La Lettre de l'URFIST de Toulouse, n°26 - 2000

Jo Link-Pezet, Élisabeth Lacombe-Carraud, Olivier Ertzscheid

▶ **To cite this version:**

Jo Link-Pezet, Élisabeth Lacombe-Carraud, Olivier Ertzscheid. La Lettre de l'URFIST de Toulouse, n°26 - 2000. 2000. sic_03250468

HAL Id: sic_03250468

https://archivesic.ccsd.cnrs.fr/sic_03250468

Submitted on 4 Jun 2021

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.

La Lettre de l'URFIST de Toulouse

Editorial

Les médias nous servent désormais d'abondants reportages à propos d'Internet et du multimédia : journaux, télévisions hertziennes ou câblées ont leurs rubriques quotidiennes ou hebdomadaires.
Récemment la nécessité de pourvoir cinq sièges au conseil d'administration de ICANN (Internet Corporation for Assigned Names and Numbers), organisme chargé d'attribuer les extensions et les désinences géographiques du Net, a eu une large couverture médiatique. Pourquoi?
Parce qu'à travers cette attribution s'organise le trafic intense de clés d'accès qui ont déjà permis entre autres le blocage de noms de marques célèbres, mais aussi parce que les internautes inscrits pouvaient voter pour un candidat de leur choix.
S'agissait-il de l'ouverture à la société civile, d'une des formes possibles de cyberdémocratie? Qui sur Internet possède un véritable pouvoir? S'agit-il de ceux qui ont le contrôle global des marchés financiers, des autoroutes de la communication, des industries informatiques, des médias : logique essentiellement fondée sur la concurrence et l'efficacité de la net économie. Qu'est devenu face à la pensée unique l'esprit d'échange des débuts?

Dernier avatar : les paillettes envahissent le net. La revue "ELLE" de Juillet 2000 parlait des soirées organisées par les start-up, où le "show-businet remplace le show-business" au cours de fêtes baroques, où se pressent des invités au nom prestigieux, avec pour objectif principal et unique : impressionner l'esprit des faiseurs d'opinion, des concurrents, des investisseurs, voire des clients au cours de soirées où co-habitent des maîtres de cérémonie et des prescripteurs d'études de tendances.
A quoi reconnaît-on une fête réussie? Quand on refuse du monde ....
La fête de start-up est un évènement social très recherché avec son rituel : d'abord des urnes pour glisser sa carte de visite (identifiant obligatoire), un écran géant pour visionner le site de la start up, un open bar et un cocktail dînatoire pour plaire, amuser, fidéliser, s'enrichir.

Nous, à l'URFIST, nous prônons la gestion des connaissances et le fonctionnement en réseau. Nous avons un site, nous savons glisser nos cartes de visites dans des urnes. Certes, jusqu'alors nous les glissions dans des urnes pour gagner un agenda ou un stylo sur le stand Dialog ou celui de la British Library à On Line.
Nous aussi, nous demandons à nos partenaires de remplir des fiches, ce qu'ils font avec plus ou moins de difficultés d'ailleurs (mais c'est la faute à Lotus Notes). Nous

écrivons une lettre, peut-être un peu trop sérieuse : dans ce numéro 26 il y est question d'indexation et d'ontologies (à la rubrique formation), de mémoires et d'intelligence collective (rubrique recherche), de comptes rendus de congrès (IDT et On-line) et de stages, enfin d'accès à l'information sur Internet.

On aime aussi organiser des pots, voire des cocktails. On se rapproche mais on sent bien que ce n'est pas tout à fait ça.
On n'est pas assez décalé. On a les ingrédients, mais on manque d'entraînement : nous devrions ajouter quelques paillettes à notre logo.
Tiens une bonne idée, on pourrait proposer une "fête.com" pour la réouverture de la bibliothèque de Toulouse I et célébrer les retrouvailles avec nos bureaux.
Nos solutions, et peut-être un reportage sur cet événement, dans notre prochaine lettre... Bonne lecture.
Bonnes fêtes à tous.

Jo Link-Pezet / Elisabeth Lacombe-Carraud

Représentation des compétences des formateurs dans FORSIC
Olivier Ertzscheid - Etudiant en thèse Toulouse II

Y être ou ne pas y être (sur le Net)
Jo Link-Pezet, Maître de conférences - URFIST de Toulouse et Olivier Ertzscheid - Etudiant en thèse Toulouse II

Mémoires et intelligence collective
Jo Link-Pezet, Maître de conférences - URFIST de Toulouse

Bibliographie sur l'indexation

Online 99 (fin)
Elisabeth Lacombe, Conservateur en Chef - URFIST de Toulouse

IDT 2000
Olivier Ertzscheid - Etudiant en thèse Toulouse II

Formation de formateurs
Elisabeth Lacombe, Conservateur en Chef - URFIST de Toulouse

Représentation des connaissances dans FORSIC

FORSIC

L'un des objectifs de FORSIC est de déterminer comment une communauté spécifique (en l'occurrence des formateurs) face à un outil de capitalisation, peut formaliser un cadre de recueil de connaissances et quel effet cela a-t-il sur cette communauté ? Cela produit-il des connaissances nouvelles ?
A cette fin, une réflexion est menée sur les attendus de la problématique de l'indexation à l'ère du "tout numérique". Face aux constantes avérées de l'univers de la recherche documentaire que sont, entre autres, l'explosion de la masse documentaire et la diversification des ressources et leur hétérogénéité, se dessinent de nouveaux enjeux pour lesquels les techniques documentaires classiques doivent

être redéfinies. Au travers de l'exemple de la mise en place d'un dispositif de formation à la recherche documentaire - qui allie des techniques d'indexation classiques, des outils innovants de représentation et de cartographie de l'information et une technologie multi-agent devant permettre l'émergence - et dont le but est de mettre à la disposition d'une communauté de chercheurs et d'étudiants des ressources pédagogiques de formation dans le champ de la recherche documentaire, FORSIC se veut un dispositif fonctionnel d'ingénierie de la connaissance dont l'indexation constitue le noyau dur autour duquel gravitent des ressources (textes, images ...), des compétences, et des individus, trois maillons essentiels de la chaîne du savoir qui doivent permettre d'esquisser les lignes de force d'une pragmatique de la connaissance.

Indexer l'inindexable ...

Tout semble a priori indexable. Il paraît toujours possible d'établir un code, de soumettre à un filtre, de faire entrer dans une grille n'importe quel type de données hétérogènes. Ce qui rend l'indexation des contenus hyper/multimedias (c'est à dire toute démarche d'indexation sur Internet) particulièrement problématique pour ne pas dire insoluble, c'est l'homogénéité de la masse informationnelle, et non pas, comme on pourrait le penser, son hétérogénéité.
En effet, ce qui pose véritablement problème et constitue *de facto* l'enjeu majeur de l'indexation est le résultat de la réticulation de ces ressources au sein d'un univers informationnel que sa masse rend insaisissable pour l'esprit humain. Il n'est qu'à voir fleurir les néologismes ("docuverse", "hypercortex", "multivers" ...) pour se rendre compte de l'effort de la pensée pour mettre en mots - c'est à dire pour rendre accessible - une réalité encore fuyante et insaisissable.
Cette constante remise en mots qui s'applique et se pratique à des niveaux perceptuels différents, place l'individu au cœur même du processus de recherche documentaire. Subsumant toutes les techniques de filtrage existantes, il demeure plus qu'il ne devient le premier et l'ultime "filtre informationnel". Les dernières génération d'interface ou d'algorithmes lui redonnent cette place que l'on crût un temps pouvoir être aliénée par la technique. Le "profil utilisateur" occupe aujourd'hui une place majeure parmi les divers aspects procéduraux que met en œuvre la recherche documentaire. Ces thématiques sont celles de la cognition distribuée, notamment développée par Hutchins [Hutchins 1995].
L'identification de modes d'indexation transverses constitue donc aujourd'hui une problématique de tout premier plan, tant pour l'information retrieval (IR) que pour la mise en place d'une ontologie seule apte à rendre compte des processus complexes de constitution de ce nouvel espace réticulé de savoir qui constitue d'ores et déjà - et de manière opérante - notre environnement cognitif privilégié, c'est-à-dire notre mode d'appropriation privilégié de la réalité qui nous entoure.
Le champ de la recherche documentaire apparaît comme le creuset d'une pluridisciplinarité qui fait de la transversalité des savoirs la pierre de touche de l'édifice de l'univers informationnel actuel. Tout en demeurant un domaine de spécialité en pleine expansion et en constitution permanente, **la recherche documentaire se positionne naturellement comme le point le plus haut pour conduire une réflexion théorique sur l'indexation**. Elle met en effet en jeu des savoirs diversifiés produits par des individus aux compétences variées sous la forme de ressources documentaires dans tous les formats et sur tous les supports. Le "facteur humain" que les approches de la cognition distribuée imposent comme paradigme fondamen
tal de la science documentaire est étroitement dépendant du "facteur temps". Ce dernier doit également occuper une place de premier plan permettant d'atteindre un "hic et nunc" de l'indexation, seule manière d'approcher et de rendre compte de manière structurée de la nouvelle organisation de la connaissance qui se met en

place sur le réseau. La mise en place d'une indexation transverse et centrée sur les contingences temporelles inhérentes au facteur humain, est la condition de réalisation première de cette expérience que nous allons relater.

Quel savoir dans FORSIC ?

Nous posons comme postulat que le savoir se définit (dans le cadre d'un dispositif de formation) par :

- l'émergence et la mise en évidence de structures (au sens de topiques);
- l'identification des compétences des acteurs;
- le positionnement de ces mêmes acteurs vis à vis de ces compétences, des ressources associées et de la communauté concernée.

En cela nous reprenons le schéma classique de l'indexation conceptuelle et structurelle qui, au moyen de traitements informatiques et calculatoires automatisés ou semi-automatisés commence par extraire des documents un certain nombre de structures et/ou de concepts devant permettre dans un deuxième temps l'exploitation individuelle de ces documents.
A l'exploitation individuelle nous ajoutons l'exploitation individualisée et l'exploitation collective. Nous prenons également en compte le fait suivant qui nous paraît être déterminant en même temps que caractéristique de l'ère numérique : l'objet de l'indexation dépasse le champ du savoir au sens où la démarche encyclopédique définissait ce terme. Avec l'apparition avérée de la bibliothèque de Babel mise à disposition d'un hypercortex qui se veut la manifestation d'une modalité collective de l'intelligence, le savoir "traditionnel" contenu dans les textes s'enrichit et se complexifie jusqu'à prendre en compte la dimension hyper ou multi-media des documents qui constituent son objet, mais également des individus, des compétences, des problèmes, etc ... Ce sont ces nouvelles entités qui sont au cœur de notre réflexion théorique sur l'indexation.

Les objectifs de l'indexation

L'une de nos interrogations porte sur ce qui fait un objet documentaire. Le public auquel il s'adresse, le type de support sur lequel il sera publié, le format dans lequel il le sera, les connaissances requises pour une compréhension optimale, sont autant d'éléments non-exhaustifs qui vont déterminer son style, l'étendue de son lexique, sa longueur, son ton, etc ... Chaque élément de cette chaîne peut à tout moment influer de manière déterminante et parfaitement réciproque sur un autre élément de la chaîne. Un objet documentaire est donc constitué du régime de causalités réciproques internes qui le fondent.
Il apparaît aujourd'hui clairement que le savoir se décline en entités complexes et hétérogènes (individus, textes, images, compétences ...). Ces entités continuent cependant de s'organiser en " informations " qui sont rendues disponibles sous des formes encore traditionnelles bien que numériques.
A travers la mise en place d'un modèle de représentation des connaissances dans FORSIC, nous postulons que le futur de l'indexation comme technique (tant conceptuelle que structurelle), passera nécessairement par la transversalité des technologies et des méthodologies liées, puisque cette transversalité est la composante essentielle et originelle des objets qu'elle prétend désormais étudier.
Cinq objectifs principaux sont identifiés pour parvenir à une indexation apte à rendre compte de l'hétérogénéité des entités qui composent le projet FORSIC, à savoir des ressources pédagogiques (sous la forme de documents dans tous les formats (html, powerpoint, word, ...), des individus et des compétences en formation pour ce qui est du champ de la recherche documentaire. Ces objectifs sont les suivants :

- l'indexation doit être **transversale**, c'est à dire qu'elle doit pouvoir rendre compte simultanément et de manière transverse des différentes entités qui constituent son objet. C'est le point le plus délicat à atteindre parce qu'il implique que les routines d'indexation soient à la fois génériques (ouvertes) pour ce qui est du moteur d'indexation (traitement informatique et calculatoire) et plurielles (spécifiques) au niveau du traitement des objets;
- l'indexation doit être **décontextualisée** : elle doit pour ce faire être le plus possible éloignée du contenu;
- l'indexation doit pouvoir être **automatisée**, et pour cela pouvoir se reposer sur des modèles de documents;
- l'indexation doit être **interactive** pour permettre a FORSIC de mettre en œuvre ses aspects collaboratifs;
- l'indexation doit être **rétroactive**, et intégrer ainsi les effets feedback nécessaires à la mise à jour du système.

Indexer les compétences

-> Sujets de l'étude : panel de formateurs ayant une implication forte et une expérience concrète et avérée de la formation en recherche documentaire,

-> Objectifs de l'étude : définir des universaux d'actes de formation.

-> Originalité de la démarche : elle postule un engagement ontologique dont la force polémique n'est pas le moindre atout. Tout engagement de ce type se décline comme suit : il existe des concepts (des idées pures, des substances) qui permettent de reconstruire le monde ou un de ses aspects par un jeu combinatoire de ces concepts premiers. L'enjeu de notre travail de recherche est de questionner le champ de la formation à la recherche documentaire en quête des ces improbables universaux.

-> Rôle des index de compétences :

- identifier les compétences et les distinguer (sans les décrire)
- cerner les contours d'une zone de compétence
- élaborer un langage commun à tous les acteurs (formateurs, institutionnels, apprenants, ...)

Pour y parvenir, ces index doivent être distincts et indépendants. La logique de définition de ces index est donc distinctive et non descriptive.

Philosophie de l'indexation

La modélisation des compétences en formation doit permettre d'identifier des domaines de connaisance spécifiques en les reliant à des individualités. La philosophie à l'origine de cette logique est celle développée par P. Lévy et M. Authier dans "Les arbres de connaissances" [Authier M., Lévy P. 1993]. M. Authier dirige actuellement la société **Trivium**, partenaire du projet et éditrice de la plate-forme logicielle "**SEE-K**" destinée au Knowledge Management.
Au sens mathématique, un arbre de connaissance correspond à une somme de Fourier, à un espace topologique pré-métrique, où la notion de proximité est indépendante du calcul de mesure.
On parlera de "compétence" dès lors que l'on pourra mobiliser des connaissances et des qualités pour faire face à un problème (de formation). Une compétence se veut une connaissance en action.

On considèrera qu'il ne peut y avoir de partage effectif de connaissances si l'on n'a pas été individuellement reconnu dans une ou plusieurs compétences.

La méthode 4 cinquième : fondements méthodologiques

La méthodologie (fournie par l'outil logiciel qui permettra de traiter l'indexation des compétences : SEE-K© de la société TRIVIUM©) qui permet d'établir une liste de compétence sur lesquelles pourront se positioner les acteurs du système (ici, les formateurs) s'organise selon quatre catégories lesquelles contiennent chacune cinq sous-catégories. Celles-ci s'inspirent des des quatre causalités aristotéliciennes :

- *"causalité efficiente : quoi faire ?"* C'est la catégorie de l'action, qui sert à déterminer d'où part le mouvement.
- *"causalité matérielle : avec quoi ?"* Cette catégorie a trait à l'environnement, c'est à dire à la condition de l'action. L'environnement est entendu comme "ce sans quoi je ne peux pas agir".
- *"causalité finale : pour qui ?"* Il s'agit ici de la finalité subjective qui vise à identifier le bénéficiaire de l'action de formation entreprise.
- *"causalité formelle : pour quoi ?"* Cette finalité objective doit permettre d'identifier un résultat concret et finalisé sous la forme, par exemple, d'un support de cours.

Pour chacune de ces quatre catégories, une liste de cinq items a été définie, au sortir d'un long travail de décontextualisation visant à désambigüiser chacun d'entre eux. Une feuille de calcul Excel a ensuite été utilisée pour permettre de croiser les items de chaque catégorie, donnant ainsi une liste de 625 compétences documentaires. Les contraintes d'élaboration retenues pour la formalisation de cette ontologie sont les suivantes :

- Dans une même dimension, chaque terme est indépendant des autres.
- Dans une même dimension, chaque terme doit être différent pour permettre de distinguer les compétences.
- Les termes d'une dimension doivent être indépendants entre eux et avec ceux des autres dimensions.

Il s'agit d'identifier les compétences et de les distinguer, **sans les décrire** . Elles doivent être génériques et décontextualisées pour permettre de distinguer le **repérage** de la compétence et sa **description** (ce qui permet une gestion de l' incertitude forte et autorise la comparaison de structures différentes dans un même espace de représentation). Le rôle joué par cet index des compétences dans le projet **FORSIC** est primordial puisqu'il constitue le langage commun entre les différents types d'acteurs.

ACTION	ENVIRONNEMENT	BENEFICIAIRE	DELIVRABLE
Identifier, acquérir	Information, données	Pour soi-même	Techniques documentaires
Concevoir, étudier	Savoir, discipline	Pour des collaborateurs, des formateurs	Méthodologies documentaires
Réaliser, produire	Réseau, humain	Pour l'institution, l'université	Théories documentaires
Enseigner, informer	Informatique, technique	Pour des étudiants, des apprenants	Expertise, conseil

Valider, maintenir	Conditions matérielles	Pour des partenaires, des fournisseurs	Norme, diplôme

Ontologie des compétences en formation à la recherche documentaire

Ainsi formulées, les compétences peuvent sembler abstraites et difficilement appréhendables par les individus. Mais il ne s'agit que d'un point de départ qui se doit d'être le plus générique et le plus ouvert possible. En se positionnant sur une ou plusieurs compétences dès leur entrée dans le système, les formateurs doivent apporter un **commentaire sur la compétence** choisie, pour la préciser ou l'interpréter. De plus, pour que cette compétence soit prise en compte, elle doit être assortie d'une ressource de formation correspondante, sur support informatique quel qu'en soit le format (.ppt, .doc, .html, .rtf ...), r**essource qui est elle-même commentée** afin de préciser ses mots-clés, le public auquel elle s'adresse, et le niveau de compétence requis pour en tirer parti. Ces deux entrées de commentaire se font en texte libre et sont récupérées par l'autre composante logicielle de SEE-K (**UMAP**) qui permet de les indexer et de représenter cette indexation sous la forme non plus d'un arbre mais d'une île. Les résultats de cette indexation peuvent alors être récupérés et analysés pour affiner et stabiliser la base de compétences initialement définie.

Les situations de communication dans le dispositif FORSIC et leur rôle dans l'indexation

Ces situations sont celles rencontrées dans la phase de prototypage.
-> échange de savoirs (sous des formes finalisées ou non) avec des pairs : cette situation initiale de communication doit permettre la constitution d'un socle référentiel et théorique commun. Son objectif sous-jacent est de permettre au système d'atteindre une certaine masse critique. Elle correspond concrètement à la collecte des documents et ressources et à la validation d'un panel initial de formateurs.
-> échange de savoirs avec des pairs "experts" : il s'agit ici, dans un deuxième temps, d'optimiser ce socle collectif par des procédures d'échange, de feed-back et d'élimination d'éventuels doublons. Concrètement, nous introduisons une première validation par l'usage qui aura uniquement valeur de pré-requis et non pas de finalité comme dans les systèmes traditionnels. Cette étape ne constituera en effet en rien un aboutissement. Son objectif est de nous permettre de disposer d'une base optimisée de ressources de formation se prêtant à une indexation plus fine et pouvant servir de base de départ au module " hypermedia adaptatif " développé par un étudiant du LIHS. Cette situation fera l'objet d'un protocole expérimental mené avec un groupe témoin de formateurs.

Perspectives ouvertes par l'indexation dans FORSIC

La stabilisation progressive par la masse (du point de vue des ressources, des formateurs, et des usagers finaux - étudiants de campus) fait la force des représentations présentées dans FORSIC. Elle s'applique simultanément à l'indexation des ressources en formation et des compétences des formateurs et permet d'obtenir une image - au sens d'une représentation globale et homogène - dynamique (évolutive) des compétences qui délimitent le champ des SIC. A terme, grâce à l'augmentation des ressources du système et au bout d'une phase d'expérimentation, nous espérons être en mesure d'isoler un certain nombre d'invariants permettant de parler d'**universaux d'actes de formation**.

La conjugaison au sein de FORSIC de modèles formels de haut niveau (modélisation du domaine de la recherche documentaire), de modèles d'indexation fortement calibrés qui permettent et intègrent la gestion de l'erreur (ontologie selon méthode 4/5), ainsi que de niveaux de collaboration "hybrides" (conjuguant humain et machine, notamment à travers l'architecture multi-agent) et clairement identifiés permettent d'entrevoir les contours d'une **pragmatique de la connaissance**.

Y être ou ne pas y être (sur le net)

Indexing, spamming, cloaking.

Jo Link-Pezet, Maître de conférences - URFIST de Toulouse
et Olivier Ertzscheid - Etudiant en thèse Toulouse II

Dans le cadre de notre rubrique habituelle sur la société de l'information, nous allons nous pencher sur deux thèmes qui nous sont chers : la recherche d'information et l'indexation. Nous avons utilisé diverses notes de lectures concernant en particulier :
1-la conférence "Search Engine Strategies" http://www.internet.com/, qui s'est tenue en 99 à San Francisco,
2- un article publié dans la revue Computer, Janvier 2000,
3- quelques réflexions ramenées du colloque SAS (World Multiconference on Systems, Cybernetics and Informatics) d'Orlando 2000 http://www.iiis.org/

L'accès à l'information sur Internet

Les moteurs de recherche représentent la clé d'accès principale au Web, ainsi que la porte par laquelle les populations qui veulent s'y exprimer ou offrir de l'information peuvent le faire. C'est aussi par là que ceux qui veulent tout simplement aller à la pêche à l'information parcourent le Web. Nous en avons déjà abondamment parlé.

L'indexation

Le point d'entrée, d'accès pour être repéré sur le Web, c'est l'indexation. Il existe, bien sûr, d'autres moyens. Par exemple, on peut suivre les liens ou tout simplement connaître l'adresse, mais le moyen le plus simple, c'est de chercher par les moteurs de recherche, ce qui revient à dire qu'on a été indexé (soit par des humains, soit par des robots). En effet, certains moteurs comme AltaVista, Lycos, Hotbot ont des robots, des spiders qui parcourent le net à la recherche de sites. D'autres, comme Yahoo préfèrent la touche humaine : les pages sont soumises à un éditeur qui juge de la qualité et de la pertinence du site en fonction de critères parfois mal connus; quelquefois, la quantité d'informations est insuffisante dans une catégorie, dans ce cas l'éditeur baissera la barrière. Ici, ce sont les humains qui détiennent les clés d'entrée.
Dans le cas où les robots visitent automatiquement les sites la question n'est pas plus simple. Encore faut-il recevoir la visite de ce robot. En général, le robot commence ses visites à partir d'une liste d'adresses historiques [ou des demandes de webmasters qui remplissent un formulaire sur le site de moteur], en général ils favorisent les sites qui comprennent de nombreux liens, comme les listes des serveurs par exemple. Les robots traversent le web en utilisant la structure de l'hypertexte. Ils trouvent un document. Ensuite, ils trouvent tous les documents qui

le citent. Toute information concernant le chemin d'un robot, les sites visités, la fréquence des visites et autres détails pratiques et concrets sont considérés comme secrets par les moteurs de recherche. Une des clés de contrôle offerte à l'internaute pour guider au moins partiellement l'indexation de son site réside dans les balises qui se trouvent dans l'en-tête de la page web. Il y aussi tout ce qui concerne le DUBLIN CORE.

Les quatre facteurs de reconnaissance d'une page

Le document réalisé par Junghoo Cho et ses collègues permet d'inférer les algorithmes de recherche des moteurs à partir de leurs comportements. Pour plus d'informations, on peut se reporter au texte Efficient crawling through URL ordering qu'ils ont produit et qui concerne la métrique utilisée par un robot pour déterminer l'importance d'un site ou d'une page.
Ils ont isolé certains facteurs :

- la requête : le robot n'a pas besoin de décider de l'importance d'une page parce que la recherche elle-même dirige le robot,
- la valeur des liens dans une page : plus une page est liée à d'autres, plus cette page est importante. Ce principe utilisé par Excite, Hotbot, Lycos, est moins vrai pour Altavista.

Il utilise la co-citation très fréquente dans le monde académique pour déterminer parmi les ouvrages cités les travaux dits canoniques ou d'autorité. Essayons d'imaginer le nombre de fois où un document doit être cité pour atteindre le niveau canonique.

- la technique des backlinks : qui recherche des liens à propos de grands sujets et de grands domaines. C'est à travers le nombre de liens qui pointent vers un page ou un site, appelés "backlinks" que les grands domaines déterminent la mesure du seuil d'importance d'un thème. En conséquence dans la plupart des cas, les petits sites de moindre importance vont être exclus.
- le rang de la page : cet aspect exacerbe le problème. En effet, au lieu de traiter tous les sites de façon égale, cette heuristique donne une préférence aux pages qui ont elles-mêmes un nombre de backlinks élevés : ce qui revient aux bonnes vieilles méthodes qui tendent à favoriser les travaux dits d'autorité. En effet, on pré-suppose que les backlinks sont des indicateurs d'importance ou de pertinence d'une page. Ce type de démarche peut aboutir à des situations très problématiques : certains vont jusqu'à offrir des cadeaux, des réductions, des logiciels gratuits et de l'accès à de l'information exclusive.
- la métrique de location : l'importance d'une page est liée à son adresse et pas à son contenu. Par exemple, on peut estimer qu'une URL qui se termine par .com est plus utile qu'une autre. Certaines adresses qui contiennent moins de barres obliques peuvent être considérées comme plus utiles que d'autres.

Pour faire simple, nous dirons que la situation n'est ni simple, ni claire. Un fait réel existe : les moteurs de recherche les plus importants ou les plus connus ou les plus riches donnent leur préférence aux sites les plus populaires, les plus riches ou les plus puissants, ce qui revient à donner la préférence (et la parole) à ceux qui ont le plus de moyens. Bien sûr, il faut se garder de généraliser trop vite. D' autres heuristiques peuvent guider les robots comme la largeur ou la profondeur de la représentation, le format. Certains robots admettent une déclaration volontaire. Mais ils sont rares et l'efficacité du procédé limité.

Le classement

Une fois que le site est indexé, il reste d'autres obstacles à franchir : comment obtenir un bon classement (sachant qu'un utilisateur parcourt environ deux pages, mais pas plus). Autant dire que les dix premières places sont chères. Le classement de pertinence (algorithme) est très difficile à justifier. Prend-on en compte la position et la fréquence des mots-clés? Il faut savoir que plus un mot est cité, meilleure est sa place. Tous les moteurs possèdent des algorithmes pour traiter le spamming et autres Cloaking .... Sur Lycos, un lien de popularité reflète le score de popularité d'un site en comptant les sites qui le citent. L'indice de clic est un système de type DirectHit qui se souvient des liens choisis par les internautes sur les moteurs et qui mesure le temps passé entre un clic et un éventuel retour de l'internaute sur la page de résultats pour en mesure la pertinence. A peu prés tous les moteurs sérieux en tiennent compte actuellement, soit avec la technologie DirectHit (Lycos, Hotnot), soit avec développé leur propre technologie similaire (Inktomi, Altavista).
On assiste à la naissance de nouveaux métiers comme webréférenceur, webmarketer, etc et d'une nouvelle discipline qui concerne la manière de créer des pages pour obtenir une bonne reconnaissance de la part des robots. Les principes utilisés pour améliorer le classement peuvent être associés à l'utilisation d'un logiciel particulier.
La conférence "Search Engines Strategies", s'est tenue en Novembre 99 à San Francisco et réunissait des créateurs de sites, des commerciaux, des promoteurs qui ont affronté les grands moteurs de recherches. Est apparue une troisième force représentée par les opérateurs de sites sans éthique, tenants du spamdexing (indexation vide), qui créent des millions de pages et mettent en œuvre des stratégies particulières pour attirer l'attention sur eux de façon peu scrupuleuse et très manipulatoire. Qu'est le spamming? Le mot "spam" vient d'un sketch des Monthy Python dans lequel le sous-titrage du discours d'un personnage était rendu illisible par l'affichage permanent d'une marque de jambon (la marque SPAM bien entendu ...). Il s'agit de fausser l'algorithme en utilisant des faux mots-clés, des textes invisibles au public, visibles aux seuls robots, qui en repérant ces mots-clés élève le score du site dans une guerre pour obtenir un bon score de classement (peu de robots disent comment ils fonctionnent pour éviter d'être trompés par les spammers. Prenons un exemple de spam utilisé par un moteur de recherche :
Si on regarde le code html, de la page d'accueil de FAST on s'aperçoit que le texte suivant est présent dans une balise :
* Fast Search & Transfer is the world leader in search and multimedia technology for the Internet and corporate market. *
A quoi correspond cette balise "teaser"? C'est un appel à une feuille de style de type css. Le navigateur voit que ce paragraphe est de la classe "teaser" (défini par le webmaster --- ça pourrait être "toto" ou "pipo"), il le met en forme : police, taille, couleur, ... de la façon dont "teaser" a été défini dans la feuille de style. Le lien vers la feuille de style est dans l'en-tête du document.
<link REL="STYLESHEET" HREF="/standard.css" TYPE="text/css">
A tous les coups, le standard.css définit que chaque fois qu'un paragraphe est de la classe "teaser" :

- la couleur de police est #0066A (idem bgcolor)
- la taille de la police est égale à 1

Résultat : du texte blanc sur fond blanc indétectable par les moteurs. Fallait y penser !!!!!!

Le cloaking (petit frère du spamming) est un bidouillage à double tranchant utilisé par les référenceurs aux Etats-Unis. Pour optimiser le classement d'un site web par des moteurs de recherche, certains webmestres utilisent la technique du masquage,

qui consiste à produire une page web personnalisée visible par lui seul pour chaque robot qui scanne le site. Les robots sont repérés grâce à leur adressage IP, en principe unique et infalsifiable. Les algorithmes d'indexation variant beaucoup selon les moteurs, cette technique permet de mieux optimiser le positionnement dans une multitude de moteurs. De plus, le cloaking permet de cacher ces astuces de référencement aux yeux de la concurrence. Mais les AltaVista, Inktomi et Northern Light jugent cette technique frauduleuse (le cloaking délivre une fausse page et autorise toutes les appropriations frauduleuses de contenus) et entendent bannir définitivement tout site qui l'utilise. Autre revers : le cloaking produit des pages invisibles vers lesquelles aucun lien ne peut pointer, ce qui réduit à zéro les critères de popularité qu'utilisent les principaux moteurs...Un groupe de chercheurs australiens s'est attaché à trouver des moyens pour lutter contre cet aspect: What is a tall poppy among Web pages?.

Guerre froide chez les cartographes du Web

Comment faire enregistrer un site et le placer favorablement dans les résultats de recherche. L'indexation et les métatags, les mot-clés et les tags de description sont de bons moyens techniques qui apparaissent dans le code source et agissent comme une loupe. Mais attention, métatags et mots clés sont les instruments favoris des spammers. Même si on nous dit que les 5 règles d'or d'un site sont : facile à lire, facile à naviguer, facile à trouver, consistent et facile à télécharger, il n'en reste pas moins qu'il faut créer des sites suffisamment technique pour surmonter les effets pervers du jeu du classement.
Pour faire simple, certains producteurs de sites tentent d'acheter la position la plus élevée. Yahoo par exemple vend de la place indirectement en offrant la possibilité de payer pour une indexation rapide. D'autres permettent d'acheter des mots clés qui affichent les résultats et la bannières des producteurs. Amazon a un arrangement de ce type avec Yahoo, Barnes et Nobles avec Lycos, NorthernLight.
Ces achats de mots-clés ne se font pas au hasard ...Les principaux moteurs de recherche disposent de "voyeurs" qui donnent au curieux en direct un échantillon des équations posées par les internautes :

- metacrawler t http://www.metaspy.com/top.html
- Excite taper http://www.excite.com/search/voyeur
- Abondance http://www.abondance.com/ressources/motcle.html (renvoi vers les voyeurs les plus connus ou les mots clés les plus utilises sur les moteurs)
- Lycos http://50.lycos.com/ pour site de Mister Sullivan http://searchenginewatch.com/facts/searches.html
- AltaVista http://tools.altavista.com/s?spage=AV/s12.htm
- système d'audit de la concurrence / mots-clés sur www.matowebo.com

Quel avenir pour les moteurs de recherche et l'indexation

On peut supposer qu'il va se jouer dans trois directions :

1. les algorithmes de type Google (plus il y a de liens vers lui, plus le site est bon) avec le danger du nivellement par le haut (manque de visibilité des sites alternatifs, prédominance de l'idéologie dominante ...)
2. les systèmes de bookmarks partagés (idée de base de Quiver et son QBar repose sur le partage des bookmarks des utilisateurs d'une communauté (ce qui est un plus par rapport à un éditeur comme Yahoo pour lequel l'expert ne

représente que lui-même) pour donner les sites leaders, et les tendances dans une communauté d'intérêt. Cela permet de classer en haut de liste, les sites les plus "populaires" (en principe les meilleurs) pour une communauté de spécialistes.

3. les moteurs humains (actuellement on ne parle que de ça - ou presque - dans les listes de discussion et news groups : il s'agit de sites où les demandes sont traitées par des internautes spécialistes du domaine. Il y en a beaucoup et on constate déjà quelques pratiques inquiétantes (par exemple, la plupart de ces spécialistes travaillent "on-line" dans des pays réputés offrir une main d'œuvre à des prix défiant toute concurrence ...)

Que deviendront les petits ?

Un fait semble clair : les moteurs de recherche les plus importants ou les plus connus ou les plus riches donnent leur préférence aux sites les plus populaires, les plus riches ou les plus puissants. La pertinence, le classement, les critères de pertinence sont menacés. Que vont devenir les petits sites, seront-ils de plus en plus difficiles à trouver?
L'"Open Directory Project" (servi par AOL Search et Hotbot) fonctionne sur le principe des éditeurs volontaires. Les utilisateurs sont appelés à devenir éditeurs (liberté d'expression et de contribution). Mais la question principale concerne la philosophie qui sous-tend le réseau? Est-ce que la dynamique du marché devrait être dominante sur le Web? Qu'en est-il de la liberté: libre-marché et utopie du libre-échange?
Les riches et les puissants influencent les moteurs qui vivent de produits dérivés et de publicité, qui sont tous en concurrences. La valeur de l'accès au Web dépend de catégories de biens que la politologue Elizabeth Anderson définit comme éthique de l'économie et qui ne doivent pas être laissés aux mains du seul marché. La connaissance, la culture ne peuvent pas être distribués comme de la marchandise, comment maintenir les principes publics, comment éviter de laisser les moteurs aux prises avec le marché. En introduction du colloque d'Orlando (Juillet 2000), Ananda Mitra (Wake Forest University) a rappelé qu'Internet est caractérisé par :

- la disparition des frontières, une forme rhizomatique qu'il s'agit de problématiser,
- un espace de textes et de discours dans lequel de nombreuses voix se font entendre et communiquent leur présence à travers des représentations diverses.

On ne peut nier que l'établissement de liens crée un espace politique, qui remet en cause et fait paraître obsolètes les hiérarchies traditionnelles et l'ordre précédent. Le pouvoir est lié à la façon dont ce discours est produit. Mais de façon têtue, se posent la question du contrôle et la nécessité de renégocier les différents modèles en réintroduisant la dimension éthique dans les discours afin d'éviter de voir les potentialités énormes que nous offre la mise en réseau reléguées au rang de marchandises utilisables en façon des contraintes du marché. C'est un devoir politique fondamental de préserver un espace public, libre, ouvert, par définition a-centré sur Internet en tant que réservoir d'information, forum universel, promoteur des valeurs humanistes, éthiques et écologiques.
Dans un sursaut de web citoyenneté, les personnalités les plus en vue du référencement viennent de créer le site Internet Positioning European Association une association à but non lucratif (loi 1901)
Quelques bonnes adresses

- le site de searchenginewatch

- le site d'Olivier Andrieu
- la liste de diffusion incontournable : motrech
  La liste Motrech consacrée aux moteurs de recherche sur Internet est un lieu d'échanges sur les problématiques, techniques, développements et évaluations/comparaisons des outils de recherche d'information sur Internet. Veuillez noter que la thématique du référencement sur les moteurs de recherche, ne doit pas être abordée sur la liste Motrech, puisqu'il existe une liste dédiée également sur egroups.fr (referencement@)

Les différents moteurs de recherche : fonctionnement

- http://www.searchenginewatch.com/
- http://www.metanews.net
- http://websearch.about.com
- http://www.searchengine-news.com/
- http://www.abondance.com
- http://www.abc-netmarketing.com/manuel/trafic/referenc/principe.htm

Le Baromètre de l'indexation

Délais d'indexation des principaux moteurs et annuaires.

- Yahoo.fr : de 1 jours à 3 mois
- Nomade : de 1 jours à 1 mois
- Francité : immédiat
- MSN.fr : indexation aléatoire
- Open Directory : 1 semaine à 1 mois
- Yahoo.com : de 1 jours à 3 mois:
- Altavista.fr : pas d'indexation ou très très peu ....
- Voilà : 1 jour pour la home page, le reste du site moins d'1 mois
- Altavista.com : 10-15 jours normalement 7 jours
- Hotbot : environ 7 jours
- Northern Light : à peu près 1 mois

Nos dernières découvertes :

- Karnak : même principe que "the informant" mais plus pertinent (plus de réponses, réponses classées selon la racine de l'url, ...)
- Guidebeam : qui utilise les services de voyeurs couplés sur la base de Google et qui permet d'affiner efficacement une recherche ...

Mémoires et intelligence collective

Jo Link-Pezet, Maître de conférences - URFIST de Toulouse

Pour moi, la réponse à la question : "Pourquoi faire de la recherche en sciences de l'information à l'heure actuelle", est simple. Parce que la numérisation, la mise en réseau, les processus socio-techniques en jeu exacerbent la question des mémoires numériques, du traitement et de l'accès à l'information dans des environnements distribués. En effet, l'évolution des technologies affecte l'organisation du travail, modifie les pratiques intellectuelles et dès lors, il nous paraît utile d'intégrer dans un même espace de recherche : l'information, l'information mise en mémoire, l'information en tant que flux, la connaissance, les savoirs, les savoir-faire, les représentations émergentes dans des environnements et des situations, où co-habitent l'individu et le collectif auquel il appartient ainsi que leurs trajectoires respectives et communes. Dans ce contexte d'une complexité accrue, il appartient au chercheur d'explorer ces nouveaux concepts, ces pratiques, de tenter de combler les lacunes des modèles existants et de comprendre les phénomènes en cours, les enjeux.. économiques, sociaux, politiques et humains.

Information et mémoires

On peut considérer qu'il y a :

- Trois espèces de mémoire : La mémoire de l'espèce (la mémoire génétique); la mémoire de l'individu (la mémoire "épi génétique"); et la mémoire sociale
- Deux lieux de mémoire : La mémoire interne est ce dont on se souvient, ou, peut-être, ce qu'on sait. La mémoire interne ne réside que dans le cerveau d'un individu. Par contre, toute forme de document constitue la mémoire externe. Toutes les formes d'inscription, d'enregistrement, de communication, d'aide -mémoire servent de mémoire externe.

Les mémoires documentaires s'articulent autour de deux axes majeurs : le langage, en tant que vecteur d'expression et le système d'écriture en tant que vecteur de production de l'information. Il est important de mettre l'accent sur la culture, car c'est par elle que s'exprime la mise en relation de la mémoire et du corps.
A l'époque de la transmission orale de la connaissance, le collectif faisait corps, faisait un avec sa mémoire. Avec l'écriture, une autre relation s'instaure semi-extérieure, semi-objectivée, relation différente au savoir, plus théorique et plus conceptuelle. L'imprimerie accélère le processus et invente les stocks, les catégories, les classifications. Il faut se soucier de l'organisation physique des contenus et du traitement (de la représentation) des contenus. Le rapport au temps change, apparition des délais de gestion et traitement. Lorsque l'informatique apparaît en tant que support, la mémoire externe est désormais totalement séparée du corps, car la mémoire interne et la mémoire externe sont disjointes. Nous voici confrontés au paradigme de la représentation de l'information. Le traitement de l'information est séquentiel et les approches théoriques correspondent à des modèles rationnels, formels tant du point de vue de l'architecture des systèmes que du point de vue cognitif et du découpage de l'activité en tâches et sous-tâches. Les conditions de traitements et de communication de l'information se transforment avec l'émergence d'un monde idéalement rationnel: raccourcissement du temps, mythe de la planification et de la rationalité absolues dans un univers peuplés de spécialistes et d'experts.
La mise en réseau des mémoires informatiques et la numérisation nous mettent au

cœur d'un dispositif dont nous sommes à la fois les témoins et les acteurs car se trouvent réintroduits le social, le collectif, l'autre, l'individu en tant que site de l'information. L'univers des réseaux est caractérisé par la complexité, la vitesse, la vitesse de la pensée. Les mémoires numériques mises en réseau exacerbent la question du traitement et de l'accès à l'information dans ces dispositifs d'écriture dynamiques distribués et posent la question du lien. L'Hypertexte , dont la figure est encore à explorer devient le point clé du dispositif. La mise en relation multiple devient une obligation qu'il faut apprivoiser en faisant des représentations cartographiques qui donnent à lire des formes synthétiques, graphiques, en mettant en évidence les itinéraires (réseaux, réseaux d'acteurs, réseaux-d'actants).

Statut des mémoires documentaires

Quel est le statut des mémoires documentaires ? Elles ne sont pas du côté des choses en soi (au sens de B. Latour), mais plutôt du côté des choses fabriquées (par l'homme). Nous avons opté pour une vision dialectique du système d'information associant les modèles formels (architecture informatique, systèmes classificatoires et protocoles de communication) et des modèles plus immanentismes (aspects cognitifs qui concerne l'usager) pour tenter de saisir la dynamique du changement et la place de l'individu dans le processus d'acquisition des connaissances.
Hybrides par définition, ces mémoires mettent en scène plusieurs technologies, différentes interfaces, donnent lieu à différents usages. Dès lors, on peut parler la co-spécification entre information, organisation du stockage, supports de stockage, modalités de traitement et usages. Le paradigme de l'émergence et de l'auto organisation nous conduit à associer dans un même regard information et connaissance, ce qui est une véritable rupture épistémologique. L'individu et le système retrouvent la circularité qui avait marqué le temps des récits et des traditions orales. Le souvenir émerge des connexions internes et des couplages structurels avec l'environnement dans un fonctionnement a-centré qui est celui de la mémoire en milieu ouvert.
La signification émerge de couplages successifs et de l'ensemble des relations et des connexions plus ou moins fortes, plus ou moins denses dans un environnement, où il est impossible de déterminer a priori le répertoire complet des comportements du système. Elle émerge d'agencements collectifs (distribution sociale de la connaissance) et de la capacité à comprendre les actions de l'autre dans des circonstances dont aucun ne perçoit la globalité. Le sens émerge a posteriori du contexte par propagation d'états de connaissance (internes et externes) converties en connaissance collective.

Une habilitation à diriger des travaux de recherche

" Mémoires et intelligence collective " : voilà le titre de la synthèse des travaux que j'ai menés ces vingt dernières années qui ont porté sur l'évolution des paradigmes de traitements de l'information, l'évolution des supports, des usages. Cette synthèse de travaux a été encadrée par M. F. Barthet, professeur d'Informatique et R. Boure, professeur en Sciences de l'information et de la Communication et je l'ai soutenue le 18 juillet 2000 à l'Université des Sciences Sociales devant le jury composé de :

- **Mme Marie-France Barthet** Professeur d'Informatique à l'Université de Toulouse I, France
- **M. Robert Boure** Professeur de Sciences de l'Information et de la Communication à l'Université de Toulouse III, France
- **M. Michael Buckland** Professeur en Information and Library Science à l'Université de Berkeley, Etats-Unis

- **M. Jean-Paul Metzger** Professeur de Sciences de l'Information et de la Communication à l'Université de Lyon III, France
- **M. Jean-Luc Soubie** Directeur de Recherche à l'Université de Toulouse III (IRIT), France
- **Mme Christiane Volant** Professeur de Sciences de l'Information et de la Communication à l'Université François Rabelais de Tours, France

Une évolution : les différents modèles de traitement de l'information

La représentation de l'information : modèles descendants

Les approches théoriques sont fondées sur des modèles rationnels, formels. Les modes de traitement de l'information en jeu sont séquentiels, symboliques, basés sur des représentations. Les modèles cognitifs reflètent cette démarche : on découpe l'activité à réaliser en tâches et sous-tâches : tout se passe à un niveau conceptuel et formel, statique. L'observation des modes de recherche d'experts en recherche documentaire montre que les experts doivent se soumettre à l'architecture imposée par les concepteurs de systèmes informatiques, mais qu'ils s'emploient à les contourner.
Leurs pratiques professionnelles les incitent à prendre en compte l'aléa, la perturbation, l'incertitude, à inclure le but, l'intention, l'action à accomplir dans la stratégie de recherche mise en oeuvre. La dialectique des moyens et des buts permet d'appréhender les processus cognitifs (déclaratif et procédural) de la recherche d'information et d'introduire la dimension stratégique et intentionnelle de la recherche d'information et de la décision pour l'action, ainsi que la relation à l'expérience et à la pratique afin de prendre en compte les différentes dimensions, les différents critères et scénarios nécessaires pour appréhender le futur.
Cette approche a été illustrée par une thèse, un inventaire des banques de données en Midi-Pyrénées et la participation à un projet européen d'aide à l'interrogation des bases de données en ligne.

Modèles et algorithmes

La recherche d'information restant extrêmement normée, il est difficile de faire émerger l'information contingente non visible à l'œil nu sur de grandes masses d'information. L'application de logiciels statistiques à de larges corpus de données permet de relier entre eux de larges ensembles d'informations, de les agréger, de présenter l'information sous une forme dynamique, synthétique, graphique (cartes ou tableaux), plus facile à lire et à interpréter pour faire des prévisions et d'appréhender le futur par le biais de représentations dynamiques. Cette approche utilisée dans une activité de veille et l'aide à la décision est une tentative pour accéder plus rapidement à la connaissance en mettant en évidence les liens et les relations qui existent entre des individus, des groupes, des objets, les itinéraires . On peut résumer cette approche sous la forme suivante :

Aide à la décision, Aide à l'interprétation	Modèles multidimensionnels	Temps, expérience
Théorie de la traduction	Représentations dynamiques	Stratégie, intention, culture
	Représentation info graphique : cartographie	Modèle de l'interprétation : mesure, trajectoire, processus

Cette approche a donné lieu à différentes recherches et publications concernant le rôle de l'information dans la décision stratégique

Le paradigme de l'émergence.

La mise en réseau des mémoires documentaires est une évolution remarquable de la situation et crée un environnement imprévisible dont personne ne peut percevoir la globalité où tout se passe par apprentissage, adaptation, communication à partir de ce qu'on sait (l'inscription corporelle de nos connaissances). La transformation dynamique des connaissances est constante. La recherche et l'usage de l'information sont liés à l' en-action (action incarnée dans un contexte social), qui reflète la constitution ontologique du sujet par rapport aux objets et qui nous renvoie à la capacité à que nous avons de comprendre les actions de l'autre et de communiquer avec lui. Si la signification émerge d'agencements collectifs d'énonciation, la connaissance se propage et émerge à travers la coopération et l'interaction. La réussite d'une situation coopérative est due aux capacités sensorielles, perceptives et motrices des acteurs qui la composent. Le sens émerge a posteriori du contexte par propagation d'états de connaissance (internes et externes) qui sont progressivement converties en connaissances collectives.
Le développement des supports informatiques permet de travailler ensemble (collecticiels, systèmes de gestion de plus en plus intégrés , hypertexte Objet-lien-lieu, technologie multi-agent). Autour de cette thématique, nous avons étudié les caractéristiques de l'hypertexte et de la coopération : travail coopératif appliqué à l'intelligence collective, animation de collectif intelligent, gestion collective des connaissances.

FORSIC : un collectif intelligent

L'idée de gérer collectivement des connaissances d'une communauté de formateurs à la recherche documentaire est liée à la notion d'organisation hyper textuelle. Il s'agit de collecter et de mettre en mouvement les connaissances d'un collectif de formateurs pour faciliter la formation à la recherche documentaire en faisant émerger un langage commun, un référentiel documentaire, en repérant les compétences des individus de façon aussi distinctive que possible. Toutes les personnes concernées propagent des états de représentation de leur connaissance dans le dispositif et favorisent l'adaptation du système en fonction des besoins exprimés et de leur évolution. Dans ce système coopèrent des individus, des outils, des modèles formels et des systèmes auto-organisateurs, des outils de gestion et d'indexation de connaissances (bases de connaissances), des outils de représentation cartographique, différents niveaux de finalités et de fonctionnalités. Il s'agit d'un dispositif d'échanges, de communication, de capitalisation de connaissances au sens de: création de la valeur pour les individus à travers la représentation d'eux-mêmes et leur communauté. La notion de coopération à tous les sens du termes et plus particulièrement celle d'indexation des compétences individuelles sont au cœur du dispositif, ainsi que le tissage d'expériences, des récits des savoirs et des savoir-faire d'hommes et de femmes.

Les sources sur l'indexation

Articles et Monographies autour de l'indexation

- The American Society For Information, Knowledge : Creation, Organization and Use, ASIS 1999 Annual Conference, October 31 - November 4, 1999 | JW Mariott Hotel | Washington, DC.
- Plaunt C. & Norgard B.A. (1998), An association based method for automatic indexing with a controlled vocabulary, Journal of the American Society for Information Science, 49, 888-902.
- Proceedings of the 20th Annual Meeting of the American Society of Indexers.
- Wellish Hans W., Indexing from A to Z, Bronx, NY : H.W. Wilson, 1991, pp. 357 - 358.
- Fox K.L., Frieder O., Knepper M.M., Snowberg E.J., " SENTINEL : A Multiple Engine Information Retrieval and Visualization System ", in JASIS, Vol.50, Number 7, 15 May 1999, 616-625.
- Davis C.H. & McKim G.W., " Systematic Weighting and Ranking : Cutting the Gordian Knot ", in JASIS, Vol.50, Number 7, 15 May 1999, 626-628.
- Sprague R.H. Jr. (1965) A Comparison of systems for selectively disseminating information. Bloomington. IN : Bureau of Business Research, Graduate School of Business, and Indiana University.
- Vignaux G., Piotowski D., Kieu Q., " Lexicographie et hypertextes ", in Hypermédias et Apprentissages, 4ème Colloque, Poitiers, 15-16-17 Octobre 1998, Université de Poitiers, Maison des Sciences de l'Homme et de la Société, pp. 97 à 101.
- Courtial J.-P., Introduction à la scientométrie - De la bibliométrie à la veille technologique., Paris, Anthropos - Economica, 1990.

On se réfèrera notamment à l'excellente bibliographie pour touts les problèmes concernant les différentes techniques infométriques (co-citations, ...)

- Nogier J.-F., Génération automatique de langage et graphes conceptuels, Paris, Hermès, 1991.
- CBMI'99, Proceedings on the first European Workshop on Content-Based Multimedia Indexing, IRIT, Toulouse, France, October 25-27, 1999

Concernant plus spécifiquement l'indexation des images ...

- Lowe H. (et al.), Towards Knowledge-based Retrieval of Medical Images : A Multidimensional Model using Semantic Indexing, Image content Representation, and Knowlege Based Retrieval, 1998, AMIA.
- Niblack W., Barber R., Equitz W. (et al.), " The QBIC Project : querying images by content using color, texture and shape. ", IBM Research Report, 1993, RJ 9203
- Popescu M. & Gader P., " Image content retrieval from image databases using feature integration by choquet integral ", Presented at The International Society for Optical Engineering (SPIE), Photonics West, Electronic Imaging '99, San Jose, January 1999.

- Rasmussen E.M. (1997), Indexing Images, Annual Review of Information Science and Technology, 32, 167-196.
- Répertoire de ressources Internet sur l'indexation image. WEBLIOGRAPHIE : Répertoire de ressources internet sur l'indexation des images, page personnelle de Henri Defago.
- Groupement de recherche sur l'indexation multimedia

Projets Européens d'indexation multimédia

- EUROMEDIA
- DIVAN (Distributed Video Archive Network). Construction d'une architecture distribuée pour la gestion - et l'accès - à de vastes archives télévisuelles. Dans DIVAN, une session d'indexation vidéo est répartie sur plusieurs machines, utilisant des connexions vers deux serveurs CORBA isolés - un serveur d'analyse vidéo et un serveur contenant une librairie de modèles.
- AVIR, (Audio-Visual Indexing and retrieval for non IT expert users) / porte essentiellement sur la vidéo
- IMDEX
- FORMULA
- COBWEB, porte essentiellement sur l'image
- CUIDAD , porte essentiellement sur le son.
- Projet STRETCH méthode d 'indexation d 'images et de données multimedia (CEA)

Articles et monographies concernant la construction et/ou l'expérimentation d'ontologies.

Ouvrages généraux

- Carnap R. (1967) The logical structure of the World : pseudoproblems in philosophy. University of California Press.
- Hirst G. (1989), Ontological Assumptions in Knowledge representation. Proceedings First International Conference on Principles of Knowledge Representation and Reasoning, Toronto, 157-169.

Sur des points plus précis :

- Chung L. (1993), Representing and Using Non-Functionnal Requirements : a process-oriented approach, PhD Thesis, Department of Computer Science, Univ. Of Toronto.
- Dardenne A., Van Lamsweerde A., Fickas S. (1993), Goal Directed Requirements Acquisition, in Science of Computer Programming, 20, 3-50.
- Feather M. (1987), Language Support for the specification and derivation of concurrent systems. ACM transactions on programming Languages, 9(2), 198-234.
- Maida A., Shapiro S. (1982), Intensional Concepts in Propositional Semantic Networks. Cognitive Science, 6, 291-330.
- Galbraith J. (1973), Designing Complex Organisations. Addison Wesley.

- Mintzberg H. (1979), The Structuring of Organizations, Prentice Hall.
- Scott W. (1987), Organizations : Rational, Natural or Open Systems. Prentice Hall, second edition.
- Yu E. (1993) Modeling organizations for information systems requirements engineering. Proceedings IEEE, San Diego, IEEE Computer Society press, pp. 33-41.

A propos de l'outil OntoLingua :

- Gruber T.R. (1992), Ontolingua : a mechanism to support portable ontologies. Technical Report KSL-91-66, Stanford University, Knowledge Systems Laboratory.
- Gruber T.R. (1993), A translation approach to portable ontology specification. Knowledge Acquisition, 5(2): 199-220

Les ontologies sur le web ...

- OntoServer homepage. L'un des projets les plus aboutis - développé par l'université de Karlsurhe. Intègre (entre autre ...) un répertoire d'ontologies, un moteur de recherche dédié, différents outils de traduciton, des application clientes, etc ...
- 1998 ACM Computing Classification System, ontologie de référence dans le domaine de l'informatique
- Richard's Home Site, page personnelle - Nombreux articles sur les modèles formels destinés au Knowledge Management
- Resources for Interaction Design: Articles and Speeches.Page personnelle - exemples de taxonomies dans divers domaines et articles scientifiques sur l'interaction et le savoir-partagé.
- "What is an ontology ?" Tony Gruber
- "Task ontology design for intelligent educational / training systems", Mizoguchi R., Sinitsa K., Ikeda M.,
- James R. Schoening, " A case and strategy for developing standardized educational domain modules and ontologies "

Tous les liens sont vérifiés et actifs au 06.12.2000

Extraits de Online Information 99 Proceedings :

Learned Information Europe , Online Information 99 - 7/9 décembre 1999, Londres.

http://www.online-information.co.uk

Elisabeth Lacombe, Conservateur en Chef - URFIST de Toulouse

D'autres extraits ont été résumés dans la Lettre n°24/25 : I. Les portails et II. Les moteurs.

III - LE LANGAGE XML (eXtensible Markup Language)

Extraits de :
HEMRICHM, SCHAFERU, XML-based concepts.
GRENN D, The Evolution of web searching.
Afin de pallier les limites de HTML, le consortium W3C a développé un sous-produit de SGML (trop complexe pour le web) permettant de prendre en considération les problèmes sémantiques et structurels de la recherche d'information.

DEFINITION

Il s'agit d'une technologie ouverte permettant d'améliorer la publication électronique, le commerce électronique, la recherche d'information et les échanges de données.

XML décrit l'information avec des doubles tags à de nombreux niveaux, créant ainsi un arbre hiérarchique. Cela permet l'accès direct à l'information cherchée.
XML décrit non seulement le contenu de la page, mais aussi les règles de manipulation des données. L'information utile peut être ainsi récupérée par un programme (comme Java script) sur le disque dur, ce qui entraîne un allègement du trafic. Le standard ouvert sur lequel est basé XML permet des échanges de données entre systèmes différents et entre pays différents (grâce à Unicode).

CARACTERISTIQUES DE L'HYPERTEXTE : LE WEB "EN SPIRALE"

Selon Vannevar Bush, "c'est comme si les éléments physiques avaient été rassemblés à partir de sources séparées et reliés pour former un nouveau livre".
L'hypertexte est modulaire, non linéaire, à plusieurs dimensions.
Modularité :
La connaissance ne s'organise qu'à partir du moment où des liens mentaux ont été construits entre des unités, dans un certain contexte.
Non-linéarité :
La présentation des textes doit trouver des médias plus pertinents pour représenter la connaissance . Celle-ci est un réservoir d'informations formidable, constitué d'unités d'information interconnectées, basées sur les associations faites par le cerveau. Il existe maintenant une technologie permettant aux auteurs et aux lecteurs de représenter et communiquer la connaissance de manière créative et non-linéaire, ce qui permet l'interactivité.
Multidimension :
Organiser la connaissance à partir des unités d'information implique plusieurs modes de recherche, de relier des réseaux existant au-dessus du texte : nous faisons des classifications et des catégories, nous préparons notre réservoir d'informations mentales à la recherche et à la navigation.
On peut représenter cette gestion de la connaissance sur du papier à l'aide de sommaires, d'index, de citations, de notes et commentaires en adoptant ainsi une représentation non linéaire de l'information. L'ordinateur a apporté la rapidité (les capacités du hardware) et l'automatisation (grâce aux moteurs de recherche). Mais ce n'est pas suffisant pour satisfaire le fonctionnement complexe du cerveau : il faut imaginer des concepts de liens intelligents pour représenter la structure de la connaissance et construire des outils de recherche d'information intelligents ainsi que

des supports de navigation.
Des systèmes hypertexte utilisant des concepts de liens plus compliqués que le Web existent :
par exemple Augment, Xanadu, Fress (File Retrieval Editing System), KMS, SEPIA, et Hypercard.
Ils s'utilisent dans un environnement restreint disposant de systèmes hypertexte d'information intelligents. Les produits basiques qui sont couramment développés sont notamment XLink, XPointer, Topic Maps.

TYPES ET SITUATIONS D'INFORMATION HYPERTEXTUELLE

Le web devrait être considéré comme une plateforme pour des applications hypertexte plutôt que comme un système hypertexte véritable. Voici quelques applications de type hypertextuel :

- applications commerciales : textes de droit et travaux de référence (encyclopédies, dictionnaires) ; manuels et documentation technique;
- applications informatiques : des concepts de liens complexes combinent la documentation technique avec des aides en ligne, de la documentation logicielle, pour donner des interfaces utilisateur constituant un immense hypertexte;
- des environnements d'apprentissage hypertextuels qui peuvent être coopératifs (interactivité étudiants/enseignants).

LES COMPOSANTS DES CONCEPTS QUI REGISSENT LES LIENS

Il faut prendre en compte les concepts individuels, les langages de liens, les concepts généraux et les systèmes hypertextuels afin de créer, générer, éditer et maintenir les divers sortes de liens dans un ensemble d'informations.
Les concepts individuels font référence aux questions : quels problèmes, quels besoins existent dans les situations réelles de publication.
Les langages de liens : quels sont les langages disponibles et quelles réponses apportent-ils aux scénarios réels. Les concepts généraux et les systèmes hypertextuels ne sont pas examinés ici.

LES CONCEPTS DE LIENS INDIVIDUELS

Typologie des liens.

La plupart des liens dans un ouvrage de référence sont des liens "*contenu-objet*" (lien d'une partie d'un document à un autre document). D'autres types de liens existent dans d'autres contextes : liens "*contenu-contenu*" (par exemple définition d'un terme) ou "*objet-contenu*" (d'un article vers une entrée d'index) ou "*objet-objet*" (lien entre deux articles dans un dictionnaire).
On peut créer différents types de médias à partir du même document source, en utilisant les liens déjà élaborés dans ce document. Les liens sont plus explicites dans les versions électroniques que dans les versions imprimées.
Toutes les occurrences d'un terme devraient être connectées à l'index dans les deux directions, deux liens étant alors nécessaires ; il est plus facile pour la maintenance d'utiliser un lien bidirectionnel. Il est encore plus utile de construire des liens multidirectionnels, chaque occurrence d'un terme étant liée à toutes les autres occurrences. L'existence de ces liens multidirectionnels n'a de sens que si les cibles sont décrites et catégorisées, afin que le lecteur puisse faire un choix.

Le concept de "variant".

Ce concept s'avère nécessaire, selon la dépendance du lien à un media ou un produit (par exemple, un lien peut être établi avec un cédérom ou une impression papier). On pourrait marquer les variants par des attributs dans le langage XML. Notons qu'il faut éviter de pointer d'un variant vers un autre, car les liens sont valides soit avec un variant, soit avec un autre.

Les données factuelles.

Données qui changent, sont incluses dans une base de données grâce à des liens d'interrogation. Elles peuvent être modifiées au niveau de cette base, et seulement une fois, au lieu de modifier chaque occurrence. Plusieurs possibilités permettent " d' adresser " un lien : dans le cas de données factuelles, on peut fixer comme cible : "addressing by query".

Quelques scénarios.

- Dans les publications en droit, les variants jouent un rôle important : par exemple la validité d'un projet de loi dépend de la date, des versions...L'éditeur (ou le système éditorial) devra s 'assurer que les variants sont toujours interconnectés.
- Les variants sont également importants dans la documentation technique : un produit peut être modifié plusieurs fois, la documentation variera sur certains points ; la documentation en ligne devra tenir compte des variants que constituent les niveaux des usagers.

  Les liens multidirectionnels sont essentiels dans les aides en ligne : le lecteur devrait pouvoir choisir à partir d'une liste toutes les occurrences correspondant à son sujet.
- Les liens doivent être particulièrement riches pour le bénéfice des éditeurs commerciaux et professionnels du web, car il leur faut offrir aux usagers des interconnexions diversifiées tenant compte de leurs besoins et de leurs intérêts différents.

Spécifications, moyens nécessaires.

Le standard permettant de construire un réseau de liens consistant et efficace est SGML (Standard Generalized Markup Language) ou XML.
La plupart des langages standard possèdent les structures de liens, qui sont indépendantes des systèmes et des applications. Les architectures de liens comme Topic Maps ou RDF (Resource Description Framework) devraient intégrer des possibilités de bâtir des réseaux d'information et de connaissance sur des liens.
Les concepts de liens doivent être compatibles avec les concepts de variants et de versions. Les outils doivent savoir adapter le système aux spécifications individuelles.

LE LANGAGE XML

Le langage HTML ne permet pas d'établir des liens suffisants pour les scenarios professionnels et commerciaux. Hytime, langage de lien puissant (sur support SGML) n'est pas adaptable sur de nombreux systèmes. Le W3 Consortium a développé XLink et les concepts Xpath et Xpointer pour fournir les structures de liens utilisées dans les documents XML.
Nous allons voir quels liens peuvent être exprimés par XLink et adressés par XPointer et XPath.

La sémantique de liens

On peut décrire la sémantique d'un lien ainsi que celle des ressources dans Xlink, par exemple, la relation entre un article sur une personne et l'information relative à cette personne contenue dans d'autres documents (biographie, travaux...)
Exemple :
Manuel de Falla -> Nuits dans les jardins d'Espagne(occurrence) Compositeur -> travaux (description)

```
<lemma-moreinfo>
xml :link= "simple"
rôle= "oeuvres"
title= "Nuits dans les jardins d'Espagne"
href= "rhapsodie/nuits-dans-les-jardins-d-Espagne.xml"
content-rôle= "compositeur"
Manuel de falla
</lemma-moreinfo>
```

Les liens multidirectionnels.

Toutes les œuvres composées peuvent être stockées comme une collection de liens. En activant un des liens sources, toutes les autres sources sont listées comme cibles. La liste des cibles contient la signification des liens, à savoir les "titres". Exemple :

```
<œuvres xml :link= "extended"
<loc
xml :link= "locator"
role= "composition"
title= "composition :West Side Story"
href= "musicals/west-side-story.xml"/>
<loc
xml :link= "locator"
rôle= "composition"
title= "composition :Chichester Psalms"
href= "liturgy/20thcentury.xml#id(bernstein-chichester)"/>
<loc
...etc
```

L'ensemble de ces liens corrélés peut être stocké dans un document séparé (maintenance plus facile)

L'adressage par navigation.

On peut décrire une cible de lien avec XLink combiné à Xpointer.
Exemple : la cible peut être une information sur le contexte de la production musicale.

```
<background xml :link= "extended">
<loc
xml :link= "locator"
role= "production"
title= "production of videotape"
href= "http://www....."/>
</background>
```

Cette structure permet de donner l'adresse de la structure interne d'un document.

Les réseaux de liens multiples.

Il est possible de créer plusieurs réseaux de liens, utilisés comme "couche" sur un ensemble d'objets d'information (comme des documents XML). Il faut définir le lien de groupes (link group) qui spécifie quelles sources contiennent les liens pertinents, par exemple les oeuvres des compositeurs :

```
<works xml :link= "extended">
```

<loc
xml :link= " locator "
role= " composition "
.....etc
</works>
Au début de l'article sur les oeuvres de référence, le fichier de liens est déclaré dans la déclaration du "link group", pour que le système hypertexte utilise le réseau de liens :
<link-network xml :link= "group"
<link-list xml :link= "document" href= "linknetwork-compositions.xml"/>
</link-network>

CONCLUSION

XLink fournit des structures de liens beaucoup plus complexes que HTML. Combiné aux langages Xpath et Xpointer, il permet :

- de spécifier l'adresse d'une ressource
- de décrire les ressources liées et leurs relations
- de pratiquer le traitement des ressources
- d'adresser de la même façon l'ancre et la cible (ce qui permet l'adressage des données qu'on ne peut que lire)

Après les recommandations du W3 Consortium, ce sont les fournisseurs de systèmes qui doivent maintenant implémenter ces langages grâce à des outils et des applications adaptées.

Bibliographie :

- XPointer
- XLink
- Topic Maps

17ème salon professionnel de l'information électronique et de l'Internet

Compte rendu

Olivier Ertzscheid - Etudiant en thèse Toulouse II

Allocution d'ouverture : L'Europe dans la société de l'information
Yves Berthelot – Secrétaire exécutif de la commission économique des Nations-Unies pour l'Europe.

- "L'information c'est dangereux."
Le secrétaire rappelle que les derniers rapports font état pour l'année 2000 de 20 pays "ennemis d'Internet", de 35 journalistes tués et de 87 journalistes retenus en otage.

- "L'information accompagne la démocratie"
Il faut remettre les citoyens au cœur des réseaux numériques
- "L'information sans la connaissance n'est rien."

Session n° 1 "La gestion des connaissances au service des performances de l'entreprise".

- **"Gestion des connaissances et veille technologique". Philippe Benhamou – Knowledge Manager – Mission Vie – ONERA**
- **"Partager le knowledge avec ses clients : l'expérience d'Ernst & Young". Carine Causse – Directeur de mission en commerce électronique – Ernst & Young.**
- **"La gestion des connaissances au service de l'innovation : applications à l'Institut Français du Pétrole." Daniel Champlon – Directeur de l'information à l'Institut Français du Pétrole.**

Session n° 2 "Les consortiums documentaires : état des lieux en France et à l'étranger".

- **"Le point sur les consortiums documentaires aux États-Unis" Ann Okerson – Bibliothécaire – Yale University. (conférence lue suite à un problème d'avion et à l'absence de la conférencière)**
- **"Analyse de la situation française concernant les consortiums documentaires". Ghislaine Chartron – Maître de Conférences – Urfist-Paris / Ecole des Chartes.**

Session n° 3 "Donner un cadre juridique à la société de l'information".

- **"Régulation d'Internet ?" Christian Paul – Député de la Nièvre – Chargé de mission auprès du gouvernement.**
- **"Internet et l'inflation législative". Pierre-Yves Gauthier – Professeur – Université Paris II.**

Session 5 "Webmestre : un métier ?"

- **"Vous avez dit Webmestre ?" Laurent Bernat – Directeur information et stratégie – Projetweb**
- **"Du documentaliste au webmestre" Documentaliste de la Caisse des Dépôts et Consignations.**
- **"Webmestre : de l'informaticien vers l'intégrateur" Laurent Prevosto – Ingénieur chercheur – EDF**

Atelier "Nouveaux besoins des organisations, nouvelles compétence professionnelles". Organisé par l'ADBS

- **"Les compétences professionnelles recherchées par les entreprises". Danielle Thiriet – Cox Consultant.**
- **"Comment les formations s'adaptent-elles ? La licence professionnelle est-elle une des réponses possibles ?"- Christiane Volant, IUT de Tours - Hélène Perennou, IUT de Paris V**

Session n° 8 : "L'intelligence économique offensive"

- **"Le marketing viral sur Internet : stratégies d'attaque et de défense" – Arnaud Dupui-Casteres – Directeur Léo-Burnett Consultants - Bruno Walther – Directeur associé – L'enchanteur des nouveaux médias**
- **"L'I.E.O. : nouveaux enjeux, nouvelles stratégies". Christian Harbulot – Directeur de l'Ecole de Guerre Economique – Cofondateur d'Infoguerre Consultants.**
- **"Présentation des principes de 4 logiciels d'I.E.O". Claude Michel & Christophe Bourgeret, Ingénieurs chargés d'étude guerre de l'information – Thompson CSF-Communication.**
- **"La guerre du sens : les procédés d'action dans les champs psychologiques" Général Loup Francart – Directeur de recherche – Institut des Relations Internationales et Stratégiques.**
- **Nos commentaires**

Les liens du salon ...

Session n° 1 "La gestion des connaissances au service des performances de l'entreprise".

"Gestion des connaissances et veille technologique." Philippe Benhamou – Knowledge Manager – Mission Vie – ONERA

Le projet développé par un doctorant de l'ONERA a pour objet de mettre au jour et d'analyser les interactions entre la capitalisation de connaissances et la veille scientifique et technique. L'objectif est de faire de la veille en partant des modélisations de la connaissance.
Suit une présentation du concept de **"Livre de connaissances"** dont la finalité est double : permettre à l'individu de mieux se connaître et aux autres de reconstruire des connaissances. Il s'agit donc tout à la fois de capitalisation et de transmission effective de savoir-faire.
Ce livre de connaissance s'articule autour d'une série de domaines :

- modèle de domaine
- modèle d'activité (qui décrit un enchaînement d'activités, suite à une question du type "comment faites-vous pour ?")
- modèle de tâche (savoir-faire maîtrisé par l'expert)
- modèle de concept (avec une terminologie associée)

A titre d'exemple, un livre peut comporter 10 modèles de domaine, 20 modèles d'activité, 30 modèles de tâches.

"Partager le knowledge avec ses clients : l'expérience d'Ernst & Young". Carine Causse – Directeur de mission en commerce électronique – Ernst & Young.

Présentation d'un outil permettant de faire du Knowledge Management avec un réseau client : "ERNIE". A destination des PME, cet outil offre une panoplie de services à accès payant dont :

- un service de F.A.Q. dans les domaines du management, de la comptabilité et de la fiscalité

- des outils on-line d'auto diagnostic ou d'aide à la décision
- un accès gratuit à la base constituée par les Previously Asked Questions (PAQ) du service.

Voici le cycle Knowledge au cœur du service ERNIE, tel que décrit par l'intervenante :

1. Le client identifie une problématique de gestion
2. Il soumet une question au service
3. Il reçoit une réponse individuelle
4. Sa question et la réponse sont archivées dans la base des PAQ
5. La base des PAQ est utilisée pour identifier des tendances
6. Lesquelles identifient les besoins
7. Lesquels conduisent à la création d'outils
8. Ces outils génèrent des méthodologies
9. Ces méthodologies sont mises à la disposition de la communauté des abonnés ERNIE

Adresse du service : http://ernie.ey.com
Correspondante du réseau ERNIE en France : anne_gabet@ernst-young.fr

"La gestion des connaissances au service de l'innovation : applications à l'institut français du pétrole." Daniel Champlon – Directeur de l'information à l'Institut Français du Pétrole.

Présentation de l'intranet de l'IFP construit dans le cadre du projet KAPPA dont la philosophie est la suivante : tout ce qui est produit à l'IFP (note, brouillon, documents finalisés ...) doit pouvoir être capitalisé pour innover. Un gros effort est donc fait pour arriver à l'uniformisation (via des couches successives de normalisation) des documents.
A l'issue de ce projet, l'IFP est arrivé à la conclusion suivante : la capitalisation des connaissances est un objectif réalisable en interne, pour le reste et au-delà d'une certaine masse critique, il devient nécessaire de sous-traiter les aspects traitement (dans ce cas, un contrat est passé avec l'INIST).

Session n° 2 "Les consortiums documentaires : état des lieux en France et à l'étranger".

"Le point sur les consortiums documentaires aux Etats-Unis" Ann Okerson – Bibliothécaire – Yale University.

La tendance qui veut que "les bibliothèques soient plus près des monastères que du commerce" est aujourd'hui passée de mode. L'impossibilité de l'exhaustivité dans le recensement, la collecte et le traitement de l'information a été à l'origine de la naissance des consortiums, dans une optique de création d'un fond documentaire national distribué. Ces consortiums peuvent être aussi bien régionaux que thématique. Leur point commun – et leur finalité – est de permettre de négocier des contrats avantageux pour les fournisseurs et les utilisateurs. Côté vendeur, il s'agit d'une nouvelle stratégie commerciale, côté acheteur, c'est une force réactive de négociation. Il est à noter que la plupart des bibliothèques des Etats-Unis font partie de plusieurs consortiums.
La plupart des références sur ces consortiums sont regroupés sur le site www.library.yale.edu

On y trouvera notamment des informations sur :

- **TRLN** (Triangle Research Libraries Network)
- **NERL** (NorthEast Research Libraries Consortium), créé et géré par Ann Okerson (Ann.Okerson@yale.edu) .
- **LIBLICENCE** : centralise des informations pratiques sur les licences consortiales permettant de gérer l'accès à ces ressources gigantesques et qui doivent être négociées avec les fournisseurs. On trouvera ici tous les détails ainsi qu'un certain nombre de conventions types et des logiciels gratuits.
- **ICOLC** (International Consortium of Library Consortiums) : fut créé spontanément en 1997 par des représentants de consortiums et joue aujourd'hui le rôle d'organisme fédérateur de l'ensemble du mouvement. La prochaine réunion du chapitre Europe de l'ICOLC se tiendra à Berlin en Décembre 2000.

"Analyse de la situation française concernant les consortiums documentaires." Ghislaine Chartron – Maître de Conférences – Urfist-Paris / Ecole des Chartes.

G. Chartron, s'inscrivant dans le contexte de l'économie des réseaux, a mené une enquête sur les trois derniers mois auprès de structures diverses (EPST, SCD, Grandes Ecoles, Pôles Européens ...) pour un échantillon de 56 questionnaires et/ou entretiens de Mars à Mai 2000. De cette enquête il ressort que :

- les produits concernés sont essentiellement des bases de données bibliographiques, des données factuelles et des revues électroniques, majoritairement dans le domaine scientifique et bio-médical.
- Parmi les principaux acteurs de ces consortiums on retrouve les SCD, les laboratoires et les associations.
- La structure juridique choisie est soit associative soit celle d'un GIP (Groupement d'Intérêt Public)

Le mouvement des consortiums documentaires, initié en France en 1997 est actuellement en plein essor. G. Chartron conclut en plaidant pour une carte documentaire numérique nationale (autour – par exemple – d'un site web fédérateur), ainsi que pour la prise en compte nécessaire de la dynamique des modèles alternatifs de l'édition scientifique.

Session n° 3 "Donner un cadre juridique à la société de l'information".

"Régulation d'Internet ?" Christian Paul – Député de la Nièvre – Chargé de mission auprès du gouvernement.

L'idée d'un modèle français de l'Internet autour d'une co-régulation (public-privé) fut le point de départ de la mission lancée à Hourtin par Lionel Jospin. C. Paul précise que la co-régulation n'est ni une privatisation du droit, ni la co-gestion d'un pouvoir pré-judiciaire et réaffirme que le choix français est qu'il n'y ait pas de "gendarme de l'Internet". Il rappelle "qu'une bonne loi libère et protège, elle n'encadre pas"... Pour plus d'informations ... rendez-vous à la remise du rapport au Premier Ministre d'ici 2 à 3 semaines (Mi-Juin) ...

"Internet et l'inflation législative." Pierre-Yves Gauthier – Professeur – Université Paris II.

P.Y Gauthier rappelle les principes énoncés par d'éminents juristes comme Kersen pour qui "La norme, ce n'est pas seulement la loi, la règle, c'est aussi l'accord

collectif et individuel".
En dénonçant "l'inflation législative", il plaide pour un minimum de lois mais sur des points essentiels, laissant le reste aux bons soins du principe de "co-régulation" précédemment évoqué. Il s'inscrit dans une conception du droit historique (de Savigny à Rippert) selon laquelle le droit se constitue par strates, et qui veut que les strates les plus anciennes soient aussi les plus sages.
Une question dans la salle rappelle qu'en dehors de "toute fantasmagorie techno-libertaire", le droit du cinéma par exemple, s'est constitué progressivement, après avoir été assimilé dans un premier temps au droit du théâtre, et qu'il sera nécessaire voire inévitable de légiférer sur les spécificités d'un média tel qu'Internet.

Session 5 "Webmestre : un métier ?"

"Vous avez dit Webmestre ?" Laurent Bernat – Directeur information et stratégie – Projetweb

L'objectif de cette intervention est de définir un profil dans une optique de recrutement. Le webmestre est encore trop souvent un homme-orchestre alors qu'il devrait être un chef d'orchestre.
La notion de site web est présentée comme se structurant autour de 3 "briques" :

- brique "contenu"
- brique "éditorial"
- brique "infrastructure"

C'est la seconde (comprenant l'architecture, l'ergonomie et le graphisme) qui paraît définir au plus juste la fonction de webmestre.

"Du documentaliste au webmestre" Documentaliste de la Caisse des Dépôts et Consignations.

Cet exposé est un compte rendu d'expérience. L'oratrice entend dresser les portraits croisés (en termes de compétences requises) du documentaliste et du webmestre. Pour cela, elle dresse une liste des tâches à accomplir dans l'élaboration d'un site web (exemple du site de la Caisse des Dépôts, élaboré dans un contexte de communication de crise anticipant les effets du passage à l'an 2000) :

- Tâche "contenus": il s'agit de la fonction classique de la documentation qui consiste à établir un plan de classement.
- Tâche "navigation – ergonomie" : là encore, on dresse un thésaurus de termes génériques que l'on dispose dans une frame.
- Tâche "création de liens hypertextes" : il s'agit d'un travail visant à repérer et à isoler des mots-clés.

Parmi les compétences ne relevant pas directement de la documentation, on retient :

- la réalisation de la maquette et de la charte graphique : cette partie a été sous-traitée.
- La mise à jour informatique et la bascule des textes : ces compétences ont été –facilement – acquises.

A l'issu de cette expérience, une définition du webmestre pour la base de profil de poste de la caisse des dépôts est décrite en ces termes : "met en œuvre, gère et anime au quotidien en veillant au respect des règles". Il s'agit donc davantage d'un gestionnaire de l'information que d'un producteur d'information. En termes d'organisation de contenus et de mise au point d'interfaces d'accès à l'information, les fonctions de webmestre et de documentaliste semblent donc parfaitement

assimilables.

"Webmestre : de l'informaticien vers l'intégrateur" Laurent Prevosto – Ingénieur chercheur – EDF

L'orateur distingue deux phases essentielles dans l'évolution des technologies associées au web : nous sommes au cœur du passage d'un web dynamique à un web dit "de phase industrielle". Parmi les causes de cet essor du web dynamique on trouve :

- de nouvelles technologies côté serveur :
  - API serveurs
  - SSL (technologies de chiffrement)
  - Banalisations des sessions persistantes (qui permettent, grâce à l'utilisation de cookies, de garder en mémoire un certain nombre de renseignements concernant un utilisateur quand on passe d'une page web à une autre)
- De nouvelles technologies côté client :
  - Javascript, Vbscript
  - Java

L'ensemble de ces technologies engendre de nouveaux outils (forums, passerelles webmail, informatique documentaire ...) qui permettent au web et à ses applications de passer du statique au groupware. On trouve ainsi de plus en plus de passerelles web-SGBD avec des architectures soit légères (de type ASP, PHP fonctionnant avec du scripting in-line, c'est à dire que les ordres dynamiques sont au niveau de la page web) soit lourdes (avec des applications du type WebObject, Oracle Application Server ...)
Le web paraît donc s'engager délibérément dans une phase de développement de type industriel dont les deux axes moteurs d'innovation sont la mise au point de nouveaux contenus (multimédia, flash, XML ...) et le rendu en termes d'accessibilité de ces contenus dans de nouvelles interfaces (PDA – Palm Pilot – , téléphonie WAP, navigateurs vocaux ...).
On peut alors isoler deux types de webmestre :

- le webmestre de site
- le webmestre d'entreprise (ou "responsable Internet", ou "directeur technique")

Atelier "Nouveaux besoins des organisations, nouvelles compétence professionnelles." Organisé par l'ADBS

"Les compétences professionnelles recherchées par les entreprises." Danielle Thiriet – Cox Consultant.

Les nouveaux profils sont ceux de veilleurs, de webmestres et de cyberdocumentalistes. En terme de demande, les entreprises recherchent une spécialisation très pointue dans le domaine (on demande de moins en moins de généralistes), des compétences techniques (informatique, numérisation, Internet) et des compétences pédagogiques (être capable de former une équipe). Les principaux problèmes rencontrés avec les candidats concernent le manque de pratique des logiciels documentaires, la méconnaissance des outils linguistiques et un décalage croissant entre leur niveau de formation et leur pratique professionnelle.
L'intervenante terminera par un "message" aux établissements d'enseignement en leur suggérant de proposer des séminaires sur le comportement professionnel et d'accentuer l'étude de l'informatique et des NTIC, de développer l'approche

marketing ainsi que la pratique des langues. Les candidats à l'embauche doivent de leur côté accepter de nouvelles méthodes de travail axées sur la production collaborative et le partage de l'information et des méthodes.

Kristy Lepage de Gemini Consulting reprendra le même argumentaire en l'articulant autour des 4 axes des métiers de l'information chez Gemini :

- information services,
- business analyse,
- internal knowledge management,
- knowledge network.

Nota-Bene : la plupart des transparents utilisés par les intervenants seront disponibles sur le site de l'ADBS (www.adbs.fr)

"Comment les formations s'adaptent-elles ? La licence professionnelle est-elle une des réponses possibles ?"
- Christiane Volant, IUT de Tours - Hélène Perennou, IUT de Paris V

Les intervenantes rappellent que la licence professionnelle a été créée par l'arrêté du 17 Novembre 1999 avec comme objectifs l'insertion professionnelle, la validation des connaissances et des compétences des salariés dans le cadre de la formation continue et la nécessité de faire face aux exigences du marché de l'emploi. Dans ce cadre, 2 projets de licence professionnelle ont été déposés début février.

Projet de licence professionnelle "Veille IST dans le secteur des sciences de la vie", Hélène Perennou, IUT Paris V.

- Licence par capitalisation d'unités de valeur
- Pré-requis : diplôme universitaire de premier cycle ou 5 ans d'expérience professionnelle.
- Parcours pédagogiques différenciés.
- Semestrialisation de la formation
- Enseignement :
  - Module 1 : Recherche d'information (méthodologie, outils documentaires ...)
  - Module 2 : Traitement de l'information (indexation, management des connaissances, validation des contenus, analyse de sites web ...)
  - Module 3 : Technologie de l'information (conception de sites web ...)
  - Module 4 : Anglais (anglais scientifique, anglais Internet, anglais pour les professions de l'information)
  - Module 5 : Communication écrite et orale
  - Module 6 : Droit de l'information et de la propriété industrielle.
  - Modules 7, 8 et 9 : culture générale, professionnelle, scientifique (selon le profil des candidats)
- 1 projet tutoré sur 4 mois en binôme
- 1 stage de 12 semaines aboutissant sur un mémoire.

Projet de licence professionnelle "Management de l'information" Christiane Volant, Présidente de la commission formation de l'ADBS.
L'ouverture de ce projet est demandée pour Septembre 2001. Il s'adresse plus particulièrement aux PME.

Après avoir fait le constat selon lequel l'émergence des nouvelles fonctions (veille, etc. ...) se conjugue avec les actions de communication des entreprises, favorisant l'émergence de mémoires d'entreprises, l'information est plus que jamais un matériau stratégique pour les organisations (qu'elles soient publiques ou privées) et les Systèmes d'Information sont la clé du pilotage de ces organisations. La nécessité se fait donc chaque jour plus grande de se rapprocher des organisations et du langage de l'entreprise pour être enfin capable de dialoguer avec ses diverses composantes (DRH ...). Pour mener à bien ce projet, un partenariat a été mis en place avec la région (ARIST) et avec les UFR gestion et histoire de l'université de Tours.

- Pré-requis :
  - DUT "Information et Documentation d'Entreprise"
  - DUT "Services et Réseaux de Communication"
- ==> l'objectif de ce recrutement est d'obtenir de nouveaux profils par synergie.
- Enseignements du tronc commun :
  - Information et société (enjeux de la société de l'information)
  - Management et organisation d'entreprise
  - Veille et intelligence économique
  - Management des connaissances et mémoire d'entreprise
  - Anglais
  - Communication interne et externe
  - Réseaux (90 heures)
  - Hypermédias et multimédias.
- Stage de trois mois.
- Cadre horaire : 450 "heures-étudiants"
- Licence ouverte en formation continue.

Session n° 8 : "L'intelligence économique offensive"

"Le marketing viral sur Internet : stratégies d'attaque et de défense" – Arnaud Dupui-Casteres – Directeur Léo-Burnett Consultants / Bruno Walther – Directeur associé – L'enchanteur des nouveaux médias

Trois grands types de champs sensibles sont identifiés comme cibles privilégiées du "marketing viral" :

- les technologies liées à l'industrie
- les activités militantes
  - liées à l'environnement
  - liées à la sécurité alimentaire
- les activités politiques

Au rang des stratégies dites "offensives" on trouve :

- création de communautés d'influence à visage masqué
- instrumentalisation des forums (et diverses techniques d'infiltration consistant à lancer des rumeurs – ou pour reprendre une terminologie plus marketing : "faire de la communication d'influence", ou à valider de fausses informations)

- instrumentalisation des sites prescripteurs (et notamment les sites des associations consuméristes)

Au rang des stratégies "défensives":

- identifier d'où vient l'information
- diffuser un contre-message
- créer une fausse cible (créer par exemple un site web faussement et/ou faiblement protégé qui sera la cible privilégiée des hackers).

"L'I.E.O. : nouveaux enjeux, nouvelles stratégies." Christian Harbulot – Directeur de l'Ecole de Guerre Economique – Co-fondateur d'Infoguerre Consultants.

La France n'a pas l'habitude du partage d'information. L'arrivée d'Internet inaugure cependant de nouvelles réalités économiques concurrentielles générant des affrontements d'un nouveau type qui vont au-delà de la simple veille économique ou stratégique. En effet, les logiques d'affrontement prennent chaque jour le pas sur les logiques d'innovation. Il est plus facile, plus rapide et moins cher de déstabiliser une entreprise par l'information en l'attaquant par des sources ouvertes (c'est à dire légales). Dans ce contexte, l'ancienne terminologie du "renseignement" et de "l'espionnage industriel" devient caduque, et, dès 1994, le rapport Martre devient l'origine de l'I.E.O.
"Ce n'est pas celui qui a la plus grosse bombe qui l'emportera dans les conflits de demain, mais celui qui racontera la meilleure histoire." J. Arquila & D. Ronfeldt. Les exemples de ces propos de 2 spécialistes de l'infoguerre ne manquent pas, qu'il s'agisse de logiques de rumeur dans l'actualité boursière (occasionnant en 1998 la spectaculaire chute d'Alcatel), ou de vrai-faux sites Internet fonctionnant comme des leurres d'opinion.
Les principes fondateurs de l'attaque indirecte par l'information sont les suivants :

- créer le doute.
  - trouver la faille et jeter le discrédit.
- entretenir le doute.
  - en temps réel et au niveau planétaire.
- jouer sur l'opinion publique.
  - par caisse de résonance et par ricochets sur les médias.

Parmi de nombreux autres principes et stratagèmes figure en bonne place le concept de "perception management", inventé (!) par le département de la défense des Etats-Unis en lieu et place de celui de "communication de crise", qui consiste à fournir des informations et des indices à des audiences étrangères pour influencer leurs émotions, leurs motivations et leurs raisonnements objectifs. Après quoi il faut encore faire œuvre de pilotage stratégique et de gestion des menaces, conditions nécessaires à l'I.E.O. mais auxquelles la mentalité française n'est pas du tout adaptée puisqu'il s'agit – entre autres – de donner un rôle clé à l'anticipation, d'assurer une transmission rapide et décentralisée des informations, d'envisager une prise de risque relative à l'interprétation de signaux faibles.
L'avance des Etats-Unis dans tous ces domaines paraît essentiellement due à la rotation des élites entre le Politique, le Militaire, le Stratégique et l'Economique, qui permet d'éviter la dichotomie française avec d'un côté les "gentils", tenants de "l'intelligence-économique-Haré-Krishna -qui-partagent-tout-en-se-donnant-forces-embrassades" et les "méchants-militaires-qui-voient-de-la-guerre-partout".

Un site de référence : www.infoguerre.com
Et si décidément vous n'aimez pas la guerre : www.walt-disney.com

"Présentation des principes de 4 logiciels d'I.E.O." Claude Michel & Christophe Bourgeret, Ingénieurs chargés d'étude guerre de l'information – Thompson CSF-Communication.

L'intervenant commence par rappeler quelques concepts clés de la terminologie ayant trait à l'I.E.O. comme celui de "War-room", inventé au Pentagone pendant le seconde guerre mondiale et décliné de nos jours en "Virtual War-Room" (1994) faisant usage de logiciels de types SIAD appliqués à la guerre économique et consistant à réunir des experts derrière plusieurs ordinateurs en appliquant une méthodologie guidée de type MRT (Méthode – Raisonnement – Tactique). A cette fin, plusieurs projets logiciels sont à l'étude.
Le conférencier précise également, non sans humour, que le concept de "déception" en langage militaire signifie "tromperie", le débarquement allié en Normandie étant la première illustration de ce type de stratagème.

- **Logiciel "Aristote"** (en référence à la "Rhétorique")
  - Système informatisé d'aide à la conception de messages persuasifs en fonction de l'auditoire.
- **Logiciel "Schopenhauer"** (en référence à "L'art d'avoir toujours raison "de Schopenhauer" et reposant également sur les principes des "Topiques" d'Aristote)
  - Système informatisé d'aide à la réfutation des arguments adverses.
    - Techniques directes (attaque des hypothèses ou des conclusions)
    - Techniques indirectes (ad hominem)
- **Logiciel "Isocrate" :**
  - Système informatisé d'aide à la détection de la désinformation/déception
- **Logiciel "Gorgias" :**
  - Système informatisé d'aide à la création de la dissuasion, au renversement des croyances des adversaires.

Le projet "Aristote" fonctionne avec une base de données contenant la hiérarchie des valeurs de l'auditoire. Trois couches d'algorithmie doivent permettre dans un premier temps de structurer ces valeurs, dans un deuxième temps de gérer la présentation des faits, et enfin de rechercher l'argument percutant. Pour ce qui est des arguments, l'algorithmie repose sur des distinctions entre arguments "quasi-logiques" (transitivité, réciprocité, contradiction, identité, définition, tautologie, justice, inclusion, faux-dilemme ...) et arguments "empiriques" (pragmatique, fin-moyen, autorité, gaspillage, direction ou encore corax – le plus vieux de tous les arguments dont on ait trace en rhétorique – qui fonctionne sur le modèle : "il y a trop de preuves contre lui, il doit être innocent !"). A ces 2 catégories, on en ajoute une troisième, les arguments de "dissociation des notions" (qui fonctionnent sur un mode inverse à celui de l'amalgame). Il ne s'agit encore que de phases de recherche et développement avec sortie d'ici 3 à 4 ans du fait de la complexité mise en jeu. Mais du fait du haut intérêt stratégique de ces recherches et de l'avance des Etats-Unis en ce domaine qui possèdent déjà ce type de logiciels et les utilisent en phase bêta, il paraît assez délicat d'estimer la réelle avancée de ces projets ... Du moins les intervenants soulignent-ils que les prochaines étapes de la R&D dans ce secteur devront s'attaquer aux problèmes de :

- SGBD d'informations liées à la psychologie
- Prise en compte de l'éthique pour l'aide à la décision

- Aide à l'argumentation scientifique et technique (projet Wittgenstein).

Pour pouvoir valider les différents stratagèmes qui sont au cœur de l'algorithmie de ces différents logiciels, il est fait appel à l'environnement de la théorie de jeux ainsi qu'à diverses techniques de psychologie sociale.

Techniques de psychologie sociale pour la persuasion :

- effet mimesis (on désire le désir de l'autre – Girard)
- effet pied dans la porte (proposer moins pour avoir plus)
- effet porte au nez (on annonce très cher et on baisse)
- effet de dotation
- effet nahum ("ça ne mange pas de pain ...", "après tout, j'ai rien à perdre ...")
- effet de dépassement
- effet de capitonnage argenté (si l'on isole les gains et qu'on les regroupe ou si l'on isole les pertes et qu'on les regroupe, l'effet produit ne sera pas le même)
- effet du spectre des possibles
- effet donner pour recevoir (pots de vin)

L'orientation de la recherche française est celle de la psychologie sociale, alors que les américains préfèrent la PNL (programmation neuro-linguistique)

"La guerre du sens : les procédés d'action dans les champs psychologiques" Général Loup Francart – Directeur de recherche – Institut des Relations Internationales et Stratégiques.

En termes de stratégie militaire, on distingue 4 grands types de communication :

- communication signifiante
- communication informationnelle
- communication promotionnelle
- communication comportementale,

et l'on prête un intérêt tout particulier à la première catégorie qui permet de prendre en compte un contexte existant (qu'il s'agisse du contexte historique et normatif de communautés ou du contexte expressif et personnel des leaders d'opinion, le tout recadré dans une logique des situations), pour le modifier et construire un nouveau contexte. Les forces d'intervention de type "casques-bleus" utilisent ces modèles.

Qu'il s'agisse de sens signifiant (qui donne de la cohérence à l'action), de sens signifié (qui donne de la pertinence à l'action) ou de sens significatif, la "guerre du sens" vise à la maîtrise des attitudes et comportements (tels que définis par la psychosociologie) de l'adversaire et/ou de l'opinion. Cette dernière est constituée de plusieurs strates : groupes primaires ? guides d'opinion ? médias ? leaders d'opinion ? décideurs, selon une structure pyramidale.
La stratégie militaire a isolé toute une batterie (sic) de moyens d'action permettant de faire évoluer les attitudes et les comportements. Cette phase que l'on peut qualifier de "manipulation" n'est pas "que" de la manipulation selon l'orateur, puisqu'elle fait partie intégrante de toute forme de communication. Le général isole 4 grandes catégories ayant chacune leurs stratégies propres :

- Communication :
  - information
  - argumentation

    - suggestion
- Persuasion :
    - suggestion
    - persuasion
    - obédience
- Mystification :
    - Déception
    - Intoxication
    - Désinformation
    - Stratagèmes
- Aliénation :
    - Propagande
    - Subversion
    - Terreur
    - Endoctrinement

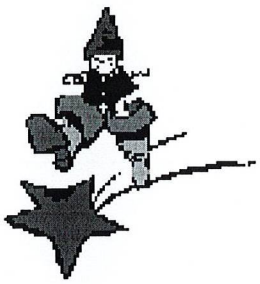
La dernière catégorie n'est là, selon l'orateur, que pour l'exhaustivité de l'inventaire, et toute armée intervenant au nom d'un pays démocratique se refuse à user de ces moyens.
Toutes ces techniques peuvent à leur tour être développées et étudiées plus finement. Ainsi la persuasion peut être "induite" (et ne pas attenter au sentiment de liberté d'autrui), "consentie" (avec un effet d'engagement - qui permet d'amener quelqu'un à en faire plus que ce qu'on lui demandait au départ – ou un effet de gel – plus vous passez de temps à attendre un but, et plus vous vous dites qu'il va finir par arriver -), ou bien encore "provoquée" (et amener quelqu'un à avoir un comportement contraire à ses habitudes).
L'obédience elle, suit un schéma pyramidal qui va de la campagne d'information au contrôle en passant par la participation et la réglementation.
Pour ce qui est des techniques de tromperie (ou mystification), qui constituent les fondements de "l'art de la guerre", elles se définissent en fonction du référentiel qu'elles corrompent : si elles ont pour cible une vérité matérielle, on parlera de "déception", s'il s'agit de vérité formelle on parlera "d'intoxication", enfin dans le cas d'une vérité informationnelle, il s'agira de "désinformation".

La plupart des questions dans la salle font référence à des préoccupations d'ordre éthique. Le général précise alors que l'éthique se situe entre la morale et l' action, et qu'à ce titre elle procède des deux. Il énonce alors un certain nombre de principes "éthiques" qui régissent l'utilisation de ces techniques d'action psychologiques et "qui sont respectées quand on intervient au nom d'un pays démocratique" (sic). On y trouve :

- principe de liberté de pensée et d'expression
- principe de véracité des données et des informations communiquées
- principe de cohérence
- principe de crédibilité
- principe de l'implication mesurée (qui consiste à garder la tête froide)
- principe de l'adaptabilité.

Bibliographie : "La zizanie" Astérix & Obélix (bonne approche de la guerre psychologique ;-))

Nos commentaires : Et maintenant, imaginez ...



Imaginez que les logiciels ARISTOTE (aide à la conception de messages persuasifs en fonction de l'auditoire) et GORGIAS (aide au renversement des croyances adverses) soient opérationnels et fonctionnent en modules couplés. Imaginez ensuite que l'on relie ces logiciels à des serveurs de news ... Imaginez qu'un ancien directeur du G.I.G.N. dirige une société qui proposerait non pas de vous vendre du service telle une start-up lambda, mais de vous aider à faire du contre-lobbying ... Imaginez que parmi les services offerts par cette société, on propose à l'aide de vrai-faux-stagiaires infiltrés dans des administrations étatiques de récupérer des rapports à haut taux de confidentialité. Imaginez que cette société ait la possibilité de créer de fausses sociétés (de type société-écran) juste le temps de collecter des documents primaires (des fax par exemple) qui serviront de preuves dans votre stratagème de contre-lobbying, imaginez que cette société pousse la malice – ou le vice – jusqu'à commander à une autre société, elle aussi spécialisée dans le renseignement, une étude visant à démontrer l'existence de la société-écran qu'elle a créée de toutes pièces, pour mieux confondre la cible visée ... Imaginez que cela réussisse.
Imaginez ? Non, téléphonez simplement à la société "Atlantic Intelligence"et elle vous expliquera comment elle a suivi à la lettre ce plan d'action suite à la demande des producteurs de foie gras de l'ouest de la France afin qu'ils puissent être labellisés "produit du Sud-Ouest".
Bien sûr, Aristote et Gorgias ne sont pas encore au point nous a-t-on dit ... Pas en France en tout cas ...
Allez savoir ... Mais il régnait une ambiance étrange dans les couloirs du Palais des Congrès au sortir de cette dernière intervention. En écoutant les réactions de l'auditoire, certains avaient cru lire dans le regard bleu acier du conférencier et dans l'assurance de sa voix posée, comme une part de mystère, de non-dit ... Certains ont même évoqué l'image de Skinner, vous savez, le chef du FBI, mais si ... dans X-Files ...

Bienvenue dans le mode de l'Intelligence économique offensive ...

Les liens du Salon.

Parmi les différents stands, nous avons repéré :

- Société Sinequa - "La net linguistic company" www.sinequa.com et leur logiciel www.intuitif.com (moteur de recherche sémantique)
- Société Lexiquest (logiciels d'ingénierie linguistique) www.lexiquest.com
- L'Association des professionnels de l'information et de la documentation www.adbs.fr
- Association Nationale de la Recherche Technique www.anrt.asso.fr
- L'Institut National de la Propriété Industrielle www.inpi.fr
- Société Intelligence Process (bibliothèque intelligente) www.intelligence-process.com
- Veille Magazine, le magazine professionnel de l'intelligence économique www.veillemag.com
- Société Archimed (solutions de Gestion Electronique de Documents)

www.archimed.fr

- The British Library (Inside : veille documentaire et recherche bibliographique sur 20 000 périodiques) www.bl.uk/inside
- Société Arisem (agents de collecte, croisement dynamique d'arbres logiques sur fond documentaire ...) www.arisem.com
- Société Archimag (magazine des nouvelles technologies) www.archimag.presse.fr
- Ministère de la Recherche www.recherche.gouv.fr
- et pour le reste des exposants, le site du salon www.idt.fr

Formation de formateurs des 13 avril et 9 mai 2000

Elisabeth Lacombe, Conservateur en Chef - URFIST de Toulouse

Cette formation vise à aider les formateurs à mettre en place des formations à la recherche documentaire pour les étudiants des campus, par cycle et discipline. Elle a été animée par E. Lacombe, J. Link-Pezet (URFIST) et O. Nunge (Croissance Formation).
Lors de la première journée de formation, ont été traités les aspects politiques et organisationnels d'une formation ainsi que l'élaboration d'une enveloppe culturelle minimale à l'intention des étudiants de 1er cycle.
La méthode de travail utilisée a consisté en travaux de groupes avec mise en commun, discussions et résolution des problèmes.

Organisation de la formation

Problèmes politiques, statutaires et matériels.

Les formateurs jouent un rôle important lors de la mise en place des formations. Des problèmes existent dans les rapports avec l'institution ainsi que sur les points suivants: la visibilité de la formation, l'intégration dans un cursus, l'implication des enseignants disciplinaires, les négociations entre ministère et institution locale, la méconnaissance des enjeux de la maîtrise de l'information de la part des enseignants disciplinaires.
Le statut du formateur doit être clarifié : quelle est sa légitimité de formateur, sa disponibilité, quel rôle lui est attribué par l'institution.Il faut aussi considérer ses motivations personnelles; décider si son rôle est de former aussi ses collègues.
L'organisation matérielle de la formation peut se décliner ainsi : nécessité d'une salle de formation et adéquations des moyens et des actions.

Définition des objectifs.

Pour les formateurs, il est important que l'institution soit claire dans ses objectifs. Il s'agit d'aider les étudiants à lutter contre l'échec, à s'intégrer à l'université, à décoder la culture universitaire. Pour cela, il faut viser à leur donner une méthode de travail, leur apprendre l'autonomie, la maîtrise des outils documentaires, à analyser

les types d'information et à évaluer la qualité de l'information.

Enveloppe culturelle minimale

(à destination des étudiants de 1er cycle)

Contenu d'un cours de recherche documentaire de 1er niveau

Les éléments de base qui semblent devoir figurer dans un cours de 1er cycle sont :

- l'apprentissage des outils et des méthodes : identification des lieux d'information/documentation et règles d'usage
- les types de documents et types de services dans les bibliothèques
- la manipulation des outils (catalogues, ouvrages, périodiques, Internet...)
- les différents supports d'information
- les références (auteurs, matière, titre, cote...)
- les systèmes de classement.

Il faut également aborder le vocabulaire de la documentation (traitement linguistique de l'information) et l'organisation de l'information en champs dans une référence.

Commentaires :
Les besoins de l'étudiant, pourtant cités dans les objectifs, ont disparu de l'élaboration du cours. L'objectif "valeur" n'est pas retraduit en "objectifs étapes".
Il semble important de travailler avant l'élaboration du cours sur la prise en compte des besoins de l'étudiant.

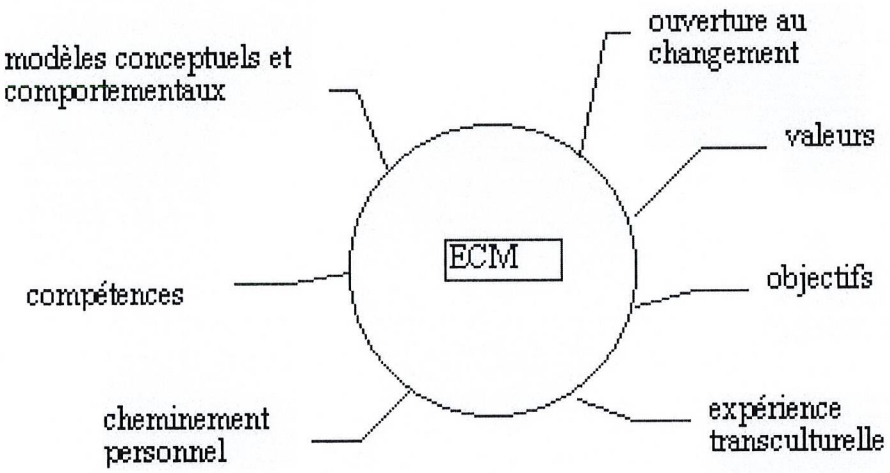
Comment prendre en compte les besoins de l'étudiant pour adapter son cours?

Un exercice est proposé aux formateurs : lister ces besoins à partir de la Pyramide de Maslow (en travail de groupe). La pyramide classe les besoins de bas en haut en besoins de sécurité, d'appartenance, de considération, de réalisation et d'accomplissement de soi.
Le besoin de sécurité :
La méthodologie documentaire est un cours nouveau, un avantage dont les étudiants vont bénéficier, une méthode. Ce n'est pas une matière supplémentaire.
Elle se veut une aide contre l'échec : il serait bien de donner un exemple de réussite grâce à l'acquisition de cette méthode. Il faut rassurer les étudiants sur les conditions matérielles (prise en compte des contraintes de temps) et leur fournir un support de cours.
Le besoin d'appartenance :
Il s'agit de bien montrer le lien entre la discipline de base et la recherche documentaire : construire son cours sur des cas pris dans le contexte des étudiants (par exemple en expliquant l'OPAC, on cherchera les noms de leurs enseignants, de juristes connus ...) Il faut qu'ils soient concernés par le discours. Si les niveaux sont hétérogènes, il faut profiter de la richesse du groupe (échanges, participation).
L'appartenance à l'Université, l'UFR doit être valorisée.
Le besoin de considération :
Il s'agit de stimuler les compétences et les capacités, de reconnaître les savoirs, les acquis ou l'identité des étudiants.
Le besoin de réalisation, d'accomplissement de soi :
Il faudrait les amener à imaginer leur devenir quel qu'il soit (ingénieur, juriste, etc...), leur montrer l'utilité de la formation pour leur culture générale.

Quelques apports théoriques sur l'enveloppe culturelle minimale (O. Nunge)

Tous les outils s'insèrent dans des systèmes qui constituent pour l'être humain l'enveloppe culturelle minimale. Elle peut être schématisée dans un cercle (cycle) où interviennent les éléments suivants :

- l'ouverture au changement : le formateur aura intérêt à consacrer son énergie aux étudiants motivés qui ne constituent pas la totalité des étudiants de 1ere année, dont une partie se comporte en "touristes"
- les valeurs: le formateur propose certaines valeurs aux apprenants
- les objectifs : il est important de se centrer sur les objectifs but
- l'expérience transculturelle : le public étudiant a déjà fait ce genre d'expérience à son arrivée à l'université
- le cheminement personnel : la formation est un lieu de parole, un espace pour s'exprimer
- l'acquisition des compétences, objectif étape de la formation
- les modèles conceptuels et comportementaux : l'étudiant devra absorber et intégrer des modèles, qui lui permettront de s'ouvrir au changement.



Grille d'ingénierie d'un cours

(élaborée par l'Urfist)

Après cette analyse détaillée de la partie organisationnelle d'une formation, de ses objectifs et des besoins des étudiants, nous avons travaillé ensemble à remplir systématiquement les autres parties de la grille d'ingénierie de formation correspondant au savoir sur lequel repose l'enseignement et à l'apprentissage des sources et des outils.

Questionnement de l'étudiant

L'étudiant est pragmatique et se pose des questions du genre : à quoi ça me sert, quel temps, quelle énergie vais-je mobiliser, est-ce que ça en vaut la peine? Il essaie d'établir des liens: quel rapport a cette formation avec la bibliographie donnée par les enseignants de la discipline?
Il sait que cette formation fait partie intégrante de la méthodologie du travail

universitaire de la première année. Le formateur pourra exploiter la situation pédagogique : travailler à partir des bibliographies de l'enseignant disciplinaire, parler du coût des livres, de la manière de se les procurer, de la mise à jour des listes bibliographiques etc...
Le formateur amènera ensuite l'étudiant vers un questionnement plus élaboré, plus précis, toujours en partant de son vécu et de son contexte.
D'après les formateurs, les étudiants sont généralement peu bavards. C'est parce qu'ils sont en situation de dépendance; le formateur devra les amener par étapes vers l'autonomie (la formation à l'autonomie a fait l'objet d'un précédent stage); ils n'ont pas obligatoirement à ce stade un besoin d'information.

Compétences à acquérir

- Savoir reconnaître et identifier les différents types de documents et les différents types de supports.
- Savoir les localiser.
- Savoir les utiliser.
- Savoir les intégrer dans sa discipline et sa pratique.

Raisonnement, mode d'apprentissage, attention

Cadrage et synthèse sont indispensables de la part du formateur, afin d'aplanir les difficultés : difficultés à gérer l'imprévu, en raison des outils imparfaits, défaillants.
La répétition est nécessaire (reformulation). Le pédagogue doit admettre que l'apprentissage se fait aussi par l'erreur.
Le formateur doit motiver l'étudiant, capter son attention en cultivant son goût du plaisir, en rendant visible l'objectif but (exemple: le mémoire, la synthèse).
L'étudiant recherche la facilité : il faudrait travailler sur les représentations (afin d'en changer).
Un rappel est fait des différents types cognitifs: visuels, auditifs, kinesthésiques, tactiles, montrant l'impact que cela peut avoir sur la pédagogie : alternance de tableaux, supports, de mise en situation, d'exemples et d'histoires.
Le seuil d'attention doit être respecté : alternance d'attitudes, dosage, pauses (besoins physiologiques).

Choix pédagogiques du formateur

Certains aimeraient prendre comme option de travail "l' autonomie contrôlée" : il s'agit de travaux dirigés durant lesquels le formateur peut prendre la main sur les postes des étudiants.
Les autres choix envisagés sont :

- l'apprentissage par l'exemple, la mise en situation (avec une recherche spécifique)
- le suivi individuel hors formation (état d'interdépendance)
- la reformulation, même s'il y a redite.

Si le cours magistral est nécessaire, il y a un travail à faire sur une forme plus attractive (audiovisuel, scénarios). Le cours magistral pose des problèmes aux formateurs. Il est rappelé qu'il faut bien définir les objectifs, le déroulement, le contenu (les invariants de la recherche documentaire).

Obstacles

Outre les obstacles financiers et structurels déjà discutés, ils sont d'ordre technique et pédagogique :

- la motivation des étudiants et la motivation des enseignants : une solution peut être l'accompagnement, la co-animation
- la validation de la formation
- les divergences d'objectifs entre formateurs et étudiants (ex le formateur veut rendre l'étudiant autonome dans sa recherche, l'étudiant veut simplement obtenir des informations pour son mémoire) : il faut là encore être clair sur les objectifs
- la gestion des niveaux hétérogènes, des différences culturelles et ...des provocateurs (pathologiques ou leaders).

Présentation d'un schéma de cours (1er cycle)

- Introduction :
- Les objectifs de la formation ; à quoi cela va vous servir
- Contenu :
  1. Rappel des pré-requis (expérience du CDI) :
     - Les types de documents : livres, périodiques
     - Les types de supports : cédéroms, Internet
  2. La production éditoriale et scientifique
  3. La référence bibliographique :
     - Définition, structure (champs)
     - Traitement documentaire : du document primaire au document secondaire
     - (Penser à faire un glossaire du vocabulaire utilisé)
  4. Les lieux d'information et les catalogues :
     - Mode d'emploi.
     - Définitions.
  5. Utilisation de l'information et validation.

Etape opérationnelle : l'apprentissage des sources et des outils

Pour préparer cette étape, les formateurs se sont engagés à élaborer un tableau pour la prochaine séance sur un outil de base : l'OPAC, sur le modèle Docthèses ci-dessous :

Caractéristiques de l'outil	Tâches de l'étudiant	Sous-tâches	Niveau cognitif
Interface Netman Interface Docthèses	Chercher une thèse	Choisir dans le menu : Cd-Rom multidisciplinaires puis Thèses	Choisir une source d'informations

Positionnement automatique sur le champ " Recherche " et le champ " Auteur	Rechercher les informations	Choisir le champ d'interrogation : auteur, mots du titre, directeur, discipline, etc…	Positionnement automatique sur le champ " Recherche " et le champ " Auteur " Rechercher les informations Choisir le champ d'interrogation : auteur, mots du titre, directeur, discipline, etc… Définir les critères d'interrogation : auteur, terme, lieu de soutenance …Connaître la structuration en champs
Localisation Bibliothèque	Rechercher les informations	Localiser la thèse pour se la procurer	Différencier le lieu de soutenance et la localisation de la thèse
Etc…			

On liste les connaissances que l'on a de l'outil, les connaissances que l'on voudra enseigner, et ensuite la manière de les enseigner, en spécifiant le niveau cognitif de connaissances visé chez l'apprenant. Le fait de lister les caractéristiques de l'outil permet au formateur de vérifier ses connaissances; la liste des tâches à accomplir par l'étudiant correspond aux connaissances à enseigner; le critère du niveau cognitif de l'étudiant constitue une base intéressante pour la pédagogie. Chaque outil utilisé pour la formation pourra ainsi être soumis à l'expertise du formateur et faire l'objet d'une réflexion pédagogique.

Par rapport à la grille d'ingénierie, il reste aussi à travailler avec les formateurs la partie "exploitation de l'information par l'étudiant dans sa discipline". Puis il faudra mettre en pratique pour chacun le cours qu'il aura construit personnellement, c'est-à-dire présenter sa formation devant le groupe de formateurs, exercice pédagogique indispensable et ...formateur.