La Lettre de l'URFIST de Toulouse, n°24-25

Jo Link-Pezet, Élisabeth Lacombe-Carraud, Anita Largouet, Claire Panijel-Bonvalot, Jean-Max Noyer, Anne Sanouillet, Viviane Couzinet, Jean-Émile Tosello-Bancal, Christophe Boudry, Ghislaine Chartron, et al.

▶ **To cite this version:**

Jo Link-Pezet, Élisabeth Lacombe-Carraud, Anita Largouet, Claire Panijel-Bonvalot, Jean-Max Noyer, et al.. La Lettre de l'URFIST de Toulouse, n°24-25. 1999. sic_03250452

HAL Id: sic_03250452

https://archivesic.ccsd.cnrs.fr/sic_03250452

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

Voici le joli mois de Mai et son cortège de nouveautés. Voilà l'envol de la lettre de l'Urfist sur le Web.
Voici que le réseau des Urfist participe à la rédaction de ce numéro élargi , étendu, par sa conception, sa rédaction, son volume, l'audience qu'il pourra rencontrer sur Internet, mondialisation oblige!!

Nous avons voulu en associant toutes les Urfist à cette entreprise, montrer qu'en intégrant les analyses de chacun on mettait en évidence l'originalité de chaque approche, de chaque individualité, de chaque contexte régional. Nous avons voulu montrer la diversité et la richesse de ces réflexions qui ont toutes un point commun: l'impact des nouvelles technologies de l'information sur la formation et la vie universitaire, professionnelle, politique, sur les organisations et sur les représentations sociales. Nous espérons donner un juste reflet de la variété des sujets d'intérêts, des compétences de nos collègues par la simple juxtaposition de leurs contributions, qu'ils en soient tous remerciés!

Ce numéro s'est construit facilement, presque naturellement autour des thèmes que nous développons habituellement.
A l'issue de cette coopération, la lettre n° 24-25, numéro spécial et double, s'articule autour des thèmes suivants :

- La formation de formateurs à l'information documentaire
- La gestion collectives des connaissances et la formation
- Impact d'Internet: évolution technologique et conséquences sur le travail intellectuel.
- L'information sur Internet: les guides et les moteurs
- Recherche en science de l'information et de la communication

- Synthèse sur les sources documentaires pour l'information
- Compte rendus de congrès et de journées d'études
- Echos des stages.

La formation d'enseignants à l'IST à l'URFIST de Bordeaux

Anita LARGOUET, Conservateur, responsable de l'URFIST de Bordeaux

CERISE - Conseils aux Etudiants pour une Recherche d'Information Spécialisée Efficace

Claire Panijel-Bonvalot, Conservateur Général - URFIST de Paris-ENC

Formation à la recherche documentaire : adresses

Quelques remarques épistémologiques sur l'évolution du "Procés de Travail"

J.Max Noyer, Maître de conférences à l'URFIST de Rennes

La surcharge informationnelle

Anne Sanouillet, Professeur en Sciences de l'Information et de la Communication - URFIST de Nice

Forsic

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

Médiations hybrides

Viviane Couzinet - Maître de conférences LERASS-ICC Toulouse

Jurisguide

Jean-Emile Tosello-Bancal, Conservateur en Chef - URFIST de Paris-ENC

Bioguide

C. Boudry, Maître de conférences et G. Charton, Maître de conférences URFIST de Paris-ENC

Métadonnées

Julia Jumeau, Professeur de Documentation - URFIST de Paris-ENC

Evaluation de la recherche d'informations

Lise Herzhaft, Conservateur en Chef - URFIST de Lyon

Méthodes de tri des résultats des moteurs de recherche

Jean-Pierre Lardy, Maître de conférences - URFIST de Lyon

Quelques adresses

François Magnan, Documentaliste

Extraits de Online Information 99

Elisabeth Lacombe, Conservateur en Chef - URFIST de Toulouse

ASIS 1999 Annual Conference - Knowledge : Creation, Organization and Use

Olivier Ertzscheid - Etudiant en thèse Toulouse II

Publication électronique des résultats de la recherche

Annaïg Mahé, Doctorante GRESI/Enssib

Publication électronique des résultats de la recherche - Synthèse

Ghislaine Chartron, Maître de conférences - URFIST de Paris-ENC

A propos des moteurs de recherche

François Magnan, Documentaliste

La formation d'enseignants à l'IST à l'URFIST de Bordeaux :

Compte rendu d'expérience et bilan.

Anita LARGOUET - Conservateur, responsable de l'URFIST de Bordeaux, avec la collaboration d'Henri BLANC (professeur agrégé, ingénieur arts et métiers ENSAM) et de Geneviève HERMET (documentaliste ENSAM, professeur certifié)

Dans le cadre de leurs missions, les URFIST sont chargées de proposer des formations aux étudiants avancés. Une mission de formation de formateurs leur est également impartie. En fonction du public à desservir, ces actions peuvent être proposées à deux types de personnels : les personnels de bibliothèque et de documentation chargés de former les étudiants d'une part, les personnels enseignants d'autre part.

Compte tenu de ses moyens humains, l'URFIST de Bordeaux souhaite développer les formations de formateurs afin de démultiplier les formations en direction des étudiants. Le public ciblé est essentiellement celui des enseignants car il nous paraît indispensable que les formations documentaires trouvent leur ancrage dans les enseignements disciplinaires.

A la rentrée universitaire 1998/1999, l'URFIST fut sollicitée par Mr Henri BLANC, enseignant à l'ENSAM de Bordeaux pour la mise en place et la participation à des formations à l'information scientifique et technique. L'ENSAM souhaitait intégrer dans le cursus de ses étudiants un module de "Formation à l'information scientifique et technique", intégré au sein de l'UV de communication. La partie relative à la veille technologique et la propriété industrielle n'était pas concernée car elle était prise en charge par l'Agence Régionale d'Information Stratégique et Technologique (ARIST) Aquitaine.

L'URFIST a proposé son aide sous trois formes :

1. La consolidation du programme de formation à dispenser aux étudiants à partir de celui proposé par la Direction Générale de l'ENSAM,
2. Des interventions directes sous forme d'enseignement et d'encadrement de TP (2h de cours théoriques, 4 heures de TP sur Internet et les banques de données),
3. La mise en place d'une action de formation des enseignants de l'ENSAM. Une formation initiale avait été assurée par l'ARIST Aquitaine un an auparavant.

Sur le plan chronologique, les deux derniers points ont été réalisés en parallèle, l'objectif étant que la majorité de la formation soit prise en charge par l'ENSAM à la rentrée universitaire suivante. Il est également important de souligner la participation active de la documentaliste de

l'ENSAM, Geneviève HERMET, qui fut étroitement associée aux différentes actions et réunions organisées.

Une démarche très pragmatique a été retenue pour répondre au mieux aux besoins de l'ENSAM. Après discussion et bilan des ressources, l'URFIST a proposé de prendre en charge pour l'année universitaire l'enseignement théorique. Pour les séances de TP, l'encadrement fut assuré en binôme : l'URFIST et un enseignant ou la documentaliste de l'ENSAM et un enseignant. Dans chaque cas, c'est l'enseignant qui proposait, en fonction de sa discipline, les sujets de recherche qui étaient traités durant la séance.

La participation des enseignants aux séances de TP, à la fois active et protégée, permettait de préparer leurs interventions futures tout en favorisant l'ancrage disciplinaire de la formation à l'IST.

En parallèle, pendant l'année universitaire 1998/99, une action de formation a été proposée aux enseignants :

L'objectif de la formation a été défini en tenant compte des travaux qui seraient demandés aux élèves de l'ENSAM, à savoir produire un rapport décrivant la démarche et la méthodologie de recherche d'informations à partir d'un sujet proposé par l'enseignant, présenter les résultats de la recherche de façon normalisée et analyser quelques articles.

Néanmoins, les contenus informationnels furent plus difficiles à définir. En effet, chaque enseignant dispose d'un corpus de connaissances variable dans le domaine de l'IST . S'il semble important de resituer les notions essentielles et le vocabulaire, les besoins peuvent être différents pour chacun. D'autant plus que l'investissement en temps que les enseignants peuvent consacrer à ce type de formation est variable.

La première réunion associant des enseignants et la documentaliste fut donc consacrée à l'élaboration du programme. Sept séances de formation d'une demi journée ont été organisées dans le courant de l'année, soit 21 heures de formation. Le programme fut le suivant :

- Présentation du panorama informationnel (objectif : fournir les notions fondamentales et le vocabulaire associé)
- Méthodologie de la recherche documentaire (objectif : pouvoir transmettre la méthode aux élèves de l'ENSAM) et normalisation documentaire
- Internet (objectif : savoir utiliser les outils de recherche d'information)
- Les banques de données (objectif : maîtriser l'utilisation des banques de données : Méca-CD, INSPEC et l'interrogation du serveur DIALOG)

La formation a touché une dizaine d'enseignants ainsi que la documentaliste de l'ENSAM. Il est à noter que ses besoins différaient légèrement de ceux des enseignants et portaient davantage sur les aspects pédagogiques et la transmission de connaissances. Il n'a pas été possible de lui proposer une formation spécifique. Nous avons donc tenté d'y répondre par du conseil. De plus, les enseignants de l'ENSAM ont été invités à suivre les enseignements théoriques proposés aux élèves.

Lors de cette rentrée universitaire, 1999/2000, la formation a été définitivement mise en place. * Le premier semestre, les étudiants nouvellement entrés à l'ENSAM suivent :

- un cours théorique de 2 heures en amphi sur le panorama informationnel (assuré par l'URFIST)
- un cours de méthodologie documentaire d'1H30, assuré par un enseignant
- une présentation de la bibliothèque de l'ENSAM suivie d'un TP (1H), assuré par la documentaliste.

Les étudiants doivent ensuite réaliser, en autonomie, un travail de recherche à partir de sujets fournis par les enseignants ; l'objectif étant de les amener à exploiter au moins les ressources documentaires de proximité et de mettre en œuvre rapidement les informations transmises dans cette première phase de la formation. Notons que cette activité engage une démarche volontariste qui devra être recommencée dans d'autres circonstances notamment durant les Bureaux d'Etudes et les projets qui se placeront, pour certains, avant l'approfondissement du deuxième ou du troisième trimestre. Sans aucun doute, la nécessité d'approfondir sera aussi mise en évidence et elle préparera la seconde phase de cette formation. * Lors du deuxième ou du troisième semestre, les étudiants suivent :

- un cours théorique de 4 heures présentant les principes de recherche d'information sur INTERNET et sur les banques de données (assuré par l'URFIST avec la collaboration de la documentaliste). Le cours est démonstratif. Réalisé sous powerpoint, il s'accompagne de connexions à INTERNET et aux banques de données qui seront ensuite utilisées en séance de TP.
- des séances de TP (4H) encadrées par un enseignant de l'équipe IST ou la documentaliste. L'accent est mis sur la structuration de la méthode de recherche qui s'effectue sur l'Internet et les banques de données (Méca CD, INSPEC et, Pascal)

Quel bilan retirer de cette expérience ?

- Du point de vue de l'URFIST :
  - En ce qui concerne la formation de formateurs Former des enseignants n'est pas chose facile. On se heurte tout d'abord à un problème de disponibilité des enseignants pour se former dans ce domaine spécifique qu'est l'IST et la maîtrise de l'information. L'intérêt et la motivation sont essentiels pour que cet obstacle soit surmonté. Cela suppose également d'être convaincu de la nécessité d'intégrer à la pédagogie les technologies de l'information et de la communication. L'énergie et la volonté d'un enseignant : Henri BLANC, soutenu par la direction de l'ENSAM, ont permis de surmonter cet obstacle. Compte tenu du temps qui peut être consacré, il faut ensuite adapter les contenus de la formation aux besoins informationnels des enseignants. Cela suppose en amont une évaluation des besoins. Le nombre d'heures dispensées dans le cadre de la formation de formateurs réalisée nous semble être un minimum.
  - En ce qui concerne le résultat final, : - L'expérience nous paraît tout à fait positive dans la mesure où les étudiants bénéficient désormais d'un enseignement à l'IST intégré dans le cursus. Il faut préciser que les élèves de l'ENSAM possèdent la maîtrise de l'outil informatique et de la logique booléenne. Ils acquièrent donc très facilement les réflexes utiles à la recherche d'information sur support informatisé. La validation finale du module d'IST complète parfaitement le dispositif et évite que cet enseignement ne soit considéré comme annexe ou en marge des enseignements disciplinaires. - Le centre de documentation et ses ressources sont valorisés et exploités. La documentation trouve sa place dans la pédagogie de façon tout à fait naturelle. Plus largement, les étudiants sont amenés à exploiter les ressources documentaires de l'université et des coopérations entre la bibliothèque de l'ENSAM et le SCD de l'Université Bordeaux1 sont aujourd'hui évoquées.
- Du point de vue des enseignants

  Durant leur scolarité à l'ENSAM, les étudiants réalisent 3 projets - Projet d'Initiation Technologique, Projet de Dominante Métier et Projet de Fin d'Etude - demandant un investissement personnel et durant lesquels leur aptitude à l'innovation est sollicitée. Dans ce contexte, la recherche d'informations est une activité importante et récurrente qui justifie une formation spécifique. Cette analyse est assez largement partagée par les enseignants de l'ENSAM qui pour certains d'entre eux sont aussi membres d'un des deux laboratoires de recherche rattachés au centre de Bordeaux.
  Malgré tout, l'introduction d'une nouvelle discipline demande des

moyens et une organisation qui permettent d'assurer les fonctions suivantes :

- définir les objectifs à atteindre,
- former aux nouveaux concepts,
- montrer l'intérêt réel pour chaque acteur (étudiants et enseignants).

Dans cette démarche, l'URFIST apporta une aide déterminante, d'abord grâce aux discussions préliminaires sur l'objectif de la formation, ensuite avec les séances de formation qu'elle a assurées, enfin avec les échanges qu'elle a suscités sur la manière la plus efficace d'organiser la formation des étudiants en respectant le cahier des charges défini par la Direction Générale de l'ENSAM. De plus, la possibilité d'utiliser les moyens de connexions disponibles à l'URFIST a permis de montrer rapidement l'intérêt et l'efficacité de recherches dans la mesure où elles sont menées de manière structurée et en accord avec l'organisation des bases d'information consultées. L'accompagnement durant les premières séances de TP a contribué favorablement à la décision pour l'engagement final des enseignants. La dizaine d'enseignants actuellement impliqués dans cette formation montrent que pratiquement toutes les disciplines enseignées à l'ENSAM sont concernées ; c'est la preuve de l'importance de la Formation à l'Information Scientifique et Technique.

- La documentaliste de l'ENSAM nous fait part de son bilan :

  " L 'expérience est en effet positive. En première année, les élèves, avec leur formation scientifique, n'ont aucun problème pour comprendre les notions de classification décimale, d'indice et de cote . Ils comprennent vite l'interrogation du catalogue informatisé et s'orientent donc assez rapidement sur les rayonnages . Les étudiants de première année consultent assez volontiers le catalogue informatisé avant d'aller dans les rayons, depuis que cette présentation de la bibliothèque et de ces outils a eu lieu.

  Avec les TP "internet" et "banques de données", les élèves comprennent vite l'avantage et la puissance des banques de données. J'ai le sentiment que cet enseignement dans toutes ces composantes (cours, TP, dossier final) leur convient, et les intéresse . Pour ma part , je suis aussi assez satisfaite de réussir à mener des TP de 4 h sans de trop grandes difficultés malgré mon manque d'expérience pédagogique. Je me sens par contre très à l'aise pendant la séance de présentation de la bibliothèque, du catalogue informatisé et les TP qui suivent, j'utilise pour cela un vidéo projecteur, ce qui rend la présentation très vivante "

Il nous manque le point de vue des élèves. Il peut être apprécié au travers des rapports produits en 1999. Ils semblent avoir acquis pour la plupart les éléments méthodologiques suffisants pour continuer soit pendant leur scolarité, soit pendant leur vie professionnelle à utiliser au mieux les supports d'information.

CERISE

Conseils aux Etudiants pour une Recherche d'Information Spécialisée Efficace

Claire Panijel-Bonvalot, Conservateur Général - URFIST de Paris-ENC

CERISE est un guide en ligne d'initiation méthodologique à la recherche d'information pour un travail universitaire (encore en cours de réalisation), destiné plus particulièrement aux étudiants de premier cycle universitaire en lettres et sciences humaines, mais néanmoins utile aux étudiants de filières scientifiques et à toute personne en situation de chercher des informations pour des raisons professionnelles ou personnelles. Il est disponible et régulièrement mis à jour sur le site web de l'Urfist de Paris à l'adresse: http://www.ccr.jussieu.fr/urfist/cerise/

Les auteurs, Martine Duhamel, Maître de Conférence à l'Université de Paris 4, et Claire Panijel, Conservateur Général des Bibliothèques à l'Urfist de Paris-Ecole Nationale des Chartes, ont travaillé en collaboration avec des enseignants et des professionnels de bibliothèques universitaires.

Objectifs du projet

Offrir sur un site web un guide convivial à disposition des enseignants assurant des formations méthodologiques dans le premier cycle universitaire, particulièrement dans le cadre des unités de "méthodologie du travail universitaire" prévues par la réforme d'avril 1997.
Ce guide peut intéresser aussi bien des enseignants du champ disciplinaire dans lequel s'incrit le cursus que des bibliothécaires ou d'autres spécialistes de l'information, chacun adaptant ce document à son propos.
Il se présente comme un ensemble de conseils de méthode et d'indications pratiques qui peuvent être intégrés à un dispositif de

formation, mais il est aussi utilisable comme outil d'auto-formation sur lequel peuvent pointer les sites web des bibliothèques et que l'on peut librement explorer de n'importe quel point d'accès.

Contenu et modes d'utilisation

L'objectif de formation visé est celui d'une **initiation à la maîtrise de l'information**, dans une définition conforme aux recommandations de la brochure "Former les étudiants à la maîtrise de l'information"(1) éditée par le Ministère de l'Education Nationale, de la Recherche et de la Technologie.
La page d'accueil part de thèmes simples, proches des questions que se posent les étudiants débutants, et orientes pour apprendre

- à trouver des repères conceptuels,
- à chercher et sélectionner l'information utile,
- à traiter et exploiter l'information recueillie.

Des exemples, pris dans différentes disciplines, complèteront peu à peu les indications de méthode, à la fois pour "donner à voir" de manière plus intuitive les démarches possibles, mais aussi pour illustrer à l'occasion les difficultés rencontrées, tirer profit des erreurs, évaluer méthodes et résultats.

L'organisation des différentes rubriques essaie de répondre à une logique d'apprentissage, tout en en préservant la cohérence des connaissances nécessaires. Il y a donc un intérêt à suivre l'ordre suggéré, mais, en fonction des connaissances déjà acquises ou de besoins ponctuels, les pages peuvent être exploitées de façon non linéaire sans effet de labyrinthe : en bas de chaque page, le sommaire du chapitre dont elle fait partie est détaillé avec des liens hypertextes, de manière à situer l'information dans son domaine et éventuellement s'y reporter.

CERISE peut encore évoluer et s'étoffer en fonction des remarques qui seront faites. Il est mis à disposition de tous pour pouvoir être testé et complèté par ses utilisateurs; leurs contributions, particulièrement pour les exemples de "parcours" seront bienvenues.
Contact : Claire PANIJEL clpa@ccr.jussieu.fr

---

(1)"**Former les étudiants à la maîtrise de l'information, repères pour l'élaboration d'un programme**"
Guide réalisé par un groupe de travail réuni à l'initiative du Ministère de l'Education Nationale. Coordination et rédaction finale : Bruno Deshoullières,
Maître de conférences à l'Université de Poitiers, Françoise Dubreuil,

Ingénieur d'études à l'Université de Paris 7, Claire Panijel, Conservateur Général de bibliothèque à L'URFIST Paris - Ecole Nationale des Chartes. Paris, 1999. (Diffusion Sous-Direction des Bibliothèques et de la documentation). Une version html disponible sur le site de Formist : http://formist.enssib.fr/reperes/

Edudoc

Formation à la recherche documentaire

L'adresse WEB du Groupe Formation des Utilisateurs :

http://recoda.fsagx.ac.be/gfu/

http://www.bib.fsagx.ac.be/bib/info/met.html (informations sur la nouvelle version du cours organisé à Gembloux. Soit un cours de 10 h Dans la rubrique "support", on trouve un lien vers le manuel utilisé pour le cours (au format pdf) et un lien vers les "slides".

Quelques remarques épistémologiques sur l'évolution du "Procés de Travail"

Réseaux, Associationnisme, Auto-organisation dans le contexte des nouvelles mémoires numériques

J.Max Noyer, Maître de conférences à l'URFIST de Rennes

(La totalité de l'article dont nous publions ici des extraits (avec la permission de l'auteur) a été proposée pour publication à la revue du Lerass)

"Le capitalisme représenterait donc une forme paroxystique d'intégration de divers types de machinismes: machines techniques, machines d'écriture économique, mais aussi machines conceptuelles, machines religieuses, machines esthétiques, machines perceptives, machines désirantes... Son mode de sémiotisation-la méthode du Capital-constituerait tout à la fois une sorte d'ordinateur collectif' du socius et de la production, et une "tête chercheuse" des innovations

adaptées à ses pulsions internes.
Dans ces conditions, sa matière première, sa nourriture de base ne serait pas directement le travail humain ou le travail machinique, mais l'ensemble des moyens de pilotage sémiotique relatifs à l'instrumentation, à l'insertion dans le socius, à la reproduction, à la circulation des multiples composantes concernées par ce processus d'intégration machinique.
Ce que capitalise le capital c'est du pouvoir sémiotique. Seulement pas n'importe quel pouvoir-car, à ce compte, il n'y aurait pas lieu de le démarquer des formes antérieures d'exploitation-mais un pouvoir sémiotiquement déterritorialisé.
Le capitalisme confère à certains sous-ensembles sociaux une capacité de contrôle sélectif du socius et de la production par le biais d'un système de sémiotisation collective. Ce qui le spécifie historiquement, c'est qu'il ne s'efforce de contrôler que les diverses composantes qui concourent au maintien de son caractère processuel."
(Félix Guattari, Éric Alliez: " Le Capital en fin de compte: Systèmes, structures et processus capitalistiques ", in Change International, N°1, Paris, 1983)

Dans cet article nous mettons en évidence un certain nombre de caractéristiques concernant l'évolution du "Procès de Travail". A partir d'une réflexion qui prend sa source au plus prés de travaux menés sur la question politico-cognitive des nouvelles mémoires numériques en réseau ainsi que de recherches sur le processus de différenciation des dispositifs collectifs de production des savoirs, nous en proposons une mise en perspective, autant que faire se peut, critique.
La montée en puissance de ce que l'on appelle "l'ontologie réticulaire", des schèmes "auto-organisationnels", le passage de plus en plus marqué, des sociétés disciplinaires aux sociétés de contrôle,(1) la mise en mouvement de nouveaux types de légitimation, d'épreuves de formation-sélection, visant les individus, les savoirs, leur "mobilisation" renouvellée et l'émergence de nouvelles temporalités et subjectivités, sont autant de marques que nous sommes face à une actualisation-différenciation du procés de travail majeure.
Nous accordons dans cette réflexion une place importante au développement des réseaux électroniques distribués, au processus de numérisation du signe, au développement de pro-grammes informatiques complexes suscitant l'émergence de nouvelles formes et pratiques organisationnelles.

Depuis la fin de la seconde guerrre mondiale se sont mis en place, progressivement mais sûrement, un certain nombre de projets, d'expérimentations concernant l'organisation du procès de travail. Ces projets se sont développés à partir d'un ensemble de constats, de schèmes conceptuels et politiques concernant ce que l'on peut appeller, d'une façon trés générale, "les moyens de pilotage sémiotique"

nécessaires au fonctionnement de sociétés de marché(s) technologiquement avancées, des multiples organisations et institutions qui les constituent. Dans ce cadre l'évolution des intelligences collectives, le développement de nouvelles formes d'intelligence distribuée occupent une place importante.

Ces constats sont les suivants.
Il y a un mouvement profond, tout à la fois d'intégration et de différenciation-fragmentation des dispositifs de production des savoirs. Ce mouvement affectant les processus innovants. Deuxièmement, il y a une croissance exponentielle de la quantité de documents, d'informations. Qu'il s'agisse des secteurs privés ou publics, des instances étatiques ou non, cette croissance est bien entendu extensive mais aussi intensive. Une hétérogénéité croissantede "traces" étant produites et collectées au terme de cascades de trans-formations toujours plus élaborées, impliquant des collectifs hybrides de systèmes homme(s)-machine(s), de mémoires trés différenciés. Trés fortement engagée au cours du XIX° siècle la co-émergence de l'état, de la production-collecte de traces et de la statistique n'a fait que se développer depuis. (2)
Enfin il y a, sinon une espèce de déliquescence, du moins une contestation croissante fondée sur des motifs tant politiques que cognitifs, des sytèmes traditionnels de contrôle et de régulation centralisés, de type hiérarchiques et descendants. Cela se traduit entre autres par des demandes accrues, souvent contradictoires dans leurs dimensions politiques et économiques de mises en visibilité toujours plus poussées de ce qui nous lient, attachent, nous fait être éléments de communautés diverses et variées.
Cette contestation voire cette déliquescence des systèmes de régulation et de contrôles est en vérité plus complexe qu 'il n'y paraît. Là encore différenciation et fragmentation sont à l'oeuvre et la multiplication de ces systèmes, dispositifs va de pair avec les nouveaux actants (3) générant des nouveaux modes de production, de transmission et dissémination des savoirs. Ces modes dynamiquement et hiérarchiquement enchevètrés définissent de nouveaux rapports différentiels entre les modes hérités et entre ces derniers et les nouveaux.
Parmi ces nouveaux dispositifs certains visent des strates locales , d'autres encore opérent à plusieurs niveaux d'échelles voir à tous les niveaux. Ces dispositifs de régulation et de contrôle peuvent prendre des formes trés variés et mettre en oeuvre des processus de normalisation multiples. Pour aller à l'essentiel, nous avons à faire face à un enchevêtrement complexe de formes centralisées et a-centrées, de formes descendantes et ascendantes, arborescentes et "rhizomatiques ", dirigistes et auto-organisationelles. Mais n'en a t-il pas été toujours ainsi? (4)
Certains dispositifs opèrent comme boucles de réfléxivité,

instances-boîtes noires complexes, centrales et surplombantes dont la performativité repose sur la connexion et la mise en procés de longues chaînes d'actants hétérogènes, d'autres opérent localement par contagion et proximité, voisinage. Ce sont des contraintes locales distribuées, parfois de manière relativement homogène, parfois de manière inégale. Ces contraintes sont en fait des normes venant au terme toujours provisoire, parce que susceptibles d' être remises en cause par le double procés de technogénèse-sociogénèse, de processus de normalisation plus ou moins conflictuels.

Les questions ouvertes de "L'intellectualité de masse" (pour reprendre l'expression de Karl Marx) , des "Intelligences Collectives" (Pierre Levy...), des "Infrastructures Informationnnelles-Communicationnnelles" (Castells, Bowker, Hutchins, Larosières, Latour, Callon...), des "Mémoires Hypertextuelles, Hypermédias" (Nelson, Hayles, Berners-Lee, Eilola, Balpe, Stiegler...), du Capital Cognitif, Social (Negri, Guattari, Vincent, Burt...), sont au coeur de nouveaux enjeux, de nouveaux affrontements. Toutes ces évolutions sont l'expression et l'exprimé de la montée de nouvelles représentations et formes d'organisation réticulaires du monde en général et du monde socio-politique en particulier. Cette montée en puissance se fait sur l'affaiblissement des modes traditionnels de dissémination des savoirs, des processus de normalisation, des modes traditionnels des moyens de contrôle et de régulation, de formation.

Du collectif: en dehors du grand partage

Dans la perspective qui est la nôtre ici, la question du collectif est double. Tout d'abord il s'agit, afin de prendre la mesure de la croissance de complexité, de sortir de la sépération traditionnelle individu-collectif. Puis d'étendre, aussi loin qu'il est possible, en toute rigueur, les effets du constat de G. Tarde que " toute chose, est une société, que tout phénomène est un fait social ". (5)

Pour ce qui est du premier problème, il s'inscrit dans la lignée des travaux qui tentent de sortir de ce que Bruno Latour, à la suite de quelques autres, à appeller le "grand partage".

Associationnisme, processus de numérisation et procés de travail

On l'a déjà fortement pressenti, les mémoires numériques, les nouveaux logiciels et les nouvelles formes organisationnelles, affectent sérieusement ce procés.

Parmi les textes et prgrammes de recherches fondateurs qui expriment l'émergence aux États-Unis d'une nouvelle économie- politique de la mémoire et la mise place progressive de réseaux d'information-communication distribués quatre répères nous semblent importants. Au sortir de la seconde guerre mondiale, Vannevar Bush, qui

avait au milieu des années trente proposé avec le Memex un premier dispositif de séléction-association de documents mis sur micro-film et dirigé le projet Mahattan, publie un texte programmatique dans la revue "Atlantic Monthly Review". Ce texte ("As we may think") (6) à partir des deux premiers constats relevés en introduction, exprime l'idée que nous sommes mis en demeure d'avoir à inventer de nouveaux outils, techniques intellectuelles capables " d'accroître les capacités analogiques et associatives du cerveau humains ". Et il inscrit d'emblée sa démarche à la suite des grandes traditions humiennne et leibniziennne. Ce texte d'une vingtaine de page, largement diffusé, à partir de mai 1945, va irriguer la réflexion d'un nombre important de chercheurs, de politiques, de dirigeants d'entreprises, de documentalistes et susciter un certain nombre de programmes de recherche et d'expérimentation. Ces programmes concernent, entre autres, la réorganisation des programmes R$D, civils ou militaires, publics ou privés, des grands fonds documentaires, des dispositifs innovants. Les constats et propositions de V.Bush co-habitent avec la naissance de l'informatique, l'aube de la guerre froide et des grands programmes liés à la course aux armements, à la machinerie de la dissuasion nucléaire en cours de développement. Ils et elles sont repris sur fond de développement de la première cybernétique qui va triompher lors du colloque de Hixon en 1948. La décennie 50 verra la croissance informatique prendre de l'ampleur, la question du time-sharing occupant un des centres de la scène. A la fin des années 50 et au tout début des années 60, les grandes bases de données de données numérisées (20) commencent à s'imposer et dans le domaine scientifique et technique en particulier, la mise ne service de l'ISI (7) et de ses deux bases principales le SCI et le SSCI expriment la conscience qu'ont une partie des élites américaines de se doter de nouveaux moyens de mise en visibilité des activités de recherche. Cette mise en visibilité étant destinée à une meilleure évaluation-gestion de la recherche, tant d'un point de vue cognitif que du point de vue d'une politique (impériale?) de la technoscience. C'est aussi à cette époque que le programme d'une science de la science va se développer, programme dédié à la mise en visibilité des fronts de recherche, des collèges invisibles et qu'exprimera dans ses ouvrages par exemple, avec brio mais avec un point de vue élitiste, D. de Solla Price. (8)

Ce sont dans ces équipes que se mettent donc en place, en n'en pas douter, les fondements des réflexions et expérimentations actuelles concernant le caractére coopératif, collectif du procés de travail, les nouvelles intelligences collectives, ce que l'on appelle à présent la cognition, l'intelligence distribuée.
En parallèle, les réflexions engagées autour du document électronique, numérique et de la textualité vont converger et s'incarner dans les programmes de recherche de T. Nelson qui au milieu des années 60 va proposer un nouveau schème d'écriture et de gestion non-linéaire,

asssociatif, des mémoires numèriques émérgentes, schème qu'il va appeler "hypertextuel' (9). Cette notion d'hypertextualité qui va dans les 25 ans suivants imprégner nombres de recherches, resurgira avec succés au tout début des années 90 en s'incarnant dans le WEB, développé par les chercheurs du CERN, à Genève. L'hyertextualité dans un contexte numérique s'inscrivant au plus prés du programme de Bush, a savoir permettre le développement de systèmes d'écritures capables de favoriser les modes fondamentalement associationnistes, analogiques, de l'activitité intellectuelle.
Du côté des réseaux, le rapport Baran (10) va ébranler, contre "l'establishment" des grandes entreprises de communication, en particulier téléphonique, les schèmes traditionnels et introduire la notion de réseau distribué, acentré, c'est à dire fonctionnant indépendamment d'une instance centrale.

Les stratégies politico-cognitives de la mémoire et des interfaces-médiations dans le contexte numérique

Au terme de ces quarantes années, nous sommes pour partie dans cette postérité là. Sous l'égide de Tim Berners Lee (24) la convergence de ces quatre "virtualisations" va s'incarner dans l' avènement du second réseau internet à savoir le World Wide Web. Les technologies et les systèmes d'écriture, de transmission, (HTTP, SGML-HTML, URL...) développées pour créer le W3, l'ont été dans le cadre conceptuel, technique voire philosophique hérité, que nous venons brièvement de décrire. Le premier internet et plus encore le second, peut être considéré comme un nouveau système d'écriture, un nouveau système distribué de la mémoire, un nouveau mode de transmission distribué et en tant que tel porteur de technologies favorisant des nouveaux modes de travail coopératif, de nouveaux modes de pensée associationnistes, de traitements documentaires, enfin de nouvelles "traces " et des modes de représentation des activités cognitives en général et des activités socio-politiques en particulier.

Dans l'analyse des procés de travail aujourd'hui, le traitement des documents numériques est à cet égard révelateur. Il ne s'agit là que d'un point d'entrée possible. Il est toutefois majeur et ce d'autant plus que les processsus de normalisation qui l'acompagnent, montrent certaines des dimensions essentielles des transformations et enjeux en cours.

En effet ces processus de normalisation ont comme points de visée, les dimensions socio-cognitives et politiques des activités intellectuelles en général. Ils fournissent un certain nombre de repères, principalement autour et à partir de la notion de document numérique "textuel", dédiés à la relance du débat critique sur les nouvelles formes d'intelligences collectives, les pratiques qui les habitent.

. Plus particulièrement ils mettent en évidence, au delà de l'évolution documentaire stricto-sensu, les effets de formatage et de performation que l'utilisation des nouvelles normes du document numérique textuel engendrent et ce au sein de multiples communautés. Nous pensons précisément ici à l'émergence de nouveaux dispositifs homme(s)-machine(s) prenant en compte les caractères de plus en plus distribués des intelligences collectives.

Dans les domaines de la formation, de la recherche, de la diffusion, l'appropriation des univers socio-cognitifs émergents est relativement avancée. Ces communautés, en s'appuyant sur les nouvelles normes, ont développé et continuent de développer des nouveaux outils, de nouvelles mémoires, ont exploité et continuent d'exploiter diverses dimensions liées à l'hypertextualité numérique, aux modes de représentations émergents des caractères de plus en plus distribués, coopératifs, différenciés des dispositifs de production, circulation des savoirs.
Il apparaît de plus en plus nettement à présent que les communautés d'oeuvres, de chercheurs, (pour ne parler que d'elles, cherchent à prendre en compte plus fortement que par le passé, les multiples " incomplétudes en procès de production " dont elles sont l'expression et l'exprimé. Et ce en particulier, au moyen de la prise en compte des nouvelles traces (noeuds, liens, hyperliens...) exprimant de manière plus précise et complète que par le passé, la vie des univers de référence, de co-production, répétition, altération-création, dissémination des oeuvres, travaux divers... Mais aussi au moyen des nouvelles pratiques cartographes permettant la mise en visibilité des modes collectifs de co-émergence des activités, des productions de telle ou telle communauté. Au coeur de ces nouvelles pratiques, les possibilités offertes par les normes et standards émergents affectent ainsi l'évolution de l'édition électronique scientifique, des bases de connaissances en réseaux.
De plus ces normes ouvrent des possibilités d'exploitation renouvellées de la "plasticité numérique", et ce de manière toujours plus singulière, en favorisant la différenciation des pratiques d'écritures-(s)-lecture(s), des "éthologies" socio-cognitives. Il vaut donc la peine de s'y attarder.

La montée en puissance du document numérique : nouvelles caractéristiques et différenciation(s).

Le processus de numérisation du signe engagé depuis ces dernières décennies a atteint aujourd'hui un niveau de développement tel que l'ensemble des activités, des productions humaines s'en trouve grandement affecté, du moins dans le monde "technologiquement développé". Parmi ces productions subissant les effets de ce processus, ce que l'on appelle un document et en amont, les notions de trace(s),

d'écriture(s) et de mémoire(s).

Ce que l'on rassemble sous le terme " document " est non seulement devenu plus vaste, mais nous avons à faire à présent à une population de plus en plus dynamique, ouverte, qui ne cesse de se différencier et de participer à d'autres différenciations à l'oeuvre, à des niveaux d'échelles variés, par exemple, au coeur des pratiques socio-cognitives les plus diverses. Déjà P. Otlet (11) avait en son temps étendu trés loin, ce qui sous ce terme de document devait être subsumé: hyperdocument. "L'évolution de la Documentation se développe en six étapes. (...) Au sixième stade, un stade de plus et tous les sens ayant donné lieu à un développement propre, une instrumentation enregistreuse ayant été établie pour chacun, de nouveaux sens étant sortis de l'homogénéité primitive et s'étant spécifiés, tandis que l'esprit perfectionne sa conception, s'entrevoit dans ces conditions, l'Hyper-intelligence. "Sens-Perception-Document" sont choses, notions soudées. Les documents visuels et les documents sonores se complètent d'autres documents, les tactiles, les gustatifs, les odorants et d'autres encore. A ce stade aussi l'"insensible", l'imperceptible, deviendront sensible et perceptible par l'intermédiaire concret de l'instrument-document. L'irrationnel à son tour, tout ce qui est intransmissible et fût négligé, et qui à cause de cela se révolte et se soulève comme il advient en ces jours, l'irrationnel trouvera son " expression " par des voies encore insoupçonnées. Et ce sera vraiment alors le stade de l'Hyper-documentation."

Plus que jamais l'émergence des normes numériques nous rappelle, à la suite de bien d'autres, (12) l'importance des "humbles pratiques d'écritures et d'enregistrements" et de ce qui les affecte mais aussi des "résistances" que certains ont à "reconnaître le rôle des inscriptions, à s'intéresser au rôle de la pratique instrumentale" et à ce qui les affecte.

En ce sens, ce qui se passe aujourd'hui est aussi ancien que l'histoire de l'écriture et donc pour partie esentielle, de l'histoire de l'homme. L'émergence de l'écriture est émergence d'une stabilisation créatrice, (supposant donc divers procs de standardisation), de modes de production de traces répétables et combinables, transportables, quelque soit les milieux associés où ces traces sont répétées, dupliquées, combinées, transformées, déplacées, où quelqu'un répète, duplique, transforme... "Ainsi l'histoire de l'écriture revient à dire une double décontextualisation: des pictogrammes sumériens à l'alphabet consonantique sémitique, de là à l'alphabet grec, autrement dit à partir de l'union première entre le signe, le langage et le monde que réalise l'univers cunéiforme, via l'énigme du mot dans les alphabets sémitiques, pour aboutir à l'illusion sonore de l'alaphabet grec, l'écriture s'est détachée du contexte; dans cette distanciation, diversement réalisée par

les différentes cultures graphiques, elle a fini par rendre visible, en l'homme, les choses du langage et les choses du monde" (13)
Premier et lent processus de normalisation autour et à partir de Sumer, autour et à partir de Sarde, qui conduit respectivement à l'émergence de l'Écriture, de la Monnaie. (14)
Plus près de nous, en Europe, entre l'an mille jusqu'à nos jours, langages, systèmes d'écritures, dispositifs de réplication-dissémination, n'ont cessé d'évoluer, de se différencier selon des processus complexes, impliquant des chaînes d'actants (au sens latourien) trés variés. Ces processus, expression et exprimé du vaste creusement du couplage originaire "cortex-silex", (15) sont pour une large part de type associationnistes et autopoïétiques. Nous renvoyons, pour un exposé brillant de ces problèmes, à la deuxième partie (intitulée "Memes and Norms") de l'ouvrage de Manuel de Landa; "A thousand years of non-linear history". (16)

Parmi ces systèmes d'écritures, les plus importants, les plus connus , qui nous font nous interroger sur la nature du document numérique, sont à la suite du procès de numérisation lui-même, le code ASCII, les normes et standards de réseau et de protocoles de télécommunications (ETHERNET, ATM, TCP-IP, HTTP...) les normes de balisage des textes et documents, (SGML,HTML, XML,TEI, HyTime...), les normes d'échanges industriels, institutionnels et commerciaux, (EDI), les normes MPEG, les normes et standards spécifiant les supports et formats logiques et matériels de l'information (Disquettes, CDROM, DVD, ZIP...), les standards de traitement de textes etc...

Ces documents donc, qui viennent à l'existence aujourd'hui, présentent un certain nombre de caractéristiques générales, caractéristiques qui sont comme autant de problèmes à explorer, comme autant de questions à questionner. Examinons rapidement ces principales caractéristiques.
La première qui a été trés rapidement exploitée, renvoie au problème du stockage du document numérique. Dès lors que l'on dispose de moyens de stockage adéquats, complexes, il est en effet possible de collecter et de rassembler de vastes quantités de documents dans des espaces relativements réduits. Les bases de données émergeant à la fin des années 50,début des années 60, ont les premières, incarné cette possibilité.
La seconde caractéristique du document numérique c'est sa grande "plasticité", c'est à dire les possibilités étendues que ce dernier offre,de traitements automatiques (qu'il s'agisse de textes, d'images de sons), de types de manipulations, de transformations. La troisième caractéristique, trés vite perçue par exemple par Ted Nelson, ce sont les possibilités d'association des documents entre eux, de parties de documents entre elles et ce quelque soit la nature des documents associés, liés. On a évoqué à ce propos, selon nous de manière parfaitement justifiée, "une

nouvelle alliance" images/textes/sons. Cette caractéristique est au coeur de la problématique des écritures et des mémoires "hypertextuelles", "hypermédias" ...
D'autres caractéristiques expriment cette plasticité. Très rapidement, l'extrême transformabilité des formats de stockage, par exemple, mais aussi l'étendue des modes de recherche, d'analyse, d'interprétation des documents ainsi qu'une grande capacité de duplication, de circulation via des réseaux électroniques, dans un espace-temps pour partie déterritorialisé, et impliquant entre autres choses, des nouvelles temporalités, des nouveaux rapports de vitesse et de lenteur concernant les pratiques d'écriture(s)/lectures.

Ces caractéristiques nouvelles affectent l'histoire des traces, de l'écriture, de la pensée, les modes de répétition et de combinaison de ces traces, des signes, les modes de production et de dissémination des documents et des effets de ces modes de dissémination (ce qui hantait déjà Platon -voir le Phèdre). (17)

La plasticité du document numérique doit toutefois être encore précisée.

Au coeur des systèmes d'écritures, et des modes de production des documents, se joue en même temps que la question de la mémoire, pour une partie essentielle, la question des conditions nouvelles de la cognition. Avec le document numérique, le développement des réseaux et l'émergence de technologies intellectuelles permettant de démultiplier les capacités d'associations, les mémoires numériques sont entrées dans un procés de différenciation sans précédent dans l'histoire des mémoires externes. Ce procés de différenciation se doublant d'un procés de mise en mouvement complexe des documents les uns par rapport aux autres.

Les deux notions centrales de l'hypertextualité à savoir celle de noeud et celle de lien, exprimant la dimension processuelle et collective du document numérique. Cette dimension processuelle du document à dire vrai, est depuis toujours présente. Il n' y a pas de texte (Derrida, Latour...) qui ne soit par et au milieu d'un collectif de pratiques d'écritures-lectures, de duplication -circulation-dissémination de textes ou fragment de textes, de leur reprise, citation, altération... dans des agencements divers et ouverts. Procés sans fin ...
Ce qui est véritablement nouveau aujourd' hui c'est la mise en visibilité de ce procés, la mise en visibilité de ce travail, entre autre par les capacités à écrire et à inscrire une partie importante des processus associationnistes et collectifs de la cognition en acte, à des niveaux d'échelles potentiellement indéfiniment ouverts.

C'est là raison pour laquelle ces caractères du document numérique doivent être "creusés",dès lors que l'on considère les ensembles de

mémoires numériques associées en réseau de type hypertextuel.
Pierre Lévy a, au tout début des années 90 (18) caractérisé "afin d'en préserver les chances de multiples interprétations" le modèle hypertextuel, par six principes abstraits.

- "Le principe de métamorphose: le réseau est sans cesse en construction et en renégociation... Son extension, sa composition et son dessein sont un enjeu permanent pour les acteurs concernés, que ceux-ci soient des humains, des mots, des images... "
- "Le principe d' hétérogénéité: les noeuds et les liens d'un réseau hypertextuel sont hétérogènes... Le processus sociotechnique mettra en jeu des personnes des groupes, des artefacts, des forces naturelles de toutes tailles, avec tous les types d'association que l'on peut imaginer entre ces éléments. "
- "Le principe de multiplicité et d'emboîtement des échelles: l'hypertexte s'organise sur un mode "fractal", c'est à dire, que n'importe quel noeud ou n'importe quel lien, à l'analyse, peut lui même se révéler composé d'un réseau, et ainsi de suite indéfiniment, le long de l'échelle des dégrés de précision..."
- "Le principe d'extériorité: le réseau ne possède pas d'unité organique, ni de mot interne. Sa croissance et sa diminution, sa composition et sa recomposition permanente dépend d'un extérieur indéteminé..."
- "Le principe de topologie: dans les hypertextes, tout fonctionne à la proximité, au voisinage. le cours des phénomènes y est affaire de topologies, de chemins..."
- "Le principe de mobilité des centres : le réseau n'a pas de centre, ou plutôt, il possède en permanence plusieurs centres qui sont comme autant de pointes lumineuses perpétuellement mobiles, sautant d'un noeud à l'autre, entraînant autour d'elles une infinie ramification de radicelles, de rhizomes; fines lignes blanches esquissant un instant quelque carte aux détails exquis, puis courant dessiner plus loin d'autres paysages du sens."

Ces dimensions du document numérique, dans l'espace-temps des mémoires hypertextuelles, sont au coeur d'un vaste travail d'expérimentation et d'exploration collective.
Et la question des normes est la question de la production-appropriation ouverte, conflictuelle parfois, de ces dimensions processuelles, collectives-collaboratives, hybrides. Elle est l'expression des efforts que font des individus, des groupes, des institutions, des industriels... pour maîtriser et exploiter à leur avantages, à leurs profits des systèmes d'écritures puissants et les mémoires qui vont avec.
Avertis pour une part importante, des émergences, des gains en complexité et différenciation, qui n'ont pas manqué de suivre les grandes

révolutions graphiques de l'histoire humaine, les grandes et petites révolutions de l'écriture, des mémoires externes, ils anticipent et accompagnent les variations qui affectent les puissants dispositifs auto-organisateurs que sont les hypertextes-hypermédias numériques en réseau.

Selon les domaines socio-cognitifs d'activités, les acteurs qui les constituent, et les niveaux d'écritures considérés, mis en jeu dans la création de ces dispositifs cognitifs hypercomplexes, ces efforts convergent ou divergent, ces efforts privilégient ou négligent tel ou tel aspect de la plasticité numérique.
De telle sorte que le débat sur les normes émergentes du document numérique est le débat le plus complexe et le plus ouvert qui ait jamais eu lieu quant à la question de l'écriture, de la mémoire et quant aux conséquences de la mutation présente de cette dernière.

Si l'on reprend les évènements dans l'ordre et pour aller à l'essentiel,
a) un changement radical du mode d'écriture et d'inscription des traces et des signes (numérisation),
b) le développement de la science informatique (langages de programmation et approfondissement de la question des pro-grammes-logiciels),
c) la création de réseaux électroniques distribués et l'émergence sous la forme de la définition de protocoles, par exemple le protocole fondateur "TCP/IP" du réseau Internet, le protocole "HTTP" fondateur du deuxième Internet, W3 "hypertextuel") permettant l' échange et la circulation de données numérisées , puis
d)le développement de normes d'écritures rendant possible le partage de ressources hétérogènes sous un mode client-serveur (langages dits de "balisage", SGML, HTML, XML), sont à la base de cette strate anthropologique complexe, nouvelle où sont produits, distribués, dupliqués, échangés, commentés, bref mis en mouvement une variété considérable de documents numériques.

Ces évènements répartis sur près d'un demi siècle à présent, ont été engendrés par un ensemble d'acteurs, de réseaux d'acteurs hétérogènes, oeuvrant aux marges ou sur des fronts de recherche ayant pour but l'amélioration des intelligences collectives, les capacités innovantes, la maitrise économique et politique des sociétés complexes. Ces marges, ces fronts ont consisté et consistent toujours, à partir du procés de numérisation , à définir des modes d'écritures rendant possible l'exploitation intelligente des ressources documentaires en croissance exponentielle dès avant la seconde guerre mondiale (préoccupation forte déjà présente dans les travaux d 'Otlet, Goldberg, Bush...) la maîtrise autant que faire se peut, des processus de fragmentation-différenciation affectant les mutliples sphères de production des savoirs, d'informations nécessaires à la bonne marche de sociétés, d'organismes de plus en plus

complexes.

Les caractéristiques du document numérique , telles que nous les avons décrites et mises en perspectives ont généré et génèrent un grand nombre de problèmes. Ces problèmes concernent tous les aspects de la vie des documents, c'est à dire de leur production, de leur circulation, de leur diffusion, de leur conservation, de leur protection.
Les systèmes qui rendent possible la réalisation du plus grand nombre des modes de chacune de ces fonctions, reposent entre autres sur la création de standards, de normes (18.1), qui correspondent donc à des objectifs, des intérêts extrêmement variés et parfois variables.
Ces objectifs et intérêts peuvent être conflictuels voire contradictoires selon par exemple qu'il vise un partage le plus large possible des ressources, ou bien la protection maximun d'un certain nombre de ressources, selon qu'ils visent la production de savoirs complexes dans et par des espaces socio-cognitifs eux-mêmes complexes ou bien l'accés simple et largement partagé à un type particulier, limité de ressources.
Les dispositifs numériques , hypertextuels, en réseau étant composés d'un emboîtement complexe d'écritures, de programmes, la définition de nouvelles normes et la manière dont elles vont s'affronter, s'imposer sont affaires complexes. De plus chaque norme spécifique bien qu'opérant , au sein d'un système d'emboîtement, d'une combinatoire de systèmes de traduction et d'interopérabilité complexes, à des places-niveaux-strates spécifiques, n'en produit pas moins des effets de création-altération à des niveaux supérieurs. Le protocole HTTP, la norme HTML , par exemple, venant à la suite de normes antérieures (ces dernières en étant pour une large part conditions de possibilité) comme des axiomes supplémentaires dans le vaste système d'écritures, des normes héritées, créés des conditions de production et de créativité nouvelles en appellant à son tour l'émergence de nouvelles normes, selon des rythmes et des temporalités divers.
A partir donc des normes fondamentales que nous avons énoncées précédemment , les problèmes posés par l'appropriation cognitive, socio-économique, politique... des documents numériques en croissance rapide, prennent forme et consistance, ou bien si l'on préfére, s'actualisent à travers l'invention et la définition de normes d'écritures, d'échange et les espaces de jeu qu'elles offrent aux intérêts et passions singulières et/ou collectives. Ces inventions-définitions répondent donc à des problèmes qui répondent eux-mêmes à des séries spécifiques de conditions portées par des ensembles d'actants plus ou moins hybrides.
La plasticité du document numérique , ses principales caractéristiques, ouvrent une certaine période d'instabilité qui, tant sur le plan plan cognitif que politique et économique soulèvent des problèmes.

Notes

(1) G. Deleuze, " Postcript on the societies of control " Publié aussi dans Pourparlers, Éditions de Minuit, Paris, 1990

(2) A. Desrosières: Histoire de la Raison Statistique, Éditions de la Découverte, 1996, ainsi que le Numéro 8 de Raisons Pratiques: Cognition et Information en Société, Éditions de l'Ehess, Paris, 1997

(3) Au sens de B. Latour.

(4) Manuel De Landa: A Thousand years of non-linear History, Ed. Swerve, 1997

(5) G. Tarde: Monadologie et Sociologie, Les Empêcheurs de penser en rond, Institut Synthélabo, 1999

(6) V. Bush: " As we may Think " in The Atlantic Monthly Review, I945 Voir aussi: http://www.stg.brown.edu/projects/hypertext/landow/ht/memex.html

(7) L' ISI (Institute for Scientific Inforlation) fondée par E. Garfield

(8) D. Solla Price: Science and Suprascience, Éditions Fayard, Paris, 1972 et Solaris N°2: http://www.info.unicaen.fr/bnum/jelec/Solaris " Aux sources de la Scientométrie ", Xavier Polanco (Version papier, Presses Universitaires de Rennes II, 1995)

(9) T. Nelson: A File structure for the complex, The Changing and the indetreminate, ACM 20th National Conference, 1965 et All for One and One for All, Hypertext'87 Proceedings, November 87 Voir aussi: http://www.eastgate.com et J. Conklin, Hypertext: An Introduction and Survey, IEEE Computer, 1987

(10) P. Baran: http://www.rand.org et A. Serres: "Histoire d'un système d' information: l'émergence d'Arpanet", Thèse, Université de Rennes II, à paraître, Octobre 2000

(11) P. Otlet: Traité de Documentation, Le livre sur le Livre, Édition C.L.P.C.F, Liège,1989 - Voir aussi: "Paul Otlet and his Successors", in Jasis Volume 48, N°4, April 1997: M. Buckland and T.B. Hahn, W.B. Rayward, D. Ray et "What is a document? Rethinking the Concept in Uneasy Times", L. Shamber

(12) B. Latour: " Les vues de l'esprit " in Culture Technique, 1985 et J. Goody, La raison graphique, Éditions de Minuit, Paris, 1979

(13) C. Herrenschmidt, "Le tout, l'ènigme et l'illusion" in Le Débat, N°62,

1990

(14) C. Herrenschmidt, "Écriture, monnaie, réseaux", in Le Débat, 1999. Voir aussi: http://barthes.ens.fr/atelier/actuel/herrenschmidt.html

(15) B. Stiegler: La Technique et le Temps, T1 et T2, Éditions Galilée, Paris, 1994, 1996

(16) Manuel De Landa: A Thousand years of non-linear History, Ed. Swerve, 1997

(17) Platon: Le Phèdre et Ann Van Sevenant, Écrire à la lumière, Editions Galilée, Paris, 1999

(18) P. Lévy: Les technologies de l'intelligence, Éditions La Découverte, Paris, 1990

(18.1) Y. Marcoux, M. Sévigny, "Why SGML, Why Now?", in Jasis, July 1987 " In the world of electronic document formats, the following levels of standardization can be identified:
a) Secret proprietary format: A format definde by a specific hardware or software producer, and whose specifications are unpublished. The original software or hardware must be used to access the documents. Under users' pressure, this kind of format is starting to disappear.
b) Public proprietary format: A format defined by a spscific producer, byt whose specifications are publicly available. Others parties can develop import/export converters for the format, if they want, e.g., RTF (Microsoft's Rich Text Format), Wordperfect.
c) De facto standard: A public proprietary format that has become very popular and is recognized by a large number of producers, e.g., GIF (Graphic Interchange Format, introduced by Compuserve), Wordperfect.
d) Official standard: A standard defined and adopted by an official standardization body. Standardization bodies include a number of non-profit organization or consortia, varoius national or international associations, as well as organizations entirely devoted to standardization on a national or intenational level, e.g., SGML, ODA.
The Intenational Organization for Standardization, (ISO), Which has its headquarters in Geneva, Switzerland, is possibly the most important standardization body in the world. In particular, the General Agreement on Tariffs and Trade (GATT) recognizes the precedence of ISO standards over others. SGML is an ISO standards, ISO 8879, adapoted in 1986. It is, therefore,, a first-rate international standard."

La surcharge informationnelle

INFORMATION OVERLOAD

Anne Sanouillet, Professeur en Sciences de l'Information et de la Communication - URFIST DEH Université de Nice Sophia-Antipolis

"Information overload is not a function of the volume of information... it's a gap between the volume of information and the tools we have to assimilate that information into useful knowledge. The cost to the user is that it takes time to sort through the information, and time is money." - Reuters
"La surcharge informationnelle n'est pas fonction de la quantité d'information... c'est l'écart qui existe entre la quantité d'information et les outils dont nous disposons pour transformer cette information en connaissances utiles. Le prix à payer par l'utilisateur est le temps qu'il faut pour la trier, et le temps c'est l'argent." - Reuters

La science de l'information, discipline commune des Urfist, est en voie de devenir la science de l'élimination de l'information - du moins de la partie "bruit" qui en masque la partie "pertinence". Grâce à l'Internet les chercheurs peuvent en principe trouver facilement les textes ou les références leur permettant d'avancer dans leurs études. Cependant beaucoup d'utilisateurs se plaignent du temps qu'il faut consacrer à une recherche sur le Web - non pas pour formuler les requêtes, mais pour réceptionner les réponses, pour identifier et vérifier celles - et celles seulement - qui seront utiles.

D'où vient le problème ? Depuis l'avènement sur Internet des moteurs de recherche, puis des portails spécialisés, les chercheurs disposent des premiers outils capables de faire face aux millions de pages Web en texte intégral. De surcroît ils ont accès à des banques de données gratuites, aux annuaires, aux périodiques électroniques sans parler de la messagerie électronique, des listserv et des groupes de discussion. La préparation d'un travail de recherche n'a jamais paru si réalisable, du moins pour ceux qui ont leur propre accès à Internet. Néanmoins nos jeunes chercheurs et même leurs enseignants se plaignent d'être submergés par le nombre de réponses, de ce que la vérification de chaque piste est fastidieuse et aspirent à trouver des solutions efficaces pour atteindre des informations identifiables et directement utilisables.

Nous sommes exposés à un véritable orage informationnel. Le courant libérateur déclenché sous la Renaissance est devenu un déluge chaotique. Le nombre de livres proposés dans une bibliothèque ou dans une bonne librairie est intimidant sinon déprimant. Quand on considère la quantité de travail intellectuel accompli depuis l'ère de Gutenberg, on devrait hésiter avant d'ajouter sa propre pépite au trésor des siècles. Mais hélas ! personne ne reste trop longtemps intimidé. Le Web permet à tout un chacun d'avoir "sa" plateforme personnelle, de

dire n'importe quoi, alors pourquoi s'en priver ?

La quantité exponentielle de données échangées sur l'Internet a provoqué la création d'outils performants pour la recherche, la sélection, l'analyse, et la traduction de l'information. La quantité excessive de documents ne devrait plus nous impressionner. Pourtant nous rencontrons des difficultés pour extraire de ce magma l'information utile aux chercheurs, aux économistes, aux scientifiques.

La technologie fait-elle vraiment gagner du temps ?

Depuis un demi-siècle le style de la recherche universitaire s'est transformé. Les jeunes chercheurs sont obligés de terminer leurs DEA ou doctorats dans un délai de deux ou trois ans. Ils pourraient utilement comparer leur sort à celui des universitaires français des années cinquante. À cette époque la machine à écrire était l'unique matériel disponible. Il n'y avait ni fax, ni photocopieur, ni Minitel, ni ordinateur, ni scanner, ni banque de données, ni Internet. Un thésard consacrait de dix à quinze ans à la rédaction de sa thèse d'État (car il en fallait deux à l'époque, la thèse principale et la secondaire). Toute sa recherche se faisait dans des archives publiques ou sur des documents détenus par des particuliers. Il recopiait des documents à la main dans des bibliothèques où il se rendait en personne même si elles se trouvaient à l'étranger. Les catalogues existaient, mais chaque notice était tapée sur un petit carton de bristol. Une fois terminé, le manuscrit devait être retapé plusieurs fois avant d'être soutenu imprimé. Une thèse était un travail de fourmi, réalisé à une vitesse d'escargot, d'ailleurs.

Qu'est-ce qui est préférable, le trop-plein d'information actuel ou le manque d'information et de moyens de nos aînés ? Sans doute le chercheur d'aujourd'hui peut faire sa collecte d'information plus rapidement et plus facilement que son grand-père. Mais voilà que pointe la menace de la surcharge informationnelle, car il lui faut traiter, lire et citer une masse de ressources pour paraître crédible. Ensuite il faut taper le travail de recherche soi-même sur son ordinateur personnel, car de tels services rendus par une secrétaire finissent par coûter le prix d'un ordinateur. Il faut donc prendre le temps de maîtriser deux aspects de l'informatique, le traitement de texte, et la recherche sur Internet. Il est certain que la technologie procure des atouts certains à l'étudiant d'aujourd'hui, mais au prix de nombreuses contraintes dont une pression pour soutenir vite que ses aînés ne connaissaient pas. Les chercheurs d'aujourd'hui devraient mieux mesurer leur bonheur. Cela dit, une amélioration des outils de recherche ou de leur utilisation ne serait pas superflue.

Les solutions

Une étude (1) indique qu'en février 1999 le moteur de recherche le plus important à l'époque, Northern Light, ne couvrait que 16% des données accessibles par le public sur Internet. Les 11 plus gros moteurs combinés ne couvraient qu'approximativement 42% de ces données accessibles (à l'extérieur du Web invisible). Le Web comptait au moins 800 millions de pages, mais c'est une estimation déjà dépassée. D'ailleurs, tous ces chiffres auront changé depuis un an.

La solution : les agents intelligents qui réagissent en fonction de l'environnement, savent s'adapter aux circonstances, et peuvent même prendre une décision. Leurs performances

s'améliorent d'après les observations qu'ils effectuent et les consignes qu'ils recoivent. Ils sont capables de choisir les documents et de les classer par ordre de pertinence (2).

Certains moteurs de recherche combinent un balayage de pages par catégories à une requête directe pour donner plus de pertinence aux réponses (Northern Light, Lycos, Voila, Livelink Pinstripe, etc.).

Le Web Invisible

Le Web Invisible est la partie du Web qui est accessible par le public mais qui ne peut pas être indexée par les moteurs de recherche classiques. Cette toile est composée de :

1. Pages accessibles par un formulaire de recherche.
2. Pages protégées par un mot de passe.
3. Pages interdites aux robots.
4. Pages à l'intérieur de cadres mal conçus.
5. Pages qui sont en Word, Word Perfect, Flash ou PDF (Acrobat)

Le savoir-faire des bibliothécaires et des Urfist pourrait être sollicité pour aborder ces problèmes. Des agents intelligents existent par secteur d'activité pour interroger cette partie souterraine, mais leur utilisation demande une formation encore plus poussée.

Les start-ups françaises et américaines travaillent activement sur un autre aspect de l'information overload, aspect qui est traité abondamment sur Internet. Car la bête noire des commerces et des entreprises en Amérique (et bientôt en France) est l'inondation quotidienne par le courrier électronique et l'impossibilité d'y faire face rapidement. Il existe pourtant des logiciels professionnels américains qui peuvent lire et répondre automatiquement à 90% du courrier électronique d'une entreprise, en faisant suivre à qui de droit les 10% restants qui exigent une attention personnelle (3).

Mais que proposer aux étudiants d'aujourd'hui qui doivent souvent travailler contre la montre ? La courbe d'apprentissage de ces techniques sophistiquées peut inhiber les non-spécialistes.

Les problèmes qui freinent l'efficacité du Web sont :

1. Le manque de maîtrise de la technique
2. Le manque de pertinence et la quantité des résultats
3. La difficulté dans l'évaluation de la qualité et de l'utilité d'un site Web (4).

1. Le manque de maîtrise de la technique

Pour trouver l'information sur le Web relativement rapidement, il faut savoir utiliser les outils de filtrage comme les moteurs de recherche AltaVista, Northern Light, Fast (le plus gros), Voilà (pour la France) et l'annuaire Yahoo. Google a la particularité d'élaborer un ordre de classement fondé sur la popularité des sites.

Il en existe des centaines d'autres, mais comme il faut comprendre les particularités de chacun, il vaut mieux en maîtriser quatre ou cinq, et en choisir plusieurs pour chaque recherche en fonction de la requête. Car ils ne se recoupent que très peu. Les métamoteurs (The Big Hub, MetaCrawler) peuvent servir, si la question est simple, pour gagner du temps. Copernic 2000 Pro est un logiciel très efficace qu'on peut installer sur son ordinateur, mais qui délivre essentiellement des informations américaines - pour le moment. Cependant les spécificités d'AltaVista ou de Northern Light seraient en grande partie gommées dans une métarecherche qui ne tiendrait pas compte du caractère des moteurs sophistiqués.

Pour certains de ces moteurs, il est nécessaire de savoir manier les opérateurs booléens et de proximité. La plupart des étudiants de 3e cycle à l'Université de Nice n'en ont jamais entendu parler. Sans quelques notions de base sur la formulation de la question, les résultats de leurs recherches seront forcément décevants.

L'étudiant de langue française a intérêt à doubler sa recherche en anglais, car le Web anglophone est sensiblement plus vaste que le Web français. Beaucoup de chercheurs n'ont ni la patience ni la compétence pour lire des articles en anglais. C'est là une des explications de l'appauvrissement de leurs recherches.

2. Le manque de pertinence et la quantité des résultats

Nous revenons au problème déjà évoqué : comment traiter l'énorme quantité de données sur le Web ? Au début le Web devait être plus simple que les banques de données. Dans une banque de données classique, avec thesaurus et conditionnement documentaire, l'utilisateur sommairement formé peut sortir un nombre gérable de références ou de textes pertinents. La même question sur le Web suscite beaucoup plus de rappel et infiniment moins de pertinence.

Quelques suggestions

1. Choisir un moteur en fonction de la question à poser
   · simple ou complexe
   · en France ou dans le monde.
2. Cerner exactement le sujet de la recherche
   · si l'on ne sait pas ce que l'on cherche, aucun moteur ne le trouvera.
3. Analyser le sujet et en extraire les concepts.
4. Choisir les termes de la question et leurs synonymes.
5. Penser aux troncatures si elles ne sont pas automatiques.
6. Combiner les termes avec des opérateurs booléens et/ou de proximité.
7. Limiter la question par champs pour exclure des domaines comme .com
   Ex. and not domain:com (AltaVista)
   ou pour exiger des domaines comme .edu
   Ex. and domain:edu (AltaVista)
   Cette solution ne vaut pas pour les disciplines business comme la gestion et le marketing.

8. Au cas où il y a des milliers de réponses, limiter la question avec des concepts supplémentaires, par exemple la discipline (Law ou Droit), le pays, etc.
9. Réutiliser les termes trouvés dans les résultats de la recherche ou dans les mots clés de la source HTML de la page (meta name="KeyWords" content=) pour mieux poser la question à nouveau.
10. Lire attentivement les descriptions des sites pour mieux les choisir.
11. Ne pas dépasser les premières pages de résultats - 50, 100 etc. en rapport avec le nombre total de liens proposés. En principe les premiers sont les plus pertinents.

3. La difficulté dans l'évaluation de la qualité et de l'utilité d'un site Web

Une bonne proportion de ce que l'on trouve sur le Web n'est pas utilisable. Il n'y a pas de Label Rouge, ni d'appellation contrôlée.

Il est donc sage de commencer la recherche dans des sources stables et connues. Si un article n'est pas signé, ou ne provient pas d'une source réputée (éducation, périodique, gouvernement par exemple), mieux vaut passer au suivant. Les périodiques électroniques ne manquent pas et constituent de bonnes sources fiables d'information. Pour commencer il y a des banques de données qui servent de portails aux meilleurs sites Web dans chaque discipline (5) .

Quels sont les critères qui rendent un site crédible ?

1. Le nom d'un auteur connu dans sa discipline.
2. Le nom d'une institution réputée.
3. L'absence d'une intention de vendre quelque chose ou de manipuler le lecteur.
4. La place du hasard

L'informatique fournit certes un atout formidable dans la recherche d'information. Mais il serait dommage d'éliminer le facteur "chance". Qui n'a pas feuilleté un livre ou une revue pour tomber inopinément sur un document qui va se révéler d'une grande utilité? Si nous arrivons à des systèmes qui ne trouveront que l'information pertinente, quelle place serait laissée à la possibilité de découvrir par hasard l'élément qui déclenche un processus auquel on n'avait pas songé auparavant? Le temps passé à parcourir des documents hors sujet est souvent bien récompensé. Peut-être cet excès de documents a-t-il du bon?

Cet article n'est pas long. C'est là aussi une manière de combattre la surcharge informationnelle.

Notes

(1) LAURENCE, Steve et GILES, C. Lee. Accessibility of information on the web. Nature,

juillet 1999, n° 6740, p. 107-109. Voir http://www.wwwmetrics.com

(2) DAMIANO, Jean-Pierre. La recherche d'Informations, la veille stratégique, l'intelligence économique et l'Internet. http://www.elec.unice.fr

(3) E-mail specialist, YY Software Corporation, Mountain View, Californie, http://www.yy.com/.

(4) FORNACIARI, Charles J. et LOFFREDO ROCA, Maria F. The age of clutter: Conducting effective research using the Internet. Journal of Management Education, déc. 1999, vol. 23, n° 6, p. 732-742.

(5) BUBL Link. http://www.bubl.ac.uk/link/ Ressources pour des universitaires en texte intégral. NISS - Directory of Networked Resources. http://www.niss.ac.uk/subject/index.html. Une bonne source d'information par discipline pour des universitaires. Accès à presque 4 000 ressources multidisciplinaires avec des liens. SOSIG - Social Science Information Gateway. http://sosig.esrc.bris.ac.uk/welcome.html . Pour trouver des ressources à l'aide d'un moteur de recherche et des opérateurs booléens.

FORSIC

Jo Link-Pezet, Maître de conférences à l'URFIST de Toulouse

La formation des formateurs a mis en évidence l'importance du groupe de travail. Les échanges riches et leur densité ont permis de soulever des questions multiples et intéressantes : l'ergonomie des lieux, les fonctions d'information, la diffusion des informations et la communication visuelle dans un espace public, la gestion collective des connaissances, la mémoire des savoir-faire, l'interaction des individus. La mise en réseau des mémoires met en évidence la nécessité d'acquérir de nouvelles compétences, en particulier pour faire face à des situations inédites et complexes qui nécessitent de travailler ensemble, d'interagir. La trajectoire de la connaissance d'un groupe se communique à travers plusieurs axes qui recoupent les préoccupations du travail coopératif qui n'a rien de spontané, qui s'apprend, se coordonne. Les axes suivants nous paraissent très intéressants à développer de façon plus scientifique :

- Notions de langage commun,
- Développement d'une action partagée par plusieurs et partage de contexte, de ressources, d'outils,
- Reconnaissance mutuelle, domaines cognitifs partagés,
- Intégration de buts individuels vers les buts collectifs et évaluation,
- Intégration des analyses individuelles pour élaborer une représentation du monde commune, extension de la réalité.

Le groupe qui coopère est un intellectual team work, un collectif intelligent à l'œuvre.

Les conditions qui favorisent la production de connaissances et la conduite de projet sont l'intention, la stratégie, l'autonomie qui permettent d'acquérir, traiter, relier entre eux les différents niveaux d'information pour en faire de la connaissance. Auto-organisation et coordination sont complémentaires dans un projet où plusieurs processus sont en cours simultanément et souvent se chevauchent. La fluctuation (ordre sans la récursivité) et le changement sont considérés comme des facteurs de création de connaissances et nécessaires au progrès parce qu'ils brisent les routines et cassent les habitudes. La notion de redondance est souvent intéressante. En effet, l'information peut être utile en dehors d'un usage immédiat et donner des visions multiples d'un même sujet, indiquer différentes manières de réaliser une même tâche, introduire de nouveaux éléments pour appréhender le futur.

Organisation du milieu ou organisation hypertextuelle

H. Takeuchi propose une organisation qui s'appuie sur les avantages du modèle top-down (stabilité, contrôle, standardisation, ordre), qui évite ses inconvénients (manque d'initiative personnelle, absence d'objectifs partagés, dysfonctionnements) en empruntant à l'organisation du travail par projet la flexibilité, l'adaptabilité, le caractère participatif, l'absence de hiérarchie. Ce schéma met en jeu trois entités : l'organisation, les individus, le système coopératif ainsi que les liens qui les unissent.

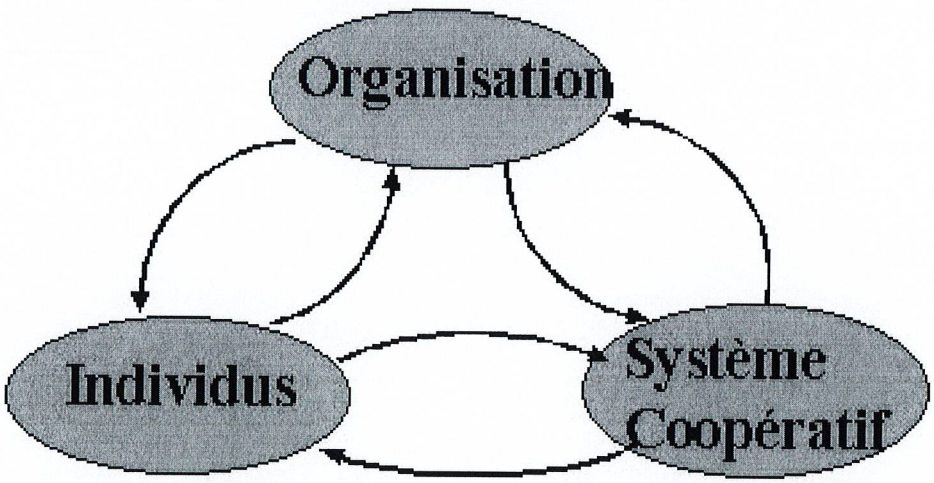

Figure 1 : Organisation hypertextuelle

Ce type d'organisation coopérative peut absorber la dynamique constante du changement et des flux. Il reconnaît les compétences technologiques et les savoirs des individus qui le composent. Il s'appuie sur la reconnaissance du travail intellectuel et de la connaissance. Non hiérarchique, autonome et auto-organisatrice, il s'appuie sur les principes de l'énaction incarnée qui considère le bien-fondé de la tâche à réaliser et du contexte. Dans des environnements complexes et évolutifs, il donne la capacité stratégique d' acquérir, créer, exploiter, accumuler de la connaissance nouvelle en permanence. Il implique la présence dans le dispositif d'un organe de décision (pilotage stratégique) et d'un groupe de projet (réalisation de la tâche). Dans le dispositif d'organisation hypertextuelle, l'organisation occupe une position importante et donne des directives, alloue des tâches et des ressources. L'organisation fournit les consignes, le projet, donne les clés d'interprétation de la situation

(pertinence).
Nous disons que cette organisation est hypertextuelle car elle permet l'accès à plusieurs couches de connaissances (espace topographique hétérogène formé de liens et de nœuds). En effet, l'hypertexte en tant que mode de traitement du texte autorise l'entrée dans l'espace au niveau de granularité souhaité ou dans l'environnement même de la ressource (information située). Il est possible de naviguer dans l'espace créant des liens entre les strates et les contextes différents.
La production de connaissance sous diverses formes et son amplification à travers l'organisation est la clé de voûte du système. Les individus impliqués dans le groupe de projet travaillent sur de la connaissance conceptuelle et sociale, ils créent de la connaissance systémique et opérationnelle. Plusieurs équipes de travail peuvent être engagées dans ce travail de création de connaissances (développement de nouveaux produits, définition de nouvelles missions). Des personnes issues de tout service et connues pour leur compétences sont affectées à la réalisation d'un projet et ce, jusqu'au bout du projet.
Le système coopératif doit leur permettre d'intégrer leurs différentes connaissances et de réaliser avec les autres membres du groupe la tâche définie par l' organisation, qu'aucun d'entre eux n'aurait été capable de réaliser seul. Il permet de communiquer et de débattre. Le débat permet de mener le travail à terme. La coopération se fait soit à distance ou par réunion. La responsabilité peut être collective ou distribuée.. .
La base de connaissances est le lieu où est la connaissance, les interprétations de tous les acteurs sont recontextualisées, recatégorisées (mémoire d'entreprise) et rendues accessibles. La base de connaissances est constituée de la connaissance de chacun, accessible à tout instant, susceptible d'être convertie et utilisable par tous. La vision de l'entreprise est alors basée sur la culture organisationnelle et la mémoire de l'organisation sur les réalisations du groupe au terme d'un cycle dynamique de création de connaissances collectives.
Ce genre de dispositif a quelques implications pratiques. Tout d'abord, il faut :

- Un projet : c'est le rôle des décideurs, ils doivent le communiquer, de manière à dégager un champ d'actions, un domaine, des missions et de l'organisation. Au-delà de la création de services et de produits, ils s'engagent vers l'avenir et créent de la stratégie signifiante à long terme,
- Avoir une politique de recherche d'innovation, de développement de produits nouveaux, qui ne peut émerger que dans la différence et la variété,
- Avoir une organisation hypertextuelle qui favorise la conversion de connaissance
- Du personnel familier avec les notions de connaissance, susceptibles de créer de la connaissance au regard de l'organisation, dans la diversité et l'autonomie, d'assumer des responsabilités,
- Construire un réseau de savoirs et d'experts,
- Un réseau d'interactions par une volonté de communication utilisant les différentes technologies disponibles, promouvoir le partage d'information et l'interaction sociale.

FORSIC, un dispositif de gestion collective des compétences

Le projet FORSIC peut se définir comme la mise en œuvre d'" une organisation capable d'apprentissage et d'adaptation ", comme un espace documentaire où l'information en réseau joue un rôle primordial en tant que ressource et en tant que moyen de transformation. Dans

FORSIC, l'objet du dispositif est la formation à la recherche documentaire des étudiants des campus et s'articule sur un modèle de la recherche documentaire en tant que discipline, les modèles cognitifs des apprenants et des formateurs basés sur la conversion des connaissances, un dispositif technologique favorisant la coopération, un dispositif socio-culturel favorisant la gestion collective des connaissances.
Il s'agit d'un dispositif pour recueillir de la connaissance utile pour réussir des formations et démultiplier les compétences des formateurs en mettant en œuvre une base de connaissances à partir de la description de celles-ci. Plusieurs technologies sont appliquées pour les mettre en mémoire, les représenter et les enrichir. Il s'agit de décrire les savoirs des formateurs dans leur vie professionnelle réelle en tant que membres participants d' une collectivité de savoir. Chaque compétence contribue à et alimente la base de connaissances. Le but recherché est la mise en œuvre de collaborations créatives basées sur la confiance mutuelle, le respect de la connaissance des gens, de nouvelles façons de travailler pour fournir un avantage égal à chaque participant : donnant- donnant. Le projet FORSIC est directement issu de la réflexion sur la formation des usagers à l'information. Le but est la création d' un référentiel professionnel et pédagogique et un outil de gestion collective des compétences des formateurs à l'information documentaire en Midi-Pyrénées, un inventaire des ressources pédagogiques. La mise en place d'un système de gestion collective s'appuie sur des contenus (la richesse du réseau), sur la communication (entre les membres du réseau) et sur des outils de collecte, de mise en forme, de cartographie, de consultation des compétences. C'est un relevé " topographique ", empirique, l'expression des savoirs présents dans la communauté, expression dynamique qui peut se transformer à tout instant.
Il s'agit d'identifier, mettre en mémoire, diffuser, partager et évaluer par l'usage la connaissance qui existe parmi les formateurs, qui leur appartient pour constituer une mémoire collective, transmissible, partageable à partir des compétences individuelles et collectives d'une communauté de pratiques spécifique. La gestion collective de connaissances des individus s'inscrit dans la culture où ils travaillent et vivent, dans leur environnement. C'est un système ouvert auquel peut participer toute personne qui s'intéresse et fait de la formation. L'indexation des personnes et des ressources est une indexation par les compétences possédées par les formateurs, attendues chez l'apprenant. L'indexation fragmente les savoirs, le contexte permet de les reconstruire : toute connaissance peut être réutilisée.

Les technologies utilisées permettent de suivre le processus pédagogique de mise en œuvre de ressources pédagogiques (création, organisation, diffusion, utilisation, exploitation) à l'aide de fiches individuelles descriptives des compétences, des savoir-faire, des parcours individuels, des formateurs, de leurs sources d'information préférées, de leurs meilleures recherches documentaires et de décrire des usages, des expériences, liées à des actions situées.
La connaissance partagée devient un capital intellectuel, un langage commun qui permet la compréhension mutuelle, la conscience du savoir des autres, l'apprentissage réciproque, la génération collective d'idées et la résolution collective de problèmes.

Forsic : un projet de recherche

FORSIC prend en compte les processus sociaux liés aux apprentissages. L'étude d'un collectif pensant ainsi que de l'indexation par la compétence est un bel enjeu de recherche, qui permet d'étudier conjointement le lien entre les modes de représentation et la technique, entre la communication et l'organisation sociale, les nouvelles formes sociales et le technique, les nouvelles formes du processus du travail intellectuel dans les environnements distribués et leur impact politique.

La production de ressources pédagogiques

Dans FORSIC, nous utilisons les modèles de la recherche documentaire en tant que discipline, les modèles cognitifs de l'apprenant, que nous avons développés. L'analyse des contenus des ressources pédagogiques faites avec UMAP, le profil des formateurs sont utilisés pour proposer une aide à la rédaction des ressources pédagogiques en fonction des compétences que les étudiants doivent acquérir au cours de leur cursus et des exigences des formateurs. Nous sommes en train d'établir des relations (relations typologiques et relations d'ordre) entre les différents éléments qui constituent chaque palier d'apprentissage.
A partir de la représentation des connaissances, des contenus pédagogiques et des processus cognitifs mis en jeu pour la rédaction de ces ressources pédagogiques et des ressources réalisées par les formateurs, nous allons mettre en œuvre un système d'aide à la rédaction de supports de cours (stratégie pédagogique, progression de l'apprenant et évaluation). Cet aspect fera l'objet de travail de recherche mené par un étudiant de thèse du laboratoire LIHS, Université Toulouse I.

L'animation collective des savoirs avec

L'équipe de recherche SMAC (IMIS, ARCADIA) de l'IRIT (Université Toulouse III) se propose d'utiliser la technologie multi-agent pour faciliter l'accès de tous à une information riche et pertinente sur le réseau Internet.
Une approche adaptative du référentiel pédagogique est indispensable en raison de certains aspects de la formation à la recherche :

- Evolution constante des supports et des outils d'information,
- Progression de l'enseignement des sciences de l'information,
- Renouvellement des besoins des usagers.

La dynamique complexe de FORSIC due à l'évolution des besoins des utilisateurs est liée à la création, à la modification des ressources pédagogiques. L'approche technique retenue est basée sur les principes de l'auto-organisation, les liens entre les composants du système reflètent l'état courant de celui-ci par une architecture par niveau. Un niveau représente les usagers de FORSIC et à ce niveau, les agents ont une attitude sociale de coopération qui leur permet de rechercher l'information pertinente par coopération. Un agent enverra spontanément une information à un autre agent, s'il sait que cet agent est intéressé.

Les agents de médiation (MA) représentent un ensemble de ressources pédagogiques. Ces agents ont une vision plus large bien que locale du réseau obtenue à travers les activités d'apprentissage du système basées sur les croyances issues de l'évaluation des transactions.

Chaque agent possède un système multi-agent auto-organisateur qui correspond au profil de l'utilisateur et qui décrit l'état de FORSIC selon le point de vue qu'un agent peut avoir de lui-même et des autres, mis à jour en permanence au gré des transactions. C'est ainsi que le système peut s'adapter et construire les couplages pertinents par un apprentissage constant qui lui permet d'affiner sa vision du monde. L'utilisateur peut évaluer le système en permanence.
Ce système permettra de créer des ressources pédagogiques, de nouvelles connaissances qui pourront être intégrées dans la base et cartographiées. A terme, nous espérons l'utiliser pour ouvrir le système aux étudiants des campus (auto-formation, recherche d'informations pertinentes).

Le déroulement du projet et les retombées

De Janvier à la fin du mois d'Avril, nous avons mis en place le dispositif, préparé la matrice d'indexation des compétences selon la méthode de Trivium. Nous sommes en possession d' une grille de recueil des compétences. Ce dispositif est mis en place en mai et les premières représentations cartographiques seront consultables dès la fin du mois de juin.

Parallèlement, les deux programmes de recherche développent selon un programme très serré :

- En ce qui concerne l'aide à l'élaboration de ressources personnalisées, nous travaillons activement à la modélisation de la recherche documentaire en tant que domaine, à la modélisation des domaines cognitifs des apprenants et des formateurs,
- En ce qui concerne l'animation du collectif intelligent, les premiers essais seront possibles dès que les premières ressources pédagogiques seront disponibles.

Grâce au projet FORSIC, nous sommes partenaires de la société Trivium dans un autre programme de recherche K 2001 qui vise à améliorer le logiciel SEE-K en partenariat avec le CEA qui est chargé de l'amélioration de l'algorithme et de l'indexation des ressources multimedia. Nous sommes particulièrement chargés de l'indexation des compétences et de l'évaluation des améliorations apportées au logiciel. Un thésard travaille sur la question de l'indexation.

FORSIC, de grandes espérances…

L'idée de la construction collective (tissage de liens) de ressources de formation nous séduit à plus d'un titre. D'une part, elle nous permet d'avoir une meilleure vision de notre environnement et de notre contexte : " augmentation de la réalité ", de mieux gérer l'action de L'URFIST au niveau local. D'autre part, elle nous donne une meilleure couverture des diverses disciplines à cause du caractère pluriel des connaissances des formateurs augmentées par le simple fait qu'elles sont partagées et disponibles. Nous avons le sentiment de participer à l'émergence d'une approche très importante socialement et politiquement à

l'heure de la société de l'information, qui élargit considérablement nos possibilités effectives de formation. Mais c'est surtout sur le plan scientifique que ce travail est très stimulant. Cette approche est très importante dans diverses disciplines et plus particulièrement pour les Sciences de l'Information et de la Communication ainsi que l'ont montré les colloques de la SFIC de Metz en 1998, On Line à Londres et ASIS à Washington en 1999 qui ont porté sur la gestion des connaissances (dont nous proposons un compte rendu: question au cœur des activités documentaires depuis toujours) …

Médiations hybrides

Contribution à l'étude de l'évolution des Sciences de l'information françaises à partir de la revue Documentaliste-Sciences de l'information.

Viviane Couzinet - Maître de conférences LERASS-ICC Toulouse

Quel lien communicationnel Documentaliste-Sciences de l'information (Doc-SI) établit-elle entre la profession et la recherche ?
Cette question a conduit à formuler comme première hypothèse que le trait d'union présent dans le titre de la revue marque symboliquement le lien communicationnel entre un énonciateur -la recherche- et un destinataire -la profession-.
L'étude des genres discursifs des articles produits de 1963 à 1997 par les enseignants-chercheurs montre que ceux-ci transmettent des représentations de la recherche en SI. Par des "exposés de mise au point", ou des "discours scientifique de type pédagogique" ils produisent un discours sur la recherche tendant à la rendre intelligible et présente à la "communauté" des professionnels de la documentation.

Parallèlement des articles scientifiques "originaux" ou "primaires" (Jacobi D., 1992) provenant des enseignants-chercheurs, à partir de 1989, mais en moins grande quantité énoncent des résultats de recherches. Les chercheurs s'adressent aux chercheurs et de leur côté, les professionnels produisent des textes destinés à la profession. Chaque "communauté" -celle des professionnels et celle des chercheurs- s'adresse à sa propre communauté.

La position seconde de SI, dans le titre de la revue, indique l'intérêt que les documentalistes lui portent, sans que celà soit leur préoccupation première. Cette symétrie des positions justifie l'acte de communication et légitime sa mise en oeuvre, les compétences de l'énonciateur et du destinataire sont comparables (Veron E., 1997). La cohabitation de la recherche et de la profession font de Doc-SI une revue à la fois scientifique et professionnelle, celle-ci l'emportant largement sur celle-là.

Mais pour appréhender le phénomène dans sa totalité, une autre hypothèse prenant en compte ses constituants hétérogènes, mais associés, dans leurs interactions a été privilégiée.

Il a été posé que si une communication multidirectionnelle s'établissait, contribuant à donner des représentations de la recherche en SI, mais aussi une "réalité" de cette recherche, alors Doc-SI serait une "revue hybride", l'hybridation supposant, en effet, la présence de deux constituants. La diversité des échanges, constatée dans les articles montre que nous sommes en présence d'un système de communication ouvert où profession et recherche sont en interaction, ce qui leur permet, à des degrés divers, d'y trouver chacune leur compte. La multi-directionnalité des communications se manifeste par la citation des travaux des enseignants-chercheurs -que ceux-ci publient ou non dans cette revue- par les professionnels et par la collaboration de professionnnels à certains travaux de recherche.

De ce fait on peut se demander quelle est la place réelle de Doc-SI dans l'ensemble des revues de sciences de l'information ? Les enseignants-chercheurs de SI, en mal de lieu de publication, proposent leurs articles à des revues professionnelles. Ainsi Doc-SI, de par son mode de fonctionnement et de sélection des contributions, intéresse les chercheurs. Dans ce contexte, elle se trouve, de facto, propulsée revue française de SI, comme on peut le voir dans les répertoires (Loustalet C., Chignard G., Bouquillon P., 1992) dans les thèses soutenues dans la spécialité (Couzinet V., Bouzon A., 1998) ou dans divers écrits d'enseignants-chercheurs, que ceux-ci publient ou non leurs articles dans cette revue. D'ailleurs les articles des chercheurs qu'elle publie bénéficient d'une certaine notoriété nationale et internationale, puisqu'ils sont cités dans les revues des pays francophones, comme celles du Canada ou de Belgique, et en Espagne.

Dès lors, la cohabitation d'articles scientifiques avec des articles professionnels a-t-elle des conséquences sur la recherche en SI telle qu'elle apparaît dans cette revue ? La priorité donnée aux "exposés de mise au point" permet de penser que tout se passe comme si les enseignants-chercheurs étaient plus considérés comme des enseignants que commme des chercheurs. Les thématiques qu'ils développent sont prioritairement des thématiques présentant un intérêt pour les professionnels. Mais avant que la section sciences de l'information et de la communication bénéficie d'une reconnaissance des instances universitaires, la profession, qui estimait que son niveau de compétence rapports destinés aux pouvoirs publics sa volonté de s'appuyer sur une recherche forte. Pour ce faire, au cours de son histoire, elle a établi des liens avec le monde de la recherche.

La prise en compte de cette situation récurrente mettant en cause l'utilité de la profession, notamment lors d'innovations technologiques rendant les usagers de plus en plus autonomes dans leurs recherches d'information de même que la nécessité, pour les chercheurs, de diffuser leurs recherches, en l'absence de revue certifiée dans la spécialité, ont conduit à formuler une hypothèse complémentaire : si le lien communicationnel existe entre profession et SI alors Doc-SI est le "porte-parole de l'intertraduction des acteurs". Autrement dit, l'approche communicationnelle de la revue paraissait opératoire, pour observer la manière dont deux champs interagissent pour résoudre leur problème respectif.

Le cadre de référence pour expliquer le système de relations établi à travers Doc-SI a été emprunté à la sociologie des sciences et des techniques (Callon 1975, 1986 ; Latour 1991a, 1991b). La pluralité des thématiques et la pluralité des genres discursifs, ont été abordés comme le résultat des liens tissés entre la profession de documentaliste et la recherche en sciences de l'information. La théorie de "l'acteur-réseau" et de la "traduction" a ainsi permis

d'expliquer la pérennité de cette revue et la progression de la participation des chercheurs.

Des entités humaines ont en effet contribué à créer des alliances. La place la plus remarquable est celle occupée par Jean Meyriat. "Père fondateur" de la SFSIC et de la section sciences de l'information et de la communication au sein du Conseil consultatif des universités, impliqué à la fois dans des activités de recherche -en tant que directeur d'études à l'École des hautes études en sciences sociales- et dans des activités professionnelles -en tant que directeur des services de documentation de la Fondation nationale des sciences politiques-, il joue un rôle important dans le rapprochement de la profession et de la recherche. Mobilisant autour de lui des personnes et des organisations dans lesquelles il est investi, il fait figure d"acteur-réseau", et ce d'autant plus qu'il est pendant un temps à la fois président de la SFSIC et président de l'ADBS.
Appartenant aux deux mondes en présence, il est aussi représentatif de l'hybridation entre la recherche et la profession. La profession, de son côté, met en place des "dispositifs d'intéressement" contribuant à rassembler autour d'elle des entités organisationnelles tant au niveau national qu'international. Le réseau qu'elle constitue ainsi lui permettant de s'affirmer comme ayant acquis un certain niveau de compétences.

L'originalité de Doc-SI, semble fortement dépendre de la capacité de la profession à créer de nouveaux liens avec la recherche. Ceux-ci confortent les relations déjà nouées et visent à "intéresser" les enseignants-chercheurs, encore absents de la revue. L'ouverture des éditions ADBS, puis l'ouverture, sous la direction d'un enseignant-chercheur, d'une collection chez un éditeur universitaire en sont des exemples.

Cette volonté permanente de maintenir les liens existants et d'en développer d'autres ne se fait pas sans résistances. Des "controverses" se font jour aussi bien du côté des chercheurs, qui considèrent que la revue est trop professionnelle, que des professionnels qui la trouvent trop difficile d'accès. Ces résistances et ces ruptures font bouger, au cours du temps, le tracé de la place occupée par les uns et par les autres au sein des réseaux. Cet état de tension permanente a pour résultat une production hybride à mi-chemin entre la production de recherche et l'expertise que l'on peut percevoir dans les articles de "mises au point".
La revue Doc-SI, placée au centre d'une nébuleuse de réseaux devient ainsi le "porte-parole" de la profession et de la recherche. Elle même tisse un ensemble d'alliances avec d'autres revues de la spécialité ou d'autres sciences dans un espace géographique large, comme le montre le paratexte des articles. Par leur intermédiaire, elle établit un lien, aujourd'hui encore fragile entre la recherche française et la recherche étrangère. Ses articles montrent le niveau de compétence acquis par les documentalistes en même temps qu'ils contribuent à leur perfectionnement et à leur positionnement social et professionnel.

De la même manière en proposant les résultats de leurs travaux, les chercheurs essaient de résoudre, tant bien que mal, le problème auquel ils>sont confrontés : celui de la nécessité de publier pour être reconnus.

Cependant la reconnaissance scientifique passe par la publication dans des revues certifiées. Or si l'on peut trouver dans Doc-SI les critères énumérés par Robert Boure (1993) qui font qu'une revue peut être considérée comme scientifique sa contribution au débat reste encore faible, même si ce phénomène se retrouve, à des degrés divers, dans nombre de revues.

les enseignants-chercheurs de SI se trouvent donc être dans une situation inconfortable : contraints de publier ils ne disposent pas d'un média pouvant répondre à leurs besoins eu égard aux moyens d'évaluation qu'utilise la recherche. De la sorte leurs pratiques scientifiques peuvent paraître éloignées de celles qui sont généralement en vigueur dans nombre de disciplines : la recherche en équipe est peu développée, les références aux revues étrangères du champ sont éparpillées, le recours aux communications dans des colloques professionnels est courant. L'absence de revue scientifique à comité de lecture dans la spécialité est un problème majeur car elle prive ces enseignants-chercheurs de la confrontation scientifique avec leurs pairs, confrontation qui contribue à la formation du chercheur. La revue a en effet non seulement une dimension informative - elle permet de publier les résultats des travaux entrepris- mais encore une dimension formative (Latour B., 1991c) qui permet aux chercheurs, et à la recherche dans une discipline, de continuer à progresser. Cette question récurrente qui à terme, vu la croissance du nombre de chercheurs, se posera aussi à l'ensemble de la section, pourrait trouver sa solution dans la création d'une revue, à comité de lecture, associant sciences de l'information et sciences de la communication. Celle-ci contribuerait notamment à l'enrichissement mutuel, à la collaboration fructueuse d'équipes et à mettre en visibilité, au niveau international, cette association spécifiquement française des Sciences de l'information et des Sciences de la communication.

Références.

- BOURE Robert (1993). Le territoire incertain des revues scientifiques. Réseaux, n°58, mars, avril, p. 93-105.
- CALLON Michel (1975). L'opération de traduction comme relation symbolique In Incidence des rapports sociaux sur le développement scientifique et technique, sous la dir. de Philippe Roqueplo et de Pierre Thuillier. Paris : CNRS ; CORDES, p. 105-141.
- CALLON Michel (1986). Eléments pour une sociologie de la traduction : la domestication des coquilles Saint-Jacques et des marins-pêcheurs dans la baie de Saint-Brieuc. L'année sociologique, p. 169-207.
- COUZINET Viviane, BOUZON Arlette. Les revues de SIC dans les thèses de la discipline. Réunion de travail du groupe TPS (SFSIC) septembre 1998.
- JACOBI Daniel (1992). Discours de vulgarisation. In Dictionnaire critique de la communication sous la dir. de Lucien Sfez. Paris : PUF, Tome 2, p. 1468-1474.
- LATOUR Bruno (1991a). Nous n'avons jamais été modernes : essai d'anthropologie symétrique. Paris : La Découverte, 210 p.
- LATOUR Bruno (1991b). Les représentations de la science : une cohérence improbable. In L'état des sciences et des techniques sous la dir. de Nicolas Witkowski. Paris : La Découverte, p. 98-100.
- LATOUR Bruno (1991c). Le royaume de l'écrit scientifique. In L'état des sciences et des techniques sous la dir. de Nicolas Witkowski. . Paris : La Découverte, p. 441-443.

- LOUSTALET Colette, CHIGNARD Geneviève, BOUQUILLION Philippe (1992). Petite bibliothèque de la communication. Cinémaction, mars, n°63, p. 246-254.
- SFSIC (Société française des sciences de l'information et de la communication) 1994. Les fondateurs de la SFSIC : Jean Meyriat. 16 p.
- VERON Eliseo (1997). Entre l'épistémologie et la communication. Hermès, n°21, p. 25-32

JURISGUIDE

Outil d'orientation pour la recherche d'informations juridiques

Jean-Emile Tosello-Bancal, Conservateur en Chef - URFIST de Paris-ENC

Le JURISGUIDE fait partie des projets pédagogiques nés à l'initiative de la Direction de l'Enseignement Supérieur (Sous-Direction des Bibliothèques et de la Documentation) pour faciliter l'usage des ressources documentaires dans le cadre des études universitaires. JURISGUIDE est connu comme un outil d'auto-formation à l'utilisation des ressources documentaires en droit, destiné principalement aux étudiants en premier cycle de sciences juridiques, en privilégiant la documentation accessible sur supports électroniques.

Au terme de son achèvement prévu en janvier 2001, le JURISGUIDE sera accessible sur la plate-forme FORMIST hébergée sur le serveur WEB de l'ENSSIB, une version miroir sera également disponible sur les sites WEB des institutions ayant contribué à la réalisation du projet. Dans sa phase actuelle de montée en charge, le JURISGUIDE est provisoirement hébergé sur le site de l'URFIST de Paris-Ecole Nationale des Chartes: http://www.ccr.jussieu.fr/urfist/JURISGUIDE.htm

2. OBJECTIFS DU PROJET

Le JURISGUIDE devra permettre aux étudiants en sciences juridiques

- d'identifier les différents types de ressources documentaires propres à leur discipline
- de s' initier aux méthodes de recherche d' informations sur les principaux supports opto-électroniques disponibles (cédéroms, banques de données télématiques et ressources disponibles sur l'Internet) et sur support papier
- d'apprendre à construire une stratégie de recherche documentaire à partir de modèles pré-établis (parcours-types définis en fonction d'objectifs de recherche précis).

Le JURISGUIDE doit également favoriser la mise en commun des supports de cours et des outils d'aide au repérage documentaire actuellement épars en milieu universitaire, afin de fournir à la communauté de l'enseignement supérieur un dispositif fédérateur de ressources propre à faciliter et démultiplier les activités pédagogiques et la recherche dans le domaine

de la documentation juridique.

3. CONTENU

Le JURISGUIDE propose trois types d'entrées pour la découverte des ressources documentaires juridiques:

1. une entrée par types de supports regroupés autour des *"Incontournables"* : i.e. une centaine d' "outils papier" -revues et ouvrages de références- constituant le socle des recherches documentaires en particulier au cours des premières années d'étude en droit,
   Cédéroms (plus d'une centaine de titres présentés, avec pour une trentaine d'entre eux -parmi les plus fréquemment proposés aux étudiants en bibliothèque universitaire- des modes d'emploi très détaillés sous forme de copies d'écran expliquant pas à pas les stratégies de recherche
   Banques de données: essentiellement celles proposées par les principaux diffuseurs français (ORT et Lamy)
   Ressources Internet: en proposant un guide de recherche d'information (présentation des annuaires, moteurs de recherche, etc..) couplé à une sélection commentée de sites juridiques accessibles sur le WEB.
2. une entrée par domaines du droit (droit civil, droit européen....) ou/et par sources du droit (législation/réglementation, doctrine, jurisprudence) est également possible. Elle permettra autour d'un thème donné (droit européen par exemple) de repérer les ressources documentaires disponibles tous supports confondus gr,ce à une présentation sous forme de tableaux de synthèse (avec liens sur les fiches descriptives des outils recensés)
3. une entrée par objectifs de recherche, proposant des éléments de méthodologie juridique et des conseils pour la recherche d'informations en droit, sous la forme de parcours de découverte balisés (ex : réalisation d'un dossier sur la notion de responsabilité du journaliste , comment cerner le concept de responsabilité, quelles notions lui associer, comment retrouver les textes de références, comment les exploiter?).

Pour chacune de ces trois entrées deux niveaux minimum de description seront prévus†:

1. présentation synthétique des ressources concernées, sous la forme de listes ou de tableaux à double entrée permettant de repérer un outil documentaire -Cédérom, Banque de données, etc?- pour une catégorie donnée (domaine du droit ou type de source)
2. présentation détaillée des principaux titres disponibles pour chaque type de support, sur la base d'une structure descriptive commune, avec trois niveaux minimum de profondeur:
   1. fiche d'identité du document (contenu, producteur, accessibilité, prix...),
   2. fiche technique (mode d'utilisation: explicitation des commandes),
   3. exemple d'interrogation (reproductions d'écrans ou de pages, assorties de commentaires explicatifs†précisant les modalités de traitement d'une recherche

sur l' outil concerné).
Sur ce dernier point, en complément des documents pédagogiques réalisés par les producteurs du JURISGUIDE, une coopération est actuellement à l'étude avec certains éditeurs du domaine -Dalloz, La Gazette du Palais- pour développer des modules de simulation d'interrogation (écrans animés, permettant de mieux rendre compte de l'interactivité d'une séance d'interrogation qu'une suite séquentielle de copies d'écran fixes...).

Le recours aux techniques multimedia est également prévu dans d'autres parties du JURISGUIDE pour faciliter l'approche pédagogique (animation d'écrans notamment dans la partie concernant les cours de méthodologie pour la recherche d'information, afin de permettre -par exemple- l'incrustation ponctuelle d'écrans de commentaires ou de définitions précisant les concepts abordés dans le cours).

4. REALISATION

L'organisation d'ensemble de l'outil et la production des documents sont réalisés en réseau par un groupe de spécialistes de la discipline, enseignants-chercheurs et professionnels de l'information (conservateurs de bibliothèques, bibliothécaires et documentalistes), avec l'aide d'étudiants juristes de différents niveaux (2nd et 3ème cycles), la coordination administrative et budgétaire du projet étant assurée par la bibliothèque inter-universitaire de CUJAS et la coordination éditoriale par l'URFIST de Paris-ENC.

La structure du JURISGUIDE, organisé en parties indépendantes, complémentaires les unes des autres, facilite la répartition du travail en réseau: chacune de ces parties est coordonnée par un des membres de l'équipe de projet. Le travail portant sur les liens internes entre les différentes entrées du JURISGUIDE de mÍme que le lissage des avancées du projet -variables selon les entrées considérées- est assuré lors de réunions bi-mensuelles de l'ensemble du groupe de projet, l'essentiel de la production étant effectué par transferts de données sur l' Internet.

JURISGUIDE propose également de nombreux liens externes sur d' autres projets pédagogiques complémentaires (CERISE en particulier, dévolu à l'acquisition de connaissances documentaires de base pour les étudiants de toutes disciplines...) et sur des sites remarquables dans le domaine juridique en privilégiant les sites francophones et européens (notamment les sites institutionnels pouvant fournir des renseignements aux étudiants juristes pour l'organisation de leurs parcours d'études universitaires).

Renseignements complémentaires:
(présentation de la maquette du projet, information sur son avancement actuel, coopération éventuelle...)
Jean-Emile Tosello-Bancal, URFIST de Paris (jetb@moka.ccr.jussieu.fr)
ou Anny Maximin, BIU CUJAS (amaximin@univ-paris1.fr)

Co-producteurs du JURISGUIDE:

- Michèle BATTISTI (ADBS)
- Lucienne DEDIEU-URIOS (Bibliothèques de l'Université des Sciences sociales Toulouse 1)
- Marie-Cécile ESCANDE-VARNIOL (Université LYON 3)
- Sabine GUILHOT DE LAGARDE (SCD LYON 3)
- Lise HERZHAFT (URFIST LYON)
- Anita LARGOUET (URFIST BORDEAUX)
- Annie LEON (Mediadix-Université Paris X)
- Jean-Jacques LEVIVE (BU PARIS 11)
- Anny MAXIMIN (BIU CUJAS)
- Christine POITEVIN (Université PARIS 8)
- Jean-Paul SCARANO (Université LYON 2)
- Jean-Emile TOSELLO-BANCAL (URFIST PARIS-ECOLE NATIONALE DES CHARTES)

BIOGUIDE

Un dispositif d'autoformation pour l'Information Scientifique et Technique en Biologie : bilan après 16 mois de mise ligne

C. Boudry, Maître de conférences et G. Charton, Maître de conférences URFIST de Paris-ENC

La recherche scientifique est fortement associée au processus de collecte et de diffusion d'informations scientifiques et techniques. La croissance importante du nombre de sources d'informations amène au constat suivant: les chercheurs ont de plus en plus besoin d'être guidés et formés pour mener à bien la recherche et la diffusion des informations indispensables à leurs travaux de recherche.
Dans ce contexte, cinq raisons au moins justifient la mise en place d'un dispositif pédagogique sur internet visant à faciliter la recherche et la diffusion d'informations en biologie:
La première vient du fait que, dans leur grande majorité, les étudiants en 3ème cycle universitaire et les chercheurs en biologie n'ont pas bénéficié dans leur cursus universitaire d'enseignements spécifiques visant à la maîtrise des usages de l'information scientifique et technique, les efforts ayant d'abord été globalement tourné vers les premiers cycles universitaires [1]. La deuxième est que la formation à l'information scientifique des étudiants en troisième cycle mais également des chercheurs en exercice s'effectue le plus souvent au fur et à mesure de leurs besoins, la priorité étant par ailleurs donnée aux formations relatives à la maîtrise des outils et techniques complexes utilisés spécifiquement

en biologie. La troisième raison vient de la difficulté de former par les méthodes d'enseignement classique les étudiants de troisième cycle (difficulté due à leur éparpillement dans les laboratoires de recherche) et les chercheurs. La quatrième est le faible nombre de supports spécifiquement dédiés à la recherche d'informations scientifiques et techniques en biologie [2,3], un grand nombre étant dédiés à la médecine [4,5] ou non rattachés à un champ disciplinaire particulier [6,7], et ceci malgré la très forte croissance des sources d'informations disponibles en biologie [8]. La cinquième est la souplesse inhérente à ce type de support qui permet à chaque chercheur de se former selon ses propres besoins en fonction des connaissances qu'il a déjà acquises sur le sujet.

Toutes ces raisons nous ont amené à concevoir "Bioguide" [9] dont l'objet est de faciliter l'accès des chercheurs et étudiants de 3ème cycle universitaire en biologie à l'information scientifique et technique http://www.ccr.jussieu.fr/urfist/biolo/frame.htm. Le but du présent article est de présenter ce support, de donner quelques éléments d'appréciation de son usage mais également de discuter, de l'intérêt de la mise en place d'un tel support, et des difficultés que génère sa maintenance et son suivi sur une longue période.

CONTENU DU SUPPORT POUR LA RECHERCHE ET LA DIFFUSION D'INFORMATIONS EN BIOLOGIE

Afin de cerner au mieux les processus de collecte et de diffusion d'informations auxquels doivent faire face chercheurs et étudiants de 3ème cycle en biologie lors de leurs travaux, les différentes étapes du travail de recherche en biologie ont été analysées (observation de l'existant, travail de recherche, diffusion du travail de recherche). Les flux d'informations associés à ces étapes ont ainsi pu être clairement identifiés. Des outils, conseils ou ressources ont été regroupés pour que les chercheurs en biologie puissent maîtriser au mieux ces flux d'informations.
Le contenu de ce support de cours est aussi bien d'ordre méthodologique (par exemple des conseils pour gérer sa bibliographie ou pour maîtriser les outils de communication utiles aux chercheurs tels les forums de discussion) que technique (obtenir des logiciels ou matériels de laboratoire utiles en biologie) et pratique (trouver l'adresse électronique d'un correspondant dont on ne connaît que le nom).

Ce support est basé en grande partie sur l'utilisation des outils et ressources internet (mais pas exclusivement). Il ne s'agit pas d'un répertoire de ressources (dont le but est classiquement de présenter un grand nombre de ressources), mais de présenter des ressources "à valeur ajoutée" dans le sens où chaque ressource présentée a été sélectionnée pour sa pertinence, ses points faibles et forts sont commentés.

Cinq modules ont été mis en place et sont directement accessibles à partir de la page d'accueil.

1. "Réunir les informations nécessaires avant de débuter un travail de recherche"
   Il s'agit là de donner des conseils mais également de proposer une méthodologie pour

effectuer une recherche bibliographique rétrospective pour collecter toute information (quel qu'en soit le support: livre, article scientifique, littérature grise...) se rapportant au travail de recherche que le chercheur doit effectuer. Le contenu de 17 bases de données bibliographiques utiles aux biologistes est détaillé et une initiation à l'interrogation de ces bases est proposée. Un guide d'utilisation est également disponible pour les plus connues d'entre elles (Medline, article@inist, Uncover). Cette partie propose également des conseils pour gérer sa bibliographie.

2. "Suivre les actualités dans son domaine de recherche"
   Ce module met l'accent sur les services de veille informationnelle qui permettent aux chercheurs de suivre l'actualité dans leurs domaines scientifiques. Par exemple, des liens vers les ressources internet sont proposés pour repérer de façon systématique de nouveaux ouvrages ou articles scientifiques nouvellement parus. De même, des liens vers les pages "Nouveautés" des principaux organismes de recherche sont proposés.
3. "Obtenir une information précise ou des données en rapport avec son travail de recherche"
   Ce module a pour objectif de proposer des pistes concrètes pour effectuer son travail de recherche au quotidien : se procurer du matériel de laboratoire, des logiciels, des images ... Des liens vers des bases de données factuelles ou des sites les répertoriant sont également proposés.
4. "Avoir des contacts avec les autres chercheurs"
   Ce module concerne l'aspect relationnel du travail de recherche. Il propose des outils spécifiques pour localiser des correspondants ou laboratoires de recherche, présente les principaux outils de communication et d'échanges de données utiles aux chercheurs et propose également des pistes pour localiser des congrès.
5. "Diffuser son travail de recherche"
   Ce module suggère des solutions pour valoriser son travail de recherche par le biais des congrès, publications scientifiques ou bien encore par le réseau internet.

Les utilisateurs disposent également d'un "guide" pour tenter d'optimiser leur utilisation de bioguide, ainsi que d'un index des termes utilisés. Chaque utilisateur a ainsi la possibilité de parcourir de façon séquentielle les 5 modules pour suivre l'intégralité du cours, ou de se rendre de façon spécifique à la page qui traite du sujet qui l'intéresse en utilisant l'index.

MOYENS TECHNIQUES MIS EN OEUVRE

Bioguide comporte un certain nombre de fonctionnalités dont l'objet est non seulement de créer une certaine interactivité mais également d'en augmenter la lisibilité et la compréhension. Pour aller dans ce sens, le langage javascript a été utilisé. Il faut cependant noter que l'incompatibilité du langage javascript et les navigateurs d'ancienne génération (versions 3 et en dessous) a nécessité le développement en parallèle d'une version sans javascript (de nombreuses remarques d'utilisateurs allaient dans ce sens). Cela alourdit d'autant le travail de mise à jour.
La réalisation d'icônes spécifiques a également permis la création d'une identité propre au

support et évité d'être confronté aux problèmes de droits d'auteurs concernant les images.

ELEMENTS D'APPRECIATION DES USAGES ET DE LA SATISFACTION DES USAGERS

Des statistiques de fréquentation ont été exploitées sur deux périodes. Il faut préciser que les chiffres concernant la fréquentation de ce support présentés sont bien entendu à mettre en perspective avec la grande spécificité de ce support, ce qui en limite d'emblée grandement l'audience. Les premières données couvrent la période du 15 novembre 1998 (date de la première mise en ligne du support) au 11 février 1999 et ont pour objet de dénombrer le nombre de connexions sur bioguide sur cette période et de déterminer le nombre moyen de visites par utilisateur. Pour cela, des scripts en Shell (Unix) ont été implémentés sur le serveur Web accueillant la page d'accueil de bioguide, afin de recueillir des informations sur la provenance des utilisateurs via le numéro IP de leur machine, l'heure et la date de leurs connexions. Plusieurs précautions ont été prises, notamment celle de ne pas tenir compte des connexions "parasites" provenant de l'Urfist de Paris (en tout 249 sur la période) et les connexions à la page d'accueil espacées de moins d'une heure pour un même utilisateur, ceci afin d'éviter de compter chaque passage par la page d'accueil lors d'une même session de travail comme autant de visites distinctes de la part d'un utilisateur (823 connexions ont ainsi été écartées de notre étude).

Ces précautions prises, 1323 connexions ont été comptabilisées lors la période d'étude (sur un total brut de 2395, ce qui représente près de 45 % de connexions "parasites"), soit 15 connexions par jour.

En ce qui concerne le nombre moyen de visites par utilisateur, il est en moyenne de 1,61. Près d'un quart des utilisateurs (25,71 %) sont revenus sur bioguide au moins 1 fois après leur première visite tandis que près de 74,29 % ne sont pas revenus sur bioguide pendant la période d'étude (Figure 1).

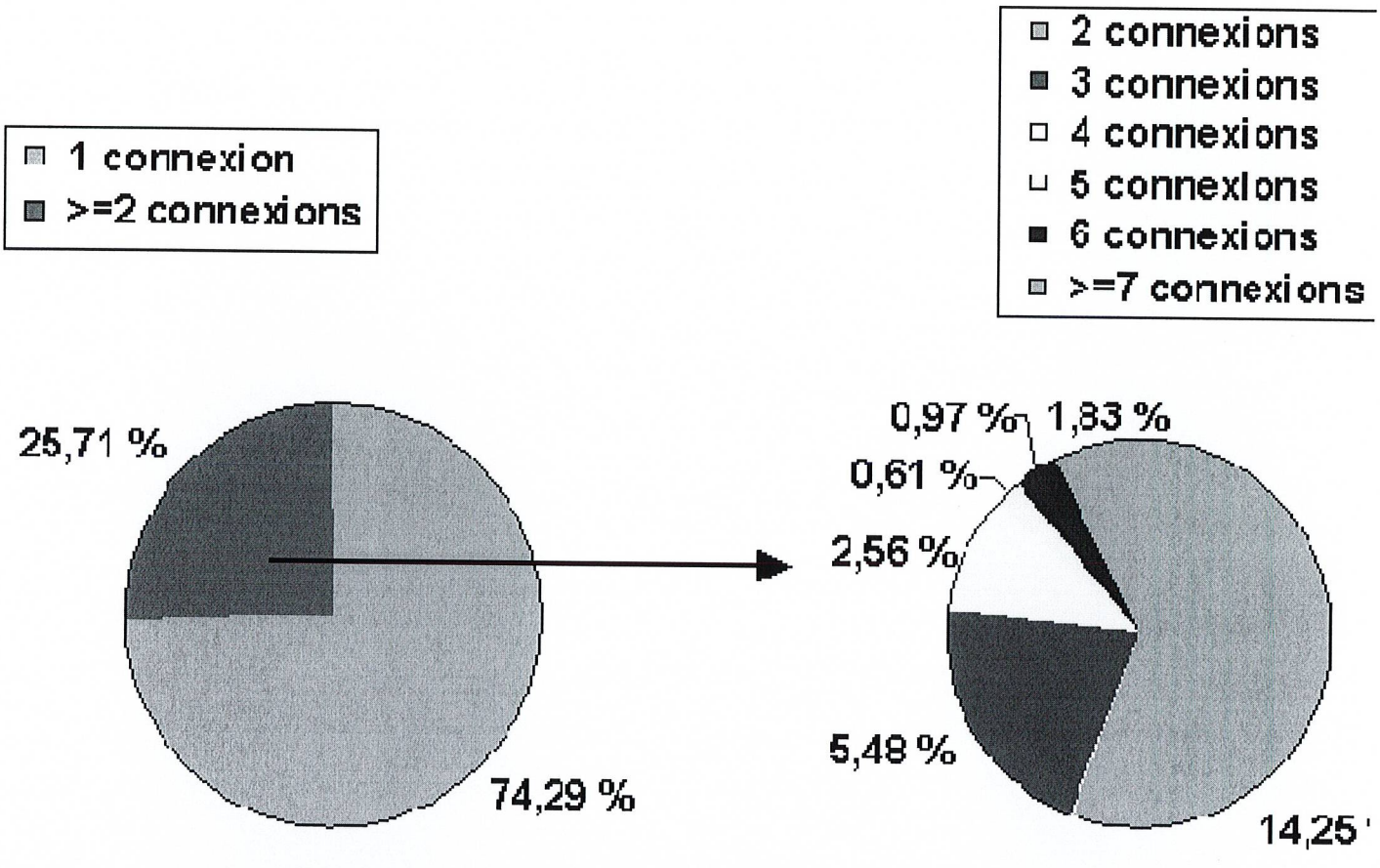

Figure 1 : Pourcentage d'utilisateurs ayant généré de 1 à plus de 7 connexions.

Les autres données proviennent de l'étude de la fréquentation sur la période du 14 janvier 2000 au 30 mars 2000, soit de 14 à 16 mois après la première mise en ligne du support. Pour des raisons pratiques et notamment pour éviter une fastidieuse exploitation des résultats, nous avons fait appel à un site spécialisé dans la fourniture de statistiques [10]. 1271 connexions ont été enregistrées durant cette période, soient près de 17 connexions par jour. La fréquentation enregistrée pour cette période ne diffère donc pas significativement des chiffres obtenus 16 mois auparavant, ne mettant pas en évidence un délaissement du support avec le temps.

Concernant le degré de satisfaction des usagers, il faut noter qu'un très faible nombre de suggestions ou de remarques visant à critiquer le contenu ou la forme de bioguide ont été recueillies via le formulaire de satisfaction mis en place à cet effet. Ceci ne signifie pas qu'il n'existe pas d'utilisateurs insatisfaits, mais est inhérent au principe même du réseau internet, qui fait que si un utilisateur ne trouve pas une information sur un site, il va vite la chercher ailleurs, sans se donner la peine de formuler son insatisfaction au concepteur du site en question. En conséquence, il est très difficile d'obtenir un retour de la part des usagers de ce type de support.

MAINTENANCE DU SUPPORT DANS LE TEMPS

Pour maintenir un taux de fréquentation relativement constant, il est indispensable que ce

type de support soit mis à jour, tant du point de vue de sa structure que de son contenu, ce qui nécessite un investissement en temps important sur le long terme. Ainsi, par exemple, depuis la date de première mise en ligne, près de 20 % des 1100 liens externes mis en place à l'origine se sont révélés depuis "en erreur", induisant un travail de repérage et d'élimination de ces liens erronés mais également un travail afin de les remplacer par des "équivalents".

CONCLUSIONS ET PROSPECTIVE

Alors que la collecte et la diffusion d'informations devrait bénéficier d'une grande partie de l'attention des chercheurs, par nécessité, une majorité d'entre eux (essentiellement les étudiants de troisième cycle qui sont soumis à une compétition croissante en terme de production scientifique) est contrainte de réduire cette partie de son activité au profit d'activités plus directement "productives" scientifiquement. Dans ce contexte et à la condition que le contenu proposé évolue en permanence (activité de veille et réponses aux remarques des utilisateurs, ce qui signifie un investissement en temps important de la part des auteurs), il semble que le type de support décrit ici réponde de façon assez satisfaisante aux attentes des étudiants de troisième cycle et aux chercheurs en biologie (la stabilité du nombre de connexions enregistré sur 16 mois semble en attester). Ce type de support permet en effet de se former à distance et de façon personnalisée à l'usage et à la maîtrise de l'information scientifique. Il permet également de consulter des ressources commentées et sélectionnées pour leur pertinence afin de répondre aux besoins de collecte et diffusion d'informations des chercheurs en biologie, et assure ainsi un rôle de guide [11] dans la mine d'informations aujourd'hui mises à leur disposition.

REFERENCES

1. Bretelle-Desmazieres Danièle, Coulon Alain et Poitevin Christine: Apprendre à s'informer : une nécessité. Evaluation des formations à l'usage de l'information dans les universités et grandes écoles française, Université de Paris 8, Laboratoire de Recherches Ethnométhodologiques, 1999.
2. B. Didier. Mai 1999 (page consultée le 8 septembre 1999). RIBI (Recherche d'Information Bio-médicales sur Internet). [En ligne]. http://www.pasteur.fr/units/biblio/ribi.html.
3. D. Aymonin et J. Reibel. Octobre 1997 [modifié en mai 1999] (page consultée le 8 septembre 1999). Chimie et Biologie sur Internet. Panorama des ressources. [En ligne]. http://www-scd-ulp.u-strasbg.fr/urfist/internet_chimie_biologie/principa.htm.
4. F. Baudic, J. Piot et B. Thirion. Septembre 1999 (page consultée le 8 septembre 1999). Bibliothèque médicale du CHU de Rouen. [En ligne].

http://www.chu-rouen.fr/documed/bibchu.html.

5. J.P. Lardy. Mars 1999 (page consultée le 8 septembre 1999). Techniques d'accès à la documentation médicale. [En ligne]. http://urfist.univ-lyon1.fr/med_rdi.html.
6. CREPUQ. Janvier 1998 (page consultée le 8 septembre 1999). GIRI 2 - Guide des indispensables de la recherche dans Internet. [En ligne]. http://www-scd-ulp.u-strasbg.fr/urfist/DOCU/HTML/GIRI/giri2/index.html.
7. M. Jolly et J.M. Mermet. Septembre 1999 (page consultée le 8 septembre 1999). Sapristi: Sentiers d'Accès et Pistes de Recherche d'Informations Scientifiques et Techniques sur l'Internet ! . [En ligne]. http://csidoc.insa-lyon.fr/sapristi/digest.html.
8. L. Lomme. L'information électronique en biologie moléculaire. Documentaliste-Sciences de l'Information. 1998. Vol 35, 179-185.
9. C. Boudry. Juillet 1999 (page consultée le 30 mars 2000). Bioguide. Recherche d'informations en biologie. [En ligne]. http://www.ccr.jussieu.fr/urfist/biolo/frame.htm.
10. http://be.netsat.be
11. C. Boucheix : L'INSERM et les réseaux informatiques. INSERM Actualités, 1993. n°115, 9-10.

Certaines données de cet article ont été publiées dans un ouvrage intitulé "Chantiers, publics et métiers de l'enseignement à distance au seuil du troisième millénaire" édité par le Centre National d'Education à Distance dans la Collection "Recherche".

METADONNEES

Usages actuels et perspectives

Julia Jumeau, Professeur de Documentation - URFIST de Paris-ENC
sous la direction de Ghislaine Chartron

Le terme de métadonnées désigne les données incluses dans les fichiers informatiques pour fournir les informations sur des ressources électroniques. En terme de documentation, ce sont des informations secondaires apposées à des ressources primaires. Cette définition est reprise à partir de l'article "Identification des ressources sur Internet et métadonnées : diversité des standards" de Catherine Lupovici (1). Les métadonnées intégrent de plus en plus ces deux concepts. Elles ont à gérer tant l'affichage que l'indexation des ressources électroniques dans un dispositif à géométrie variable capable de satisfaire tel système de gestion de données en ligne , l'interrogation de l'information via un intranet ou l'usage de robots généralistes. Le World Wide Web Consortium (W3C), organisme international organisant le fonctionnement du web préconise l'utilisation de métadonnées destinées à faciliter les besoins commerciaux des différents acteurs du "marché de l'information", les vendeurs et fournisseurs d'information, ce qui aboutit au Resource Description Framework

(RDF). Le Dublin Core, issu de ces travaux, est un dispositif destiné plus particulièrement aux musées, bibliothèques, sites gouvernementaux. Simples et souples, ils se veulent lisibles par les machines, compréhensibles par les humains. Comprendre l'utilité des métadonnées est un premier pas vers une meilleure extraction de l'information sur internet.

1 - Les métadonnées au commencement du langage HTML

Actuellement règne le désordre : une recherche quelconque sur le web peut apporter des signalements de documents en très grande quantité et surtout de qualité disparate ; pour autant cette recherche ne portera que sur 20 à 30% des informations disponibles sur le web, d'après les études de Greg Notess (2), laissant ainsi dans l'oubli bien des documents qui auraient pu être utiles. Cependant, pour obtenir des réponses aussi adaptées que possible aux requêtes, les " concepteurs de robots de recherche n'ont de cesse de peaufiner des algorithmes pour intégrer les paramètres de pondération : indices de popularité, de pertinence, de densité, dédoublonnage des pages , etc... susceptibles de permettre l'affichage des résultats les plus appropriés. Le renseignement des zones affectées à la description des métadonnées des fichiers html, entre les balises *<head> et </head>* est une solution minimale, prévue dans le code HTML, mais peu usitée par les auteurs de pages. Une reqête sur les mots "metadata métadonnées" effectuée le 30 mars 2000 à l'aide de Copernic sur le web français a rapporté 61 réponses dont 11 sans titres. De plus les robots ne les utilisent pas ou très peu. Actuellement les métadonnées servent plus sûrement à renseigner le lecteur d'une page internet, en utilisant l'onglet " affichage, sources de la page", du navigateur utilisé (3). Il serait bien préférable, d'ailleurs, que ces renseignements (auteur, institution, dates de publication et de mise à jour, adresse url de la page), qui sont de première utilité, figurent en clair sur les pages elles-mêmes.

L'affichage d'une page suite à une question posée à un robot ne nous renseigne pas sur sa validité, son contenu, sa taille, son auteur, tous éléments d'information que les communautés savantes ont l'habitude de lire dans les catalogues non numériques des bibliothèques. Les 15 éléments de " metadata " du Dublin Core, à insérer entre les balises *<head> et </head>* constituent un début de réponse (4). Malheureusement, actuellement, seules quelques communautés scientifiques les utilisent ou les affinent, telles le Ministère de la Culture ou le CISMEF de Rouen qui y intègre le MeSH (5), thésaurus très utilisé en médecine. Même si actuellement les robots courants ne se servent pas du tout ou très peu des éléments descriptifs du Dublin Core, on peut penser que cet usage va se développer. Les métadonnées rendent de grands services dans les réseaux internes pour retrouver facilement de l'information. Pour cela, les communautés universitaires doivent faire l'effort de les insérer dans leurs pages. Les autres communautés en découvrent la nécessité avec le développement des intranets et chaque fois que leurs usagers ont à se servir des outils de recherche. Partager l'information nécessite de pouvoir la retrouver facilement, donc d'avoir recours à une indexation de l'information pertinente pour une communauté donnée.

2 - Les métadonnées et les catalogues de bibliothèques

Il faut aussi assurer une continuité avec les habitudes prises par les usagers lors des consultations des catalogues de bibliothèques - catalogues utilisant des formats de description des données non unifiés sur l'ensemble de la planète, formats MARC, USMARC, ... et des descripteurs issus de thésaurus différents, le tout en différentes langues et différents caractères. Dans cette approche, on a pu observer le projet CORC d'OCLC (6). Ce projet permet à des bibliothécaires d'évaluer des sites et de mettre en commun des ressources numériques remarquables pour leur contenu scientifique, en juxtaposant les champs de description des sites en plusieurs formats d'indexation. Il s'agit alors d'éléments descripteurs analogues aux fiches bibliographiques, mais décrivant des ressources internet et associés aux contenus. Ces éléments " multi formats", peuvent ensuite être intégrés aux ressources locales des catalogues numériques de bibliothèques, chaque bibliothèque utilisant alors le format dont elle a besoin. Une souplesse permise par la technologie XML qui est utilisée en arrière plan du projet CORC. Il faut également penser, quel que soit le format d'échange de données utilisé, quel que soit le thésaurus ou la liste d'autorité utilisés, à la possibilité de modifier, actualiser, effacer ou archiver à un autre moment.

3 - Les métadonnées : les besoins des usagers et les perspectives

On pourrait souhaiter que pour un document donné on puisse automatiser son classement de façon à ce qu'il soit retrouvé :

- lors d'une requête en se servant d'un moteur généraliste,
- en formulant une question dans une banque de données soumises à des règles d'accès particulières (droits d'accès limités à certaines catégories de personnes, informations sécurisées, etc..),
- en formulant une requête prenant en compte la validité de l'information,
- en considérant également sa viabilité dans le temps (accès à des systèmes d'archivages).

De nombreuses questions se posent :

- Comment assurer la maintenance des signalements bibliographiques, comment automatiser la mise à jour des liens après la réorganisation d'un site, sa réplication, l'archivage ou la disparition calculée des informations apportées ? D'autres éléments peuvent venir s'ajouter à ceux-ci : les normes d'affichage des caractères, des formats de données (textes, images fixes ou non, son, les taux de compression, la combinaison de ces éléments etc...), la possibilité de pouvoir extraire de l'information ciblée sur un petit téléphone portable, ou bien à l'aide d'appareils pour malentendants ou malvoyants, l'emploi de formats de données propriétaires (*.doc, *.gif, *.pdf, etc...) ou libres (*.txt, *.png, *.htm etc..), la possibilité de modifier ou non les métadonnées,

de les actualiser, voire de décider de la disparition ou de la conservation ultérieure du document après telle ou telle date (sorte de date de peremption).

- On pourrait souhaiter également une gestion plus élaborée de la localisation du document, comme le préconise le Digital Object Identifier, le DOI (7), ce qui permettrait de prendre en compte le nombre et la localisation des alias ou copies ou sites miroirs, dans l'idée de gérer au mieux les droits des auteurs et surtout des éditeurs. L'existence d'un identifiant donné pour un élément d'information déterminé permettrait le suivi, la traçabilité du document, quel que soit son support et son format. Mais qui déciderait de l'unité minimale de document à identifier ? L'auteur ? L'éditeur ? Un organe neutre ? En fonction de quels critères ?
- Ces métadonnées - qui pour être utilisables sur un maximum de configurations possibles, doivent être normalisées ou standardisées, doivent-elles être renseignées à la source, par l'auteur du document qui devra alors extrapoler toutes les utilisations ultérieures avec tous les oublis ou les exagérations que l'on peut imaginer ? Il est assez peu réaliste de le penser, étant donné qu'à l'heure actuelle on trouve encore un nombre respectable de pages "intéressantes" mais dépourvues de balise "title" (Cf les résultats des moteurs rapportant des documents ayant pour titres "untitled"). Ou bien faudra-t-il "réinventer des professionnels de l'information " qui les conçoivent et les organisent ? Pourra-t-on se fier aux nouveaux outils d'extraction du sens (comme les logiciels data-mining actuels) pour les concevoir ?

Quelles que soient la ou les solutions retenues, elles impliquent des coûts, du temps, une valeur ajoutée, comme disent les marchands. Peut-on espérer une validation un peu plus neutre que la validation à la source par le créateur ?
Mais actuellement, personne ne peut nier le besoin d'y voir plus clair dans la recherche d'informations dans internet. C'est pourquoi les " metadata " ou " métadonnées " donnent lieu à d'intenses réflexions et de nombreuses interrogations à l'heure actuelle. La consultation de la page IFLA : Digital libraries : metadata resources (8) permet de se faire une idée de l'ampleur des réflexions sur ce sujet.

Ces divers éléments ont sous-tendu certains travaux de Ghislaine Chartron à l'URFIST de Paris comme on peut le voir sur "Le Panorama recadré des différents standards"(9) , et donné lieu à une étude approfondie dans Solaris n° 6 (10).

Références

(1) Identification des ressources sur Internet et métadonnées : diversité des standards" de Catherine Lupovic. Documentaliste : sciences de l'information, 1999, vol.36, n° 6;- p.321-325

(2) Greg Notess http://notess.com/ Site remarquable pour le suivi régulier de quelques moteurs. Permet de se faire une idée du recouvrement des différentes bases de données utilisées par les robots et du nombre de réponses apportées par les différents robots sur une

même requête au même instant.

(3) Référencer son site, http://www.ccr.jussieu.fr/urfist/html/referencer.htm in Réaliser des pages web : initiation au langage HTML, par Christophe Boudry et Julia Jumeau, 1999.

(4) Le Dublin Core metadata initiative http://purl.oclc.org/dc/ Eléments sémantiques et descriptifs du Dublin Core. Compte-rendus des ateliers et publications des groupes de travail.

et Resource Description Framework (RDF) Model and Syntax Specification http://www.w3.org/TR/REC-rdf-syntax/

(5) Le Dublin Core et son application dans le projet CISMEF (Benoit Thirion, CHU-Rouen) http://www.ccr.jussieu.fr/urfist/presse/standard/dcurfist1/sld001.htm et aussi Utilisation des métadonnées Dublin Core : projets français http://www.chu-rouen.fr/documed/dcfr.html Le recensement, par Benoît Thirion des différents sites francophones utilisant les métadonnées du Dublin Core.

(6) Le projet CORC d'OCLC http://www.oclc.org/oclc/corc/ ou la présentation sur AUROC http://www.asso.univ-paris5.fr/auroc/formation.htm Recensements d'analyses de sites fédérés par thèmes

(7) DOI http://www.doi.org/using_DOIs.html

(8) IFLA : Digital libraries : metadata resources http://ifla.inist.fr/II/metadata.htm

(9) Panorama des différents standards par Ghislaine Chartron, 1999 http://www.ccr.jussieu.fr/urfist/presse/standard/coursintro.htm

(10) Solaris n° 6 : http://www.info.unicaen.fr/bnum/jelec/Solaris/d06/ Normes et documents numériques: quels changements?/Sous la direction de Ghislaine Chartron et Jean-Max Noyer, 1 février 2000

EVALUATION DE LA RECHERCHE D'INFORMATIONS

Petite bibliographie

Lise Herzhaft, Conservateur en Chef - URFIST de Lyon

Trouver de l'information utile est un premier pas dans la recherche d'informations ; mais cela ne suffit pas. A-t-on retrouvé toute l'information disponible, toute l'information possible et elle seule? Les formulaires créés pour la retrouver sont-ils adaptés, compréhensibles ? La démarche d'évaluation accompagne chacune des étapes de la recherche d'informations :

La formulation de la recherche, le choix d'une source d'information, de son support, de son 'éditeur', du logiciel qui permet l'interrogation et l'affichage des réponses, la complétude de celles-ci (avec ou sans résumé, mots-clés… ou s'il s'agit de texte intégral, la conservation des graphiques, des références bibliographiques citées…)
Si dès le début de l'existence des bases de données, ascii puis sur cédérom, des évaluations de leur qualité ont été faites, le succès de l'internet a amplifié ce besoin : comment qualifier ce qu'a-t-on trouvé, est-ce la meilleure page, est-elle fiable, est-ce la seule…
Il ne s'agit pas ici de théorie sur l'évaluation mais de ses aspects pratiques.

Si évaluer la qualité de l'interface qui permet d'interroger la source d'informations, son intuitivité, relève des ergonomes, évaluer la pertinence, la fiabilité de la source d'informations et son environnement sur le serveur qui l'héberge, évaluer la finesse de l'interface d'interrogation proposée pour chacun des types d'accès - (ascii, cédérom, sur l'internet) - évaluer la puissance du logiciel d'interrogation, évaluer le nombre acceptable de réponses, leur pertinence au premier degré, tout ceci relève des documentalistes, des professionnels de la recherche d'informations. Evaluer le résultat, le contenu des informations ressort du chercheur lui-même.

Une bibliographie abondante existe sur certains aspects de ce sujet puisqu'un thème de la WWW virtural Library lui est consacré et même un didacticiel (en anglais !). En voici une sélection :

Evaluer les bases de données du contenu à l'interface d'interrogation:

LARDY Jean-Pierre, HERZHAFT Lise. Bibliographic treatments according to bibliographic errors and data heterogeneity: the end-user point of view. In Online Information 92. 17th International Online Information Meeting, 8-10 December 1992. London: Learned Information, 1992. p. 547-556

Les critères d'évaluation des banques de données : la démarche qualité chez les professionnels de l'information électronique/ Annick Duflos. Paris: ADBS, 1995. 146 p.

Evaluation of Information Sources
Site donnant des conseils pour évaluer les sources d'information en général avec un tableau comparatif des supports (livre, périodique, page web).

Support de cours de l'URFIST de Lyon sur l'évaluation des bases de données, cédéroms et sites web.1999.

Site qui compare en 1998 les abonnements électroniques à la version papier :
type d'accès complet ou non, différences de présentation (graphiques, images oubliés…)
How Full is the Full in Full-Text? à partir de 26 titres de diverses disciplines (plutôt sciences sociales et économiques).

Celui-ci compare les différentes versions d'une même base de données (Medline), proposée par différents serveurs et exprime le point de vue des documentalistes en février 1999. De nombreux critères sont listés.

Un groupe de spécialistes de l'information (California State University) et de chercheurs

évaluent selon la même grille, notée, 28 fournisseurs de bases de données accessibles sur le web et les producteurs leur répondent (entre janvier 97 et novembre 99)
Database Reviews and Vendor Responses

Le personnel et les utilisateurs de la bibliothèque de Juneau (Alaska) évaluent en 1999 les produits de 6 fournisseurs d'information (EBSCO, UMI, OCLC, SIRS, ELECTRIC LIBRARY et IAC)
Alaska statewide database

Evaluer les cédéroms :

JACSO Peter. CD-ROM software, dataware and hardware evaluation, selection and installation. Englewood : Libraries unlimited, Inc., 1992. (Database searching series n°4).256 p. Ouvrage de base sur le sujet.
JACSO Peter Content Evaluation of Textual Cd-Rom and Web Databases (Database Searching Series) Parution prévue en 2000

CD-ROM EVALUATION CRITERIA
Liste des critères à prendre en considération, selon une progression de 3 niveaux, fonction de la précision/difficulté pour évaluer les cédéroms

Evaluer l'interface d'interrogation ; un exemple en sciences juridiques :

Mode d'usage et d'appropriation de l'information textuelle juridique observée chez les doctorants en droit. Etude en cours pour le projet "Ingénierie textuelle et bibliothèques électroniques" de l'ARASSH (Association Rhône-Alpes en Sciences sociales et humaines) 1998-2000.
Sera accessible sur le site de l'URFIST de Lyon : ARASSH

Evaluer les sites internet :

En français

Cerise
Urfist de Paris : les questions de base à se poser
Insa de Lyon :
les questions à se poser et une petite bibliographie.
Comment évaluer un site et/ou une page web ?
Critères par objectif, de l'Ecole nationale supérieure des Mines de Paris.

En anglais :

Site de base sur le sujet :
WWW virtual library evaluation
Information Quality WWW Virtual Library qui rassemble différents documents traitant de la qualité de l'information, des critères de sélection de sites et plus précisément
Smith, Alastair : Evaluation of information sources [On-line]. march 2000

Puis quelques **sites de conseils** sur l'évaluation :

Un des plus ancien et complet qui replace l'évaluation des sites internet dans le contexte global de la recherche d'informations :
Evaluating Quality on the Net par Hope N. Tillmann mis à jour en 01/2000

Une méthode d'évaluation appliquée aux besoins différenciés des utilisateurs (informations pour l'entreprise, grand public, d'actualités…)
Evaluating Web Resources par (Jan Alexander and Marsha Ann Tate, Widener University) ou selon la même logique, avec exercices :

Who Dunnit: What Kind of Web Page Is This??
Conseils très judicieux pour évaluer l'information produite sur l'internet par E. Kirk une spécialiste des bibliothèques électroniques et de l'éducation à distance.
Evaluating Information Found on the Internet.
Ou ICYouSee: T is for Thinking par John R. Henderson, un cours très dynamique et concret sur les points à considérer lorsqu'on est relié au web, avec exemple, exercice et bibliographie ...
Kathy Schrock's Guide for Educators: Critical Evaluation Information, sous forme d'enquêtes sur les sites visités puis d'articles de l'auteur et une bibliographie importante d'articles généraux sur l'évaluation ou de listes de critères à prendre en compte selon le public visé.

De très nombreuses bibliothèques universitaires américaines donnent accès à leur liste de critères de sélection/évaluation de sites :
The Good, the Bad and the Ugly: or, Why It's a Good Idea to Evaluate Web Sources (Susan Beck,New Mexico State University Library)
Evaluation of Web Pages (Beth Brin and Dan Lester, BSU)

Liste de critères d'évaluation :

Evaluating the Quality of Internet Information Sources
Ou Thinking Critically about World Wide Web Resources (Esther Grassian, UCLA College Library)
Ou Internet Information Evaluation Form grille réalisée par les étudiants de la Kentucky Academy of technological Education

Enfin **des exercices :**
Pour s'entraîner à l'évaluation : Internet Detective un tutoriel d'évaluation des ressources sur l'internet (il faut s'inscrire gratuitement. A faire en 2 heures environ)

Hoax? Scholarly Research? Personal Opinion? You Decide!
UCLA College Library

Ou How to Evaluate World Wide Web Pages
des conseils et des pages à comparer, regroupées pour un cours de la School of Library and Information Studies à l'Université d'Hawaii à Manoa, donné fin 1997. (Les pages ne sont plus toujours accessibles)

Dans cette bibliographie, on trouve à plusieurs reprises des grilles ou listes de critères à examiner.
En réalité, ces grilles s'appliquent de façon différenciée selon qu'il s'agit d'un type de support ou d'un autre (cédérom, bdd ascii, page web) ou surtout selon le public visé (chercheur, entreprise, grand public, enfant, étudiant ; utilisateur néophyte ou expérimenté) .
De plus, il ne faut pas oublier que les 'objets' évalués changent, sont améliorés, précisés : l'évaluation doit être datée sinon elle perd tout intérêt.

Ces quelques pistes voudraient permettre au chercheur d'informations de se situer/ positionner dans sa recherche . En effet, si pour l'internet, les documentalistes et bibliothécaires recensent de plus en plus les sites web intéressants pour la communauté qu'ils servent, pour les autres ressources, des sources bien identifiées par des répertoires au contenu précis, facilement et rapidement accessibles, des formulaires de requête synthétiques mais complets, adaptés aux besoins du chercheur, une visualisation des réponses adaptable, paramétrable, restent des éléments essentiels pour une recherche de qualité.

METHODES DE TRI DES RESULTATS DES MOTEURS DE RECHERCHE

Jean-Pierre Lardy, Conservateur en Chef - URFIST de Lyon

La publication sur le web continue à croître sans relâche et le milliard de pages HTML statiques vient d'être dépassé selon les estimations d'une étude de Inktomi . Si les annuaires, produits manuellement, présentent l'avantage de classer les sites par thèmes, les moteurs, produits automatiquement, sont les outils les plus utiles pour fouiller le web.
La masse d'informations disponibles conduit malheureusement à des résultats pléthoriques la plupart du temps, ce qui déroute les utilisateurs. Aussi les moteurs de recherche ont développé des méthodes de tri automatiques des résultats . Cela leur permet aussi de se distinguer les uns des autres. Dans la pratique aucune méthode de tri n'est parfaite mais cette variété offre à l'utilisateur la possibilité de traquer l'information de différentes manières et augmente donc ses chances d'améliorer ses recherches. Le but du classement est d'afficher dans les 10 à 20 premières réponses les documents répondant le mieux à la question. Si on ne trouve pas ce que l'on cherche dans les toutes premières pages de résultats, il faut reformuler la question. Pour cela il faut comprendre les mécanismes sous-jacents pour en tirer vraiment profit.

On peut considérer trois grandes méthodes de tri.

Le tri par pertinence

Cette méthode repose sur des travaux de recherche déjà anciens de Robertson et Sparckjones , mis en pratique dans le logiciel d'indexation WAIS à la fin des années 80.
Les résultats d'une requête sont affichés selon un ordre déterminé par le calcul d'un score pour chaque réponse. La pertinence est basée sur les cinq facteurs suivants appliqués aux termes de la question :

1. Le poids d'un mot dans un document est déterminé par sa place dans le document : il est maximum pour le titre et le début du texte; à l'intérieur il est plus important si le mot est en majuscule.
2. La densité est basée sur la fréquence d'occurrence dans un document par rapport à la taille du document. Si deux documents contiennent le même nombre d'occurrences, le document le plus petit sera favorisé.
3. Le poids d'un mot dans la base est basé sur la fréquence d'occurrence pour toute la base de données. Les mots peu fréquents dans le corpus sont favorisés. Les mots vides sont soit éliminés, soit sous-évalués.
4. La correspondance d'expression est basée sur la similarité entre l'expression de la question et l'expression correspondante dans un document. Un document contenant une expression identique à celle de la question reçoit le poids le plus élevé.
5. La relation de proximité est basée sur la proximité des termes de la question entre eux dans le document. Les termes proches sont favorisés.

Cette technique a montré son efficacité dans le cadre des bases de données WAIS assez homogènes et peu volumineuses. Elle a été reprise dans les moteurs de recherche apparus à partir de 1994 et basés sur les techniques d'exploration du web par les robots . Cependant l'algorithme exact n'est jamais connu car il est considéré comme secret industriel et quelquefois protégé par un brevet (cas d'Excite).
Les documents HTML peuvent contenir dans l'entête des informations concernant le contenu du document. Ces méta-données correspondent aux balises TITLE, META keywords et META description. Une étude a montré qu'elles étaient malheureusement peu utilisées. Certains moteurs de recherche en tiennent compte dans leur calcul.
Cependant le tri par pertinence présente l'inconvénient d'être facile à détourner par des auteurs désireux de placer leurs pages en tête de liste : pour cela il suffit de répéter les mots importants soit dans l'entête, soit dans le texte en utilisant des techniques de spamming (écrire le texte en blanc sur fond blanc par exemple) pour modifier à son avantage le classement. Les moteurs ont réagi en détectant ses techniques.

Cette méthode est utilisée par AltaVista, Ecila, Excite, FAST, HotBot, Inktomi, Lokace, Voila. Le résultat dépend beaucoup de la question et l'on choisira, chaque fois que cela est possible, des termes précis et non ambigus.

Le tri par popularité

Les limites du tri par pertinence ont conduit à rechercher d'autres méthodes reposant sur des principes tout à fait différents et indépendants du contenu des documents. Connues sous le nom de tri par popularité, on distingue :

La méthode basée sur la co-citation

Lancé en 1998 par deux étudiants de l'Université de Stanford, Google, http://www.google.com, classe les documents grâce à la combinaison de plusieurs facteurs dont le principal PageRank . Ce dernier utilise le nombre de liens pointant sur les pages. L'article de Page et Sergey en donne une description.
Plusieurs moteurs de recherche offrent cette fonctionnalité. Avec AltaVista il faut entrer : www.site.com -host:www.site.com
Cela permet à n'importe quel auteur de pages de découvrir les liens pointant sur son œuvre.

Google évalue l'importance d'une page par les liens qu'elle reçoit mais analyse en plus la page qui contient le lien. Les liens des pages "importantes" pèsent plus lourdement et aident à découvrir d'autres pages "importantes". Ainsi le tri est indépendant du contenu et évite les dérapages de la méthode précédente, le choix des liens étant laissé à la libre décision des millions d'auteurs de pages HTML. Il faut cependant noter que cette technique défavorise les pages récentes et donc inconnues.

La méthode basée sur la mesure d'audience

La société DirectHit a été fondée en avril 98 et propose de trier les pages en fonction du nombre de visites qu'elles reçoivent. DirectHit analyse le comportement d'un internaute dans l'utilisation d'un moteur de recherche : sur la page d'accueil, il saisit un ou plusieurs mots de recherche dans un formulaire, consulte la page de résultats classés par ordre de pertinence, choisit l'un d'entre eux, va sur le site correspondant pour le consulter. Si la page ne lui convient pas, il revient sur la page de résultats du moteur, choisit un autre lien, etc. jusqu'à ce qu'il ait trouvé un document pertinent. DirectHit enregistre ce comportement pour tenter de trouver les pages les plus "populaires" sur un moteur de recherche et ainsi améliorer leur classement. Il fonctionne, en règle générale, en tâche de fond sur un moteur existant. A chaque consultation d'un utilisateur, DirectHit note sur quel lien celui-ci a cliqué et quel était le rang de ce lien. Il mesure le temps passer sur une page avant que l'utilisateur ne revienne aux résultats. S'il ne revient pas, il en "déduit" que le site proposé était pertinent. Il sera alors mieux classé dans les résultats suivants, lors d'une interrogation sur le même mot-clé. Ainsi les interrogations et la façon d'interroger et de naviguer des internautes vont enrichir la base données de DirectHit.
Cette méthode comme la précédente pénalise les pages récentes mais évite le spamming.

DirectHit peut être interrogé directement sur son site mais alimente aussi les résultats de nombreux outils de recherche comme HotBot, LookSmart et des sites Web comme celui de ZDNet.

L'annuaire Snap utilise une technique appelée "Global brain", classant les sites selon leur

popularité auprès des internautes, afin de les inclure dans ses algorithmes de pertinence.

Le tri par calcul dynamique de catégories

NorthernLight, lancé en Août 1997, propose le classement des documents trouvés dans des dossiers (clustering) constitués automatiquement en fonction des réponses. Un dossier peut lui-même être constitué de sous-dossiers. Quatre types existent :

- Subject (e.g., hypertension, baseball, camping, expert systems, desserts)
- Type (e.g., press releases, product reviews, resumes, recipes)
- Source (e.g. commercial Web sites, personal pages, magazines, encyclopedias, databases)
- Language (e.g., English, German, French, Spanish)

Dans chaque dossier final, les réponses sont triées par pertinence.

Autres recherche

Chez IBM John Kleinberger a développé l'algorithme HITS (Hypertext-Induced Topic Search) pour identifier les sources d'autorité parmi les pages HTML en fonction du nombre de liens qu'elles possèdent. Le moteur de recherche CLEVER (Clientside Eigenvector Enhanced Retrieval) utilise cette technique pour caractériser les pages.

Conclusion

On constate donc que les méthodes de classement des résultats se sont diversifiées ce qui complique la tache des auteurs de pages mais augmente les chances de trouver de bons résultats dans la jungle que représente actuellement le web.
En fait le mieux serait sans doute de combiner ces différentes méthodes : c'est ce que vient d'annoncer NorthernLight .

Moteurs cités

AltaVista, DirectHit, Ecila, Excite, FAST, Google, HoBot, Inktomi, Lokace, NorthernLight, Snap, Voila

Bibliographie

- Lawrence S., Giles L. "Accessibility and Distribution of Information on the Web", Nature 400(6740): 107-109, 1999 http://wwwmetrics.com/
http://www.notess.com/search/stats/nature99.shtml (Summary and analysis by Greg R. Notess)

- "Web Surpasses One Billion Documents", 18 Janvier 2000
http://www.inktomi.com/new/press/billion.html

- Courtois M.P., Berry M.W. "Results Ranking in Web Search Engines", ONLINE, mai 1999

- Robertson S. E., Sparckjones K. "Relevance weighting of search terms", Journal of the American society for Information Science, 27 (3): 129-146, 1976

- Koster M. "The Web robots FAQ"
http://info.webcrawler.com/mak/projects/robots/faq.html

- Hassani Md, Lardy J. P. "Techniques de tri des moteurs de recherche" DEA SIC Université LYON I, 1998

- Page L. "PageRank: Bringing Order to the Web"
http://www-pcd.stanford.edu/~page/papers/pagerank/

- Brin S., Page L. "The Anatomy of a Large-Scale Hypertextual Web Search Engine"
http://www7.scu.edu.au/programme/fullpapers/1921/com1921.htm

- Kleinberg J. "Authoritative sources in a hyperlinked environment" Proc. 9th ACM-SIAM Symposium on Discrete Algorithms, 1998

- The CLEVER project http://www.almaden.ibm.com/cs/k53/clever.html

- Search Engines - Submission Tips, Help and Use.
http://www.sofer.com/research/engines.html

- Feldman S. "Northern Light Adds Link Popularity to Relevance Ranking Factors", 1999
http://www.infotoday.com/newsbreaks/nb1101-3.htm

Sites vérifiés le 29/03/2000

QUELQUES ADRESSES

François Magnan, Documentaliste

* **L'agent de la semaine :** Il est maintenant possible d'interroger en langage naturel COPERNIC 2000, le métamoteur de recherche que vous installez sur votre disque dur, avec la version 4.5 . Cette possibilité n'existe toutefois que pour des interrogations en langue anglaise.
Source : Copernic 2000

* **L'information de la semaine :** depuis Le 17 avril dernier, Francopholistes, qui recense les listes de diffusion francophones, n'est plus géré par le CRU (Comité de Réseau des Universités) mais par PresenceWeb, une société privée.
Source : Internet Actu
Annuaire Francopholistes

* **Le site de la semaine :** désirez-vous vous former au courrier électronique ? Allez sur CIRFFI Cliquez sur la rubrique "Nouvelles Technologies", puis "Ressources en ligne", vous trouverez un cours en ligne, interactif, sur Outlook Express.

* **La surprise de la semaine** : c'est peut-être de voir l'éditeur Gallimard se lancer dans la production de contenu sur le web, effectivement, le célèbre éditeur français annonce son intention d'acheter la société d'édition électronique et d'Internet Bibliopolis.
Source : Multimedium

* **L'interview de la semaine** n'est certainement pas celle du PDG de yahoo, Tim Koogle. Ce Monsieur parle à peine de son annuaire !
Source : Le Journal du Net.
Allons plutôt sur le site Abondance pour lire les très intéressantes interviews des responsables des annuaires francophones Yahoo!, Nomade et Voila ()

* **Le moteur de recherche de la semaine**, c'est certainement Google : (et Google version française). Grâce à cette nouvelle interface, l'internaute francophone a désormais un accès aux commandes et à l'aide en ligne en français. Saluons aussi une autre nouveauté par rapport à l'ancien google : un filtre linguistique permet de ne sélectionner que des résultats en français.
Google fut créé en 1998 par Larry Page et Sergey Brin, deux étudiants de l'Université de Stanford qui développèrent une technologie avancée pour trouver l'information sur Internet, principalement fondée sur l'analyse des liens hypertextes qui permettent d'estimer la qualité ou l'importance - et donc la pertinence d'une page web. Rappelons la particularité de ce moteur de recherche par rapport aux autres : son principal critère de classement des résultats proposés n'est plus la présence de mot-clef dans telle ou telle partie de la page Web, mais le nombre de liens qu'on aura fait envers votre page Web. De fait, google excelle dans la production immédiate du bon résultat pour les requêtes courantes, telles que les noms d'entreprises.

Quelques trucs de recherche sur google :

- tous les mots-clefs que vous rentrez seront pris en compte : ainsi, si vous rentrez : candidat mairie paris, Google ne vous retournera que les pages qui contiennent exactement ces trois mots. Autrement dit, l'opérateur AND (et) est automatiquement rajouté par Google entre chaque mot.
- Google ignore l'opérateur OR : ainsi, vous ne pouvez pas demander des pages qui contiennent soit le mot bicyclette soit le mot vélo. Il ignore aussi la troncature et recherchera les pages qui contiennent les mots de votre requête comme vous les avez écrit : ainsi, si vous demandez : auto, il ne vous donnera pas de réponses contenant le mot automobile. Pour ces deux raisons, je vous suggère de recommencer votre requête sous une autre forme si vous n'êtes pas content du résultat.
- Si vous mettez des majuscules, Google n'en tiendra aucun compte.
- Google accepte la recherche par phrase, il faut alors mettre cette phrase entre guillemets ex : "formation professionnelle continue" De la même façon tout signe placé entre plusieurs mots fera prendre ces mots pour une phrase : ainsi, belle-de-nuit est automatiquement considéré comme une phrase, comme "belle de nuit"
- Google ignore les caractères comme certains chiffres et lettres uniques et mots courants que l'on appelle mots vides. Ainsi la lettre "a" isolée n'est pas pris en compte. Si vous voulez vraiment l'inclure dans une requête, utilisez le signe + en respectant la syntaxe indiquée par l'aide en ligne.

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

Les principaux sujets traités dans les communications de Online 99 ont été les suivants :

- les nouveaux modes de travail et la gestion des connaissances
- les nouveaux services sur Internet : les portails, l'évolution des moteurs de recherche

- le langage XML : concepts, développement et applications ; la conception de sites.

Ont été aussi largement évoqués : le droit de l'information, l'information commerciale (particulièrement l'édition électronique) et l'intelligence économique, les bibliothèques virtuelles.

Dans cette lettre, nous évoquerons les nouveaux services sur Internet. Les autres sujets seront développés dans les lettres suivantes.

I - LES PORTAILS (OU GALERIES)

I.1 - Définition

Il ne s'agit pas d'un concept nouveau, plutôt d'une évolution dans le sens d'une organisation de l'accès à l'information sur Internet tendant vers l'individualisation, la spécialisation.
Ce terme est concomittent du développement du commerce électronique, qui représenterait un marché d'environ 8 millions d'utilisateurs.
Les portails concernent trois types d'entreprises sur le web, dès 1994 :

- les fournisseurs d'accès et de service Internet
- les sites de recherche
- les fournisseurs de navigateurs.

Le concept est né de la fusion des navigateurs avec les moteurs et annuaires de recherche pour fournir des services de recherche aux utilisateurs.
De nouveaux types de fournisseurs sont apparus parallèlement aux portails et au commerce électronique : les méta-moteurs de recherche, les moteurs analysant la popularité des sites, les moteurs offrant une recherche en langage naturel.
Il s'agit d'essayer de répondre à deux problèmes : le "trop" d'information (information overload) et l'absence d'information (information overlook), la solution à ces deux problèmes revenant à fournir des informations à la fois pertinentes et exhaustives.

Le portail doit être différent selon qu'il vise l'utilisateur individuel, ou une organisation entière (bibliothécaires, vendeurs d'une entreprise…) où l'aspect compétitif de l'information prend le pas sur l'aspect gain de temps et intérêt personnel.

I.2 - Portails et publication électronique.

L'arrivée d'Internet (email, listserv, newsgroups…) a apporté un changement dans l'esprit des utilisateurs par rapport à la publication : on n'est pas obligé de se limiter au monde de l'impression, on peut agir plus vite, mieux et à meilleur prix. Il semble qu'avec les formats HTML et PDF, l'impression devienne vraiment secondaire dans le commerce éditorial. Les utilisateurs se sont emparés des nouvelles technologies, témoins les réalisations Los Alamos, LANL, l'initiative SPARC, la proposition du Dr Varmas du National Institute of Health.

Les éditeurs doivent s'adapter à ces nouveaux besoins et trouver des revenus à partir d'un environnement qui n'est plus celui de l'imprimé. Ils doivent créer une valeur d'usage pour fidéliser les abonnés et faire de la prospective pour gagner les nouveaux utilisateurs du web.

I.3 - Evolution des portails

Les nouveaux utilisateurs du web, passée la découverte d'informations jusque là non accessibles, sont vite submergés par le nombre et déçus par la qualité.
Les moteurs de recherche générant des index perdent de leur efficacité avec l'augmentation des informations du web. C'est pourquoi les sites spécialisés se sont développés : concernant un sujet ou des gens réunis par un intérêt commun. Ils sont appelés "gateways" puis "portals" (parfois quand le sujet est très spécialisé, portail vertical ou "vortal"). Il existe des portails qui vivent de la publicité et sont gratuits pour les utilisateurs. Ce sont par exemple les sites de villes ou relatifs à un intérêt particulier (ex. : garden.com).
D'autres portails fonctionnent par abonnement ou redevance, tout en offrant une partie du site gratuitement. Ils versent une participation aux éditeurs qu'ils font payer à leurs utilisateurs.
Pour retenir les utilisateurs, ils doivent être capables d'offrir des informations spécifiques et couvrir uniquement le champ intéressant le client.

I.4 -Implications des éditeurs professionnels, institutionnels et scolaires par rapport aux portails.

Les exigences des utilisateurs sont les suivantes : ils veulent l'article, pas le numéro complet, ils ne veulent pas une " non réponse ". Dans ces conditions ils acceptent de payer le service.
Le modèle économique des éditeurs doit évoluer vers la granularité (information détaillée et précise). Il leur faut vendre seulement l'article, ou des abonnements couvrant moins d'une année, pour augmenter leur audience sur le web. Ils doivent trouver autant d'utilisateurs que possible achetant des produits livrés sour toutes les formes, diversifier et changer leurs offres.

Leur stratégie doit être :

- de trouver les personnes intéressées
- de diffuser les contenus de la meilleure manière : en identifiant les portails utilisés, en intégrant des contenus dans ces portails, en négociant des liens avec les portails…
- de modifier le service apporté : devancer la demande d'article, en faisant des recherches sur le web.
- d'assurer la sécurité des droits.

II - LES MOTEURS DE RECHERCHE

II.1 - Couverture - Comportements des utilisateurs.

La dimension du web est difficile à mesurer, car certaines pages ne sont pas répertoriées. Les pages dynamiques (créées par l'ordinateur, en vue d'une utilisation particulière), les sources accessibles par mot de passe, les pages situées à des niveaux " profonds " sur un site, parfois les videos, sons et images, certains formats comme Catchword, PDF...etc. D. Green l'appelle le web invisible.

L'utilisateur d'Internet commence à faire des opérations basiques et utiliser le courrier électronique, puis il pratique le " surfing ". Après ce stade de découverte, il se centre sur ses besoins auxquels il cherche à trouver une réponse plus facilement et plus rapidement : il utilise les portails, les annuaires, les moteurs de recherche, il combine les différents outils. L'utilisateur devient ensuite créateur de contenu et appréhende Internet comme un moyen de communication enrichi.

Trouver de l'information sur Internet devient une nécessité qui a mis en avant, rendu populaires certains moteurs de recherche aux USA : Yahoo, Go, Lycos, Excite, Alta Vista, Looksmart, Snap, Hotbot, Goto, Askjeeves.

II.2 - Sont-ils tous efficaces ?

Personne ne sait combien de pages il y a sur le web. On ne peut faire que des estimations. Selon Nature Magazine (1999) aucun moteur n'indexerait plus de 16 % du web.

Un rapport du NEC Research Institute (1999) a calculé que :

- Northern Light indexe environ 20 % du web (160 millions de pages),
- Fast (Search Transfer) 25 %, avec des pages non américaines,
- Excite et Lycos : 6 à 8 %,
- Altavista (150 millions de pages) et Inktomi : 13 à 18 %,
- Google et Infoseek : 9 à 12 %.

Il ne faut donc pas de limiter à l'interrogation d'un seul moteur, utiliser des méta-moteurs comme Copernic, Bull's eye, Dogpile, Mamma, 2Q, Infind, Isleuth, Surfy, Webtaxi.

Yahoo n'indexe que 500 000 sites, et privilégie l'interaction humaine avec les liens rencontrés par les moteurs (de même pour Snap).

Quelques observations :

On ne peut concevoir d'index exhaustif (il serait impraticable), le problème d'index multilingue se pose de plus en plus, les robots de recherche sont insuffisants pour donner du sens aux résultats, les interfaces permettant l'interrogation en langage naturel ne sont pas encore très développés.

Les stratégies de développement des collections et des bibliographies appliquées aux produits imprimés ne peuvent être transférées dans l'environnement du web. Les moteurs sont limités dans leur puissance d'indexation et les limites de la profondeur d'indexation pénalisent les gros sites. Cependant il est remarquable que la plupart des utilisateurs n'ont pas besoin d'un énorme index; des moteurs comme Google et Direct Hit sont très utiles en dépit du petit nombre de pages indexées. Direct Hit Index www.directhit.com sélectionne les sites en fonction de leur popularité, Google indexe les sites qui ont le plus grand nombre de liens entre les pages et les sites.

II.3 - Caractéristiques des moteurs de l'an 2000.

Ce qui est positif :

- Ils sont gratuits pour la plupart.
- Ils couvrent la plupart des grands sites américains, canadiens et européens.
- Les méta-moteurs permettent de trouver les données qui ne sont signalées que par un moteur.
- Les moteurs les plus connus s'améliorent sans cesse : la combinaison de robots et d'index humains permet de donner du sens aux résultats (Yahoo, Nerdworld).

Sur 10 outils classiques testés (Maldonado Martinez A, Fernandez Sanchez E), la plupart (en recherche avancée) utilisent des opérateurs booléens, de proximité, la troncature; la recherche par champ est plus ou moins étendue (Northern Light offre le plus grand nombre d'options, la reformulation n'est faite que par Lycos et Infoseek).

Ce qui doit être amélioré :

- l'aide à l'utilisateur et l'interface
- la couverture des pages
- l'énoncé des critères de classement et de pertinence utilisés
- la transparence, l'arbitraire dans le choix des sites
- l'indexation et une meilleure maintenance des changements
- l' indexation des gros sites (le suivi des liens)
- la maintenance des pages web non anglophones
- la recherche intelligente, à l'instar des agents intelligents (capacité de sauvegarde et de reformulation).

Les moteurs de recherche en sont à leur début, aucun ne marche parfaitement. Ils visent plutôt l'utilisateur moyen que le professionnel mais ne lui procurent cependant pas totale satisfaction.
Les améliorations permettant la personnalisation de son propre site (comme My Yahoo, My Excite) sont pour certains inadéquats ou superflus.
Les utilisateurs dans leur ensemble acceptent mieux les systèmes simples (comme Yahoo, Excite, Lycos…) que les systèmes complexes et puissants (options " recherche avancée " de plusieurs moteurs). Il n'y a pas de relation entre la popularité d'un moteur et ses capacités de recherche. Les systèmes complexes sont et seront utilisés par un public réduit et spécialisé.
On verra se développer encore l'importance des méta-moteurs. De nouveaux modèles commerciaux comme Northern Light et Powerize ne facturent que la fourniture du document.

Classement et pertinence :
Les moteurs ont leurs méthodes pour décider ce qui est le plus approprié et utilisent pour cela plusieurs critères :

- le titre : élément clé du classement (Yahoo, Nerdworld) pour les annuaires.

- la place des mots sur la page : les têtes de chapitre, mots clés, début de phrases, proximité entre eux.
- la répétition des mots clés
- les liens vers la page web.

Quand on produit une page web, il faut connaître les critères de "pertinence" établis par les moteurs. Les meilleures pages web, sélectionnées par les moteurs sont celles écrites par des auteurs du web avertis. Il est important de savoir si les moteurs reconnaissent ou non les méta-tags. Par exemple Excite, Lycos, Webcrawler, Northern Light ne les reconnaissent pas, Infoseek, Altavista, Hotbot, le font.

Evolution des moteurs depuis 1998 :

- ceux qui proposent un deuxième niveau d'analyse

Direct Hit

- ceux qui font de la recherche en langage naturel

Ask Jeeves
The Electric Monk

- ceux qui analysent les liens :

Google (seul site public ayant cette fonction)
Clever Clever examine le contexte hypertexte d'une recherche par mot-clé, à partir des pages sélectionnées par Altavista.
Focused Crawler, à l'étude chez IBM.

- les index spécifiques d'un sujet :

1Jump
informations sur les entreprises
Powerize

II.4 - Alternatives aux moteurs :

- Les professionnels, les utilisateurs avertis préfèrent utiliser les bibliothèques virtuelles pour obtenir des informations de bonne qualité rapidement.
  Exemples de Bibliothèques virtuelles :

  SOSIG
  OMNI
  Biz/Ed

  Elles permettent de se faire une idée exacte sur un sujet, grâce à une classification hiérarchique faite par des experts concernant des

champs particuliers.
Des sites comme BioMednet , et Chemweb , permettent à des communautés on-line d'échanger, de travailler pour soi et avec ses pairs.

- Les agents intelligents constituent aussi une autre piste prometteuse pour l'utilisateur car ils se basent sur les choix, les opinions des utilisateurs. On peut avoir la liste des agents à l'adresse www.botspot.com
  Citons : Mata Hari, Bull'sEye Pro, Copernic.
- Les newsgroups, porteurs de connaissances implicites, sont nécessaires pour créer de nouvelles opportunités.

Extraits de :

- ARNOLD S.E., Search engines 2000 : rough running, prone to stall
- BRADLEY P., The Great Search Engine Con-trick
- BOWES F III, Gateways, portals and publishers.
- Mc LAGAN D., Portals: all change or all the same ?
- GREEN D., The Evolution of web searching.
- MALDONADO MARTINEZ A, FERNANDEZ SANCHEZ E, Comparing Internet Search tools.

ASIS 1999 Annual Conference - Knowledge :

Creation, Organization and Use

The American Society for Information Science - ASIS 1999 Annual Conference - October 31 - November 4, 1999 | JW Mariott Hotel | Washington, DC

Cette conférence annuelle de l'American Society for Information Science se distingue par la richesse de son champ d'investigation, richesse rendue presque inévitable du fait de la transversalité des notions et des concepts nécessaires à l'analyse de la création, de l'organisation et de l'utilisation du savoir et de la connaissance.
Cette densité est au cœur des préoccupations du champ des Sciences de l'Information et de la Communication.
Nous avons choisi de rendre compte d'articles représentatifs des recherches en cours ainsi que d'articles plus théoriques qui dressent un état de l'art pour un champ de recherche donné.
Afin d'améliorer la clarté et la lisibilité de notre présentation, nous avons regroupé ces comptes rendus en trois parties qui nous semblent représentatives de la teneur générale de l'ouvrage :

1. Représentation du savoir et des connaissances
   - "Using ontologies for knowledge management : an information system perspective" Igor Jurisica, John Mylopoulos, Eric Yu - pp. 482 à 496
   - "Contextual classification in the metadata object manager (m.o.m.)" Thomas pole - pp. 194 à 202

- "Utilizing faceted structures for information systems design " Uta priss & Elin Jacob - pp. 203 à 212

2. Pratiques de l'indexation
   - "Using ordinary language to access metadata of diverse types of information resources : trade classifications and numeric data." Youngin Kim, Barbara Norgard, Aitao Chen, Fredric Gey - pp.172 à 180
   - "Topological indices of textual identity networks." Gregory H. Leazer & Jonathan Furner - pp. 345 à 358
   - "Icons as visual form of knowledge representation on the world wide web : a semiotic analysis" Yan Ma & Virgil Diodato - pp.181 à 193
   - "Text indexing of images based on graphical image content." Timothy B. Patrick, MaryEllen C. Sievert, Mihail Popescu - pp. 675 à 680
   - "The identification of index terms in natural language object descriptions." P. Bryan Heidorn - pp. 472 à 481
3. Modèles comportementaux de la recherche d'information sur le web
   - "Information seeking on the web, an integrated model of browsing and searching." Chun Wei Choo, Brian Detlor, Don Turnbull - pp. 3 à 16

USING ONTOLOGIES FOR KNOWLEDGE MANAGEMENT : AN INFORMATION SYSTEM PERSPECTIVE

Igor Jurisica, John Mylopoulos, Eric Yu - pp. 482 à 496

Il s'agit d'une réflexion théorique sur la construction d'ontologies dans le cadre du Knowledge Management qui isole 4 catégories ontologiques génériques :

- ontologie statique : aspects statiques du monde, quelles sont les choses qui existent (entités), quelles sont leurs attributs et leurs relations …
- ontologie dynamique : aspects changeants du monde en terme d'états, d'états transitionnels et de processus.
- ontologie intentionnelle : croyances et opinions diverses des agents de ce monde (je crois, je veux, j'approuve ou désapprouve …), tâches et sous-tâches …
- ontologie sociale : couvre tous les aspects de type sociétal (positions, rôles, autorité, permanence des structures d'organisation, réseaux d'alliances …)

Ces 4 catégories doivent permettre de rendre compte de tous les travaux sur la constitution d'ontologies en les classifiant. Le corrélat étant que toute représentation du Savoir peut s'appuyer sur les concepts premiers qui structurent ces 4 ontologies. Cette approche se justifie par le fait que les technologies des systèmes d'information n'ont plus pour seul but de traiter ou de transmettre des données ou de l'information, mais sont amenées à jouer un rôle déterminant dans la gestion des connaissances.

CONTEXTUAL CLASSIFICATION IN THE METADATA OBJECT MANAGER (M.O.M.)

Thomas Pole - pp. 194 à 202

L'auteur présente un modèle de métadonnées fondé sur une classification contextuelle qui doit permettre d'optimiser la représentation ainsi que la gestion du savoir, développé pour une filiale du groupe Thompson.
Ce modèle de classification contextuelle se veut la synthèse de plusieurs modèles traditionnels (classifications hiérarchiques, classifications à facettes ...). La nécessité de ce mode de classification est justifiée par l'insuffisance des systèmes booléens (" free text ") et hiérarchiques dont les fondements technologiques n'ont que peu évolués (analyse statistique).

Les attentes des usagers de l'information se structurent autour des points suivants :

- possibilité d'utiliser plusieurs types d'interrogations (texte libre, arborescence ...)
- possibilité d'itérations de la recherche
- possibilité de "mixer" les procédés d'interrogation au cours des différentes phases d'itération
- possibilité d'interrogation par l'exemple (" trouver de l'information qui ressemble au document que j'ai déjà")
- possibilité d'ajuster la recherche sans avoir à s'intéresser à l'algorithme utilisé
- possibilité de chercher dans des sources d'information multiples depuis un point d'accès unique.

Après une description des spécificités des modèles de classification traditionnels (modèles d'index, modèles de classifications hiérarchiques, modèles de thesaurus à facettes), est défini le modèle de classification contextuelle qui se distingue des modèles hypertextuels (HTML), à deux dimensions (SQL) ou hiérarchiques (Microsoft Site Server/Knowledge Manager), dans la mesure où le paradigme utilisé ici ne situe pas le savoir (" data ") et sa classification (" metadata ") dans un même lieu : la classification de l'information ainsi que toutes les métadonnées concernant le savoir et les inter-relations qu'il développe sont situées à l'extérieur du savoir lui-même.

UTILIZING FACETED STRUCTURES FOR INFORMATION SYSTEMS DESIGN

Uta Priss & Elin Jacob - pp. 203 à 212

Partant du postulat selon lequel un site web bien structuré permet de faciliter l'accès à l'information, les auteurs recommandent l'utilisation de modèles dits " à facette ", approche

qu'ils considèrent comme déterminante dans l'organisation du savoir puisqu'elle permet de rendre moins aléatoires et plus gérables les procédés d'organisation de l'information.
Ils définissent ensuite les deux modes de création de facettes (par agrégation ou par composition de termes) et les possibilités qu'ils offrent dans la gestion de hiérarchies conceptuelles qui, même réduites à peu de termes, basculent très vite dans la complexité.
L'utilisation de thesaurus à facette, par le contrôle du vocabulaire de descripteurs et des relations, permet de représenter le contenu conceptuel d'un document. Ce modèle s'avère tout particulièrement efficace dans des applications de type W3, où, en incluant les descripteurs appropriés dans les métadonnées de chaque page, la structure hypertextuelle du site peut être générée de manière interactive par la manipulation des diagrammes de relations entre descripteurs. Toutes les facettes ayant une localisation précise dans le thesaurus, la cohérence et la consistance du site sont assurées. La flexibilité de cette approche peut également permettre d'optimiser la navigation en fonction des attentes de chaque catégorie d'usager de l'information.

USING ORDINARY LANGUAGE TO ACCESS METADATA OF DIVERSE TYPES OF INFORMATION RESOURCES : TRADE CLASSIFICATIONS AND NUMERIC DATA

Youngin Kim, Barbara Norgard, Aitao Chen, Fredric Gey - pp.172 à 180

Description d'une méthode d'accès à des types spécifiques de données (numériques) rendues difficilement accessibles du fait de l'utilisation de métadonnées réclamant une hyperspécialisation dans le domaine (vocabulaire spécifique).
Approche EVM (Entry Vocabulary Module) pour rendre plus transparent l'accès à ce vocabulaire de spécialité et faciliter la recherche d'information.
Le cadre de l'étude porte sur la base Foreign Trade Import-Export qui comporte des donnes spécifiques concernant l'import et d'autres concernant l'export toutes deux consultables selon un système de codage et de description appelé Harmonized System (HS) comprenant un codage numérique associé à un vocabulaire de spécialité. Il s'agit de mettre au point une possibilité d'interrogation en langage " quasi-naturel " ne faisant plus appel à une terminologie technique abstruse.
[D'autres prototypes EVM ont été développés dans le cadre de la base United States Patent Classification et de la base International Patent Classification ainsi que pour BIOSIS et INSPEC http://www.sims.berkeley.edu/research/metadata/oasis.html.]

Pour résoudre ce problème, les auteurs utilisent une autre classification qui traite de données commerciales (the Standard Industrial Classification - SIC -), abondante en données textuelles, servant de passerelle (" switching language ") entre leur système EVM et l'organisation HS, passerelle autorisée du fait de l'organisation hiérarchique des deux systèmes.

Le module EVM fonctionne comme un dictionnaire d'associations - utilisant un algorithme

statistique - mettant en correspondance des items lexicaux basiques (titres, auteurs, abstracts …) et des métadonnées (classifications numériques, vocabulaire de spécialité …) permettant de prévoir quelle entrée de métadonnée correspondra le mieux à l'interrogation en langage naturel entrée par un utilisateur.

C'est la coexistence de 2 approches, de 2 techniques, de 2 pratiques d'indexation - l'une pour organiser et classifier des données à des fins de collecte statistique, l'autre pour faciliter l'accès à des objets informationnels différenciés - qui a rendu nécessaire l'approche EVM, qui n'est viable que dans la mesure où existe un nombre suffisant de " training data " pour un vocabulaire de spécialité donné auquel il s'agit de donner accès en langage naturel.

TOPOLOGICAL INDICES OF TEXTUAL IDENTITY NETWORKS

Gregory H. Leazer & Jonathan Furner - pp. 345 à 358

Se fondant sur la notion de réseaux d'identité textuelle (textual identity networks) en tant qu'ensemble de documents partageant un contenu sémantique ou linguistique commun, cet article propose une méthode d'analyse topologique pouvant optimiser la recherche d'information.

Après avoir rappelé les principaux acquis de ce champ d'investigation (la modélisation et le contrôle de ces réseaux d'identité textuelle est un des principaux objectifs du catalogage), les axes de recherche principaux sont caractérisés comme suit :

- compréhension de la manière dont les documents (au sens d'entité bibliographique, d'artefact de savoir enregistré) dérivent les uns des autres et de la structure logique des associations qui les lient.
- représentation de la structure de ces réseaux intertextuels
- exploration du potentiel offert par l'analyse de ces représentation dans la recherche d'information.

C'est la théorie des graphes qui est ici choisie pour représenter ces structures, les graphes pouvant être à leur tour représentés sous la forme de matrices numériques qui peuvent faire ressortir des relations de type " adjacence ", " distance ", … A partir d'exemples concrets, et en s'inspirant des acquis dans le champ de la chimie (qui utilisa la théorie des graphes pour modéliser atomes et molécules) les auteurs calculent une série d'indices topologiques, pouvant être rassemblés en index topologiques. Les notions mathématiques et métriques utilisées dans le calcul des indices topologiques ont déjà fait leurs preuves dans le champ de la recherche d'information, c'est leur application à l'analyse des réseaux intertextuels qui est entièrement nouvelle et doit permettre, à terme, de mieux paramétrer les systèmes de recherche d'information par une meilleure compréhension de la nature et de la structure des réseaux.

ICONS AS VISUAL FORM OF KNOWLEDGE REPRESENTATION ON THE WORLD WIDE WEB : A SEMIOTIC ANALYSIS

Yan Ma & Virgil Diodato - pp.181 à 193

Compare les principes d'indexation traditionnels avec les modalités d'indexation des icônes sur le WWW en essayant de voir si les premiers sont adaptés à l'analyse de ces icônes. L'étude porte sur un échantillon des pages d'accueil de bibliothèques américaines en isolant l'information icônique de la structure de ces pages. Partant du postulat selon lequel ces icônes sont une forme visuelle d'une représentation du savoir, il s'agit d'examiner leur structure pour déterminer en quoi elle est représentative de l'information qui leur est associée. Le terme d'icône englobe ici les trois types de signes définis par Peirce : index, icône et symbole.
La méthodologie choisie repose sur les modèles d'analyse sémiotique d'Eco, Peirce, Saussure, Barthes et Metz ainsi que sur celui développée par Ann De Vaney pour l'étude de la télévision éducative, qui comprend 5 phases et qui est augmenté pour l'occasion d'une sixième phase :

- identification du format (pages d'accueil de bibliothèques)
- identification du segment (icônes utilisées comme des pointeurs d'index vers 5 catégories de services : informations générales, catalogues, services, ressources, recherche)
- structure du segment : permet d'examiner la structure d'indexation des icônes et de la comparer avec les techniques traditionnelles.
- identification des éléments de construction : doit permettre d'identifier les codes et la syntaxe sur lesquels sont construites ces icônes.
- Analyse quantitative : rapports entre la structure icônique utilisée et les buts d'indexation poursuivis.
- Analyse qualitative : analyse de l'interaction entre la structure des icônes et les ressources liées en termes de connotations.

Des résultats sont avancés en termes de :

- persistance des icônes au travers de l'arborescence (67% n'apparaissent qu'au premier niveau, celui de la page d'accueil)
- utilisation d'animations (javascript, dhtml ...)
- incorporation d'éléments textuels dans la structure de l'icône.
- ...

Les auteurs parlent d'icônes de troisième ou quatrième génération depuis les débuts du WWW, se fondant sur une sophistication croissante des aspects fonctionnels, artistiques, photographiques et ergonomiques. Ils mettent en évidence les codes culturels fortement connotés dans le choix de la symbolique de ces icônes et concluent sur la médiocrité du

niveau d'indexation atteint par le recours à l'iconographie.

TEXT INDEXING OF IMAGES BASED ON GRAPHICAL IMAGE CONTENT

Timothy B. Patrick, MaryEllen C. Sievert, Mihail Popescu - pp. 675 à 680

Aux trois méthodes traditionnelles pour l'indexation des images (assignation manuelle de termes aux images, indexation basée sur du texte associé, indexation à partir du contenu graphique), cet article propose une méthode alternative : il s'agit d'indexer manuellement un ensemble " d'images de référence " sélectionnées et d'utiliser ces images comme filtre d'interrogation ou sous forme d'algorithme de recherche sur contenu graphique pour transférer les termes d'indexation de ces images vers les autres images à indexer, en utilisant les scores de similarité produits par l'algorithme de recherche et en les pondérant.

THE IDENTIFICATION OF INDEX TERMS IN NATURAL LANGUAGE OBJECT DESCRIPTIONS

P. Bryan Heidorn - pp. 472 à 481

Le problème de la recherche d'images est ici abordé sous l'angle de la résolution de problèmes où les images sont les objets. Il s'agit d'étudier les "construits" (en terme de vocabulaire et de communication) d'experts et de novices dans un domaine - ici la botanique - utilisés pour décrire des objets dans le cadre d'un modèle de tâche de type "identification d'objets".
Si le genre et l'espèce suffisent habituellement aux experts pour identifier une plante, les non-experts (pour qui la connaissance du domaine est moins partagée) utilisent des descriptions plus "parlantes" ("verbose"). Un premier résultat fait ressortir la fréquence du recours à des analogies visuelles dans le domaine ("cette plante ressemble à une tulipe") ou en dehors du domaine ("… ressemble à un papillon") qui nécessiterait pour être prise en compte une optimisation des mécanismes d'analyse sémantique.
Les conditions expérimentales de cette étude se placent sous l'angle d'une approche de type " agent intelligent ", où un agent " speaker " interroge en langage naturel un autre agent "listener" afin de retrouver une image.

INFORMATION SEEKING ON THE WEB - AN INTEGRATED MODEL OF BROWSING AND SEARCHING.

Chun Wei Choo, Brian Detlor, Don Turnbull - pp. 3 à 16

Etude sur les habitudes de navigation d'une trentaine d'utilisateurs venant de sept entreprises différentes, tous spécialistes de l'information, sur une période de 2 semaines à l'aide :

- d' un questionnaire détaillé
- d' un browser - mouchard qui enregistre les actions de navigation (avant, arrière, bookmark ...) et les Url afin de reconstruire et d'analyser les séquences de recherche
- d' une entrevue individuelle

A l'issue de la phase de collecte des données, 61 épisodes significatifs de recherche d'information sont identifiés.

Après un rapide historique des études inaugurales dans ce domaine (Catledge et Pitkow [1995] : utilisateurs divisés en "browsers" et "searchers" - Tauscher et Greenberg [1997] : identifient sept "web browsing patterns" - Huberman, Pirolli, Pitkow et Lukose [1998] : développent une loi mathématique du "surf informationnel" en mettant au jour des régularités fortes dans les habitudes de navigation, régularité fondée sur la profondeur de l'arborescence et ses implications en terme de coûts pour l'utilisateur), un modèle comportemental de la recherche d'information dans un environnement web est établi. Il repose sur :
- Marchionini [1995] : distingue trois types généraux de browsing selon l'objet de la recherche et la systématicité des techniques utilisées :

- Directed browsing
- Semidirected browsing
- Undirected browsing

- Wilson [1997] : quatre catégories de recherche et d'acquisition de l'information :

- Passive attention
- Passive search
- Active search
- Ongoing search

Ce premier modèle est recadré au travers des acquis sur les modèles organisationnels de Aguilar [1967] et de Weick et Daft [1983] qui distinguent 4 modes distincts :

- Undirect viewing
- Conditioned viewing
- Informal search
- Formal search

Les auteurs ajoutent à cela le modèle comportemental de recherche d'information de Ellis

[1989, 1993, 1997] qui étudia les comportements de chercheurs en sciences sociales, de physiciens, de chimistes et d'ingénieurs et isola 6 catégories dans les activités de recherche d'information :

- starting
- chaining
- browsing
- differentiating
- monitoring
- extracting

... modèle assez semblable à celui que développe Marchionini [1995] décomposant le procédé de recherche d'information en 8 sous-procédés qui se développent en parallèle :

1. recognize and accept information problem
2. define and understand the problem
3. choose a search system
4. formulate a query
5. execute search
6. examine results
7. extract information
8. reflect/iterate/stop

Il s'agit dès lors d'intégrer les acquis de ces différentes formalisations et de croiser les modèles de Ellis et de Aguilar, Weick et Daft pour proposer un cadre générique de modèle comportemental et d'itinéraires de recherche d'information adapté au WWW et composé comme suit :

- Undirect viewing = starting + chaining = 12 épisodes / 61
- Conditioned viewing = browsing + differentiating + monitoring = 18 / 61
- Informal search = differentiating + extracting + monitoring = 23 / 61
- Formal search = extracting + monitoring = 8 / 61

Celui-ci permet d'établir une typologie extrêmement fine des modèles comportementaux en usage sur le web en matière de recherche d'information en intégrant et en pondérant les trois paramètres fondamentaux de toute recherche d'information que sont : la nature de l'information recherchée, les techniques de recherche d'information et la finalité d'utilisation de l'information recherchée.

Pour plus d'information : http://choo.fis.utoronto.ca/esproject/

Publication électronique des résultats de la recherche

Journée INRA-INSERM du 29 mars 2000

Annaïg Mahé, doctorante GRESI/Enssib

Enjeux de la publication électronique des résultats de la recherche,
M-F Chevallier - Le Guyader (Direction de l'information et de la communication, Inra) et S. Mouchet (Département de l'information scientifique et de la communication, Inserm)

En 1992-93, ces deux organismes avaient déjà proposé la mise en place d'un serveur de prépublications mais cette initiative n'a alors reçu aucun écho de la communauté des biologistes.
Depuis peu des initiatives se développent telles que e-biomed (US - NIH), e-biosci (Europe, EMBO) ou l'Open Archives Initiative.
L'Inra et l'Inserm sont particulièrement conscients du rôle et de l'importance de l'IST pour la recherche : les budgets consacrés sont importants (la majeure partie va aux périodiques : Inra = 65% sur un budget de 19 MF). L'Inserm met en place des initiatives d'aide au développement de collaboratoires afin de favoriser le partage de l'IST et de transformer l'information brute en connaissance (des plateformes sont mises en place sur la demande spécifique des communautés scientifiques et des colloques électroniques sont organisés)

Questions que ces organismes se posent quand à l'évolution de l'IST et qui ont donné lieu à cette journée de débat :

- les publications doivent-elles être payées par les organismes finançant la recherche ?
- des serveurs de preprints doivent-ils être développés au sein de ces institutions ?
- quelles doivent être les modalités d'évaluation et quelle en sera leur évolution ?
- quel sera l'impact sur le droit moral, le droit patrimonial et le copyright ?

Ghislaine Chartron, animatrice de la matinée, situe le débat dans son contexte historique.
Trois axes majeurs en tissent la trame :

- le développement des nouvelles technologies
- le développement des champs scientifiques
- les politiques publiques comme soutien de l'IST

Dans les années 60 a débuté l'informatisation des bases de données bibliographiques sous l'impulsion des agences fédérales américaines qui financent la recherche. Puis se développent les réseaux (Transpac, TCP/IP) dans les années 80 où l'on peut alors constater l'existence d'une vraie politique de l'IST en France. Apparaissent ensuite les cédéroms ainsi que les bases de données factuelles. Dans la dernière décennie, ce sont les autoroutes de l'information qui entrent en scène ainsi que le web avec les modèles alternatifs qu'il permet tels que le développement des revues électroniques, les bases de données de preprints et les

bibliothèques électroniques, particulièrement aux Etats-Unis et au Royaume-Uni.
Certains chercheurs pionniers ont alors pris conscience du rôle qu'ils pouvaient jouer : Ginsparg, Odlyzko ou Harnad. Ce dernier, chercheur en sciences cognitives, a développé les concepts de "scholarly skywriting", d'"open peer commentary" et de "self-archiving" et un certain nombre de projets comme Cogprints (eLib), Psycholoquy ou le Behavioral and Brain Sciences Journal (BBS), journal qu'il dirige depuis 20 ans, et qui intègre la notion d' "open peer commentary".

Repenser la communication scientifique

Steven Harnad

Le but principal d'un chercheur est la diffusion de ses résultats, qu'ils rencontrent les yeux de ses pairs actuels et à venir. La différence avec un auteur "normal" c'est qu'il ne reçoit pas de revenus pour ce travail (et aussi qu'il y a très peu de lecteurs pour un article). Pour faire évoluer la publication scientifique, il est nécessaire de mettre en place un complot contre la technologie papier ("papyrocentrique") qui est coûteuse et qui n'est plus nécessaire actuellement. Selon Harnad, 90% des usages de la littérature sont possibles sur écran (lecture des sommaires, choix d'un titre, lecture rapide d'un résumé, recherche d'une référence ou d'une phrase ou mot précis, etc.), l'usage du papier pour une lecture de fond étant tout à fait normal (dans ce cas là on imprime). Il existe actuellement des coupe-feu financiers (financial fire-walls) qui barrent l'accès à l'IST : abonnements, licences de sites, pay-per-view. Il est donc nécessaire de libérer la littérature des journaux en ligne à comité de lecture et d'en permettre un accès le plus large possible.
Le problème n'est pas lié à la crise des périodiques mais il est intrinsèque à la littérature scientifique elle-même : les chercheurs ne veulent pas vendre leurs rapports mais les offrir. Le pacte obligatoire avec les éditeurs pour la publication n'est plus nécessaire aujourd'hui. L'argument de la protection de l'œuvre ne vise que la perennité des coupe-feu financiers. Selon Harnad, les revues papiers vont continuer à exister ainsi que la publication dans des revues à fort facteur d'impact et l'expertise des articles par les pairs. Cependant, les institutions doivent aujourd'hui fournir des logiciels d'auto-archivage et les chercheurs stocker leurs articles sur ces serveurs, le but recherché étant la disparition totale des revues, pas seulement la diminution des prix, permettant ainsi un accès direct et à peu de frais pour tous les chercheurs à tous les articles. L'expertise des articles ne représente qu'une faible partie des coûts. Les revues continueront d'exister comme " étiquette " de qualité. Si ce système ne convient pas aux éditeurs, il sera repris par les chercheurs eux-mêmes.

Commentaires et débat :
On constate une certaine contradiction entre la volonté de faire disparaître les revues scientifiques et l'opinion que celles-ci continueront à exister en tant qu' " étiquettes ". Un chercheur constate que ce système est très séduisant mais qu'il est impossible à mettre en place du fait du transfert des droits d'auteurs aux éditeurs. A cela Harnad réplique que sa revue BBS existe effectivement toujours en format papier parce que les auteurs sont trop lents à utiliser la version électronique et il conseille aux auteurs de garder leurs droits d'offrir l'article sur le réseau comme ils l'offrent aux éditeurs. Un autre chercheur constate que les comités d'évaluation ne prennent en compte que les articles publiés dans des revues

à fort facteur d'impact. Susy Mouchet (Inserm) indique que la présence d'informations supplémentaires sur les revues en ligne peut permettre d'augmenter la qualité de l'expertise des articles.

PubMed Central des NIH

Pierre Oudet, Inserm

Pierre Oudet constate que l'état d'avancement des projets américains est en grand décalage avec celui de l'Europe, qui a donc de grandes avancées à faire.
Il présente trois grands projets américains : PubMed Central, BioMed Central (version privée du premier) et HighWire Press. Selon Pierre Oudet, la publication électronique est un autre mode de fonctionnement et non un transfert sur support électronique du support papier. Les résultats de la recherche sont composés de données quantitatives, de graphes, d'images, de données brutes, etc. Le concept de publication doit donc évoluer et les publications actuelles ne sont qu'un indicateur de localisation des connaissances. En biologie par exemple, les publications ne supportent pas les connaissances permettant de reproduire les expériences et les lecteurs intéressés peuvent ensuite prendre contact avec les auteurs via l'adresse électronique pour en savoir plus. Par ailleurs, les banques de produits biologiques qui se développent sont financées par les organismes de recherche. On constate donc une grande variété des supports de connaissances.
Le concept de publication évolue dans l'ère électronique à travers les collaboratoires et les interactions, les échanges d'informations privilégiées dans des cercles fermés. Cela génère des contraintes qui doivent être prises en compte pour la diffusion électronique des résultats :

- le manuscrit doit contenir les références et les liens à toutes les connaissances, réseaux, bases de données connexes
- cela dépasse la forme papier et le format doit être unique pour permettre la mise en place et l'utilisation d'outils de stockage et de recherche performants, cela devra s'imposer de façon progressive.

Il donne un exemple portant sur la recherche sur le cancer dont la communauté a développé une base (CGAP - NCBI) qui contient toutes les informations connues dans le domaine (avec revues ("reviews") dans le domaine, commentaires), les publications devant être directement liées à ces bases de données.

L'accès et le traitement informatique font partie intégrante de la recherche biomédicale et les structures qui financent la recherche doivent y participer activement dans le cadre de coordinations internationales.

PubMed Central
e-biomed, le projet lancé début 99 est devenu PubMed Central et PubMed Central Express. En mai 99, Harnold Varmus fait une annonce publique concernant la mise en place d'un serveur destiné à la mise à disposition des rapports de recherche, libre d'accès et sans coûts pour l'utilisateur. Ce serveur serait pris en charge par les NIH et les instituts de recherche et permettrait un système à double-vitesse avec des documents évalués et d'autres non-évalués. Les buts recherchés sont la diminution des coûts de la diffusion de la recherche et la

nécessité d'un système homogène pour accroître la visibilité de tous les résultats. Pour cela, le serveur doit permettre :

- une mise à disposition immédiate et l'archivage
- des échanges et des collaborations : les informations deviennent des sujets vivants
- des liens vers différentes sortes de données et de bases
- l'organisation et la préservation de l'évaluation scientifique et des politiques de recherche des organismes
- la préservation de la propriété intellectuelle

Les NIH consacrent 100M$ / an pour les abonnements, la publication électronique est évaluée à 2-3M$. Des critiques et des objections sont faites quant à la qualité scientifique des documents, la diffusion des informations cliniques, la faillite des sociétés savantes implicite dans ce projet, l'identité des journaux et leur hiérarchie et la création et l'évolution de la structure.
PubMed Central est mis en place : le serveur a pour but de donner un accès libre à toutes les informations publiées dans le domaine des sciences de la vie. Les informations transitent d'abord par BioMed Central qui transmet ensuite les données validées. Le deuxième site, PubMed Central Express, qui doit contenir les informations non validées, reçoit peu de soutien de la communauté scientifique. Actuellement le serveur est quasiment vide, un travail expérimental est en cours sur deux journaux : Molecular Biology of the Cell et PNAS. Le serveur existe de façon expérimentale en local mais il n'est pas encore accessible internationalement sur le web. La question du format unique n'est pas encore résolue. Le serveur devrait être opérationnel en mai-juin 2000. Son budget pour l'année 2000 est de 2,5M$ (salaires et équipements).

BioMed Central
Il s'agit d'une structure commerciale (Current Science Group) consistant en une association d'éditeurs afin de créer une nouvelle revue complètement électronique avec une structure d'évaluation et des services associés (tels que le développement informatique pour la structuration de données ou la traduction d'articles). Tout article évalué et structuré est transféré à PMC.

HighWire Press C'est une structure financée sur contrat qui fournit l'accès au texte intégral et aux archives (celles-ci étant coûteuses pour les éditeurs, ils acceptent de les transférer sur des serveurs de ce type. Ex : Blood depuis 1997). L'accès au texte intégral est à titre expérimental pour certains journaux, d'autres ne sont plus disponibles gratuitement pour les versions courantes.

Commentaires et débat :
Selon Harnad, si le site de PMC est actuellement vide c'est qu'il n'a pas obtenu les accords des éditeurs. Quant à BMC, il s'agit d'une structure concurrente qui veut prendre la place des éditeurs. Pierre Oudet indique qu'il s'agit de la mise à disposition du texte ainsi que des liens vers les bases de données concernées et que ce travail est plus complexe que celui d'un éditeur. Il ne s'agit pas d'un serveur de preprint classique, les publications n'étant qu'un élément mineur de la recherche. De tels liens par exemple ne sont pas présents chez

HighWire et la démarche de PMC est plus large car il s'agit d'intégrer les spécificités de la recherche biomédicale. Il est en effet important pour le chercheur de pouvoir accéder à toute l'information disponible autour de la publication.
La question de la garantie de l'accès libre est posée, ainsi que le besoin d'un certain débit pour pouvoir accéder à des serveurs situés aux Etats-Unis. Pierre Oudet répond que les Américains ont effectivement la volonté d'ouvrir le système et qu'ils ne souhaitent pas être taxés de monopole. La France entre autres doit donc prendre activement part à ce type d'initiative. Le problème et la solution ont été les mêmes pour les bases de données de séquences : le système n'a pu être verrouillé qu'avec l'entrée de l'Europe.

E-Biosci : projet européen de diffusion des résultats de la recherche,
Franck Gannon, EMBO

Selon Franck Gannon, il n'est pas question ici de bouleversement total mais d'une évolution. Les journaux scientifiques continueront mais avec des changements. Le projet américain e-biomed a des avantages mais comporte aussi des limites :

- celles du domaine : certains journaux ne sont pas inclus et la décision appartient aux NIH
- il existe une certaine mafia des rédacteurs
- le monopole des Etats-Unis pour le stockage de l'information est dangereux pour la sélection de l'information et pour son accès
- BMC pose un problème : que se passe-t-il si d'autres groupes (Inserm, Embo, etc) veulent aussi proposer des articles validés à PMC ?
- si rien ne se fait en Europe, il y a obligation de suivre les Etats-Unis

Le problème qui se posait alors était de savoir qui pouvait parler pour l'Europe. Le rapport Bangemann demandait déjà des initiatives mais c'est le projet américain qui a entraîné une réaction. L'EMBO couvrant un certain nombre de domaines en sciences de la vie, il était une voix possible.
Le projet NIH est un modèle centralisé. En Europe il est possible d'utiliser la richesse de sa diversité en mettant plusieurs sites en place en réseau partagé. La validation reste nécessaire actuellement. BMC peut aussi y soumettre des documents. Le problème actuellement c'est que le contenu n'est pas accessible gratuitement mais une mobilisation des chercheurs pourrait le rendre possible. PubMed de son côté a effectivement du contenu mais les résumés des articles ne sont pas suffisants : il faut pouvoir accéder au texte intégral et pour cela, selon Franck Gannon, il est nécessaire de travailler avec les éditeurs. Le projet consisterait aussi en un comité de pilotage et un comité technique pour la mise en place et la maintenance du réseau partagé.

Commentaires et débat :
S. Harnad : si ce n'est pas gratuit, à quoi ça sert ? Franck Gannon répond qu'une adresse unique pour accéder à du texte intégral, c'est une valeur ajoutée certaine, que l'accès payant peut évoluer et que selon lui les journaux électroniques gratuits vont prendre de l'importance. Le fait que le projet prenne en compte les éditeurs est une façon réaliste de voir les choses et que cela ne signifie qu'il est contrôlé par les éditeurs.

Pierre Oudet indique alors que BMC offre la possibilité d'intégrer toutes les sociétés de tous les domaines souhaitant y participer et que les chercheurs peuvent mettre en place des journaux électroniques s'ils le souhaitent. Faut-il donc prendre le système BMC ou développer un système européen ?
Une personne signale la complexité du problème européen qui entraîne un discours plus conservateur, tandis que les Etats-Unis, plus homogènes, ont plus de facilité pour ce type de projet.
Selon Franck Gannon, il est intéressant que plusieurs projets co-existent afin de voir ce qui peut fonctionner par la suite. Une autre personne pose la question des modes de financements en soulignant le rôle de la publicité. Franck Gannon considère que la publicité n'est pas un désavantage en soi et que les chercheurs peuvent cliquer si cela les intéresse (la publicité existe aussi dans les revues papier et est appréciée par certains lecteurs).
Une personne signale que les grands organismes de recherche pourraient reprendre les revues données aux éditeurs commerciaux. Selon Franck Gannon, cela deviendra inévitable si un projet tel que e-biosci existe. Il faut qu'une structure de ce type existe au préalable pour que cela se fasse. Par ailleurs, un projet européen permettrait aussi de préserver la diversité des langues. Un biostatisticien demande pourquoi les domaines doivent être cloisonnés, n'y-a-t-il pas de possibilité pour un accès universel à tous les domaines. S. Harnad renchérit en demandant l'utilité de ces " territoires ". Susy Mouchet indique qu'il est effectivement intéressant qu'un tel système soit ouvert à tous les domaines, des sciences de la vie aux sciences de l'homme. Par ailleurs, une personne demande si un tel projet se doit de n'être qu'européen et pourquoi. Franck Gannon répond que la Communauté Européenne est très intéressée pour que quelqu'un prenne en charge le projet et il est certain que des financements seraient disponibles.

EDP-Sciences : 5 ans d'édition " online ", bilan et perspectives

Jean-Marc Quilbé, directeur

En 1994, la société a été tirée vers l'éditon électronique par les communautés scientifiques elles-mêmes (physiciens et astronomes). Le serveur a été ouvert en 1995 avec 4 titres sur 9 en ligne. En 1997, des liens ont été mis en place de A&A vers la base ADS de la NASA et vers le CDS de Strasbourg. En 1998, 12 revues sur 13 sont en ligne, la chaîne de production est entièrement internalisée. En 1999, 18 revues sur 19 sont en ligne, des liens d'articles à articles entre the European Physical Journal et Physical Review (APS) sont mis en place, un logiciel de gestion des URL à long terme est développé (Link Manager), les résumés sont gratuits et deux serveurs miroirs sont installés (US, Japon). En 2000, 26 revues sur 28 sont en ligne, le DOI va être implémenté et des compléments multimédia seront rajoutés aux articles (vidéos, simulations).
La question des formats de fichiers stables n'est pas résolue (formules mathématiques encore mal gérées).
C'est surtout en 98-99 que le nombre de connexions/mois a décollé. Actuellement, on compte 80 000 utilisateurs par mois pour 20 000 institutions différentes. En mars 95, 80% des utilisateurs étaient en France et 20% aux USA. Actuellement, la France représente 10%, les USA 30%, le reste de l'Europe varie de 20 à 40%.

Bilan

- la mise en ligne des journaux renforcent la visibilité des articles
- les bases de résumés mettent à égalité les revues européennes et américaines
- les serveurs de preprints coexistent avec les journaux
- les auteurs rejettent les journaux électroniques sans version papier (plusieurs expériences purement électroniques ont été des échecs) : le papier est encore très demandé
- le métier de l'éditeur se transforme profondément :
- en 2001, 50% des effectifs seront directement impliqués dans l'édition électronique contre 10% en 1995
- le serveur web est de 20 Giga octets, il va passer à 35 fin 2000
- des lignes spécialisées pour Internet ont été mises en place et des investissements lourds sont prévus sur le serveur
- les délais de publications ont été réduits : système éditorial électronique (MMS : Manuscript Mangement System)

Bilan commercial ? : pour l'instant, pas de retour sur investissement en vue, pas d'abonné gagné mais aucun de perdu non plus

- il serait intéressant d'attirer d'autres propriétaires de contenus pour offrir des services
- l'édition en ligne est un outil de promotion des revues
- le contrôle par les numéros IP est de moins en moins possible mais par mot de passe il est peu apprécié

Avenir

- le peer-review est la base de la publication scientifique
- convergence entre base de préprints et journaux : les domaines concernés ont besoin du papier, et l'antériorité des informations nécessaires peut se situer à l'échelle d'un siècle

Témoignages de chercheurs sur les nouvelles pratiques de diffusion des résultats de la recherche

Alexandra Louis, Laboratoire génome et informatique, Université de Versailles - Saint Quentin-en-Yvelines

Présente la mise en place d'un système d'information permettant de gérer les co-citations de gènes, les co-occurrences ainsi que la clusterisation pour mettre en évidence des familles de gènes avec des liens vers les bibliographies concernées (voir résumé)

Philippe Bessières, Mathématiques, Informatique et Génôme, Inra, Jouy-en-Josas

L'efficacité des outils en Biologie Moléculaire s'est développée de façon foudroyante et cela donne en retour des masses de données à croiser. Le laboratoire est une toute nouvelle entité qui montre l'évolution en cours dans le domaine. Il s'agit d'une démarche d'encyclopédiste. L'information est assez bien structurée en Biologie Moléculaire (ce qui n'est pas le cas pour d'autres domaines en Biologie) : la richesse des annotations textuelles donne un réseau de

connaissances riche. Les entités et les concepts décrits sont systématiquement associés à des références bibliographiques. Pour qu'ils puissent être intégrés, les résultats doivent être reliés aux connaissances déjà disponibles. Par ailleurs, des mêmes entités sont développées selon différents points de vue qu'il faut pouvoir confronter (génétique, chimie, physiologie, etc). Un navigateur de cartes du génôme a été développé, Micado, qui permet une navigation graphique. Un projet (Quadrige) est en cours entre l'Inra et la faculté d'Orsay, qui permet la catégorisation des co-citations et co-occurrences mais aussi le sens de ces co-citations (travail linguistique).

Les universités face au développement de l'édition électronique

Claude Jolly, sous-directeur des bibliothèque et de la documentation, MENRT

Les universités sont diverses ainsi que les situations en édition électronique. Les établissements universitaires sont des producteurs et des consommateurs d'informations. L'explosion électronique a eu pour conséquences :

- l'obligation pour les différents métiers de travailler ensemble
- des changements dans l'organigramme des universités

En ce qui concerne les thèses, de nouvelles propositions vont être faites afin de faire évoluer le dispositif actuel qui ne correspond pas du tout à la réalité des besoins. Les universités ont clairement un rôle de diffusion de l'information. En ce qui concerne l'évolution des services de documentation, les fonctions de bases n'ont pas changé mais les conditions d'exercice oui. L'émergence des ressources électroniques a permis un nouveau positionnement des SCD dans l'université. Le papier privilégie un système éclaté mais l'électronique entraîne une politique d'établissement qui a rapproché les arbitrages documentaires de la Présidence de l'Université. La sélection, la validation et l'évaluation de l'information sont devenues déterminantes. La mutualisation des ressources est devenue nécessaire. Des groupes de veille thématiques ont été mis en place et on a vu l'apparition de groupements d'achats en vue de l'acquisition des ressources électroniques. La formation des usagers prend encore plus d'importance. Le problème crucial reste celui de l'archivage et de la conservation et pour cela il faut prendre en compte les différences disciplinaires en ce domaine, certaines disciplines étant accumulatives, d'autres non.

Publication électronique des résultats de la recherche : quel impact sur l'évaluation ?,

Jean Zinn-Justin, SPhT, CEA

Rédacteur de plusieurs revues, co-éditeur d'un journal international (Elsevier), collaborateur d'EDP-Sciences, collaboration avec Ginsparg lors du démarrage du serveur de preprints. Actuellement, le serveur de Los Alamos a accumulé 127 000 articles depuis 1991, et le taux de soumission augmente tous les mois. Ses avantages sont la rapidité, l'accessibilité, la gratuité du texte intégral, l'adresse unique, la collaboration avec le SLAC pour les citations, la garantie de la perennité de l'information. Ses limites sont l'aspect rudimentaire des outils de recherche, l'absence de validation rapide et explicite (elle est implicite dans le temps via les références croisées et les citations mais elle doit être rapide et explicite pour l'évaluation de jeunes chercheurs et pour l'obtention de financements), l'absence de hiérarchisation de

l'information. Certaines tâches ne peuvent être automatisées entièrement : elles nécessitent l'organisation et la collaboration entre chercheurs.
Les archives remettent en cause le rôle, le fonctionnement et l'existence des journaux scientifiques. L'abonnement est une barrière pour l'accès à l'article et constitue une parcellisation de l'information. Mais en même temps ces revues sont le seul moyen de validation du travail des chercheurs (il n'est pas nécessaire qu'il soit en ainsi mais il l'est de fait). La validation n'est pas le rôle primaire des revues (diffusion) mais elle est devenue nécessaire. Le rôle primaire leur a été enlevé et il y a actuellement une insistance sur la validation.
Les résistances au changement sont dues au fait que les technologies progressent trop rapidement pour qu'on ait le temps de prendre des habitudes. Par ailleurs, les maisons de publications résistent aussi car elles doivent s'adapter de leur côté. En ce sens, les chercheurs ont certainement une influence sur les sociétés savantes dont ils sont membres.
Le changement de validation doit être fait par toutes les communautés à un niveau international. Le paradoxe est que les chercheurs continuent de publier dans les revues auxquelles ils déconseillent aux laboratoires de s'abonner. De plus, les chercheurs sont rarement conscients des enjeux économiques. Dans la situation actuelle ce sont malheureusement les revues européennes qui sont les plus menacées. En effet, le marché américain étant plus grand, les coûts fixes sont plus stables. Il est difficile pour un chercheur français d'abandonner de préférence les revues européennes, même si celles-ci sont les plus chères !
Le but n'est pas de forcer le changement, le chercheur ne changera que lorsqu'il en aura envie. L'évaluation des publications scientifiques est importante à la fois pour le lecteur et pour l'auteur. Au lecteur, elle permet une validation immédiate : une information scientifique, pertinente et correcte (sur le serveur de Los Alamos, l'adresse du laboratoire est obligatoire, c'est un premier niveau de sélection). Pour l'auteur, elle permet une reconnaissance de l'originalité de son travail.
La publication électronique permet d'améliorer cette validation par l'ajout de commentaires. A l'évaluation directe par les comités de lecture, il existe des alternatives : la validation spontanée, les commentaires explicites. Ce sont des modèles dynamiques d'évaluation. Pour cela, ils ne sont pas visibles à court et moyen terme. Par ailleurs, la validation par la citation peut être fausse car un article faux peut être cité souvent justement parce qu'il est faux.
L'électronique permet la souplesse d'une nouvelle validation, avec une organisation plus rationnelle, plus efficace et moins coûteuse.

Claude Kordon, Inserm, Paris

Rédacteur en chef d'un journal international, ancien chef de laboratoire.
Selon lui, le salut est dans un cadre européen. L'autonomie scientifique de la communication est très importante : il faut éviter le monopole américain. Le système de validation scientifique est paradoxal car il entraîne une pression à la publication très forte, ce qui peut engendrer des fraudes. L'électronique donne une piste majeure pour assainir la situation mais cela ne marchera que si c'est gratifiant pour le chercheur. Il faut que son travail soit protégé contre le pillage et que son travail soit reconnu. Actuellement, l'impact des publications électroniques est quasi nul dans la plupart des institutions de recherche.
Lui-même a eu l'occasion de proposer une version électronique du journal mais les

Américains ont réagi très fortement, considérant que cette version serait la poubelle de la version papier.
Il paraît donc indispensable d'associer au niveau européen les sociétés savantes, les rédacteurs européens de revues, les éditeurs et les organismes de recherche. La fondation européenne pour la science serait un cadre possible. Cela permettrait de négocier les conditions de validation des données de journaux électroniques et de changer les règles. Une déontologie des rapporteurs est nécessaire mais il faut des moyens volontaristes pour cela.
Une des meilleures solutions est sans doute de mettre en place des expériences pilotes dans certaines disciplines afin de pouvoir avoir une macro-perspective. En effet, les habitudes de publications varient d'une discipline à une autre et les chercheurs n'auront confiance que si le fonctionnement est convivial et adapté par les pairs.
Il faut aussi définir des produits originaux. Il ne pense pas que les journaux électroniques vont faire disparaître le papier mais ils doivent être originaux pour avoir leur place. Il faudrait effectivement faire une recherche sur les moteurs de recherche qui reproduise davantage la logique du chercheur afin d'aider le chercheur dans sa recherche et en même temps de le laisser dans un terrain familier, cela permettant des recoupements. Une expérience avait débuté à l'Inserm en neurosciences : il s'agissait d'un système de graphes avec un graphe maître en neuroanatomie et une superposition de plans différents selon des points de vue disciplinaires différents. Il y a une forte connotation disciplinaire mais c'est effectivement une des voix prometteuses pour obtenir une validation interne.
En conclusion, Claude Kordon propose de mettre en place des stratégies complémentaires mais différentes des Américains afin de devenir un interlocuteur privilégié.

Synthèse

Ghislaine Chartron

La matinée a été consacrée à l'usager final, à l'accès via un point unique et une norme du web.
Trois propositions ont été faites :

- l'auto-archivage et la libération de la littérature scientifique par Steve Harnad. il s'agit d'une mouvance très forte dans certaines communautés
- l'organisation d'archives libres à partir des éditeurs, NIH, EMBO
- l'achat de l'offre des éditeurs par les bibliothèques, un système assez conservateur et qui se verrouille actuellement (cf ScienceDirect)

L'après-midi a porté sur les possibilités électroniques de mise en place de systèmes de connaissances et non plus seulement sur la gestion des publications. Il s'agit là de l'intégration de toutes les données de la recherche. Le deuxième thème était l'évaluation des processus de validation. En conclusion, certaines options pourraient en fait reconduire le système actuel sans grand changement structurel.

Commentaires et débat

Une hypothèse nouvelle a été faite : celle de la validation ouverte. Harnad indique que cela existe déjà dans sa revue BBS. Zinn-Justin l'a proposé pendant 6 ans sans succès à son

comité de rédaction.
Une personne indique que la seule façon de convaincre le chercheur, c'est d'avoir une caution institutionnelle minimum. Une autre dit que le problème de la publication n'est pas lié à l'électronique mais à la pollution des publications. Deux solutions sont possibles : le putsch ou une initiative semblable à celle des Etats-Unis. Claude Kordon déclare qu'il est plutôt pour les putschistes mais qu'il vaut mieux être reconnu pour cela. Il serait possible de mettre en place des revues électroniques spécialisées pour les résultats négatifs ou curieux que les bonnes revues ne veulent pas publier. Par ailleurs il est nécessaire d'obtenir une modification des processus d'évaluation dans les universités. Susy Mouchet, en guise de conclusion, indique qu'il y a une distinction à faire entre la publication et les données autour et que les liens entre eux sont importants à faire. Un développement disciplinaire de la culture de preprint est à mettre en place. Au cours de cette journée n'a été qu'effleuré ce qui va évoluer de fait. Les besoins et la mobilisation des chercheurs seront déterminants.

Publication électronique des résultats de la recherche

Synthèse

Journée INRA-INSERM du 29 mars 2000

Ghislaine Chartron, Maître de conférences - URFIST de Paris-ENC

Cette journée organisée à l'initiative des Directions de l'Information et de la Communication des deux organismes de recherche (INRA et INSERM) avait pour objectif majeur de recueillir l'avis des chercheurs concernant les projets à mener dans le cadre des services électroniques dédiés à la circulation de l'IST. La journée a rassemblé 220 personnes, moitié chercheurs et moitié professionnels de l'IST dans les anciens locaux de Polytechnique, au ministère de l'Education nationale, de la Recherche et de la Technologie.

En complément du compte rendu détaillé rédigé par Annaïg Mahé, nous voudrions souligner ici certains points qui ont attiré notre attention :

- *L'économie actuelle du système de publication des résultats de la recherche est contestée par plusieurs chercheurs, au-delà des "pionniers".*

  Le système actuel est qualifié d'anti-économique par S.Harnad et Z. Justin qui dénoncent la situation paradoxale où l'auteur d'un article ne demande aucune rétribution pour diffuser ses résultats mais où le lecteur se trouve confronté à des barrières économiques grandissantes en terme de coûts d'abonnements. La démonstration est simple et elle interroge aujourd'hui un certain nombre de

chercheurs, vu les réactions de plusieurs participants à cette journée.
L'enjeu d'une économie symbolique d'échanges des résultats de la recherche au niveau international est présent dans la réaction de certains chercheurs.
Des chiffres sont avancés par Z. Justin : une revue électronique coûterait 70% du prix du papier, on pourrait estimer à 10% du prix de la revue la part concernant le travail d'évaluation.
Avec les potentialités de diffusion des réseaux, une proposition forte défendue par S. Harnad est justement de découpler la fonction de diffusion de la fonction de validation, les revues devant garder selon lui uniquement la fonction de validation, la diffusion étant notamment assurée par l'auto-archivage sur des sites Internet (à des niveaux qui restent à définir). La question est alors de savoir si l'ensemble des revues doivent être maintenues, si le recentrage vers les sociétés savantes ne suffirait pas ? Une tendance prévisible serait le déplacement des coûts vers l'auteur et son laboratoire, payant un ensemble de services associés à la publication électronique qui lui seraient proposés ; c'est dans cette orientation que semble s'inscrire le projet BioMedCentral, comparable à un éditeur d'une méga-revue associée à PubMedCentral qui, lui, est un archivage des articles organisés par les NIH.

- *Au niveau politique, se renouvellent les débats de fonds public/privé et les débats sur le risque d'un monopole américain :*

  Ces débats ont déjà eu lieu dans les années 70-80 principalement autour du développement des banques de données bibliographiques et avec pour conséquence majeure la mise en place de différentes politiques d'IST dans de nombreux pays. L'enjeu actuel de l'accès au texte intégral renouvelle ces débats avec une autre dimension, celle d'une politique européenne coordonnée face aux projets américains. La peur d'une dépendance américaine est présente dans le discours de plusieurs intervenants, l'éclatement des centres décisionnels en Europe est souligné comme un handicap important face à la situation américaine. Le modèle rééquilibré des bases mondiales de séquences génétiques est mis en avant.
  Quant à la répartition des initiatives entre le secteur privé et les institutions publiques finançant la recherche, les avis sont très partagés :
  Pierre Oudet défend une politique de l'édition intégrée à une politique de la recherche, Franck Gannon conçoit le projet E-Biosci en partenariat avec les éditeurs, dans la ligne politique qui semble aussi être défendue par la commission européenne. Rappelons aussi pour l'histoire que le serveur de pre-prints de Ginsparg s'est vu également refuser les subventions de la National Science Foundation à ses débuts, c'est le rapport de force instauré par la

communauté des chercheurs qui a changé l'orientation politique. De toute évidence, les revendications des chercheurs pèseront de tout leur poids dans les appuis aux inititatives publiques.

- *La stratégie de la publication électronique par rapport à la publication papier dans le domaine biomédical :*

  Les avis exprimés lors de cette journée sont divers. Pour S. Harnad, l'avenir est uniquement électronique (abandon du papyrocentrisme, ce qui n'empêche pas une impression papier pour la lecture approfondie d'un article). Pour P. Oudet et S. Mouchet, l'avenir est aussi électronique mais essentiellement pour des raisons cognitives : l'enjeu est celui d'une " publication évoluée ", publication non plus d'un texte mais d'un ensemble de connaissances que permet l'électronique et qui correspondrait à la réalité de nombreuses recherches dans le domaine biomédical ; l'enjeu est ici de permettre réellement la reproductibilité des expériences à partir de l'acte de publication. Pour C. Kordon, la vision est plus une complémentarité des deux supports : l'électronique pouvant notamment publier les informations utiles mais non prises en charge actuellement par l'édition papier : les matériaux, les résultats négatifs, les informations cliniques...

- *L'électronique met en exergue certains dysfonctionnements du système de publication.*

  L'inflation du nombre de revues a été évoquée par plusieurs intervenants, C. Kordon en particulier relève que parmi les 4000 revues indexées en sciences de la vie, 75% ont un facteur d'impact inférieur à 1.
  L'organisation de la fonction d'évaluation est mise en doute également par plusieurs chercheurs : mafia, piratage, fraude sont évoqués. L'électronique pourrait permettre, d'après certains, d'assainir cette procédure d'évaluation, d'augmenter la qualité de l'expertise d'une part en permettant un traçage et d'autre part en donnant les possibilités d'une évaluation dynamique plus ouverte.

- *Quels projets innovants ?*

  Le ministère de l'Education nationale, de la Recherche et de la Technologie vise à faire avancer très rapidement l'accès électronique aux thèses en accord avec leur auteur, c'est un début notable annoncé par C. Jolly, sous-directeur des Bibliothèques et de la Documentation. De son côté, l'Inserm va également favoriser le développement de serveurs dédiés à la valorisation des publications de ses chercheurs, 3 IFR seraient initialement impliquées. Une coordination EPST-Universités apparaît cependant de plus en plus indispensable.
  Par ailleurs, une réflexion sur le transfert actuel des droits de

l'auteur à l'éditeur semble être amorcé, point de départ incontournable pour penser les dispositifs d'accès à l'IST. Concernant le projet européen E-Biosci, celui-ci semble miser uniquement sur un archivage à long terme des revues des éditeurs, avec une fraîcheur décidée par l'éditeur, ne remettant que partiellement en cause l'économie actuelle du système.

En conclusion, autour de la publication électronique des résultats de la recherche, on assiste aujourd'hui à des débats, des initiatives qui peuvent paraître contradictoires : dans un souci de mise à disposition de nouveaux services, les bibliothèques signent des contrats pour offrir l'accès aux bibliothèques numériques des éditeurs à leurs lecteurs, l'Inserm vient notamment de signer une convention sur 3 ans avec Elsevier pour rendre accessibles 100 revues de l'offre ScienceDirect à partir des 4000 postes de consultation situés au sein de l'institution scientifique. Par ailleurs, des initiatives commencent à être prises pour encourager la libre diffusion des résultats de la recherche…

La conduite d'actions semble en fait très pragmatique : assurer une continuité de services, expérimenter en parallèle d'autres services et décider à court terme des choix à reconduire ou non. Mais le risque sera peut-être alors de perdre une vision politique au profit d'une dépendance grandissante à de nouveaux services éditoriaux commerciaux à forte valeur ajoutée.

A propos des moteurs de recherche

Cours sur la recherche d'informations sur Internet (février et avril 2000, URFIST de Toulouse)

François Magnan, Documentaliste

Imaginez-vous recherchant de l'information dans la plus grande bibliothèque du monde, où les documents seraient disposés sans rangement particulier ni fichier central : cela serait un cauchemar pour le bibliothécaire et pour l'usager. Peut-on comparer le web avec ce type de bibliothèque ? Pas exactement. Si le web ne dispose pas d'un catalogue central, vous y trouverez des dizaines d'outils de recherche différents : chacun a sa propre base de donnée, son propre langage d'interrogation, ses capacités en terme de recherche et de présentation des résultats. Pour celui qui désire profiter des ressources disponibles sur le réseau Internet sans y passer de trop nombreuses heures, il est alors nécessaire de se familiariser avec plusieurs outils de recherche et de développer des techniques efficaces de recherche d'informations. Il sera ici question des moteurs de recherche du Web, mais aussi des répertoires de sites et des métamoteurs de recherche. Après avoir exposé les principaux traits de ces outils et quelques conseils d'utilisation, nous verrons que cette typologie est menacée.

Outils et méthodes

Les répertoires (parfois appelés annuaires) signalent des sites sélectionnés, classés et commentés par des équipes de "surfeurs". La version française deYahoo, (www.yahoo.fr) le plus célèbre d'entre eux, répertoriait 85.000 sites au mois de décembre 1999. On peut rechercher l'information dans les répertoires :

- soit en parcourant les catégories;
- soit en faisant de la recherche par mots-clefs. En général cette recherche se limitera au champ "commentaire" des sites sélectionnés remplis par le webmaster du site ou par un "surfeur"

Particulièrement appréciés par les débutants pour leur simplicité d'utilisation, les répertoires sont de bons outils lors d'une recherche d'information d'ordre général, comme :

- quels sont les principaux sites sur l'emploi ?
- quels sont les sites des Universités parisiennes ?

A la différence des répertoires, les moteurs de recherche du Web sont des outils purement automatiques. Ils servent à faire de la recherche par mots - clefs. Un moteur de recherche tel qu'Altavista (www.altavista.com) fonctionne grâce à trois outils :

- un robot dit "spider", ou "araignée", qui parcourt le Web afin de capturer les informations contenues dans les pages du Web. Celui d'Altavista répond au doux nom de "scooter";
- un index qui réunit ces informations;
- Le serveur que vous interrogez compare vos mots-clefs avec l'index puis donne en réponse des liens hypertextes pointant vers des pages web.

Les moteurs de recherches indexent des millions de pages et seront les meilleurs outils lorsque vous rechercherez des documents sur un sujet très spécifique. Si la quantité de pages indexées par les moteurs de recherche est bien plus importante que le nombre de sites relevés par les répertoires, cette profusion se fait au détriment de la qualité de la sélection, et selon une étude parue dans "Nature" en juillet 1999, aucun moteur de recherche n'indexe plus de 16% des documents du réseau Internet.

Les métamoteurs interrogent plusieurs outils de recherche en même temps, qu'il s'agisse de répertoires ou de moteurs de recherche. Ils seront donc utilisés avec profit face à un problème de silence documentaire. Les plus performants éliminent les doublons contenus dans les réponses et vous indiquent leurs provenance, moteur par moteur. Comme les métamoteurs regroupent les possibilités d'outils dont les fonctionnalités et les syntaxes varient, ils n'accepteront que l'emploi des opérateurs : AND, OR, AND NOT et la recherche pas phrase.

Répertoires de sites, moteurs de recherches, métamoteurs : cette typologie est commode et permet au chercheur d'information d'adapter ses techniques à l'outil qu'il rencontre ; or il est de plus en plus difficile d'identifier l'outil auquel on est confronté.

Les évolutions :

Les moteurs de recherche et les répertoires de site sont soumis à une course à l'audience : être un simple outil de recherche n'est plus suffisant. Tous se sont orientés vers ce que l'on appelle un "portail" et offrent beaucoup d'autres services : courrier électronique gratuit, météo, dépêches d'actualité. Allez sur le portail de Voila http://www.voila.fr, vous pourrez rechercher de l'information sur le moteur de recherche Voilà ou des sites sur le répertoire QuiQuoiOù , créer et gérer une liste de diffusion, vous informer sur l'actualité récente, consulter des cartes géographiques, participer à des séances de "chat". Le paroxysme de cette situation est peut-être atteint par Dejanews http://www.deja.com, à l'origine une base de donnée archivant les messages envoyés dans différents forums, où les possibilités de recherche sont de plus en plus dissimulés derrière l'information commerciale.
Lors de leurs débuts sur le Net, les répertoires étaient réellement des répertoires et les moteurs de recherche des moteurs. Aujourd'hui cela se complique : afin de combiner les avantages propres à chaque mode de recherche, les outils sont de plus en plus hybrides. Ainsi, un bref panorama de quelques grands outils francophones nous montre que :

- Le moteur de recherche AltaVista est lié au répertoire de Lokace, mais il ne fait jamais appel à ce dernier lors d'une requête.
- Le répertoire de Yahoo France est lié au moteur de recherche Inktomi, mais il ne passe la main à ce dernier que lorsqu'il ne trouve pas de réponse à une requête dans les sites du répertoire.
- On trouvera sur Voila le moteur de recherche éponyme et le répertoire de site QuiQuoiOù. Après une requête, les résultats du répertoire, jugés plus pertinent que ceux du moteur, sont proposés en premier, distingués par une icône représentant une boussole, puis viennent les résultats du moteur de recherche, accompagnés d'une loupe.

Doit-on pour autant annoncer la disparition de cette distinction entre annuaires et moteurs de recherche ? En fait, il convient de différencier l'interface des différentes bases de données auxquelles elle donne accès. Ainsi, le portail de Yahoo France permet d'accéder aux bases de données suivantes :

- celle de son répertoire de sites
- celle du moteur de recherche fournit par Inktomi
- celle d'une agence de presse

Si la spécifité de chaque source d'information reste valable, le sens de l'observation devient donc une qualité primordiale du bon chercheur d'information : il ne suffit plus de connaître les techniques appropriées à chaque outil, il doit savoir les distinguer les uns des autres. Or cette observation se complique car le monde de l'Internet évolue à une vitesse particulièrement rapide. A titre d'exemple, voici quelques unes des nouveautés apparues dans le paysage du surfeur francophone entre le mois de février 2000 et le début du mois d'avril 2000 :

- Une version française du moteur de recherche Altavista est apparue : il ne s'agit pas d'une simple interface en français, mais bien d'une adaptation : ainsi la base de donnée d'Altavista France est différente de celle d'Altavista Etats-Unis
- Un nouveau moteur de recherche conçu en France, Albert, permet de poser des questions en langage naturel
- Un nouveau moteur de recherche français, Souany est apparu
- Un nouveau moteur de recherche, Caloweb : semble n'être qu'une réplique française du moteur de recherche américain Google, sans adaptation particulière ..... pour l'instant
- Un nouvel annuaire de sites francais : www.eclairnet.com est apparu et propose des pages régionales

Il n'est bien sûr pas nécessaire de connaître chaque nouvel outil pour rechercher de l'information sur le réseau, mais toutes ces nouveautés bouleversent régulièrement l'univers des outils de recherche et amènent de nouvelles possibilités utiles aux professionnels de l'information qui ont donc intérêt à s'informer régulièrement des évolutions.

Quelques sources d'informations :

1 / Situation actuelle et perpectives des grand outils de recherche :

Jean-François Vincent : "Scies, limes et marteaux: chronique des outils de recherche par internet : ouragan sur la typologie?". La Lettre du bibliothécaire québécois, n°19, Août-septembre 1999.
http://www.sciencepresse.qc.ca/lbq/lbq19.1.html
Greg R. Notess : "A Multiplicity of Databases on Search Engines". EContent, octobre 1999.
http://www.ecmag.net/EC1999/net10.html
François Bourdoncle et Patrice Bertin : "Recherche d'aiguilles dans une botte de liens". La recherche, n°328, pp. 66-71, Février 2000.

2 / Mesurer le web :

Steve Lawrence and Lee Giles : "Accessibility of information on the web". Nature, Vol. 400, pp. 107-109, juillet1999. http://www.wwwmetrics.com A cette adresse, on trouvera les principales conclusions de l'étude et la possibilité de se procurer gratuitement l'intégralité de l'article par courrier électronique.
Pour trouver le compte-rendu d'une étude mesurant le web faite conjointement par Inktomi et le NEC Research Institute en janvier 2000.
http://www.inktomi.com/webmap/

3 / Se former à la recherche d'information sur Internet :

CREPUQ (Canada) :"GIRI : Guide d'initiation à la recherche dans Internet", édition révisée en 1999.

www.bibl.ulaval.ca/vitrine/giri/

Jean-Pierre Lardy : "RISI : Recherche d'Information sur l'Internet", édition révisée en 2000.
www.adbs.fr/adbs/sitespro/lardy/toc.htm

Alfred et Emily Glossbrener : "Moteurs de recherche pour le web", Ed. First Interactive, 303 p, avril 1999.

4 / S'informer sur les évolutions des outils de recherche :

La liste de diffusion "Moteurs de recherche" ou "Motrech", consacrée aux moteurs de recherche sur Internet. liste non modérée, administrée par Jérôme Charron.
http://www.veille.com A cette adresse, on trouvera les instructions pour s'inscrire.

"Search Engine Watch", un site d'information sur les moteurs de recherche, géré par Danny Sullivan.
http://www.searchenginewatch.com

"Abondance", un autre site d'information sur les moteurs de recherche, géré par Olivier Andrieu.
http://www.abondance.com

5 / Des sources pour se former, s'informer et former les autres :

- Les sites des URFIST

Revue de Presse :

Netsurf n°48, mars 2000

* Un dossier : recherches d'informations sur Internet p50-62

- Mieux utiliser les moteurs de recherche : les règles de base : un comparatif des métamoteurs à installer sur votre disque dur : Sherlock, Webferret, Copernic 2000, Inforian Quest et Subject Search Spider - Une interview d'Olivier Andrieu : l'avenir serait aux outils thématiques
- Un article : les veilleurs de l'internet, avec une interview de Guillaume Lory, directeur technique de Cybion *
- "Un moteur pour votre site" : les solutions pour installer un moteur de recherche sur votre site Internet
- "Classez vos e-mail avec Outlook Express 5.0" : classez, trier et filtrer votre courrier électronique
- "Les éditeurs HTML wysiwyg" : quelques éditeurs HTML professionnels au banc d'essai : Dreamweaver, Golive, Coldfusion,
- Frontpage : Sur un cd-rom joint avec le magazine, l'agent intelligent u-map de trivium en version d'évaluation