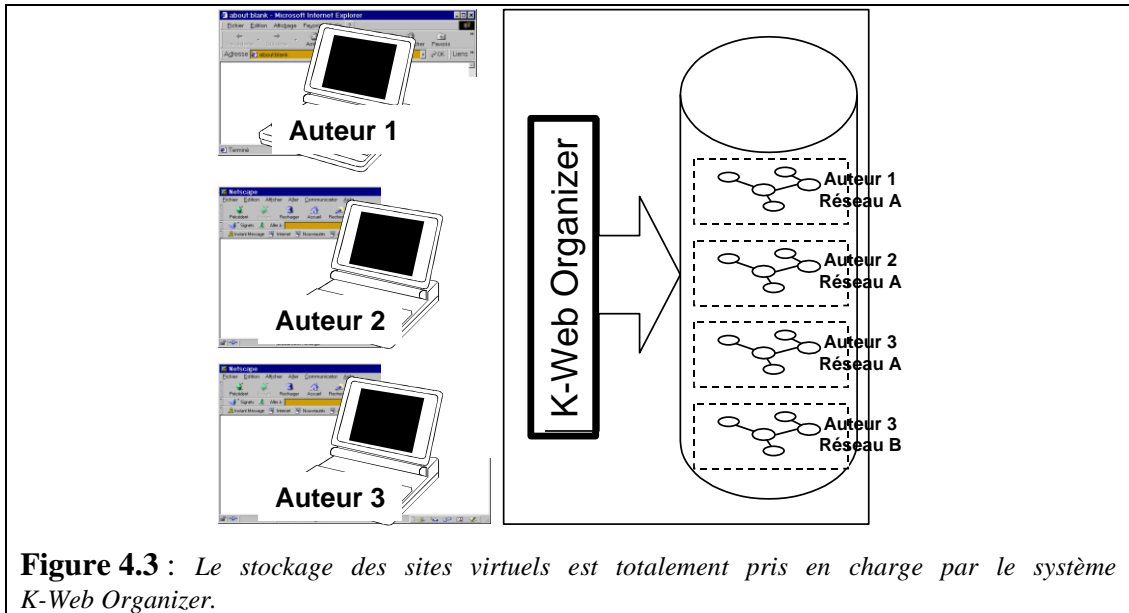


Figure 4.3 : *Le stockage des sites virtuels est totalement pris en charge par le système K-Web Organizer.*

Des principes de fonctionnement collaboratifs

La création de réseaux virtuels repose sur les principes mis en évidence dans le logiciel HyWebMap. Les mécanismes liés à l'affectation de ressources documentaires à des nœuds libres, les nœuds et liens nommés, l'actualisation par agent intelligent, etc., sont toujours présents au sein de K-Web Organizer. Les fonctions déterminantes qui viennent s'ajouter concernent toutes les interactions collaboratives liées à la création ou à l'utilisation de réseaux virtuels :

- Quelles sont les "communautés" existantes et quels en sont les centres d'intérêt ?
- Un même réseau peut-il être soumis à plusieurs communautés ?
- Comment se gère l'émergence d'une nouvelle communauté ?
- Comment être informé de l'utilisation d'un réseau personnel par des tiers ?
- Peut-on suivre les évolutions d'un réseau personnel par les révisions d'autres auteurs ?
- Comment s'assurer de la propagation des modifications apportées à un réseau HyWebMap au sein d'une communauté en ayant fait usage ?
- Comment élaborer un document agrégatif à partir d'une communauté ?
- etc.

Un document numérique personnalisable : Le Document Réticulaire Structuré

Dès l'origine, la conception d'HyWebMap a été marquée par la volonté d'intégrer au sein d'un environnement unique de travail l'ensemble des fonctionnalités relatives à la recherche, la description et la réorganisation d'informations collectées sur le Web. L'aspect documentaire était peu présent, les données traitées par ce logiciel auteur étant stockées selon un formalisme beaucoup plus proche de celui des bases de données relationnelles (approche donnée) que des systèmes de gestion documentaire (approche document).

Cette situation rendant l'échange et le partage de réseaux HyWebMap difficile, une re-formulation XML s'est rapidement imposée, transformant les "réseaux-bases de données" générés par HyWebMap en document numérique structuré plus aisément partageables (cf. tableau 1 et tableau 2). Cette "Document Type Definition" propose une structure évolutive qui met en évidence la possibilité d'exprimer un document dont l'instance sera agrégative (associer aux contenus de départ d'autres sources d'informations), nécessairement individualisée (les compléments d'informations sont déterminés par chaque usager en fonction de son contexte) et partageables (entre utilisateurs d'une même communauté de pratique par exemple).

L'expression de cette reformulation XML de la structure initiale des données HyWebMap correspond véritablement aux impératifs de la gestion de l'information (prise ici au sens large) à l'heure de la généralisation des réseaux et des espaces documentaires ouverts. Maintenir une indépendance des documents par rapport aux environnements informatiques, permettre une réutilisabilité des données produites, rendre possible leur intégration dans des systèmes d'informations différents, sont autant de points forts des technologies XML qui se retrouvent en filigrane dans les documents numériques.

Agréger des informations autour d'un noyau composé par exemple de contenus pédagogiques élaborés par des enseignants, adapter cette agrégation afin qu'elle serve à l'enrichissement personnel au gré des interactions au sein d'une communauté d'enseignants, de tuteurs et d'apprenants constituent les invariants du document

numérique des Formations A Distance [Papy 03c] [France 01] [Senteni 01] [Moeglin 99] [Moeglin 98]. Rendre la lisibilité de la conception d'un thésaurus fait tout autant appel à ce principe d'agrégation.

Un thésaurus au-delà de la définition⁵⁷ proposée par la norme ISO Z39.19-1993 "*Guidelines for the Construction, Format and Management of Monolingual Thesauri*", est une expression modélisée de la réalité d'un système d'information (pris au sens large) et repose sur l'identification de descripteurs caractéristiques (langage contrôlé) et des relations de dépendances fonctionnelles s'établissant entre eux. En ce sens un thésaurus est un document d'une grande utilité pour parcourir sémantiquement les concepts d'un univers modélisé. A ce titre, l'élaboration d'un thésaurus est un processus long et exigeant de recensement documentaire et d'analyse bibliographique [Lefèvre 00] [Aitchinson 92] [Cattenat 92], [Cattenat 77] [Chaumier 77].

Il existe bien sûr des sites Internet "thésaurus"⁵⁸, forme de transposition électronique des documents papier où la justification de cette transposition réside essentiellement dans la capacité à rendre actif les liaisons sémantiques entre les descripteurs. Cette transposition électronique à l'interactivité limitée ne restitue pas toutes les phases préliminaires qui conduisent à la finalisation du thésaurus. En effet, les descripteurs du langage naturel, les expressions du langage contrôlé, les définitions, les relations sémantiques entre les termes ne sont que la partie apparente qui masquent totalement le dépouillement terminologique et bibliographique qui ont abouti au résultat final.

Thésaurus : un document structuré isolé de ses sources

C'est ainsi que l'expression finale du thésaurus aboutit à la dé-contextualisation des descripteurs retenus et des relations impliquées. On obtient une véritable abstraction souvent difficile à lire et à exploiter pour les non-experts (le résultat est assez similaire à ce que l'on peut observer dans les Modèles Conceptuels de Données accompagnant toute modélisation de bases de données relationnelles).

⁵⁷ " ...A thesaurus is a controlled vocabulary arranged in a known order and structured so that equivalence, homographic, hierarchical, and associative relationships among terms are displayed clearly and identified by standardized relationship indicators that are employed reciprocally... "

⁵⁸ www.nlc-bnc.ca/8/4/r4-280-f.html, www.culture.gouv.fr/documentation/thesarch/pres.htm

Thésaurus / DRS : des démarches similaires, des résultats exclusifs

C'est à ce stade de l'élaboration du thésaurus, lorsque toutes les ressources bibliographiques sont disponibles que le Document Réticulaire Structuré (DRS) prend toute sa dimension. En effet, il va permettre de constituer véritablement le document final mais également de préserver l'ensemble des références grâce auxquelles ce résultat est obtenu.

La Figure 4.4 illustre cette phase synthétique de mise en relation des descripteurs retenus (cf. Tableau 1). Ce résultat est obtenu par les documentalistes à partir d'une sélection de références bibliographiques jugées pertinentes. Dans cette figure, les références n'ont pas été reportées car les modalités pratiques de rédaction excluent à ce stade ce type de rapprochement. En revanche dans le document réticulaire structuré correspondant élaboré avec HyWebMap (cf. Figure 4.5), les relations sémantiques (cf. Tableau 2 et Tableau 3) et les sources bibliographiques sont préservées. Ce sont les annotations qui vont permettre de maintenir ce rapprochement.

Pour assurer une plus grande visibilité, le résultat final correspond au thésaurus traditionnel conforme à la norme ISO Z39.19-1993 a été traduit en document XML⁵⁹ (cf. Tableau 4) répondant à la DTD du Tableau 5.

Le document Réticulaire Structuré (partiel) correspondant est quant à lui instancié dans le tableau 6 et a été validé par la DTD du tableau 7.

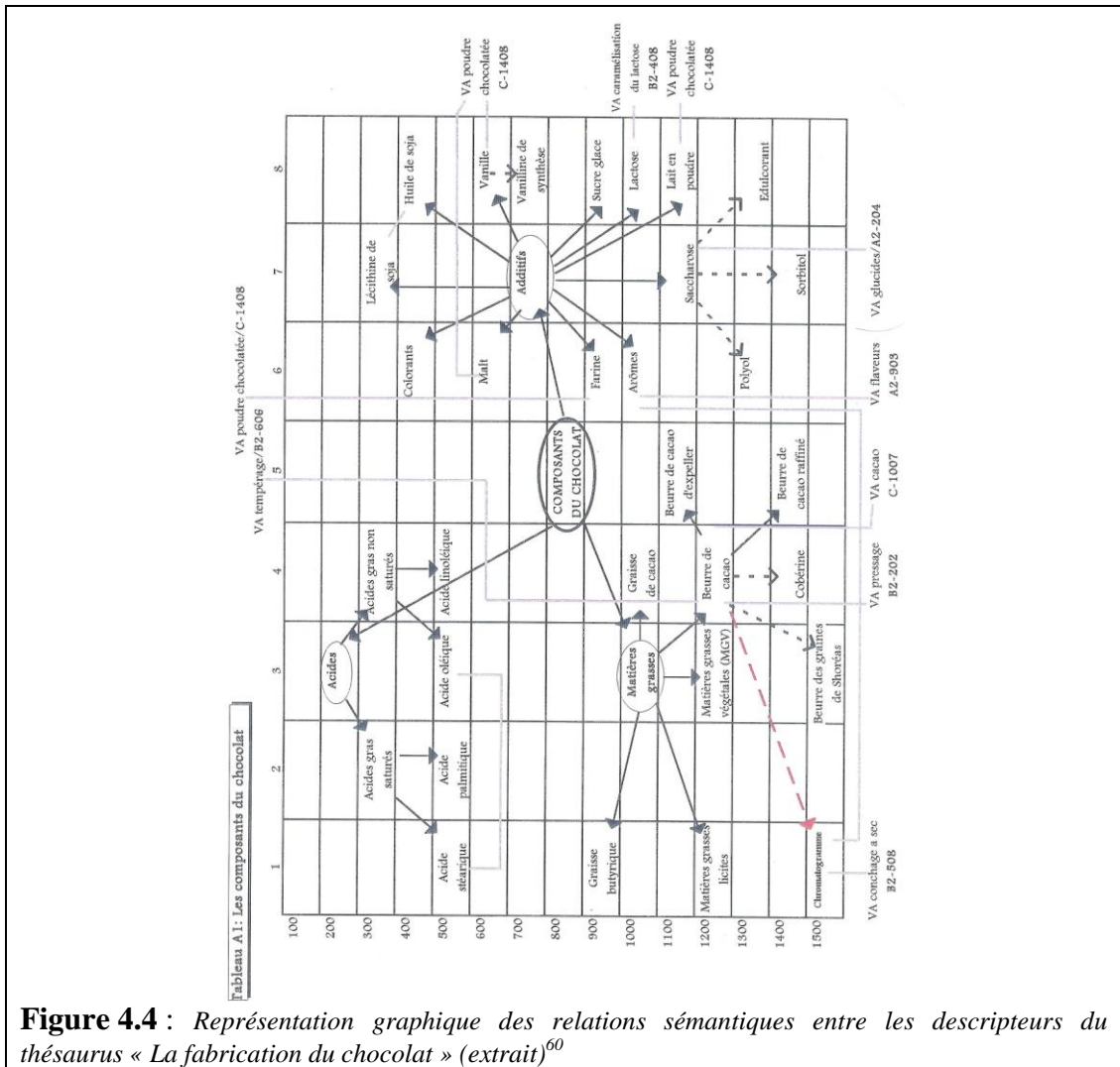
Au travers de cette DTD, on s'aperçoit que le thésaurus en tant que document final exploitable est effectivement disponible mais qu'il est toujours associé, aux niveaux les plus fins (celui des descripteurs), à toutes les références documentaires nécessaires.

Le thésaurus transposé sous forme de DRS propose :

1) un document structuré où chaque élément d'informations est clairement identifié et balisé [Bui Thi 03] [Bradley 92] [Newcomb 91] [Dachelet 90] [Furuta 89],

⁵⁹ Avec le XML, ce n'est donc que très récemment que l'orientation consensuelle [XML 98] de séparer le fond et la forme est devenue réalité. En son temps, le SGML a bien représenté une tentative peu éloignée de ce que le XML propose aujourd'hui, mais les conditions d'élaboration de cette norme orientée édition, les exigences conceptuelles des DTD, et l'environnement matériel et logiciel n'ont pas permis de généraliser ce principe de séparation du fond et de la forme.

2)- une réticulation évolutive de par les ressources qui sont associées à ces éléments au travers des URL, des annotations et des mots-clefs.



60 Ce thésaurus de 70 pages portant sur « La fabrication du Chocolat » a été conçu par 3 étudiantes en maîtrise de Documentation à l'Université Paris 8. Destiné au centre de documentation de Nestlé France, il a été supervisé par la responsable du centre de ressources documentaires de l'entreprise et par l'enseignante spécialiste des langages documentaires.

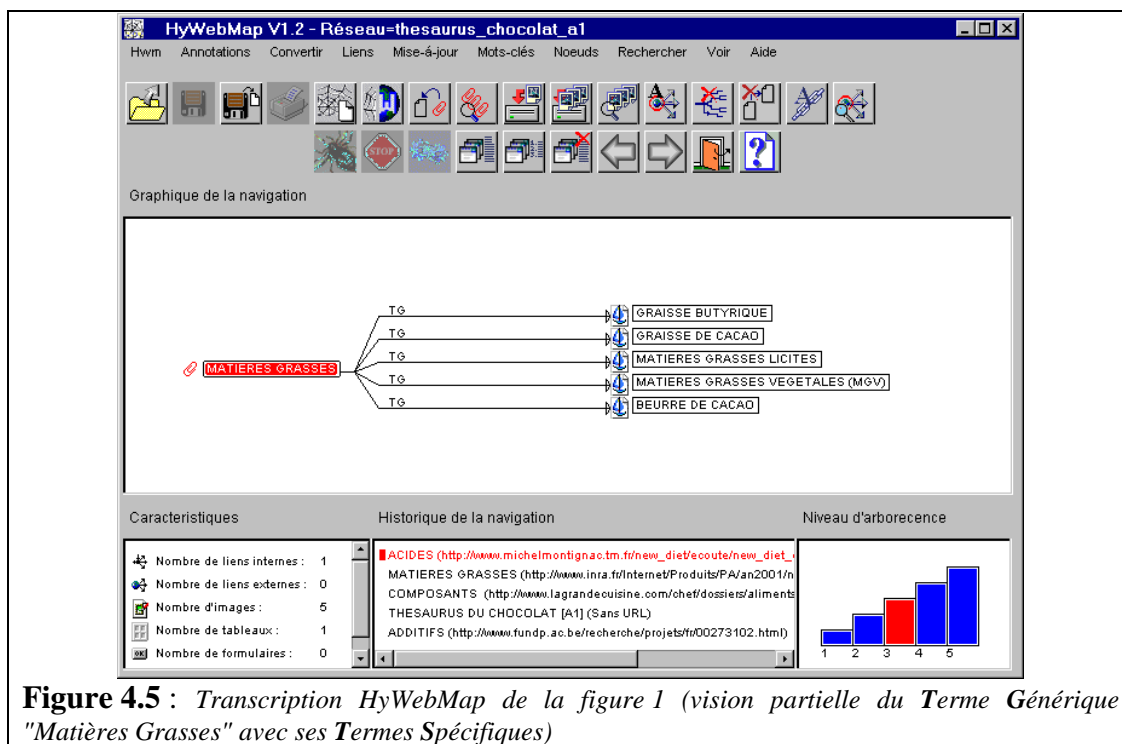


Figure 4.5 : Transcription HyWebMap de la figure 1 (vision partielle du Terme Générique "Matières Grasses" avec ses Termes Spécifiques)

acide linoléique	A1-504	Farine	A1-906
Acide oléique	A1-503	Graisse butyrique	A1-906
Acide palmique	A1-502	Graisse de cacao	A1-1004
Acides	A1-203	Huile de soja	A1-408
Acides gras non saturés	A1-304	Lactose	A1-1008
Acide stéarique	A1-302	Lait en poudre	A1-1108
Additifs	A1-501	Lécithine de soja	A1-307
Arômes	A1-1006	Malt	A1-606
Beurre de cacao	A1-1204	Matières grasses	A1-1003

Tableau 1 : Extrait des 270 termes en langage naturel (228 langage contrôlé) du thésaurus réalisé pour Nestlé France.

Relation	Nature
VA	Voir Aussi (Deux termes reliés par VA sont associés)
EP	Employ Pour (Deux termes reliés par EP sont synonymes)
TG	Terme Générique (ADDITIF est le terme générique de MALT)
TS	Terme Spécifique (GRAISSE DE CACAO est le terme spécifique de matières grasses)

Tableau 2 : Quelques-unes des relations sémantiques exprimées dans un thésaurus

Langage naturel	Définitions	Langage contrôlé	Relations
ACIDE LINOLEIQUE	Un des acides gras non saturés contenu dans les fèves de cacao (3%)	ACIDE LINOLEIQUE	TS d'acides gras non saturés
ACIDE OLEIQUE	Un des acides gras non saturés contenu dans les fèves de cacao (35%)	ACIDE OLEIQUE	VA acide stéarique TS acides gras non saturés
ACIDE PALMITIQUE	Un des acides gras saturés contenu dans les fèves de cacao (28%)	ACIDE PALMITIQUE	TS acides gras saturés
ACIDE STEARIQUE	Un des acides gras saturés contenu dans les fèves de cacao (34%). Dans l'organisme, l'acide stéarique se transforme rapidement en acide oléique insaturé	ACIDE STEARIQUE	VA acide oléique TS acides gras saturés
ACIDES	Contenus dans la fève de cacao	ACIDES	TS acides TG acide oléique TG acide linoléique

Tableau 3 : Extrait des relations entre les termes du langage contrôlé.

```

<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE thesaurus SYSTEM "thesaurus_chocolat.dtd">
<thesaurus subject="La Fabrication du chocolat">
  <TABLE title="Les composants du chocolat" code="A1">
    <TERM>
      <DESCRIPTOR id="A1-203" content="ACIDES"/>
      <LINK idref="A1-302" type="TS"/>
      <LINK idref="A1-304" type="TS"/>
    </TERM>
    <TERM>
      <DESCRIPTOR id="A1-302" content="Acides gras
saturés"/>
      <LINK idref="A1-501" type="TS"/>
      <LINK idref="A1-502" type="TS"/>
    </TERM>
    <TERM>
      <DESCRIPTOR id="A1-501" content="Acide
stéarique"/>
      <LINK idref="A1-503" type="RV"/>
    </TERM>
    <TERM>
      <DESCRIPTOR id="A1-502" content="Acide
palmitique"/>
    </TERM>
    <TERM>
      <DESCRIPTOR id="A1-304" content="Acides gras non
saturés"/>
      <LINK idref="A1-503" type="TS"/>
      <LINK idref="A1-504" type="TS"/>
    </TERM>
  </TABLE>

```

```

<TERM>
  <DESCRIPTOR id="A1-503" content="Acide oléique"/>
  <LINK idref="A1-501" type="RV"/>
</TERM>
<TERM>
  <DESCRIPTOR id="A1-504" content="Acide
linoléique"/>
</TERM>
<TERM>
  <DESCRIPTOR id="A1-1003" content="Matières
grasses"/>
  <LINK idref="A1-901" type="TS"/>
  <LINK idref="A1-1004" type="TS"/>
  <LINK idref="A1-1201" type="TS"/>
  <LINK idref="A1-1203" type="TS"/>
  <LINK idref="A1-1204" type="TS"/>
</TERM>
<TERM>
  <DESCRIPTOR id="A1-901" content="Graisse
butyrique"/>
</TERM>
<TERM>
  <DESCRIPTOR id="A1-1004" content="Graisse de
cacao"/>
</TERM>
<TERM>
  <DESCRIPTOR id="A1-1201" content="Matières
grasses licites"/>
</TERM>
<TERM>
  <DESCRIPTOR id="A1-1203" content="Matières
grasses végétales (MGV)"/>
</TERM>
<TERM>
  <DESCRIPTOR id="A1-1204" content="Beurre de
cacao"/>
  <LINK idref="A1-1105" type="TS"/>
  <LINK idref="A1-1405" type="TS"/>
  <LINK idref="A1-1404" type="S"/>
  <LINK idref="A1-1503" type="S"/>
  <LINK idref="A1-1501" type="NC"/>
</TERM>
</TABLE>
<TABLE title="Les composants du chocolat" subtitle="Les étapes
communes" code="A2">
</TABLE>
<TABLE title="Les processus de fabrication de la poudre de cacao et du
chocolat" subtitle="Les étapes propres à chaque fabrication" code="B1">
</TABLE>
<TABLE title="Les produits dérivés de la fève de cacao" code="C">
</TABLE>
</thesaurus>

```

Tableau 4 : Transcription XML de l'extrait du thésaurus " La fabrication du Chocolat " ⁶¹

⁶¹ Tous les fichiers xml et les dtd de cet article sont disponibles à l'URL : <http://www.labart.univ-paris8.fr/papy/cide5/>. Ils ont été parsés avec le logiciel Xerces (<http://xml.apache.org/>)

<pre> <?xml encoding="ISO-8859-1"?> <!ELEMENT thesaurus (TABLE+)> <!ATTLIST thesaurus subject CDATA #REQUIRED> <!ELEMENT TABLE (TERM*)> <!ATTLIST TABLE title CDATA #REQUIRED subtitle CDATA #IMPLIED code CDATA #REQUIRED> <!ELEMENT TERM (DESCRIPTOR , LINK*)> <!ELEMENT DESCRIPTOR EMPTY> <!ATTLIST DESCRIPTOR id ID #REQUIRED content CDATA #REQUIRED> <!ELEMENT LINK EMPTY> <!ATTLIST LINK idref IDREF #REQUIRED type (TS TG RV S NC VA) #REQUIRED> </pre>
<p>Tableau 5 : DTD XML du thésaurus</p>

<pre> <?xml version="1.0" encoding="iso-8859-1"?> <!DOCTYPE thesaurus_hwm SYSTEM "thesaurus_chocolat_hwm.dtd"> <drs subject="La fabrication du chocolat"> <generic_keyword_list> <gterm id="1" content="ACIDE LINOLEIQUE"/> <gterm id="2" content="ACIDE OLEIQUE"/> <gterm id="3" content="ACIDES PALMITIQUE"/> <gterm id="4" content="ACIDES"/> <gterm id="5" content="ACIDES GRAS NON SATURES"/> <gterm id="6" content="ACIDES GRAS SATURES"/> <gterm id="7" content="ACIDE STEARIQUE"/> <gterm id="8" content="ADDITIFS"/> <gterm id="9" content="AROMES"/> <gterm id="10" content="BEURRE DE CACAO"/> <gterm id="11" content="BEURRE DE CACAO D'EXPELLER"/> </generic_keyword_list> <nodes> <node id="1" title_hwm="ADDITIFS"> <annotations> <annotation id="1"> Il s'agit essentiellement d'agents de texture (pectine, dextrines, matières amylacées, gélatine..), d'agents de sapidité ou d'auxiliaires de fabrication tels que le talc, cire, huile végétale, parafine, etc destinés à préparer les surfaces de fabrication en contact avec les produits. En matière de confiserie, les additifs sont fort peu nombreux et leurs emplois strictement réglementés par le décret du 19 décembre 1910. (tiré du Glossaire des confiseurs de France) </annotation> <annotation id="2"> Langage contrôlé : ADDITIFS </annotation> <annotation id="2"> </pre>
<p>) TG de GRAISSE BUTYRIQUE, MATIERES GRASSES LICITES, MATIERES</p>

<pre> GRASSES VEGETALES (MGV), GRAISSE DE CACAO </annotation> </annotations> <specific_keyword_list> <sterm idref="8" /> <sterm id="1" content="AGENT DE TEXTURE"/> <sterm id="2" content="AGENT DE SAPIDITE"/> </specific_keyword_list> <author_keyword_list> <aterm content="AGENT DE TEXTURE"/> </author_keyword_list> <links> <internal_link from="1" to="2" type="TG vers TS"/> <internal_link from="1" to="3" type="TG vers TS"/> <external_link from="1" to="2" type="TG vers TS"/> </links> <web title="Influence d'additifs naturels sur les propriétés structurales et rhéologiques du chocolat" url="http://www.fundp.ac.be/recherche/projets/fr/00273102.html"/> </node> <node id="2" title_hwm="COLORANTS"> </node> <node id="3" title_hwm="HUILE DE SOJA"> </node> </nodes> </drs> </pre>
<p>Tableau 6 : Instance XML partielle du Document Réticulaire Structuré de la figure 3.</p>

<pre> <?xml encoding="ISO-8859-1"?> <!ELEMENT drs (generic_keyword_list?, nodes)> <!ATTLIST drs subject CDATA #REQUIRED> <!ELEMENT generic_keyword_list (gterm)*> <!ELEMENT gterm EMPTY> <!ATTLIST gterm id_gterm NMTOKEN #REQUIRED content CDATA #REQUIRED> <!ELEMENT nodes (node*)> <!ELEMENT node (annotations?,specific_keyword_list?, author_keyword_list?, links?, web?)> <!ATTLIST node id_node NMTOKEN #REQUIRED title_hwm CDATA #REQUIRED> <!ELEMENT annotations (annotation*)> <!ELEMENT annotation (#PCDATA)> <!ATTLIST annotation id_annotation NMTOKEN #REQUIRED> <!ELEMENT specific_keyword_list (sterm_global sterml_local)*> <!ELEMENT sterml_global EMPTY> </pre>
--

```

<!ATTLIST   stern_global
            idref_gterm NMTOKEN #REQUIRED>

<!ELEMENT  stern_local  EMPTY>
<!ATTLIST  stern_local
            id_stern  NMTOKEN #REQUIRED
            content  CDATA  #REQUIRED>
<!ELEMENT  author_keyword_list  (aterm)*>

<!ELEMENT  aterm  EMPTY>
<!ATTLIST  aterm
            content  CDATA  #REQUIRED>
<!ELEMENT  links  (internal_link)*>

<!ELEMENT  internal_link  EMPTY>
<!ATTLIST  internal_link
            from  NMTOKEN #REQUIRED
            to    NMTOKEN #REQUIRED
            type  CDATA  #REQUIRED>

<!ELEMENT  web  EMPTY>
<!ATTLIST  web
            title  CDATA  #REQUIRED
            url    CDATA  #REQUIRED>

```

Tableau 7 : DTD XML d'un Document Réticulaire Structuré

Conclusion

La Recherche d'Informations – dès lors qu'elle n'est pas assimilée au souhait d'une réponse unique [Ferret 01] -, ne peut en aucun cas être réduite à la saisie de quelques mots dans le formulaire de saisie d'un moteur de recherches ou d'un annuaire. La liste dérisoire du "top 20" des mots les plus fréquemment soumis à l'annuaire Nomade est là pour nous le rappeler.

JUILLET 2003	MARS 2003	OCTOBRE 2002	AVRIL 2001
Sexe Pages Jaunes SNCF ANPE Météo Michelin TF1 Tour de France Immobilier FNAC Charme Humour CAF Emploi Kazaa Jeux Chat MP3 Camping Gay	Sexe ANPE Pages Jaunes SNCF TF1 Impôts Humour Immobilier FNAC Emploi Michelin Météo Charme Kazaa Musique Chat Jeux MP3 Gay Eminem	Sexe ANPE Pages Jaunes SNCF Halloween Emploi TF1 Météo Immobilier MP3 FNAC Musique Charme Humour Chat Jeux Michelin Gay Kazaa Star Academy	Sexe Gratuit Tiberi France Jeux Location Paris Gay Photos Mp3 Immobilier Emploi Vidéo Annuaire Musique 2001 Hôtel SNCF Pokémon Webcam
<i>"Top-20" des mots-clés les plus fréquemment utilisés sur l'annuaire "Nomade"</i> (http://www.nomade.fr)			

Devant la masse croissante des informations créés et diffusés, de façon totalement anarchique, c'est notre capacité à chercher, traiter, produire, organiser, intégrer ces données virtuellement accessibles qui est interpellée.

Le problème en la matière est épineux comme le souligne Pierre Moeglin *"D'ores et déjà les dispositifs technologiques proposent des solutions de recherche que le commun des mortels ignore. Or le constat est étonnant, c'est souvent à la convergence de ces différentes solutions qu'il faudra faire appel pour s'informer qualitativement"* [Moeglin 98].

Cette convergence liée à la Recherche d'Informations relève bien de contextes, de connaissances préalables, d'objectifs, d'habilités, de stratégies mentales, de médiations techniques et humaines, d'usages, etc. somme toute d'un processus complexe qui renvoie aux notions incertaines de signification et de valeur de l'information [Moles 02].

A l'occasion de mes travaux sur l'hypertextualisation, en une période antérieure au Web, j'ai pu observer que les risques de désorientation et de surcharge cognitive étaient en définitive bien minimes puisque les réseaux hypertextes (et les documents électroniques linéaires créés à l'origine en ce qui concerne l'hypertextualisation) étaient le fruit d'un travail de conception d'auteurs, travail de conception qui garantissait (au moins) un sens à l'ouvrage réalisé. L'hypertextualisation, processus d'automatisation de réseaux hypertextes en augmentant les possibilités d'accès à des documents d'auteurs, encourageait les lecteurs occasionnels à découvrir de nouveaux sens, au gré des lectures associatives et intuitives rendues possibles par les interfaces simplistes des premiers systèmes hypertextes.

Mes travaux sur l'hypertextualisation ont introduit une séparation du fond et de la forme, ont utilisé l'un pour révéler l'autre. Les marqueurs de mise en forme disposés de sorte qu'ils mettaient en exergue la structure logique d'un document ont permis de proposer à la lecture hypertextuelle, des éléments atomiques, à partir du sens contenu dans la structure supérieure du réseau hypertexte.

Cette question latente du fond, de la forme et du sens se retrouvent avec une toute autre insistance depuis que le Web existe et depuis que de nombreux systèmes d'informations valides, pérennes et stratégiques en constituent le fond. La forme ne concerne plus uniquement la mise en forme matérielle d'un caractère, d'un paragraphe ou d'un document mais relève de ressources numériques autrement plus sophistiquées : pages HTML, fichiers texte, PDF, images, animations, vidéo,...

Des contenus distincts, élaborés par des logiques d'auteur bien distinctes, destinés à répondre à des besoins différents, vont être agrégés, organisés, commentés, distribués, discutés, etc... à l'issue de processus de recherches d'informations conduits par des usagers insérés dans des environnements sociaux, culturels et cognitifs tout aussi différents. Ce sont ces usagers, les circonstances de leurs

recherches et leur environnement d'usage qui vont donner un sens à cet assemblage d'informations éparses. Élément supérieur et fédérateur du processus de la recherche, l'utilisateur, relais de son environnement social, culturel et cognitif va déterminer le sens de l'agrégation de ressources ainsi concoctée.

C'est dans cette démarche du sens, que les Interfaces Homme-Connaissances prennent toute leur dimension. En effet, il ne s'agit pas de parasiter l'éclosion du sens par une quelconque ingérence des logiques techniques autoréférentes, étrangères au contexte de la recherche et des usagers, mais bien au contraire d'améliorer les mécanismes inférenciels, les rétroactions qui renseignent l'utilisateur sur ses orientations, ses méconnaissances, etc.. Les recherches d'informations conduites idéalement au moyen de ces interfaces Homme-Connaissances devraient permettre de trouver des éléments de réponses en laissant la plus grande marge de manœuvre aux habiletés et aux stratégies des usagers. Mais pour quels résultats ?

Les systèmes de recherche d'informations comme HyWebMap offrent aujourd'hui aux usagers la possibilité de réorganiser les informations collectées pour rendre plus explicite la dimension subjective de la recherche. Mais cette réorganisation se heurte à un nouveau problème de "forme", problème qui n'a pas pour autant abandonné la sophistication des formats de documents pour gagner celle des technologies. En effet, les technologies de la facture de celles qui sont embarquées dans HyWebMap, parce qu'elles déterminent des conditions d'utilisation, des procédures techniques arrêtées en rapport direct avec les représentations des concepteurs, vont inévitablement produire une **empreinte** sur le mode d'organisation des informations collectées. Ainsi, malgré un "fond" conforme à ses attentes informationnelles, l'utilisateur ne trouvera pas en raison même de la forme imposée par cette empreinte technologique, une expression en phase avec ce qu'il a pu rassembler.

Appliquées à la Recherche d'Informations, les Technologies de l'Information et de la Communication sont engagées dans un processus d'innovation artificiel, car à ce jour :

- les dispositifs technologiques sont dotés de mécanismes que la grande majorité des usagers potentiels ignore⁶²,
- l'évaluation des systèmes de recherche d'information en regard aux attentes des usagers n'est guère introduite dans l'offre industrielle [Chaudiron 02]. Les concours institutionnels (TREC, CLEF,...) ne traduisent que la performance technique des environnements logiciels,
- les professionnels de l'information témoignent de par leur existence même des réalités culturelle, humaine, juridique, stratégique et contextuelle de l'information. Tout en connaissant la logique et la problématique des outils informatiques, ils savent que la recherche d'informations relève également de méthodologies et de savoir-faire.

Il m'apparaît que des ingrédients permettant d'éviter les injonctions technologiques [Jeanneret 99] existent d'ores et déjà pour envisager une symbiose techno-informationnelle qui profitera à tous les usagers de systèmes d'informations rompus aux exigences méthodologiques de la Recherche d'Informations. La proposition de DTD concernant le "Document Réticulaire Structuré" constitue cette enveloppe d'inspiration technologique (XML, Technologies du Web, Consortium W3) mise au service d'une approche anthropocentrée de l'information. La capacité désormais illimitée de donner une structuration explicite aux éléments, couplée à des mises en forme dynamiquement adaptables à la diversité des dispositifs de restitution disponibles (écrans, imprimante, PDA, mobiles, ...), répondant de fait à des attentes informationnelles est, à la portée du plus grand nombre.

Le polymorphisme engendré par XML élaboré au sein d'un document indépendant de tout environnement logiciel et matériel permet de décrire les positions subjectives des demandes, des démarches suivies, des stratégies adoptées, d'amorcer des échanges et des partages enrichissants, ...sans réduire à l'implicite l'environnement local et global socio-culturel des différents acteurs. Et qu'importe le développement éventuellement exagéré des documents XML, les technologies déployées actuellement sont conçues pour y faire face :

⁶² *"Les prospects n'imaginent pas qu'une telle plate-forme industrielle de traitement existe"*.
F. Bourdoncle, "Exalead : un seul moteur pour transformer une messagerie en base de connaissances" 01 Informatique, <http://www.01net.com/article/222231.html>

"Au printemps dernier, des chercheurs de l'Inria ont donné naissance à un éditeur d'entrepôt de données XML, développé par le Commissariat à l'énergie atomique. Ses performances : un teraoctet scruté en trois secondes." Emmanuelle Delsol, 01 Informatique, le 05/12/2003, <http://www.01net.com/article/224853.html>

CONCLUSION

L'information électronique, terme qui a connu l'obsolescence prématurée d'une trop grande collusion avec les technologies de même nom, se retrouve désormais sous un vocable destiné à mieux restituer les progrès de la numérisation: "les contenus numériques". Depuis que la systématisation de la numérisation a conduit les Technologies de l'Information et de la Communication à s'emparer des médias traditionnellement analogiques, il n'est plus de données qui ne soient numériques. Ainsi, le "Plus Grand Commun Numérique" permet aujourd'hui, à partir d'un poste informatique banalisé, toutes les manipulations, toutes les associations, toutes les combinaisons et toutes les créations. Ces "contenus numériques" que les Systèmes de Gestion de Contenus (Content Management System) rendent disponibles et aisément diffusables ont abandonné une forme encore jadis conditionnée par le fichier, conteneur privilégié des applications informatiques. Le fichier n'est plus désormais qu'une circonstance technique temporelle, éventuelle même, qui ne présage pas des conditions réelles d'utilisation des données qu'il enferme. Désormais, les données numériques n'ayant plus de conteneur significatif se retrouvent "in-formes". Les conditions d'utilisation induites par la multiplicité des "dispositifs-clients" de restitution (écrans de taille et de résolution différentes, PDA, TabletPC, mobile, équipements multimédia,...) exigent de fait une très haute façonnabilité de ces données numériques. Les formes peuvent désormais être multiples puisque les pierres angulaires de ce polymorphisme numérique sont opérationnelles : formats de stockage, systèmes de gestion et de diffusion, formalismes de référence, de liens et d'identification.

C'est au nom de l'interopérabilité maximum des systèmes d'informations et par voie de conséquence, la certitude d'une amélioration des situations de communication, que la Technique impose un point de vue aujourd'hui marqué par les "contenus numériques". Par la séparation du conteneur et du contenu, du fond et de la forme, les Technologies de l'Information et de la Communication finalisent un processus sophistiqué de dématérialisation de l'information. La modularité, l'extensibilité et

l'ouverture, caractérisant les contenus numériques, anticipent toutes les conditions d'utilisations, comme le souligne Vincent Quint⁶³ "*Le développement de XML et des technologies associées apporte de nouvelles possibilités dans le domaine de l'hypermédia. Ces technologies peuvent s'appliquer à travers le Web et sont ainsi mises à disposition de très nombreux utilisateurs. Elles permettent aussi de relier des ressources d'informations fortement distribuées. Représentées selon le modèle XML, ces ressources peuvent être l'objet de traitements très variés*". En ouvrant le "monde des possibles" en matière de traitement des données, les TIC suggèrent une augmentation des possibilités d'utilisation qui permettrait même d'étendre celles des usages.

Or la question des usages qui transitivement interpelle celle des usagers, ne peut être réduite à une dimension unilatéralement technologique. En effet, de la même manière que l'amélioration de la communication humaine n'est pas proportionnelle aux avantages de l'interopérabilité technique, il est encore trop tôt pour envisager que l'élimination de la forme des données numériques, sous-entend pour les usagers, des conditions d'usages améliorées.

Les recherches qui convergent aujourd'hui autour du "document numérique", dans lequel s'enchâssent également mes travaux de recherches, ne constituent pas un courant dissident aux avancées technologiques en matière de contenus numériques. Le "document numérique" qui n'exclue en rien la diversité combinatoire des contenus, exprime la relation certes fragile et néanmoins existante entre contenu et usage. La forme du document, de cette mise en forme qui n'est pas réductible à la mise en page ou à la mise en forme de caractères est dotée d'une valeur sémiotique d'usage. En fonction de la forme qu'elle revêtira ; article, rapport, note, document de travail, mémoire, livre etc., une information de même teneur sera porteuse de méta-informations qui renseignera l'utilisateur sur l'accessibilité intellectuelle, le degré de spécialisation, la cible ou l'état de finalisation.

Outre le fait que le document est présent dans toutes les démarches de capitalisation de connaissances, il est lié à des cycles de vie humains : il naît, évolue, et est exploité dans un temps qui est celui du contexte humain et du contexte social. Ainsi,

⁶³ "Hypermédia et Technologies Web", V. Quint, in [H2ptm'03]

différents processus entrent en action, ces processus étant organisés dans le temps par l'individu ou par la communauté : le temps de la conception (auteur, sources utilisés) et le temps de l'exploitation (consultation, référencement).

Le document numérique demeure un document, un objet qui tout en assurant ses dimensions sociale, cognitive et culturelle, bénéficie des avantages technologiques du numérique en matière de stockage, de diffusion, d'accessibilité, de partage et de ré-utilisation des contenus.

Mes travaux de recherche relatifs à l'hypertexte et l'hypertextualisation, à NeuroWeb, à HyWebMap et a fortiori à K-Web Organizer ont confirmé la nécessité de garantir pour les usagers la consultation (hypertexte) et la recherche d'information (SRI) au sein de fonds véritablement documentaires. Au cours de mes travaux, principalement ceux relatifs à l'hypertextualisation, même si le document subissait une déstructuration physique, le réseau hypertexte résultant préservait la structure logique globale du document initial. Cette idée du document reste présente dans HyWebMap (malgré l'approche orientée "donnée"), puisque au-delà des fonctionnalités d'agrégation offertes par le logiciel, les facilités de réorganisation des nœuds et des liens conduisent à élaborer un document structuré qui témoigne de la démarche de recherches de l'utilisateur.

C'est parce que le document, porteur d'usages multiples, est simultanément forme, signe et médium, que j'ai cherché à en maintenir la présence dans mes travaux de recherches. Lui associer les artifices des technologies du numérique (en particulier le XML), m'a paru essentiel pour augmenter le confort des usagers en matière de recherche, d'annotation, de partage, de création, etc. et faciliter le rapprochement des univers socio-informationnels des usagers-consommateurs et des usagers-producteurs.

La posture de l'utilisateur au sein de système d'informations (dans son acception la plus large) qui le conduit à produire de l'information et à en rechercher, constitue la cible de mes recherches actuelles. Mes travaux qui se développent sur deux champs recouvrants concernent des environnements informationnels où les pratiques documentaires des usagers-consommateurs sont aussi importantes que les pratiques documentaires, informationnelles et communicationnelles des usagers-producteurs ;

le premier concerne les environnements numériques de travail (ENT) liées aux plates-formes d'enseignement à distance (consultation, production de documents pédagogiques, parcours de lectures, communautés de pratiques, échanges de documents de travail, recherche,..) et le second concerne la bibliothèque universitaire (étudiants, enseignants, professionnels des bibliothèques, catalogue, accès aux collections, classification disciplinaire, structuration des savoirs,...).

Ces deux axes de recherches impliquent une utilisation à bon escient des Technologies de l'Information et de la Communication dans le sens où elle ne peut être que la suite d'observation et d'analyse des besoins informationnels des usagers en des circonstances où l'environnement intellectuel et organisationnel ainsi que les objectifs des usagers sont connus. Il ne s'agit pas ici de développer des logiques techniques auto-proclamées qui se fondent sur des modélisations d'utilisateur (le paradigme utilisateur) en rupture avec les réalités des pratiques informationnelles (usagers novices et experts) mais plutôt de proposer des améliorations techniques n'omettant pas que *"le système de pertinence d'un individu est un état psychologique de prédisposition mettant en cause le cognitif, l'affectif, le perceptif et le comportemental. Il est en fonction de l'ensemble des problèmes spécifiques qui préoccupent l'individu, des projets qu'il a, qui forment son orientation de vie au moment où on le considère"*⁶⁴.

Axe 1 : Communication médiatisée par ordinateur dans les environnements numériques de travail appliqués à la Formation à Distance

L'enseignement à distance et les environnements numériques de travail proposent un terrain expérimental particulièrement illustratif de la place que le document numérique est amené à prendre dans les interactions fortement médiées par les TIC. Interroger, communiquer, proposer, échanger, agréger, partager, collaborer, etc sont autant d'actes qui d'un point de vue cognitif et pédagogique sont déterminants pour que l'apprenant puisse tirer profit des solutions de production et de diffusion de connaissances mis à sa disposition. Mais cela relève d'une activité distincte et de

⁶⁴ A. Mucchieli, "Les Sciences de l'Information et de la Communication, 3^{ème} édition, Paris, Hachette, 2001.

longue haleine qui ne peut être envisagée que préalablement à toute situation réelle d'interactions fonctionnelles avec les dispositifs techniques.

Le document numérique des technologies idoines propose désormais un cadre d'expression formelle, creuset technologique où tout reste cependant à inventer, pour que ces interactions déterminantes viennent alimenter et enrichir les contenus pédagogiques.

Les problématiques majeures que soulèvent l'enseignement à distance relèvent de multiples dimensions liées aux modalités d'organisation de ces formations, aux structures administratives et institutionnelles nécessaires, à l'expression pédagogique et bien entendu aux technologies employées qui exigent des compétences pointues afin que les équipements matériels et logiciels demeurent disponibles auprès d'une population d'apprenants et d'enseignants susceptibles de les solliciter de façon totalement asynchrone selon des fuseaux horaires et des lieux couvrant la totalité du globe terrestre.

Ces environnements numériques de travail destinés à l'enseignement et la formation, de par les conditions effectives de fonctionnement et de suivi qu'ils imposent, ne convoquent plus de la même façon la question de l'autonomie de l'apprenant. En effet, longtemps considérée comme une problématique centrale à ce type de formation à distance, synonyme de frein à toute évolution positive de l'apprenant, cette dernière se trouve être tout simplement éliminée parce que placée en position de pré-requis pour tout candidat à cette offre de formation. C'est en fait un public affilié intellectuellement, assurément déterminé, doté de capacités à élaborer des stratégies évolutives d'acquisition et de constructions de connaissances au moyen de méthodologies acquises et adaptées au fil des expériences individuelles. C'est ce public d'apprenants prêts à s'impliquer dans cette "lunette pédagogique" de courte ou moyenne durée, qui va se trouver à compléter le paradigme initialement pédagogique de l'Enseignement à Distance en un paradigme où la capacité à opérer une sélection sur l'information, à proposer des stratégies mobiles d'acquisition de connaissances, de confrontation, d'échange et de collaboration, somme toute à la capacité à apprendre, va s'avérer déterminante.

Cette question est depuis longtemps connue des Sciences de l'Éducation qui ont montré que la construction de connaissances conduisant à la résolution de problèmes et à l'élaboration de concepts ne repose pas exclusivement sur du "texte mémorisé" qu'il suffirait d'apprendre. Apprendre met en jeu des modalités complexes où l'apprenant tout en étant guidé joue un rôle actif et déterminant. Au fil de son parcours, qui le conduit du monde scolaire à celui de l'Université, l'élève puis l'étudiant est amené à acquérir les moyens d'abandonner l'accompagnement pédagogique fort des premières années de l'école au profit d'une autonomie certainement plus exigeante, et plus rigoureuse aussi, qui s'impose au fur et à mesure de sa progression dans les formations de l'enseignement supérieur.

Les environnements numériques destinés à la formation à distance ne constituent pas une simple réplique électronique des formations "en présence". Les formations à Distance en gardent assurément l'aspect diplômant et qualifiant tout en encourageant l'individu impliqué à des confrontations intellectuelles constructives⁶⁵. Elles sont une invitation permanente à l'auto-évaluation, un encouragement constant à l'éloignement des zones d'imprécisions intellectuelles liées aux connaissances mal assimilées. Ce processus d'apprentissage actif s'exprime en une double confrontation :

- **confrontation Personne / Personne** (1 ↔ 1), **Personne / Communauté** (1 ↔ N). Les communications électroniques (messagerie, chat, forum) imposées et spontanées avec l'enseignant et le tuteur sont de multiples prétextes à l'interrogation, l'explication, la reformulation, la justification, l'argumentation, et la démonstration. Ces lieux de rencontre électroniques et de discussions thématiques (puisque liés à l'objet d'un cours, d'un devoir, d'un projet) constituent des conditions favorables pour l'apprenant d'évoluer.
- **confrontation Personne / Information**. Parce que ces lieux électroniques de formation à distance s'insèrent dans un environnement technologique commun au réseau des réseaux, parce l'apprenant toute liberté d'alimenter les concepts qui lui sont proposés dans les cours / contenus pédagogiques de la formation en ligne avec d'autres regards, d'autres points de vue, d'autres sources d'informations, qui tout en

⁶⁵ "...Avec la capacité désormais infinie de stocker le savoir sous toutes ses formes, le rôle de notre système d'éducation est à présent d'enseigner l'art et la manière de s'orienter et de naviguer dans le savoir." [Breton 95]

étant l'occasion d'exercer son sens critique, lui permettront de trouver son propre chemin vers l'appropriation des concepts étudiés. Ainsi, les sources d'informations électroniques multiples, sérieuses et authentifiées, disponibles sur Internet vont permettre à l'apprenant de se construire un véritable environnement personnel d'appropriation de connaissances, sorte de cartable électronique personnalisé.

D'un point de vue technologique, ce processus d'apprentissage actif va se concrétiser par des contenus pédagogiques rédigés par les enseignants mais adaptés par l'apprenant. Ce dernier va s'appuyer sur la structure initiale des contenus pédagogiques pour y agréger de nouvelles sources d'informations qui vont lui permettre de mieux comprendre certains points du "cours en ligne" demeurés obscurs ou difficiles à appréhender. Cette démarche agrégative, même si elle peut s'avérer insuffisante pour comprendre telle ou telle partie du cours, servira de base de travail aux enseignants et aux tuteurs afin de circonscrire plus rapidement les incompréhensions, les attentes et les interrogations de l'apprenant. La question de l'agrégation soulevée ici ne concerne pas exclusivement des documents "externes" extraits du Web, elle concerne également les annotations de l'apprenant, les conversations par messagerie, chat et forum qui peuvent être agrégés au document numérique personnalisé de l'apprenant au même titre que n'importe quelle autre ressource documentaire (pages HTML, images, fichiers PDF,...).

Navigateurs, moteurs, méta-moteurs, portails, agents intelligents,...outils de consultation, de distribution, d'exploration, d'analyse, de représentation, outils cartographique...l'éventail de solutions techniques destinées à couvrir la Toile afin d'en extraire la substantifique moelle informationnelle, et d'assister – voire même de suppléer - l'utilisateur dans ses recherches ne cessent de se développer. Et parallèlement à la conviction selon laquelle les TIC ne produisent pas de l'organisation, une tendance à penser qu'elles y contribuent, se développe également.

C'est ici que la manifestation de l'affiliation intellectuelle sera la plus flagrante, car il ne s'agira pas pour l'apprenant de subir les problèmes de désorientation et de surcharge cognitive inhérents à l'exploration hasardeuse du Web. L'écueil majeur réside dans un butinage d'informations aléatoire et permanent, par un sujet aux intentions floues, dans un ensemble de savoirs constamment non-structurés. Dans ce

cadre, le problème des "parcours", c'est-à-dire celui de la construction, par les instruments d'appropriation des connaissances, de savoirs personnels est un problème crucial.

C'est en cela qu'une méthodologie dans la recherche d'informations qui inclut la nécessité d'une analyse critique des sources consultées et que l'on peut observer chez les étudiants affiliés sera d'une aide précieuse à l'apprenant afin qu'il ne perde pas ses objectifs initiaux et qu'il insère bien cette recherche d'informations dans le "processus parent" de la compréhension d'un cours.

Dans ce contexte, les environnements informatiques, outre le fait qu'ils transposent des situations d'apprentissage socialement et physiquement reconnues en des situations virtualisées et stéréotypées, introduisent des logiques techniques spécifiques qui favorisent peu l'acquisition et la construction de connaissances. Or il n'est pas possible d'ignorer les dérapages technologiques qui montrent que les productions d'artefacts informatiques opèrent une rupture fondamentale au sein des groupes sociaux.

Les plates-formes d'enseignement à distance illustrent parfaitement cet état de fait puisque la plupart d'entre elles, commercialisées ou non se révèlent être de complexes boîtes à outils qui conduisent les responsables de formation en ligne et les enseignants à gérer les logiques contradictoires de l'outil dans le respect de ce qui a été défini par les équipes pédagogiques.

A partir du double constat de la croissance continue des technologies et de la persistance des difficultés d'appropriation des utilisateurs, c'est une définition renversée des technologies qui se développe, c'est-à-dire non pas depuis les possibles techniques mais depuis leur usage potentiel. En ce sens les outils matériels et symboliques, et plus largement les technologies, vont être considérés comme des artefacts c'est-à-dire des outils culturels et sociaux qui sont élaborés au travers des générations et transmis au sein des communautés. Ils sont des médiateurs de l'activité et deviennent dans l'usage des instruments au service des activités des sujets et des finalités qu'il y construisent. Trois médiations principales seront distinguées:

- les médiations à l'objet de l'activité, orientées vers la connaissance du monde, de ses caractéristiques et de ses propriétés (médiation épistémique) et vers l'action et la transformation (médiation pragmatique) ;
- les médiations orientées vers les rapports que le sujet entretient avec lui-même (médiations réflexives) ;
- les médiations orientées vers les rapports médiés à autrui (médiations interpersonnelles), inscrites dans les contextes collectifs de l'activité.

De l'usage à l'appropriation des artefacts, l'activité des sujets est un développement où les dimensions d'atteinte de buts et de réalisation de tâches de l'activité productive s'articulent aux dimensions d'élaboration de ressources externes et internes de l'activité constructive. Activité productive et constructive sont les deux faces de l'activité humaine qui s'enrichissent et se transforment réciproquement : une difficulté rencontrée au plan productif pourra donner lieu à un développement de ressources au plan constructif qui, en retour, va modifier les formes et les conditions des activités productives.

Ce champ de recherches implique une interdisciplinarité très forte qui rassemble Sciences de l'Information et de la Communication, Psychologie Ergonomique et Sciences de l'Education.

Axe 2 : La bibliothèque Universitaire. Lieux d'usages pour des communautés de pratiques distinctes

Les bibliothèques universitaires, indépendamment de toute considération bibliothéconomique relative à la performance de ce lieu ressources (fréquentation, fonds documentaires disponibles, adéquation des services aux besoins des usagers, trajets, mouvements, comportements, arrangement et localisations des collections ...) constituent un espace destiné principalement à l'enseignement supérieur et à la recherche, où des pratiques complexes liées à l'information (archivage, promotion, diffusion) et à la communication sont en jeu, directement ou indirectement, entre des communautés distinctes d'usagers, des connaissances et des informations (description des connaissances).

Les usagers dont il est question ici relèvent de deux grandes communautés elles-mêmes composites ; la première regroupant les professionnels de la bibliothèque

(conservateur, bibliothécaires, assistant, catalogueur, acquéreur...), la seconde constituée d'un public d'étudiants (formation initiale et formation permanente tous cursus) et d'enseignants.

Ce lieu-ressources a pour mission de mettre à disposition des étudiants, en cours d'affiliation intellectuelle (DEUG, licence) ou déjà rompus aux méthodes de travail du "métier d'étudiant" (maîtrise et au-delà), des ouvrages (livres, articles, encyclopédies, périodiques, bases de données on-line et *off-line*,...) en correspondance avec les objectifs des formations diplômantes dans lesquels ils sont inscrits. La particularité disciplinaire de l'université, lettres et sciences humaines, sciences exactes, sciences juridiques et économiques, sciences médicales, etc. va ainsi déterminer les propositions de collections en phase avec le secteur disciplinaire de l'établissement, dont le volume peut atteindre le million d'exemplaires⁶⁶ (Sorbonne, Cujas, Sainte-Geneviève, Jussieu, Institut de France,..). Les fonds documentaires qui constituent les collections sont alimentés par le personnel qualifié et spécialisé des bibliothèques (conservateurs, bibliothécaires) et par les propositions des enseignants-chercheurs. Ces collections, compte tenu de la variabilité des parcours universitaires doivent répondre à des besoins et des attentes extrêmement larges d'un public d'usagers les utilisant pour les enseignements (donnés, suivis) et pour la recherche (doctorants, enseignant-chercheur).

La bibliothèque universitaire regroupe un univers d'usages et de pratiques d'une rare complexité en correspondance avec la dimension de la tâche à accomplir qui consiste, ni plus ni moins à proposer une organisation et une structuration de tous les savoirs et de toutes les connaissances, tout en préservant l'accessibilité des documents conservés.

Au sein de cet espace, des systèmes de classification sont adoptés pour structurer les connaissances (CDU, Dewey, Inter Marc, ...), et des modalités pratiques de description sont définies pour établir une fiche signalétique de chaque exemplaire conservé (prêt, localisation, accessibilité, sujet-matière,...).

Les fonds bibliographiques des bibliothèques sont établis sur des modèles d'organisation du savoir telles les classifications majeures que sont la Dewey et la

⁶⁶ "Bibliothèques Universitaire : Mode d'Emploi", <http://www.sup.adc.education.fr/bib/>

CDU. Elles incarnent des archétypes de l'organisation de la connaissance. Mais cette structuration des savoirs et les grandes classifications qui sous-tendent le fonds bibliographiques sont des élaborations intellectuelles non directement liées aux usages.

Ce sont des structurations idéalement conceptuelles auxquelles les recherches bibliographiques menées communément ne font appel que partiellement, dans la mesure où les proximités sémantiques des grandes thématiques de classification ne sont pas nécessairement envisagées par les usagers des bibliothèques.

Les principes d'organisation communs à toutes les bibliothèques universitaires, les règles de fonctionnement, les systèmes de classification et de description normalisés rendent l'univers de la bibliothèque rapidement opaque pour les usagers non initiés.

Celui-ci exige des usagers, pour s'y mouvoir physiquement et intellectuellement, un degré de représentation et de conceptualisation élevé tant les objets (livres, revues, mémoires, bases de données,...), leurs propriétés (magasin, salle, PEB, exclu du prêt, support multimédia,...), les relations existant entre eux (classifications, Rameau, Thésaurus) sont loin d'être triviaux.

Aujourd'hui, l'accès le plus direct à l'ensemble des services proposés aux usagers est offert par le catalogue des notices bibliographiques. Informatisés depuis quelques années, ces catalogues sont devenus des *Online Public Access Catalog* (OPAC^o) consultables à partir d'un navigateur Internet banalisé.

Le catalogue informatisé seul dispositif opératoire, actif de l'intérieur ET de l'extérieur de l'espace physique de la bibliothèque et de l'université, constitue un point de mire et la synecdoque de la bibliothèque. Le temps de la consultation, le catalogue en ligne focalise les besoins informationnels des usagers transformés pour l'occasion en utilisateur. Le temps des sollicitations opérées au moyen de l'interface de consultation, l'utilisateur tente de concilier les objectifs de ses recherches, la médiation technique imposée par l'interface et les modes d'organisation des connaissances (souvent imaginés, fréquemment ignorés) au sein de la bibliothèque.

Ce catalogue représente de fait un objet-pivot, qui ne réalise pas uniquement une simple médiation technique entre un usager contraint d'exprimer par des mots une attente informationnelle complexe, et un système de gestion de fiches

bibliographiques. Il est le lieu de contact entre deux mondes régis par des objectifs, des règles sociales, culturelles et cognitives bien distinctes.

Aujourd'hui, les OPAC développés par des sociétés privés, sont avant tout des réalisations informatiques permettant de donner un accès aux ressources documentaires proposées par la bibliothèque, et ce, indépendamment du contexte socio-informationnel.

Les travaux de recherches que j'ai engagés avec une équipe interdisciplinaire que j'ai constituée (Sciences de l'Information, conservateurs, psychologue-ergonomes, géographe-cartographe)⁶⁷ s'appuient sur l'hypothèse que les recherches des usagers - et les services qu'ils sont susceptibles de solliciter – peuvent être améliorées si l'univers socio-informationnel de la bibliothèques et les règles d'usages des professionnels sont rendus perceptibles.

Les Interfaces Homme Connaissances exploitant navigation associative et représentation graphique dynamique⁶⁸ sont une expression des TIC qui devrait permettre de répondre à cette hypothèse par l'affirmative. Il s'agit cependant de ne pas minimiser le fait que l'enrichissement de moyen technique complexifie la tâche des usagers. Sur ce point, la participation active des psychologues-ergonomes, soucieux de transformer les situations par une amélioration des instruments devrait limiter les dérapages des logiques instrumentales.

Pour la bibliothèque, ces travaux vont permettre de mieux connaître ses publics afin d'ajuster sa politique de services et son ingénierie documentaire aux besoins nouveaux, mais aussi de saisir l'opportunité d'un travail de recherches pour faire développer des outils expérimentaux s'inscrivant dans sa politique générale de modernisation des accès à l'information d'une part, et d'évaluation des services d'autre part (fonctionnalités de gestion nouvelles en matière d'indexation CDU pour les agents de la Bibliothèque, évolution de la liste de cotes validées).

⁶⁷ <http://visualcatalog.univ-paris8.fr>

⁶⁸ Un système de visualisation est envisagé pour synthétiser au moyen d'une "métaphore graphique" les résultats de recherche résultant de l'interrogation du fonds des notices bibliographiques à partir d'un formulaire de saisie sélective (titre, auteur, sujet RAMEAU,...). Une lecture hypertextuelle des résultats de recherche organisés en liste ternaire (notice, sujet rameau, CDU) permettra de mettre en évidence les relations ternaires.

La généralisation des services de type "Bibliothèques Numériques", qui passent invariablement par des catalogues en ligne de plus en plus fournis (le SUDOC propose plus de 5 millions de références, la BNF plus de 7 millions), banalise l'accès à des modes d'organisation des connaissances qui demeure d'une rare complexité en regard à l'immensité de la tâche. Cette mécanisation logicielle systématique des systèmes d'informations, induite par des Technologies de l'Information et de la Communication héritées du Web, s'accompagne d'un "allant de soi" en matière de manipulations techniques qui tend à déteindre sur les habiletés intellectuelles sous-tendues. Il ne faut pas oublier que ces systèmes d'informations ne sont en définitive que les ingrédients destinés à agrémenter un processus long et exigeant de transformation des individus, de compréhension et de connaissances de soi, des autres, et des systèmes - quels qu'ils soient - dans lesquels ils évoluent par choix ou par contrainte. Dans ce dessein et indépendamment de toute considération technologique, la capacité de l'individu à appréhender et idéalement à s'approprier ces informations, est loin d'être acquise. Elle repose sur des habiletés, des stratégies, des savoir-faire qu'il revient à chacun d'adapter et de puiser au fil des expériences sociales, culturelles et cognitives qu'il est amené à vivre.

La capitalisation numérique de ces expériences inestimables au moyen de systèmes automatiques risque de devenir rapidement inaccessible pour le plus grand nombre, si la dimension instrumentale des systèmes continue à être envisagée comme le seul facteur de promotion.

Le risque d'une fracture définitive est à craindre entre les déposants et les dépositaires. En effet, les savoirs humains conservés par le truchement de logiques dédiées au machinisme et à la procéduralité, structurés par les exigences technologiques, vont rapidement s'avérer inexploitable pour les moins familiers d'entre nous aux logiques artefactuelles.

De tels enjeux imposent que des voies d'équilibre soient ouvertes ou préservées afin que la médiation sociale liée aux situations informationnelles, souvent absente en raison d'improbables représentations, ne soit pas pour autant assujettie à l'écrasante médiation technique.

ANNEXE A

Exercice de modélisation :

Le circuit des documents au centre de documentation de la Section Française d'Amnesty International

(M. Lomnitz, M. Siccardi)

**DESS CONCEPTION ET GESTION DES RESEAUX
D'INFORMATION ET DE DOCUMENTATION**

Cours : Ingénierie des connaissances

Vocabulaire et Abréviations :

AI : Amnesty International.

SF : Section Française d'Amnesty International (Paris)

SI : Section Internationale d'Amnesty International (siège social : Londres)

AI Index : Il s'agit du nom de la cote attribuée aux documents d'Amnesty International. Exemple : AMR 29/006/00

ÉFAI : Edition Francophones d'Amnesty International (ce sont les éditions qui traduisent les documents d'Amnesty International de l'anglais au français)

Consultants : Terme utilisé au centre de documentation pour désigner les utilisateurs externes venant consulter sur place les documents.

OFPRA : Office Français pour les Réfugiés et les Apatrides

Coter : Il s'agit de l'action consistant à attribuer une cote (AI Index), à un document.

Doc VI : Documents Version International

Doc VF : Documents Version Française

1. DESCRIPTION DE LA REALITE

1.1. Les champs de compétence et les missions

Le centre de documentation d'Amnesty International est spécialisé dans les droits de l'homme, au niveau international. La documentation d'AI porte sur l'aspect juridique des droits de l'homme mais aussi sur les défenseurs des droits de l'homme, sur l'éducation aux droits humains...Elle porte également sur le mouvement d'AI dans son intégralité. Son fonds est majoritairement composé de documentation directement produite par AI.

Le centre de documentation d'AI a pour mission de conserver et de communiquer tous les documents produits par la Section Française, une partie des documents produits par la Section Internationale, ainsi que d'autres documents hors de la production d'AI, afin de faire connaître les violations des droits de la personne humaine dans le monde, de sensibiliser tout public aux droits de l'homme et de faire découvrir AI et son action au plus grand nombre. Le centre de documentation d'AI permet aussi à l'interne de répondre à toutes les demandes d'information des salariés dans le cadre de l'exercice de leurs fonctions. Le centre de documentation d'AI travaille notamment beaucoup avec le service de communication.

1.2. Les consultants du centre de documentation d'AI

Le public du centre de documentation d'AI se compose de trois types de personnes :

- Le personnel d'AI de tous les départements et de tous les pays qui sollicitent régulièrement la documentation du centre, que ce soit pour des études de cas, des projets de manifestation, des renseignements généraux, des recherches biographiques ou autres, des demandes de dossiers spécifiques ...
- Le personnel d'autres associations, des journalistes, des avocats qui demandent des informations pour leur métier (article de presse sur la violation des droits de l'homme dans un pays, procès d'une victime défendue par AI...) ou pour faire avancer leur mouvement (associations homosexuelles qui veulent sensibiliser l'opinion sur la condition des « gays » dans un pays...).
- Des consultants qui se déplacent au centre de documentation et consultent directement les documents d'AI (le lundi et le mercredi de 14h00 à 18h00). Ce sont souvent des réfugiés politiques qui ont besoin de documentation pour monter leurs dossiers de demandeurs d'asile pour l'OFPRA, ou bien des personnes ayant subi des violations dans leurs pays et qui font des recherches sur des prisonniers ou des personnes disparues de leur entourage. Les consultants sont aussi souvent des étudiants qui ont des travaux à faire sur les droits de l'homme, la peine de mort, la torture, les réfugiés ou des membres de groupes d'AI qui ont besoin d'information pour des actions...

1.3. Le Fonds documentaire d'Amnesty International

La documentation d'AI SF est composée de plusieurs types de documents. Pour notre exercice, nous allons nous limiter à la présentation de la documentation de base d'AI: les documents pays. Il s'agit de documents concernant des zones géographiques et des thématiques spécifiques. Ces documents sont les plus nombreux et les plus représentatifs des activités et des missions d'AI.

Les documents pays sont tout d'abord conçus à Londres à la SI d'AI. Ils sont tous rédigés en anglais. Une fois les documents originaux produits, ils sont indexés et une cote leur est attribuée. Le document original doté d'une indexation et d'une cote est appelé Document Version International (Doc VI). La cote des Doc VI est appelée AI index. Elle est composée de la façon suivante : elle débute par les trois premières lettres du continent concerné dans le document, puis, elle indique l'indice du pays traité, puis elle donne le numéro du document dans la série et enfin de l'année de parution du document. Exemple : AMR 45/04/99 (document portant sur l'Amérique, plus précisément le Paraguay : pays n°45, c'est le quatrième document paru sur ce pays en 1999).

Voici une liste d'exemples de cotes de documents pays :

Type de Documents	Cote AI SI
Afrique	Exemple : AFR 12/04/01 Ce document concerne l'Afrique et plus particulièrement l'Angola et il est le 4 ^{ème} document paru sur ce pays en 2001.
Amérique	Exemple : AMR 12/23/99
Asie	Exemple : ASA 30/21/01
Europe	Exemple : EUR 43/12/98
Moyen orient et Afrique du nord	Exemple : MDE 24/36/01

En plus de leur cote, les documents pays sont systématiquement indexés selon leur thématique repris selon les termes du thésaurus d'AI. Voici la liste des différents termes d'indexation utilisés pour Indexer les documents pays :

- Homosexualité et droits humains
- Juridique
- Médicale
- Education et droits humains
- Enfants
- Entreprise
- Femmes
- Militaires sécurité pénitentiaire
- Peine de mort
- Police militaires pénitentiaires
- Philosophies et religions
- Spectacle

- Syndicats
- Torture
- Institutions et finances internationales
- Finances internationales

Une fois les documents pays produits par la SI indexés et cotés, ils sont envoyés directement à la SF ainsi qu'aux Édition Francophones d'Amnesty International. Les ÉFAI traduisent alors tous les documents et rajoutent la date de traduction et le nom du traducteur. Les ÉFAI traduisent en français le contenu ainsi que le titre des documents. Une fois traduits, les Doc VI sont alors appelés Doc VF. Une fois traduits, les Doc VF sont envoyés au centre de documentation d'AI SF.

Chaque nouveau document qui arrive au centre de documentation est systématiquement enregistré par les documentalistes. Les documents issus du « Weekly Mailing » et des ÉFAI, sont enregistrés dans une base de donnée commune nommée « Fichier central AI SF ». C'est à partir de ce fichier que les documentalistes effectuent des recherches d'information pour les consultants externes et internes. Tous les documents reçus et intégrés dans le fichier central sont appelés « Documents Finaux », car ils sont à la fin du cycle du circuit des documents d'AI, et ne seront plus modifier.

1.4. La commande de documents par les consultants internes

Les demandes de documentation se font par courrier, par téléphone, par fax ou par E Mail. Il y a plusieurs types de demandes et plusieurs types de réponses. Nous allons, pour notre devoir, nous concentrer sur les demandes effectuées par mails émanant des consultants à l'interne (ayant déjà un formulaire identifiant remplis et connus par le centre de documentation), et concernant les documents pays d'AI.

La commande d'information par des consultants internes se fait généralement selon la procédure suivante :

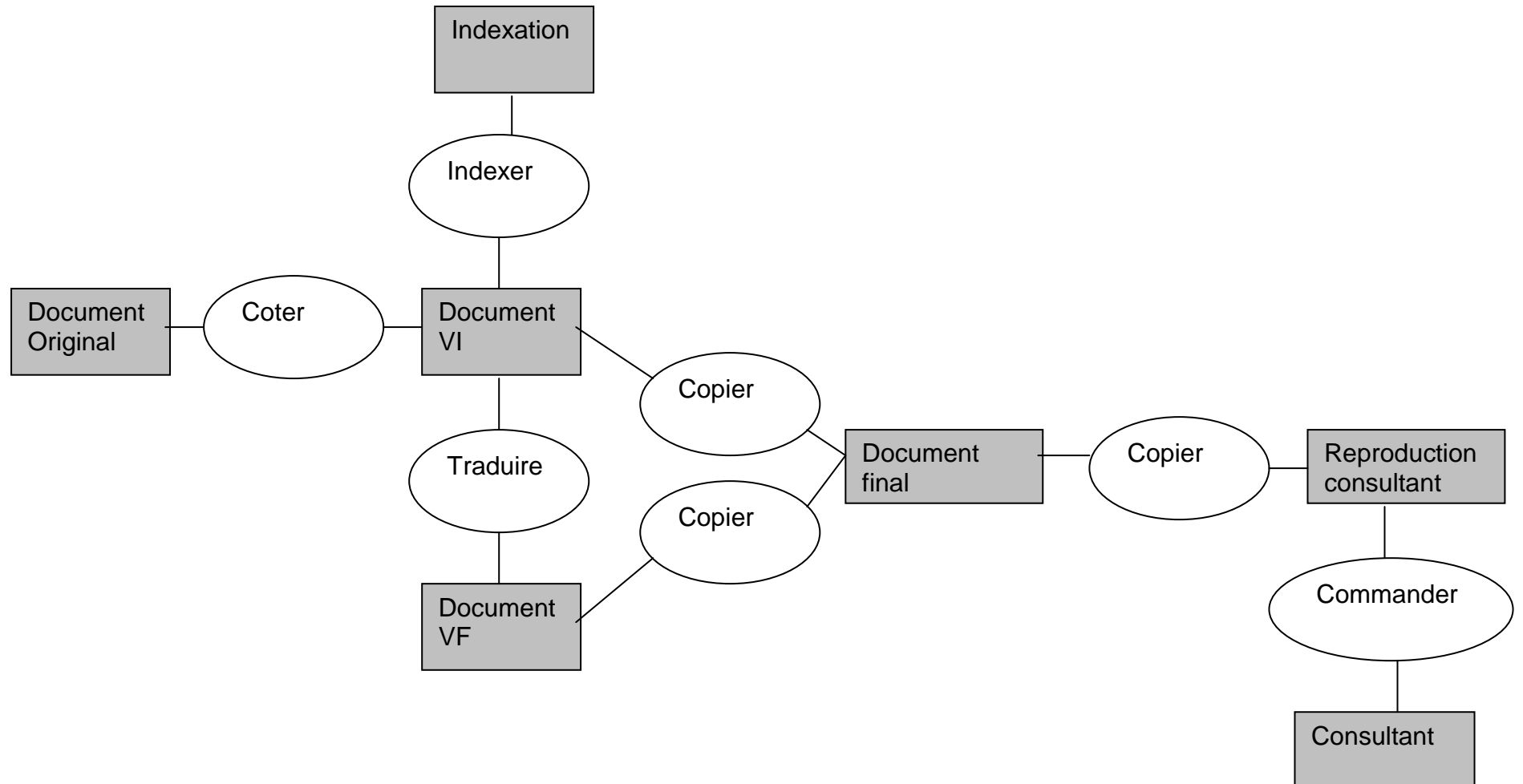
- Le consultant interne envoie un mail ou appelle le centre de documentation, et formule sa commande. Par exemple : « Je désire tous les documents concernant les enfants soldats au Rwanda entre 1994 et 1995, mais seulement en français ».
- Une fois la commande effectuée, les documentalistes vérifient le service et les coordonnées du consultant interne et s'assurent qu'il est bien enregistré et qu'il possède un identifiant (c'est utile par la suite pour les statistiques).
- Une fois la fiche identifiant des consultants vérifiée, la commande est traitée par les documentalistes, qui interrogent la « le fichier central d'AI ».
- Une fois la commande traitée, les documentalistes envoient une copie numérique des documents sélectionnés pour le consultant interne par mail.

NB : Les commandes de documents à l'interne sont gratuites et peuvent concerner tous les types de documents disponibles.

1.5. Le circuit des documents d'Amnesty International

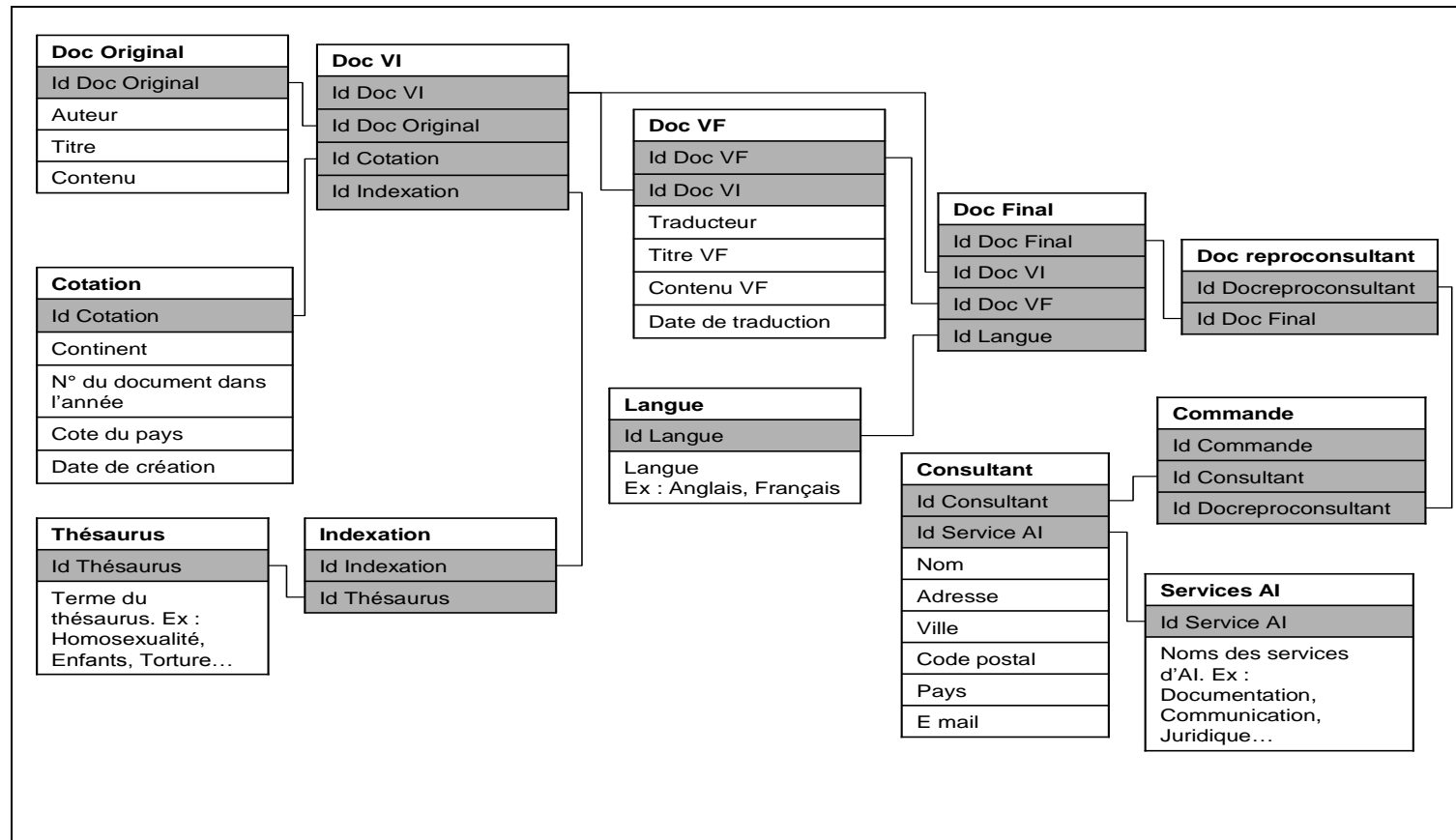
Nous avons axé la modélisation sur le cas du circuit des documents pays depuis leur création à la SI jusqu'à leur réception et traitement à la SF, puis jusqu'à la réponse à une commande pour un consultant interne.

2. MODELE CONCEPTUEL



3. MODELE RELATIONNEL

3. 1. SCHEMA RELATIONNEL



NB : En gris = les clés primaires

3.2. TABLES

DOCORIGINAL (Id-original, Auteur, Titre, Contenu)

DOCVI (Id Doc VI, Id Doc Original, Id Cotation, Id Indexation)

DOCVF (Id Doc VF, Id Doc VI, Traducteur, Titre VF, Contenu VF, Date de traduction)

DOCFINAL (Id Doc Final, Id Doc VI, Id Doc VF, Id Langue)

COTATION (Id cotation, Continent, N° du document dans l'année, Cote pays, Date de création)

INDEXATION (Id Indexation, Id Thésaurus)

LANGUE (Id langue, Langue)

THESAURUS (Id Thésaurus, Terme du thésaurus)

DOCREPROCONSULTANT (Id Docreproconsultant, Id Doc Final)

CONSULTANT (Id Consultant, Nom, Adresse, Code postal, Ville, Pays, Service AI, Email)

COMMANDE (Id Commande, Id Client, Id Docreproconsultant)

SERVICE AI (Id Service AI, Service AI)

(NB : Id = Identifiant = Clés primaires)

4. QUESTIONS

4.1. QUESTION n°1 : Quels sont les documents Version Internationale concernant l’Afrique ?

En langage naturel :

Quels sont les documents Version Internationale concernant l’Afrique ?

En algèbre relationnel :

R1 = SELECTION (COTATION ; Continent = « Afrique »)

R2 = JOINTURE (R1, DOC VI)

R3 = JOINTURE (R2, DOC ORIGINAL)

R4 = PROJECTION (R3 ; auteur, titre, contenu)

En SQL :

```
SELECT auteur, titre, contenu
FROM COTATION, DOC VI, DOC ORIGINAL
WHERE Continent = « Afrique »
```

Réponse :

Auteur	Titre	Contenu
Charles Rupert	Angola : women’s struggle against torture	Bla Bla Bla Bla Bla Bla Bla Bla
Carlos Jordan	A Rwandan journalist in jail without lawsuit	Bla Bla Bla Bla Bla Bla Bla Bla
Mourhad Benchlik	Liberia : scoffed human rights	Bla Bla Bla Bla Bla Bla Bla Bla
John Carpenter	Congo : Several attacks against a gay	Bla Bla Bla Bla Bla Bla Bla Bla

4.2. QUESTION n°2 : Quels sont tous les documents en langue française ?

En langage naturel :

Quels sont tous les documents en langue française ?

En algèbre relationnel :

R1 = SELECTION (LANGUE ; Langue = « français »)

R2 = JOINTURE (R1, DOC FINAL)

R3 = JOINTURE (R2, DOC VF)

R4 = JOINTURE (R3, DOC ORIGINAL)

R5 = PROJECTION (R4, auteur, titre, contenu, traducteur, titre VF, contenu VF)

En SQL :

SELECT Auteur, Titre, Contenu, Traducteur, Titre VF, Contenu VF

FROM DOC FINAL, DOC VF, DOC ORIGINAL

WHERE Langue = « français »

Réponse :

Auteur	Titre	Contenu	Traducteur	Titre VF	Contenu VF
Charles Rupert	Angola : women's struggle against torture	Bla Bla Bla Bla Bla Bla Bla Bla	Philippe Marais	Angola : les femmes se battent contre la torture	Bla Bla Bla Bla Bla Bla Bla Bla
Carlos Jordan	A Rwandan journalist in jail without lawsuit	Bla Bla Bla Bla Bla Bla Bla Bla	Delphine Dupont	Un journaliste Rwandais emprisonné sans procès	Bla Bla Bla Bla Bla Bla Bla Bla
Mourhad Benchlik	Liberia : scoffed human rights	Bla Bla Bla Bla Bla Bla Bla Bla	Bernadette Lamy	Libéria : Droits de l'homme bafoués	Bla Bla Bla Bla Bla Bla Bla Bla
Rebecca Smith	Children-soldiers in Colombia	Bla Bla Bla Bla Bla Bla Bla Bla	Philippe Marais	Enfants soldats en Colombie	Bla Bla Bla Bla Bla Bla Bla Bla
Robert Evans	Torture and prostitution in Thailand	Bla Bla Bla Bla Bla Bla Bla Bla	Benjamin Blin	Torture et prostitution en Thaïlande	Bla Bla Bla Bla Bla Bla Bla Bla

4.3. QUESTION n°3 : Quels sont les documents créés en 2001 ?

Il s'agit des documents dont la cote comprend la date de création 2001. Cette cote est établie au niveau des DOC VI.

En langage naturel :
Quels sont les documents VI créés en 2001 ?

En algèbre relationnel :
R1 = SELECTION (COTATION ; Date de création = « 2001 »)
R2 = JOINTURE (R1, DOC VI)
R3 = JOINTURE (R2, DOC ORIGINAL)
R3 = PROJECTION (R2, Auteur, Titre, Contenu, Date de création)

En SQL :
SELECT Auteur, Titre, Contenu, Date de création
FROM COTATION, DOC VI, DOC ORIGINAL
WHERE Date de création = « 2001 »

Réponse :

Auteur	Titre	Contenu	Date de création
John Carpenter	Congo : Several attacks against a gay	Bla Bla Bla Bla Bla Bla Bla Bla	2001
Abou N'guyen	Children tortured and rapped in Costa Rica for medical tests	Bla Bla Bla Bla Bla Bla Bla Bla	2001

4.4. QUESTION n°4 : Quels sont tous les documents concernant la torture ?

En langage naturel :
Quels sont tous les documents concernant la torture ?

En algèbre relationnel :
R1 = SELECTION (THESAURUS ; Terme du thésaurus = « torture »)
R2 = JOINTURE (R1, DOC ORIGINAL)
R3 = PROJECTION (R2, Auteur, Titre, Contenu, Terme du thésaurus)

En SQL :
 SELECT Auteur, Titre, Contenu, Terme du thésaurus
 FROM THESAURUS, DOC ORIGINAL
 WHERE Terme du thésaurus = « torture »

Réponse :

Auteur	Titre	Contenu	Terme du thésaurus
Charles Rupert	Angola : women's struggle against torture	Bla Bla Bla Bla Bla Bla Bla Bla	Torture
John Carpenter	Congo : Several attacks against a gay	Bla Bla Bla Bla Bla Bla Bla Bla	Torture
Robert Evans	Torture and prostitution in Thailand	Bla Bla Bla Bla Bla Bla Bla Bla	Torture
Abou N'guyen	Children tortured and rapted in Costa Rica for medical tests	Bla Bla Bla Bla Bla Bla Bla Bla	Torture

4.5. QUESTION n°5 : Quels sont les documents créés en 2001 concernant l'Amérique Latine ?

En langage naturel :

Quels sont les documents créés en 2001 concernant l'Amérique Latine ?

En algèbre relationnel :

R1 = SELECTION (COTATION ; Cote Pays = « AMR » ET Date de création = « 2001 »)

R2 = JOINTURE (R1, DOC ORIGINAL)

R3 = PROJECTION (R2, Auteur, Titre, Contenu, Date de création, Cote pays)

En SQL :

```
SELECT Auteur, Titre, Contenu, Date de création, Cote pays
```

```
FROM COTATION, DOC ORIGINAL
```

```
WHERE Cote Pays = « AMR » AND Date de création = « 2001 »)
```

Réponse :

Auteur	Titre	Contenu	Date de création	Cote Pays
Rebecca Smith	Enfants soldats en Colombie	Bla Bla Bla Bla Bla Bla Bla Bla	2001	AMR
Abou N'guyen	Children tortured and rapped in Costa Rica for medical tests	Bla Bla Bla Bla Bla Bla Bla Bla	2001	AMR

BIBLIOGRAPHIE

A

[ACM 88], "Hypertext 88", ACM, Vol. 31 (7), July 1988

[Acquier 02], R. Acquier, Les Amphis de France 5, série "Des méthodes pour apprendre", discipline Sciences de l'éducation, Univ. Montpellier 1, 2002, <http://www.canal-u.fr>

[Aitchinson 92], J. Aitchison., A. Gilchrist, "Construire un thesaurus : manuel pratique", ADBS, Paris, 1992. 217 p.

[Andrieu 00], O. Andrieu, "La folie Google s'empare du Web", ADIT, Technologies Internationales, n°69, novembre 2000, pp. 38 – 41

[Andrieu 01], O. Andrieu, "Quelle taille pour le Web ?", ADIT, Technologies Internationales, n°71, février 2001, pp. 25 - 28

[Aristote 02], "La Nouvelle interaction Homme-Environnement", séminaire X Aristote, Ecole Polytechnique, 6 juin 2002

B

[Ballay 99], J-F. Ballay, "Capitaliser et transmettre les savoir-faire de l'entreprise", Eyrolles, Collection de la DER d'EDF, Paris, 1999

[Balpe 03], C. Balpe, "De l'évolution des croyances et des pratiques professionnelles", Actes du colloque inter-IUFM Professionnalité et formation des enseignants des 15 au 17 avril 2002 de Bordeaux, IUFM d'Aquitaine, à paraître 2003.

[Balpe 96], J-P. Balpe, A. Lelu, I. Saleh, **F. Papy**, "Techniques avancées pour l'hypertexte", Editions Hermès, Paris 1996.

[Balpe 97], J-P. Balpe, "Technologies numériques et construction du savoir", <http://hypermedia.univ-paris8.fr/jean-pierre/articles/Technonum.html>, 1997

[Baltz 98], C. Baltz, "Une culture pour la société de l'information ? Position, théorique, définition, enjeux", Documentaliste – Sciences de l'information, Vol. 35, n°2, 1998, pp. 75-82

[Baltz 03], C. Baltz, "Quand la documentation s'éveillera...", Documentaliste – Sciences de l'information, Vol. 40, n°2, avril 2003, pp. 148-153

- [Barthes 66], R. Barthes, "L'analyse structurale du récit", École pratique des hautes études, Centre d'études des communications de masse, Paris, Seuil, 1966
- [Belisle 99], C. Belisle, "La navigation hypermédia : un défi pour la formation à distance", Journal of Distance Education, 1999
- [Ben Hamadou 01], A. Ben Hamadou., "Elaboration d'outils pour le filtrage d'informations et la génération automatique de résumés", CIDE 2001, 4ème colloque Int. sur le Document Electronique. Méthodes démarches et techniques cognitives, Toulouse, 23-25 Oct. 2001.
- [Bernat 98], L. Bernat, "La métamorphose du documentaliste", Audiovisuel Communication et Société, n°4, décembre 1998, pp 77-83
- [Bihanic 03], D. Bihanic, "Les hypermédias graphiques explorateurs, Les hypermédias cartographiques", JFT'2003.
- [Bodengraven 03], M. van Bodengraven, C. Politt "Rendre les sites Web et les OPAC accessibles", 69th IFLA General Conference and Council, 1-9 August 2003, Berlin.
- [Boucheix 01], J-M. Boucheix, "Nouvelles Technologies et aide à la compréhension de documents techniques", Hypermédias et apprentissages, Actes du 5ème colloque, Grenoble, 9, 10 et 11 avril 2001, pp 275-282
- [Bougnoux 01], D. Bougnoux, "Introduction aux sciences de la communication", La découverte, 2001, Paris.
- [Bouhai 02], N. Bouhaï, "Lire, réécrire et partager le savoir sur le web : problèmes et solutions", Thèse de doctorat en Sciences de l'Information et Communication, Université Paris 8, 2002
- [Boullier 00], D. Boullier, L. René, "Le travail d'institution dans l'enseignement à distance", actes du colloque int. TICE 2000, Troyes, 18-20 octobre 2000, pp. 245-251
- [Boullier 00b], D. Boullier, "La Loi du support : leçons de trois ans de formation à distance", in l'Université Virtuelle sous la direction de Henry Sanier, Les cahiers du numérique, volume 1 – n°2, 2000
- [Bourda 01], Y. Bourda Y., "Objets pédagogiques, vous avez dit objets pédagogiques ?", Cahiers GUTemberg, n°39-40, Mai 2001, pp. 71-79
- [Bourdoncle 00], F. Bourdoncle, P. Bertin, "Recherche d'aiguilles dans une botte de liens", La recherche, n°328, février 2000
- [Bourigault 92], D. Bourigault, "LEXTER : un Logiciel d'EXtraction de TERminologie", 2ème colloque international de TermNet, Avignon, Juin 92
- [Bourigault 94], D. Bourigault, "LEXTER : un Logiciel d'EXtraction de TERminologie. Application à l'acquisition des connaissances", Thèse de doctorat, EHESS, Juin 94.
- [Bradley 92], N Bradley, "SGML Concepts", Aslib Proc., vol. 44 (7/8), July/August 1992, pp. 271-274

[Bretelle-Desmazières 99], F. Bretelle-Desmazières, A. Coulon, C. Poitevin, "Apprendre à s'informer : une nécessité. Evaluation des formations à l'usage de l'information dans les universités et les grandes écoles françaises", Laboratoire de Recherches Ethnométhodologiques, février 1999

[Breton 95], P. Breton, "L'utopie de la communication", cahiers libres/essais, éditions La découverte, 1995

[Brown 88], P.J. Brown, "Hypertext : the way forward", éditions J.C. Van Vliet, Document manipulation and typography, Proc. of the International Conference on Electronic Publishing, Nice, avril 1988, pp.183-191.

[Brown 89a], P.J. Brown, "Do we need maps to navigate round hypertext documents ?", Electronic Publishing, Vol 2 (2), July 89, pp 91-100

[Brown 89b], P.J. Brown, "Linking and searching within hypertext", Electronic Publishing, vol. 1 (1), avril 1988, pp. 45-53

[Bui Thi 03], M.P. Bui Thi, "La structuration sémantique des contenus des documents audiovisuels selon les points de vue de la production", Thèse de doctorat en Sciences de l'Information et de la Communication, Université Paris 8, 26 juin 2003

[Buci-Glucksmann 02], C. Buci-Glucksmann, "La folie du voir : une esthétique du virtuel", Paris : Galilée, 2002, 271 p.

[Buci-Glucksmann 03], C. Buci-Glucksmann (s/dir), "L' art à l'époque du virtuel", Paris : L'Harmattan, 2003, 243 p

C

[Campos 03], M. Campos, "The progressive construction of communication. Toward a model of cognitive networked communication and knowledge communities", Canadian Journal of Communication, Vol. 28, 2003, pp. 291-322.

[Carbonell 04], N. Carbonell, " Interfaces adaptatives : modèles de l'utilisateur", journée d'étude "Interfaces adaptatives ", Laboratoire Paragraphe et CITU, 17 juin 2004. <http://h2ptm.univ-paris8.fr/ia2004/>

[Card 99], K. Card Stuart, D. Mackinlay Jock, B. Shneiderman, "Information Visualization, Using Vision to Think"; Morgan Kaufmann Publishers, Inc., 1999.

[Cartable 02], "Le cartable électronique", Ministère de la Jeunesse, de l'Éducation nationale et de la Recherche, Direction de la technologie – SDTICE, <http://www.educnet.education.fr/plan/cartel.htm>, dernière mise à jour août 2002

[Castelli 96], C. Castelli et al. "Getting lost in Hyperspace : Lessons learned and future direction", in ED-MEDIA 96/ED-TELECOM 96, 1996.

[Castells 96], M. Castells, "La société en réseaux", Fayard, 1998

[Cattenat 77], A. Cattenat, G. Paul, "Etude de quelques langages documentaires", Département Informatique, Filière documentation, Juillet 1977

[Cattenat 92], A. Cattenat, G. Paul, "Thésaurus ou non ? ou comment interroger des bases de données textuelles sans être savant", 2^{ème} conférence d'Informatique appliquée de l'Université Paris VIII, octobre 1992,

[Cerratto 03], T. Cerratto Pargman, "Collaborating with writing tools, an instrumental perspective on the problem of computer-supported collaborative activities." In *Interacting with Computers : the Interdisciplinary Journal of Human-Computer Interaction*, Vol. 15, 737-757.

[Chartron 02], G. Chartron, B. Guyot, T. Lafouge, S. Lainé-Cruzel, G. Lallich-Boidin, M-F. Peyrelong, J-M. Salaün, "Le document numérique: un objet fédérateur en sciences de l'information", *Documentaliste Sciences de l'information*, Décembre 2002

[Chaudiron 02], S. Chaudiron, "L'évaluation des systèmes de traitement de l'information textuelle : vers un changement de paradigme", *Mémoire d'Habilitation à Diriger les Recherches*, Univ. Paris X, Nanterre, Nov. 02, 300p.

[Chaumier 77], J. Chaumier, "analyse documentaire : le traitement linguistique de l'information documentaire", Paris : Entreprise moderne d'édition, 1977, 126 p.

[Chauvin 02a], S. Chauvin, **F. Papy**, "Interfaces métaphoriques", journée d'étude "Esthétique de l'Interface", Laboratoire Paragraphe et CITU, 18 juin 2002 <http://h2ptm.hymedia.univ-paris8.fr/interfaces/program.htm>

[Chauvin 02b], S. Chauvin, "Spatialisation et visualisation d'informations : design numérique des interfaces graphiques interactives", *Mémoire de DEA*, Univ. Paris 8, 2002, 96 pages.

[Chauvin 04], S. Chauvin, "VISUAL...UDC Graphical interface for exploring library online catalog", *ACM DIS 2004 (Designing Interactive System)*, August 2004, Cambridge, Massachusetts.

[Chevallet 02], J-P. Chevallet, L. Nigay, "Les interfaces pour la recherche d'informations", in *IHM et recherche d'informations*, C. Paganelli (dir), *Traité STI*, Paris, Hermes, 2002, pp 65-102

[Christoffel 02], M. Christoffel, B. Schmitt, "Accessing Digital Libraries as Easy as a Game", *Proc. of the 2nd International Workshop on Visual Interfaces for Digital Libraries*, Portland (Oregon), July 2002

[Clément 95], J. Clement, "Du texte à l'hypertexte: vers une épistémologie de la discursivité hypertextuelle", in Balpe J.-P., Lelu A., Saleh I. (coords.) in *Hypertextes et hypermédias: Réalisations, Outils, Méthodes*, Hermès, Paris, 1995.

[Cmo 01], Actes du colloque "La Communication Médiatisée par Ordinateur : un carrefour de problématiques", 69^e Congrès de l'Association canadienne-française pour l'avancement des sciences (Acfas), Université de Sherbrooke, 15 et 16 mai 2001 <http://grm.uqam.ca/activites/cmo2001/actes.html>

[Cochard 02], G-M. Cochard, "Approche théorique des aspects pédagogiques et management - le back-office de la FOAD", journée d'étude sur "Enseignement à distance : Théoriques et Pratiques", 12 Nov. 2002, Univ. Paris 8, Lab. Paragraphe, <http://h2ptm.univ-paris8.fr/foad/>

- [Cochard 03], G-M. Cochard. "Dans les coulisses de FOAD". Actes de la conférence "l'enseignement à distance, théoriques et pratiques" Novembre 2003, Université Paris 8. Collection Actes Huit.
- [Cockburn 00], A.Cockburn, B. McKenzie, "What do Web Users do ? An empirical analysis of Web Use", Int. J. Human-Computer Studies, 2000
- [Conklin 87], J. Conklin, "Hypertext : an introduction and survey", Computer, Sept 1987, pp 17-41
- [Coquin 02], R. Coquin, "La réalisation d'un site pour le service documentation du journal Ouest-France. Du support papier au support numérique", mémoire DESS, Univ. Paris 8, juin 2002, 143 pages.
- [Coulon 97], A. Coulon, "Le métier d'étudiant", Paris, PUF, 1997.
- [Coulon 99], A. Coulon, "Penser, classer, catégoriser : l'efficacité de l'enseignement de la méthodologie documentaire dans les premiers cycles universitaires", Laboratoire de Recherches Ethnométhodologiques, Université Paris 8, février 1999
- [Courbières 02], C. Courbières, "Une approche communicationnelle de l'analyse documentaire", actes du colloque "Recherches récentes en Sciences de l'information", 21-22 mars 2002, Toulouse, Adbs-Editions, V. Couzinet et G. Régimbeau (dir), pp 105- 125
- [Courbon 97], J-C. Courbon, S. Tajan, "Groupware et Intranet", Interéditions, Paris, 1997, 230 pages
- [Coutaz 90], J. Coutaz, "Interfaces homme-ordinateur : conception et réalisation ", Paris, Dunod, 1990, 455 pages.
- [Creis 1984], Creis, "Société et Informatique", ouvrage collectif du CREIS, Delagrave, 1984
- [Crosnier 95], H. Le Crosnier, "Structuration du savoir et organisation des collections dans les répertoires du Web", BBF, Paris, t. 40, n° 2, p. 23-31
- [Crosnier 98], H. Le Crosnier, " L'hypertexte en réseau. Repenser la bibliothèque ", BBF, Paris, t. 43, n° 3, p. 18-26.
- [Cubaud 98], P. Cubaud, C. Thiria, A. Topol, "Experimenting a 3d interface for the access to a digital library. In Proceedings of DL'98, ACM Conference on Digital Libraries, Pittsburgh, 1998

D

[Dachelet 90], R. Dachelet, "Hypertexte et hypermédia: Documents - Informations – Connaissances", Le document Electronique, Cours INRIA dirigé par Christian Bornes, 11 - 15 Juin 1990, pp. 135–161

[De Lièvre 01], B. De lièvre, Ch. Depover, "Apports d'une modalité de tutorat proactive ou réactive sur l'utilisation des aides dans un hypermédia de formation à distance" sous la direction de Vries E., Pernin J. -P., Peyrin J. -P. (Eds) Hypermédiat et apprentissages, 2001, pp. 323-330.

[Demaizière 96], F. Demaizière, "L'utilisation des nouvelles technologies en éducation et en formation continue", in Savoir Former - Bilans et perspectives des recherches sur l'acquisition et la transmission des savoirs, les éditions Groupe DEMOS, novembre 1996, pp 111-115

[Devaud 03], C. Devaud, "L'atlas de l'architecture et du patrimoine de la Seine-Saint-Denis : Espace numérique de diffusion et d'échanges d'informations", mémoire de DESS, Univ. Paris 8, juin 2003, 109 pages

[Dinet 02], J. Dinet, J-F Rouet, "La recherche d'informations : processus cognitifs, facteurs de difficultés et dimensions de l'expertise", in IHM et recherche d'informations, C. Paganelli (dir), Traité STI, Paris, Hermes, 2002, pp 133-162

[Dkaki 99], T. Dkaki, "Collecte, pré-traitement et traitement des informations issues du Web dans un environnement coopératif", in Solaris, n° 5, Presses Universitaires de Rennes, 1999

[Dubey 01], G. Dubey, "Le lien social à l'heure du virtuel", PUF, 2001

[Dufour 01], C. Dufour, P. Bergeron, "Lorsque systèmes d'information Web et professionnels de l'information se rencontrent", Proc. of the 29th Annual Conference of the Canadian Association for Information Science, 27-29 May, 2001, D.Grant Campbell (ed.), Université Laval, Québec

[Dujol 86], A. Dujol, "Le clair et l'obscur : usage de la classification a la bibliothèque publique d'information.", BBF, Paris, n°3, 1986

F

[Faerber 00], R. Faerber, "L'enseignement distribué, un levier pour la formation continue", Cahiers Pédagogiques, mai 2000.

[Favier 02], L. Favier, F. Matrin-Juchat, "La science de l'information face à de nouveaux paradigmes : prise en compte de la dimensions sociale de la recherche d'information et remise en cause de la figure de l'usager", actes du colloque "Recherches récentes en Sciences de l'information", 21-22 mars 2002, Toulouse, Adbs-Editions, V. Couzinet et G. Régimbeau (dir), pp 255- 268

[Fayet 99], S. Fayet, "Méthodologie documentaire", BBF, T 44, n°1, 1999, pp 50-55

[Ferret 01], O. Ferret, B. Grau, M. Huraut-Plantet, G. Illouz, C. Jacquemin, "Comment trouver LA réponse", 3ème congrès du Chapitre français de l'ISKO, 5-6 juillet 2001, Chaudiron S. et Fluhr C. (eds), Université de Nanterre – Paris X, pp159-168.

[Flake 02], G.W. Flake, S. Lawrence, C.L. Giles, F.M. Coetzee, "Self-Organization of the Web and Identification of communities", IEEE Computer, 35(3), pp. 66-71, 2002

[Flichy 02], P. Flichy, "La place de l'imaginaire dans l'action technique : le cas d'internet", Réseaux, n°109, 2002

[Foenix-Riou 01], B. Foenix-Riou, "Recherche et Veille sur le Web visible et invisible ", Editions Tec & Doc /Bases Publications, avril 2001, 234 pages

[Fondin 01], H. Fondin, "La Science de l'information : posture épistémologique et spécificité disciplinaire ", Documentaliste – Sciences de l'Information, 2001, vol. 38, n°2, pp 112-122.

[Fondin 02], H. Fondin, "La Science de l'information et la document, ou les relations entre science et technique", Documentaliste – Sciences de l'Information, 2002, vol. 39, n°3, pp 122-129

[France 01], H. France, K. Lundgren-Gayrol, "Apprentissage Collaboratif à Distance", Presses de l'Université du Québec, 2001

[Fressard 02], O. Fressard, " Les formations thématiques à la bibliothèque de Paris-8", Journée Formist, 13 juin 2002.

[Furuta 89], R. Furuta, "Concepts and models for structured document", J. André, R. Furuta, V. Quint (eds), Structured Documents, Cambridge Univ. Press, 1989, pp. 7-38

[Furuta 89b], R. Furuta, P.D. Stots, "Object Structures in Paper Documents and Hypertexts", Woodman'89, pp. 147–151

G

[Gagneux 02], A.Gagneux, H. Emptoz, "Le Document Web, de la structure à la Lisibilité", CIFED'2002, pp105-112

[Gandon 02], F. Gandon, R. Dieng-Kuntz, O. Corby, A. Giboin, "Web Sémantique et Approche Multi-Agents pour la Gestion d'une Mémoire Organisationnelle Distribuée", IC'2002, Rouen, 28-30 mai.

[Garlatti 99], S. Garlatti, S. Iksal, "Documents virtuels personnalisables pour des systèmes d'informations en ligne", IHM99, Montpellier, 1999

[Gasté 01], D.Gasté, "Navigation ou déambulation multimédia ?", Actes du colloque "La Communication Médiatisée par Ordinateur : un carrefour de problématiques", Université de Sherbrooke, 15 et 16 mai 2001

[Gaudin 00], F. Gaudin, "Quelle normalisation pour les documents numérisés en vue d'une conservation et d'une consultation à long terme ?", *L'archivage, Document numérique*, Vol 4, n°3-4/200, pp 199-218

[Ghitalla 02], F. Ghitalla, "L'âge des cartes électroniques. Outils graphiques de navigation sur le Web", *Communication et Langages*, n°131, avril 2002.

[Giffard 97], A. Giffard, "Petites introductions à l'hypertexte", *Banques de données et hypertextes pour l'étude du roman*, Nathalie Ferrand (dir), PUF écritures électroniques, 1997

[Goldfarb 91], C. Goldfarb, "The SGML Handbook", Oxford University Press, 1991

[Gros 94], C. Gros, D. Bourigault, J-L. Vuldy, "Linguistic-Based Toolbox for Hypertext Automatic Linking on Large Technical Documentation" In Proc. 3rd International Conference on Information and Knowledge, Management, Gaithersburg. 1994

[Guyot 00], B. Guyot, "Les dynamiques informationnelles", Habilitation à diriger des recherches, 2000. Université Grenoble III Stendhal. 130 p

H

[H2ptm 95], Actes de la conférence internationale "Hypertextes et Hypermédias, réalisations, outils et méthodes", coord. Jean-Pierre Balpe, Alain Lelu, Imad Saleh, Hermès, Paris, 1995

[H2ptm 97], Actes de la conférence internationale "Hypertextes et Hypermédias", réalisations, outils et méthodes, coord. par Jean-Pierre Balpe, Alain Lelu, Marc Nanard, Imad Saleh, Hermès, Paris, 1997

[H2ptm 99], Actes de la conférence internationale "Hypertextes et Hypermédias", réalisations, outils et méthodes. coord. par coord. Jean-Pierre Balpe, Alain Lelu, Stéphane Natkin, Hermès, Paris, 1999

[H2ptm 01], Actes de la conférence internationale "Hypertextes et Hypermédias", réalisations, outils et méthodes. 18-20 septembre 1999 / coord. par Jean-Pierre Balpe, Sylvie Leleu-Merviel, Imad Saleh, Jean-Marc Laubin, Hermès, Paris, 2001

[H2ptm 03], Actes de la conférence internationale "Hypertextes et Hypermédias", réalisations, outils et méthodes. coord. par J-P. Balpe, I. Saleh, **F. Papy**, D. Lepage, Hermès, Paris, 2003

[Habernas 73], J. Habernas, "La technique et la science comme idéologie", Gallimard, 1973

[Harvey 95], P-L. Harvey, "Cyberspace et communautique", Les presses Universitaires de Laval, 1995

[Harvey 01], P-L. Harvey, "La connaissance au-delà du savoir à l'ère des inforoutes : de la diffusion de l'information à la création du sens dans les bibliothèques virtuelles", congrès ASTED, nov. 01

[Hascoët 04], M. Hascoët, "Visualisation d'information et interaction", Méthodes avancées pour les systèmes de recherche d'informations, traité des Sciences et technologies de l'Information, Ihadjadene M. (sous la dir. de), ed. Hermes, Paris, 2004.

[Hénocque 02], B. Hénocque, "Appropriation des messageries électroniques dans les entreprises en réseau", Presses Universitaires de Bordeaux, décembre 2002, 189 pages.

[Herwijnen 90], E. Van Herwijnen, "Practical SGML", Kluwer Academic Publishers, 1990, 307 pages

[Holzem 99], M. Holzem, "Terminologie et documentation. Pour une meilleure circulation des savoirs", ADBS Editions, 1999, 292 pages.

[Hudon 01], M. Hudon, "Structuration du savoir et organisation des collections dans les répertoires du Web", BBF, Paris, T. 46, n°1, 2001, pp. 57-62.

[Hudon 03], M. Hudon, "Macrothésauri et systèmes d'information gouvernementaux accessibles au grand public", Actes du 4ème congrès du Chapitre français de l'ISKO, 3-4 juillet 2003, Organisation des connaissances – Approches conceptuelles

[Hudrisier 99], H. Hudrisier, "Industrie et culture du code NTIC : vers une nouvelle prospérité sociale au delà de la phase primitive du machinisme informatique", Colloque ICOMTEC, Poitiers, 18-19 mars 1999.

[Hustache-Godinet 98], H. Hustache-Godinet, "Lire-Ecrire des hypertextes", Thèse de doctorat, Université Stendhal Grenoble 3, Octobre 1998

I

[Ihadjadene 01], M. Ihadjadene, R. Bouché. "The dynamique nature of searching and browsing on web-opacs: the Cathie experience", Proc. of the 6th International ISKO conference, Beghtol C., Howarth L. et Williamson N. (dir), 10-13 July Toronto, Canada, 2001, pp 327-332.

[Isko 01], Actes du 3ème congrès du Chapitre français de l'ISKO (Information Society for Knowledge Organization), 5-6 juillet 2001, *Filtrage et résumé automatique de l'information sur les réseaux*, S. Chaudiron et C. Fluhr (dir.), Univ. de Nanterre Paris X, 283 pages

J

[Jacquesson 99], A. Jacquesson, A. Rivier, "Bibliothèques et documents numériques : concepts, composantes, techniques et enjeux", Paris : Éd. du Cercle de la librairie, 1999, 377pages.

[Jamet 98], E. Jamet, "Comment lisons-nous ?", Dossier la lecture, Sciences Humaines, n°82, avril 98, pp. 20 – 25.

[Jeanneret 99], Y. Jeanneret, "Les technologies de la pensée restent à penser", Sciences Humaines, hors-série n°34, mars/avril 1999, pp 23-25

[Jeanneret 00], Y. Jeanneret, "Y a-t-il (vraiment) des technologies de l'Information ?", presses universitaires du Septentrion, 2000, 135 pages

[Jolly 94], V. Jolly, "Mise en place d'une chaîne d'hypertextualisation automatique de documents technique," Mémoire Ingénieur CNAM en informatique fondamentale en vue des applications, Mai 1994.

K

[Kahn 00], A. Kahn, "Génétique moléculaire", Encyclopédie Universalis, 2000

[Keyes 85], E. Keyes, D. Sykes, E. Lewis, "Technology + Design + research = Information Design, Electronic Publishing Drives a New Approach to Design", in Text, Context and HyperText, E. barret (eds), MIT Press, 1985, pp. 251-265

[Kobayashi 00], M. Kobayashi, "Information retrieval on the web", ACM Computing Surveys, vol 32, n°2, pp. 144-173, 2000

L

[Lacroix 01], B. Lacroix, "L'émergence du Web Sémantique et l'exemple du site Internet du Muséum National d'Histoire Naturelle", Mémoire de DESS, Univ. Paris 8, Octobre 2001, 120 pages

[Latouche 1995], S. Latouche, "La mégamachine", série Bibliothèque du M.A.U.S.S. dirigée par Alain Caillé, éditions La découverte, Avril 1995

[Laublet 02], P. Laublet, L. Naït-Baha, A. Jackiewicz, B. Djioua, "Collecte d'informations textuelles sur le Web selon différents points de vue", in IHM et recherche d'informations, C. Paganelli (dir), Traité STI, Paris, Hermes, 2002, pp 260-296

[Lavaivre 01], H. Lavaivre, "L'appropriation des méthodes de l'information structurée dans le cadre de l'IRESO", mémoire de DESS, Univ. Paris 8, Juin 2001, 130 pages.

- [Le Bœuf 01], C. Le Bœuf (s/dir), "La fin du groupware ? Résurgence d'une dynamique organisationnelle assistée par ordinateur", Ed L'Harmattan, 2001, 226 pages
- [Le Coadic 97], Y. Le Coadic, "Usages et usagers de l'information", ADBS, Nathan, 1997, 128 pages
- [Le Quellec 98], B. Le Quellec, "Construction Automatique de documents hypermédias à caractère technique dans un cadre industriel", thèse de doctorat, Univ. Paris 8, 1998.
- [Lecolinet 01], E. Lecolinet, J-D. Fekete, S. Pook, "Bibliothèques : comparaisons entre le réel et le virtuel en 3D, 2D zoomable et 2D arborescent", ASTI 2001, pp 24-25, Paris, France, 24-27 avril 2001
- [Lefèvre 00], P. Lefèvre, "La recherche d'informations. Du texte intégral au thésaurus", Hermès, 2000, 253 pages.
- [Leleu-Merviel 97], S. Leleu-Merviel, "La conception en communication", Paris, Editions Hermès, 1997
- [Leleu-Merviel 04], S. Leleu-Merviel, "Effets de la numérisation et de la mise en réseau sur le concept de document", Revue i3 . 4(1):121-140
- [Leloup 98], C. Leloup, "Moteurs d'indexation et de recherche", Eyrolles, 1998
- [Lelu 93], A. Lelu, "Modèles Neuronaux Pour l'Analyse de Données Documentaires et Textuelles", Thèse de Doctorat, 1993, Université Paris VI
- [Lelu 98], A. Lelu, "Représentations cartographiques de corpus textuels : codages, algorithmes, ergonomie", Habilitation à Diriger les Recherches, 1998, Université Paris 8
- [Lelu 99], A. Lelu, M. Hallab, H. Rhissassi, **F. Papy**, S. Bouyahi, N. Bouhaï, H. He, C. Qi, I. Saleh, "Projet NeuroWeb : un moteur de recherche multilingue et cartographique", 5e conf. Int. H2PTM'99, 23-24 septembre 1999, Paris, France
- [Lelu 00], A. Lelu, M. Hallab, **F. Papy**, S. Bouyahi, H. Rhissassi, N. Bouhaï, F. Tang, (2000) "Textual mapping for multilingual and multwriting access to information on the Internet", Actes de RIAO 2000, coord. CID, collège de France, Paris, 12-14 avril 2000
- [Lelu 01], A. Lelu, S. Bouyahi, "Synthèse d'information en ligne : bilan du logiciel NeuroWeb", actes du chapitre français de l'ISKO, 2E et 6 juillet 2001.
- [Levan 95], S.K. Levan, A. Liebmann, "Le Groupware", Hermes, série Informatique et organisation, Paris, 1995, 146 pages
- [Linard 76], M. Linard, "Des machines et des hommes, apprendre avec les nouvelles technologies", Paris, L'Harmattan, 1996
- [Lucarella 90], D. Lucarella, "A Model for Hypertext-Based Information Retrieval", Proc. of the European Conference on Hypertext, INRIA, France, November 1990, pp. 81-93

[Lupovici 01], C. Lupovici, "Du catalogue informatisé à la navigation dans l'information numérique", *Revue de la Bibliothèque Nationale de France*, n°9, 2001, pp 80- 83

M

[Mc Knight 89], C. Mc Knight, A. Dillon, J. Richardson, "Problems in hyperland? A human factors perspective", *Hypermedia*, 1,2, (1989), pp. 167–178

[Maignien 95], Y. Maignien, "La bibliothèque virtuelle ou de l'Ars memoria à Xanadu", *BBF*, Paris, T. 40, n° 2, p. 8-17

[Maisonneuve 03], M. Maisonneuve, "Du catalogue de la bibliothèque aux ressources du Web", Paris, *ADBS*, 2003, 148 pages.

[Marchionini 88], G. Marchionini, B. Schneiderman, "Finding Facts vs. Browsing knowledge in hypertext systems", *Computer*, January 1988, pp. 70–79

[Mattelart 99], A. Mattelart, "Dangereux effets de la globalisation des réseaux", *Manière de Voir*, n°46; juillet/août 1999

[Mattelart 01], A. Mattelard, "Histoire de la société de l'information", Ed. La découverte, Paris, 2001.

[Mayère 97], A. Mayère, "La société informationnelle", collection *Communication et civilisation*, l'Harmattan, avril 1997.

[Metzger 04], J-P. Metzger (dir), "Médiation et Représentation des Savoirs. Recherches en sciences de l'information et de la communication", Harmattan, Avril 2004, 251 pages

[Michard 02], A. Michard, "XML - langage et applications (2e édition)", Eyrolles, 376 pages, 2002.

[Michel 93], J. Michel, "Les réseaux, un mode de fonctionnement à définir, caractériser et évaluer", "Prévenir", n° 27, 2^{ème} semestre 1993, pp. 11-18 <http://www.paris.enpc.fr/~michel-j/publi/JM249.html>

[Miège 98], B. Miège, "Le communicationnel et le social : déficits récurrents et nécessaires (re)-positionnements théoriques", *Loisir et société*, vol. 21, n° 1, Presses de l' Université du Québec, 1998

[Miège 03], B. Miège (dir.), "Communication personnes systèmes informationnels", *Hermes*, collection STI, Mai 2003, Paris, 196 pages.

[Mileo 96], T. Miléo, "Les réseaux de la société de l'information" - Commissariat général du Plan. Rapport de l'atelier " Nouvelles technologies et réseaux d'information ", Réseaux de la société de l'information (Les) Éditions ESKA / Éditions ASPE europe, octobre 1996, 229 p.

[Millerand 01], F. Millerand, "Le courrier électronique : artefact cognitif et dispositif de communication", Actes du colloque "La Communication Médiatisée par Ordinateur : un carrefour de problématiques", Université de Sherbrooke, 15 et 16 mai 2001

- [Milon 99], A. Milon, "La valeur de l'information", PUF, 1999, 232 pages
- [Moeglin 98], P. Moeglin, "L'industrialisation de la formation. Etat de la question", CNDP, Paris 1998
- [Moeglin 99], P. Moeglin, "Du mode d'existence des outils pour apprendre", Les Enjeux de l'information et de la communication, GRESEC, mise en ligne 17 novembre 1999, http://www.u-grenoble3.fr/les_enjeux/2000/Moeglin/Moeglin.pdf
- [Moles 72], A. Moles, "Théorie de l'information et perception esthétique", Paris, Denoël/Gonthier, 1972, 328 pages.
- [Morizio 99], C. Morizio, "Ils zappent, ils cherchent, ils lisent...des documents électroniques", BBF, Paris, T44,n°5, 1999
- [Mounier 02], E. Mounier, "Systèmes documentaires et systèmes de gestion de bibliothèques : place et rôle de l'opérateur professionnel", , in IHM et recherche d'informations, C. Paganelli (dir), Traité STI, Paris, Hermes, 2002, pp 103-132

N

- [Nanard 90], J. Nanard, "La manipulation directe en interface homme-machine", thèse de doctorat d'état, Université des Sciences et Techniques du Languedoc, Montpellier II, 1990
- [Nanard 93], J. Nanard, M. Nanard, "Should anchors be typed too? An experiment with MacWeb", Proc. of Hypertext93 (Nov.14--18, Seattle, Wash.). ACM, New York, 1993, pp. 51--62.
- [Nanard 98], J. Nanard, M. Nanard, La conception d'hypermédias, in Les Hypermédias, approches cognitives et ergonomiques / dir André Tricot, Jean-François Rouet, Paris, Hermès 1998, pp 15- 34
- [Newcomb 91], S. NewComb, N. Kipp, V.T. Newcomb, "Hyttime, the Hypermedia/Time based Document Structuring Language", ACM, Vol. 34(11), Nov. 1991, pp. 67-82
- [Nielsen 89], J. Nielsen, "Hypertext and Hypermedia", Academic Press, 1989
- [Nielsen 90], J. Nielsen, "The art of navigating through hypertext", ACM, vol.33 (3), march 1990, pp 298-310.

[Panigel 98], C. Panigel, "FOURMI – exemples de dispositifs de formation d'étudiants", Groupe de travail, "Formations Universitaires en Réseau aux Méthodologies de l'Information", Urfist de Paris/Ecole des Chartes, 1998, <http://www.ccr.jussieu.fr/urfist>

[Papy 92a], F. Papy, J-L. Vuldy, "Analyse Pertinente de Textes Electorniques (APTE) - Les étapes de l'hypertextualisation automatique", Collection de notes internes de la Direction des Etudes et Recherches, EDF, Mission Mathématiques, Informatique, Télécommunications, février 1992

[Papy 92b], F. Papy, J-L. Vuldy, "Réflexion sur la complémentarité entre les hypermédias et les SGBD", in "Les bases de données", édition spéciale, Bulletin des Utilisateurs du Centre de Calcul des Etudes et Recherches (BUCCER), EDF, Direction des Etudes et Recherches, Service Informatique et Mathématiques Appliquées, février 1992

[Papy 93], F. Papy, J-L. Vuldy, "Automatic creation of hypertext networks from technical documents", Vienna Conference on Human Computer Interaction, Vienna, Austria, 20-22 septembre 93

[Papy 94], F. Papy, J-L. Vuldy, "Actualisation de réseaux Hypertextes", Collection des Notes Internes, EDF, Direction des Etudes et Recherches, Service Informatique et Mathématiques Appliquées, février 1994

[Papy 95a], F. Papy, J-L. Vuldy, "L'hypertextualisation automatique de documents techniques", Actes du séminaire "Hypermédias, Education et Formation", Institut Blaise Pascal-Masi 1995/31, Masi, 1995

[Papy 95b], F. Papy, "Hypertextualisation de documents techniques", Thèse de doctorat, Université Paris 8, 1995

[Papy 99], F. Papy, I. Saleh, N. Bouhaï, "Chercher et réorganiser l'information sur le Web", 5ème colloque Hypermédias et Apprentissages, 9-11 avril 2001, Grenoble, pp. 49-58

[Papy 02], F. Papy, "HyWebMap et K-Web Organizer : des dispositifs de construction de connaissances pour la FOAD", actes de la journée d'étude du 22 nov. 2002, Les TIC au cœur de l'enseignement à distance, I. Saleh, D. Lepage, S. Bouyahi (coords), Collection Actes Huit, Université Paris 8

[Papy 03a], F. Papy, N. Bouhaï, "Navigation et recherche par catégorisation floue des pages HTML", Actes des JFT'2003, 30 juin au 2 juillet 2003, Ecole Polytechnique de l'Université de Tours

[Papy 03b], F. Papy, N. Bouhaï, "HyWebMap et K-Web Organizer : Dispositifs complémentaires d'organisation individuelle et communautaire de connaissances", 4ème Congrès ISKO-France, Grenoble, 3-4 Juillet 2003

[Papy 03c], F. Papy, S. Chauvin, "HyWebMap documents : an collaborative e-schoolbag for E- learning", ITHET'03, 7-9 Juillet 2003, Marrakech, Maroc.

- [Papy 03d], F. Papy, "De l'hypermédia aux contenus riches. Evolution des concepts" , actes du séminaire Aristote "De l'hypermédia aux contenus riches", Collège de Polytechnique, P. Bigeon, D.Menga (resp. Scientifique), 30 janvier 2003.
- [Papy 03e], F. Papy, "K-Web Organizer : construction d'espaces de connaissances collaboratifs ", journée d'étude "Recherche et fouille d'information sur le Web : perspectives sémantiques et statistiques", Laboratoire Paragraphe et CITU, 14 janvier 2003. <http://h2ptm.hymedia.univ-paris8.fr/webmining/>
- [Papy 04a], F. Papy, M. Sidir, "Case study : merging distance education ICT with collaborative learning scenarios", 5th Int. Conf. on Information Technology Based Higher Education and Training: ITHET'04, 31 May-2 June 2004, Istanbul
- [Papy 04b], F. Papy, "Le document numérique éducatif : objet agrégatif, adaptatif et partageable", Enseignement à distance, Saleh I. et Bouyahi S. (sous la dir. de), ed. Hermes, Paris, 2004
- [Papy 04c], F. Papy (coord), "Bibliothèques Numériques", ed. Hermes Publishing, Paris, 2004 (à paraître)
- [Papy 04d], F. Papy, V. Folcher, M. Sidir, T. Cerratto Pargman, "E-Learning et technologies pour la coopération : inadéquations artefactuelles et logiques des activités instrumentées", ERGO-IA, 17-19 novembre 2004
- [Parrochia 94], D. Parrochia, "Cosmologie de l'information : pour une nouvelle modélisation de l'univers informationnel", Paris : Hermès, 1994
- [Pavé 89], F. Pavé, "L'illusion informaticienne", l'Harmattan, logiques sociales, 1989
- [Pelfrêne 02], R. Pelfrêne, "Les postes de lecture assistée par ordinateur", in IHM et recherche d'informations, C. Paganelli (dir), Traité STI, Paris, Hermes, 2002, pp 297-330
- [Perriault 97], J. Perriault, "Apprendre à distance", Sciences Humaines, hors série n°16, mars/avril 1997.
- [Perriault 02], J. Perriault, "Education et Nouvelles Technologies", Nathan Université, 2002
- [Pincemin 99], B. Pincemin, "Diffusion ciblée automatique d'informations : conception et mise en oeuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents", Thèse de doctorat, Université Paris IV – Sorbonne, 1999
- [Pisani 03], F. Pisani, "Internet saisi par la folie des Weblogs", Le monde diplomatique, août 2003
- [Poissenot 02], C. Poissenot, "De l'objet au point de vue : les bibliothèque entre sciences de l'information et sociologie", actes du colloque "Recherches récentes en Sciences de l'information", 21-22 mars 2002, Toulouse, Adbs-Editions, V. Couzinet et G. Régimbeau (dir), pp 145- 162
- [Projet 01], Projet de Loi sur la Société de l'information, n°3143, enregistré à la présidence de l'Assemblée Nationale le 14 juin 2001

Q

[Quint 92], V. Quint, I. Vatton, "Combining Hypertext and structured documents in Grif", ACM ECHT Conference, Milano, 30 Nov- 4 Déc 1992, pp. 23-32.

R

[Rabardel 95], P. Rabardel, "Les hommes et les technologies : approche cognitive des instruments contemporains", éd. A. Colin, Paris, 1995

[Rabardel 01], P. Rabardel, N. Carlin, M. Chesnais, N. Lang, G. Le Joliff, M. Pascal, "Ergonomie, concepts et méthodes", Octares éditions, 3^{ème} édition, Toulouse, 2001

[Richard 96], N. Richard, "Les autoroutes de l'information et le multimédia : vers quelle société ?", Terminal n°71/72, été-automne 1996, pp 237-258

[Roisin 99], C. Roisin, "Documents multimédia structurés", Habilitation à Diriger les Recherches, INP Grenoble, septembre 1999

[Role 91], F. Role, "La norme SGML pour décrire la structure logique des documents", Documentaliste-Sciences de l'information, Vol. 28 (4-5), 1991, pp. 187-192

[Rostaing 01], H. Rostaing, "Le Web et ses outils d'orientation", BBF, t.46, n°1, 2001, pp 68-77

[Rouet 97], J.F. Rouet, "Le lecteur face à l'hypertexte", in apprendre avec le multimédia Où en est-on ? Sous la direction de GRIMONT A., Paris Editions Retz, 1997.

[Rouet 01], J-F. Rouet, "Les systèmes d'information en ligne : quels outils pour quels usages ?", Actes "Réseaux humains / Réseaux technologiques", Maison des Sciences de l'Homme et de la Société, Université de Poitiers, 19 mai 2001

[Roussey 02], C. Roussey, S. Calabretto, J-M Pinon, "Le thésaurus sémantique : contribution à l'ingénierie des connaissances documentaires", Actes d'IC'02, Rouen, 28 au 30 mai 2002

[Rowley 92], J. Rowley, "Organizing knowledge : an introduction to information retrieval", Ashgate, 2nd edition, 1992

[Roy 02], R. Roy, "Pour une approche conviviale de l'accès à distance aux collections des bibliothèques publiques", ISKO 2002, Grenoble.

-
- [Salaun 02], J-M. Salaun, "Documents et numérique", Contribution au rapport du Conseil d'Analyse Economique sur la Société de l'information. 01 décembre 2003
- [Saleh 99], I. Saleh, **F. Papy**, "Les bases de données pour l'Internet et l'Intranet", Paris, Hermès, 1999
- [Saleh 00], I. Saleh, N. Bouhaï, **F. Papy**, "HyWebMap, a system to create interactive networks", 3d International Conference on Human-System Learning, CAPS'3, Learning's WWW, Dec 12-14, 2000, Paris, France
- [Saleh 00b], I. Saleh, N. Bouhaï, S. Bouyahi, **F. Papy**, "Système HyWebMap : Outil de navigation et de restructuration des connaissances", Actes du colloque "Technologies de l'Information et de la Communication dans les Enseignements d'Ingénieurs et dans l'Industrie", Troyes, 18,19,20 octobre 2000
- [Saleh 01], I. Saleh, "Web Sémantique, Dynamique et Coopératif", Mémoire d'Habilitation à Diriger les Recherches, Octobre 2001
- [Saleh 01b], I. Saleh, **F. Papy**, N. Bouhaï, "HyWebMap :Structurer l'information sur le Web", 29th Annual Conference of the Canadian Association of Information Science (CAIS/ACSI), Laval, Québec, Mai 2001
- [Saleh 01c], I. Saleh I., **F. Papy**, N. Bouhaï, "De HyWebMap à K-Web Organizer. D'une application auteur à un système collaboratif", H2PTM'01, 18-20 octobre 2001, J-P. Balpe, S. Leleu-Merviel, I. Saleh, J-M. Laubin (coords), Lavoisier, 2001, pp. 205- 221.
- [Saleh 04], I. Saleh, **F. Papy**, "Navigation et recherche dans un système hypertexte", Méthodes avancées pour les systèmes de recherche d'informations, traité STI, M. Ihadjadene (sous la dir. de), Hermes, Paris, 2004
- [Samain 01], T. Samain, "Du catalogue enrichi au métacatalogue", BBF 2001, t. 46, n°2
- [Sandvad 89], E. Sandvad, "Hypertext in an Object-Oriented Programming Environment", Woodman'89, Bigre 63 - 64, pp. 30-41
- [Senteni 01], A. Senteni, M. Aubé, A. Dufresne, "Un modèle de support au travail collaboratif dans un centre virtuel d'apprentissage", 5ème colloque Hypermédias et Apprentissages, grenoble, 9-11 avril 2001, E DE VRIES, J-Ph. PERNIN, J-P. PEYRIN (eds)
- [Sidir 03], M. Sidir, G-M. Cochard, "Enseigner et apprendre autrement : nouveaux rôles du formateur, nouveaux acteurs", H2PTM'03, Paris 8, septembre 2003
- [Simon 03], C. Simon, "L'organisation des professionnels de l'information-documentation en réseau à EDF : contribuer à l'efficacité de la fonction", Mémoire de DESS, Univ. Paris 8, juin 2003, 100 pages
- [Simonnot 02], B. Simonnot, "De la pertinence à l'utilité en recherche d'information : le cas du Web", actes du colloque "Recherches récentes en Sciences

de l'information", 21-22 mars 2002, Toulouse, Adbs-Editions, V. Couzinet et G. Régimbeau (dir), pp 393- 412

[Slatin 85], J. M. Slatin, "Hypertext and the Teaching of Writing", in Text, Context and HyperText, E. barret (eds), MIT Press, 1985, pp. 111-129

[Slavic 03], A. Slavic, "Mise en oeuvre de la CDU : des rayons de bibliothèque à un langage d'indexation structuré", 69th IFLA General Conference and Council, 1-9 August 2003, Berlin.

[Smolczewska 04], A. Smolczewska, G. Lallich-Boidin, "Validation par prototypage d'un modèle de segmentation des documents techniques composites", Cide 7, 22-25 juin 2004, ed. Europia, Paris, pp75-92

[Souchier 03], E. Souchier, Y. Jeanneret, J. Le Marec, "Lire, écrire, récrire", Bibliothèque Publique d'Information, (Etudes et Recherches), 2003

[Sperber 01], D. Sperber "L'individuel sous influence du collectif", La Recherche(344):32-35, 2001.

[Stockinger 01], P. Stockinger, "Traitement et contrôle de l'information – procédures sémiotiques et textuelles", Paris Hermes Sciences Publications, 2001

T

[Tazi 99], S. Tazi, A. Chikh, "Annotations Structurés d'Aide à la Reconstruction de Documents. Le modèle ASARD", actes de CIDE'99, conf. Int. sur le document Electronique ; Damas, Syrie, 5-7 juillet 1999

[Tilley 93], S. R. Tilley, M. J. Whitney, H. A. Müller, M-A. D. Storey, "Personalized Information Structures", The 11th International Conference on Systems Documentation (SIGDOC'93), Waterloo, Ontario, 5-8 Octobre, 1993, pp 325-337

[Timimi 99], I. Timimi, J. Rouault, "La veille sur Internet. Une avancée dans la recherche de l'information stratégique", revue électronique Les enjeux, dir de publication B. Miège, GRESEC, numéro 1, novembre 1999, http://www.u-grenoble3.fr/les_enjeux/n1/Timimi-Rouault/

[Tricot 98], A. Tricot, J-F. Rouet (dir), "Les Hypermédias, approches cognitives et ergonomiques", Paris, Hermès 1998

V

[Verroust 97], G. Verroust, "Hypertexte", Encyclopédie Universalis, 1997

[Virbel 95], J. Virbel, "Annotation dynamique et lecture expérimentale : vers une nouvelle glose ?", In Littérature, volume 96, pages 91-105, 1996

[Virilio 94], P. Virilio, "La Machine de vision", Ed. Galilée, 1994, 158p.

[Vuldy 93], J-L. Vuldy, V. Jolly, "Réalisation d'une base documentaire appliquée au comportement diélectrique des isolations dans l'air", Collection des Notes Internes,

EDF, Direction des Etudes et Recherches, Service Informatique et Mathématiques Appliquées, Service Matériel Electrique, février 1993

W

[Weissberg 99], J-L Weissberg, "Présence à distance", ed. L'Harmattan, Paris, 1999

[Weissberg 00], J-L. Weissberg, "Auteur, nomination individuelle et coopération productive", Revue Solaris, déc. 2000 - janvier 2001
<http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d07/7weissberg.html#fn1>

[Wertel 99], I. Wertel-Fournier, "L'iconographe dans le labyrinthe des mots et des images : pour un imagier numérique comme espace cartographié de l'iconothèque", Thèse de doctorat en Sciences de l'information et de la communication, Univ. Paris 8, décembre 99.

[Wolton 00], D. Wolton, "Internet, et après ?", Editions Flammarion, Paris, 2000

X

[XML 98], Extensible Markup Language, XML 1.0 Recommendation published on 10th February 1998, <http://www.w3c.org/TR/1998/REC-xml-19980210>

Y

[Yee 99], M. Yee, "Guidelines for OPAC displays", 65th IFLA Council and General Conference, Bangkok, Thailand, August 20 - August 28, 1999 (<http://www.ifla.org/IV/ifla65/papers/098-131e.htm>)

Z

[Zacklad 01], M. Zacklad, M. Grundstein (s/dir), "Management des connaissances", Hermès Science Publications, 2001, Paris

