

Outils, modalités et stratégies d'accès à l'Information Scientifique et Technique sur Internet pour les besoins d'enseignement et de recherche

Mokhtar BEN HENDA

Université La Manouba, Institut Supérieur de Documentation

Avril 2001

1. Préambule : définition des concepts :

Par Internet on sous entend toutes ressources et tous services conçus et fonctionnant sur le principe du travail collaboratif, à distance, ouvert et distribué en mode client serveur. Il s'agit en fait d'un environnement de technologies communicantes qui servent à rendre des services d'information et de communication sous forme de ressources d'Information Scientifiques et Techniques.

- Les technologies communicantes : Une connectivité (Réseaux, Télécommunications) qui prend forme de réseaux Intranet, Extranet et Internet. Ces technologies communicantes ont permis l'évolution des systèmes d'information et de communication des systèmes fermés aux systèmes ouverts et distribués.
- Les services Info-Com : Un rôle de génération, de traitement, d'édition, de communication et échange de l'IST.
- IST : Toutes données susceptibles d'engendrer une information à son tours capable d'engendrer un état de connaissance ou/et de prise de décisions. C'est la matière première des universités et des centres de recherche. « *L'information savante d'aujourd'hui, recouverte par le terme "information scientifique et technique" , est définie à la fois par des contenus, constitutifs de connaissances, par des supports documentaires et des canaux de communication spécifiques.* »¹

2. Problématique :

A l'ère actuelle des réseaux planétaires, la recherche et l'accès aux ressources d'information scientifique et technique change d'approche : la surabondance des ressources réoriente la quête des données d'un paradigme de recherche et de localisation à un problème de filtrage et d'application d'ordres de pertinence.

Le contexte universitaire, consommateur par excellence de l'Information Scientifique et Technique se trouve confronté à des enjeux de qualité, de validation et de valeur ajoutée inhérente à l'information trouvée, filtrée et échangée.

Tant sur le plan de la recherche comme sur le plan de l'enseignement, les universitaires et les étudiants se trouvent contraints à faire un choix approprié des outils, des sources et des méthodes à utiliser pour entamer le processus de la recherche d'IST sans subir les effets indésirables du silence et du bruit d'information.

¹ URFIST. Paris, Information scientifique et technique. <http://www.ccr.jussieu.fr/urfist/def-ist.htm> 25/04/2001

Cette intervention aborde dans sa dimension générique, les outils, les modes et les méthodes de recherche d'IST pour les besoins des enseignants, chercheurs et étudiants.

3. Questionnements :

- Quelles sont les mutations clés dans les systèmes d'information et de communication d'une part et dans la nature des ressources d'information, des outils et des méthodes de leurs création, traitement, diffusion, recherche et accès ?
- Par quel moyens l'université adhère-t-elle à l'évolution de ces différents aspects du nouvel environnement informationnel ?
- Comment l'enseignant et l'apprenant accueillent-ils cette métamorphose structurelle dans leurs approches cognitives des systèmes d'information et de communication ?
- A quel degré y a-t-il eu une adaptation aux nouveaux modes d'usage des NTICs pour exploiter au mieux les ressources d'information et de communication à des fins de recherche et d'enseignement ?

Cet exposé essaiera de répondre sommairement à ces interrogations par l'avancement de certaines définitions qui font référence dans les milieux spécialisés. Il sera également question de passer en revue certaines expériences et études de cas conduites dans des milieux académiques multiples.

4. Internet : les polémiques de la qualité, de la fiabilité et de la rentabilité de l'information dans un processus de recherche

A ne laisser parler que les statistiques, il est facile de constater l'envergure considérable du développement des contenus et des usages d'information autour d'Internet.

Selon les dernières statistiques compilées² par l'Université de Nice³, il y aurait, selon le recensement de novembre 2000, environ 407 millions d'internautes répartis comme suit :

| | |
|-----------------|----------------------------|
| USA/Canada | 167 millions d'internautes |
| Europe | 113 millions |
| Asie/Pacifique | 104 millions |
| Amérique latine | 16 millions |
| Afrique | 3,11 millions |
| Moyen-Orient | 2,4 millions |

Selon une étude réalisée par le Benchmark Group, le Web serait constitué de quelques 2,9 milliards de pages réparties en presque 22 millions de sites.

On estime l'évolution du contenu d'Internet à un million de pages par jour avec plus d'un milliard de liens entre les pages.

² Compilées à partir de http://www.nua.ie/surveys/how_many_online/index.html

³ Université de Nice. Statistiques d'Internet. <http://www.unice.fr/BU/lettres/pstat.html> (25/04/2001)

Le développement rapide, et quelque part anarchique d'Internet, malgré les efforts de la normalisation et de la standardisation que les gardiens du temple (WC3, IETF, IAB, etc.) essaient de lui garantir, a engendré un état de confusion et de diversification incontrôlables dans les formats et les supports des contenus, dans la qualité et l'authenticité des ressources d'information et la notoriété de leurs auteurs. Les serveurs d'hébergement gratuit sont en quelque sorte un échantillon de cette vocation bric-à-brac de cette tendance que prend Internet : des sites conçus par n'importe qui présentés n'importe comment avec des contenus flous et parfois faux côtoient des sites de grande valeur scientifique, drainant des trafics considérables et des effets de valeurs ajoutées importants.

Des études de comportement dans des situations de recherche ont été menées pour observer le contexte général de la recherche sur Internet d'un point de vue masse d'information, outils et usages. Des constatations ont pu être dégagées.

Premier constat : La quantité d'information en ligne sur le réseau Internet a tellement pris de l'ampleur que toute l'activité de recherche et d'accès à l'information a pris une nouvelle tournure : alors qu'il était question de trouver une information infime soit-elle pour un sujet de recherche quelconque, il est désormais question de filtrer parmi une masse considérable de données restituées sur Internet, celles qui représentent un taux optimal de pertinence par rapport à une requête posée.

Deuxième constat : De façon quasi unanime, tous les utilisateurs sont conditionnés par le recours aux moteurs de recherche les plus connus pour rechercher presque exclusivement sur le Web. Les archives des News, les OPACS, les revues spécialisées etc. sont totalement ignorés en tant que réserves d'IST parfois plus importantes que les ressources du Web.

Troisième constat : De part leur nature restrictive quant à leur champs d'action (couvertures géographique, linguistique, thématique etc.), les outils de recherche couramment utilisés par la grande majorité des internautes, restent impuissants devant les exigences des utilisateurs qui veulent toujours localiser rapidement l'information voulue avec des marges importantes d'authenticité et de fiabilité.

Quatrième constat : Les modes d'usage dans la recherche de l'information restent aussi souvent rudimentaires par rapport aux capacités réelles des outils. La réussite d'une recherche d'information dépend plus souvent de la stratégie cognitive et du raisonnement organisé de l'utilisateur que de la maîtrise des outils technologiques elles mêmes. En effet, la plupart des personnes procèdent souvent à des méthodes simplistes et aveugles de recherche sur des outils qu'ils ont tendance à pérenniser sans recherche d'alternatives.

Or, les outils de recherche et leurs performances d'identification, de localisation, de filtrage et de traitement des documents sur Internet tendent, sous l'effet de la concurrence à améliorer leur potentiel analytique des requêtes (langage combinatoire, langage naturel, etc.), à offrir davantage de services de valeurs ajoutées (traductique, traitement graphique etc.), sans pour autant que les usages progressent vers ces nouvelles formes d'optimisation des requêtes.

Moins évidents encore sont les usages post-recherche : comment organiser l'information recueillie, comment la traiter pour en extraire une autre à des fins d'analyse, de communication et de prise de décision? Là il ne s'agit plus de trouver l'information tout court. Il est plutôt question de construire la réponse à des questions posées pour aboutir à une

décision ou entamer une action. Il s'agit, outre la procédure de recherche à laquelle Internet est destiné, de faire preuve d'intelligence de traitement, de gestion d'idées à travers des outils additifs qui ressortent du domaine de l'intelligence artificielle et des systèmes experts. Bref, il s'agit dans des niveaux avancés de la recherche de l'information de développer la stratégie globale de la collecte, de la veille et de l'anticipation. Nous en reparlerons un peu plus en avant.

Donc, une recherche méthodique se basant sur la maîtrise des options offertes par les outils devient nécessaire aux utilisateurs progressivement voués à eux mêmes suites aux effets de la désintermédiation. De plus en plus éloignés des bibliothèques et des bibliothécaires de référence, les utilisateurs ont tout à apprendre pour développer des stratégies de recherche efficaces capables de contrôler les résultats escomptés loin de la menace des deux facteurs nocifs du silence et du bruit de l'information. On observe encore une majorité écrasante d'utilisateur néophytes et même initiés, suivre les méthodes simplistes des requêtes en unitermes sans considération des effets «purificateur» des opérateurs booléens, sur des corpus de documents exclusivement Web sans se hasarder à fouiner dans les archives des news ni dans les catalogues des bibliothèques numériques ou les revues spécialisées en texte intégral.

Nous irons brosser en ce qui suit un aperçu général de ces différentes possibilités «cachées» des outils classiques de la recherche sur Internet. Nous progressons ensuite vers les outils évolués dits intelligents pour démontrer leur apport stratégique aux opérations de recherche et d'accès à l'IST.

5. Critères avancés de la recherche d'IST sur Internet

On ne se limitera pas dans cette partie aux modes d'usage et de fonctionnement des outils de recherche. Ceci est presque maîtrisé dans sa forme la plus rudimentaire par la majorité des utilisateurs Internet. La barre de l'analphabétisme technologique tend heureusement à régresser grâce à la généralisation massive d'Internet. Il est temps de conduire cette évolution, et particulièrement dans les ères de l'enseignement et de la recherche, à des horizons plus vastes.

Nous aborderons ce volet sur trois aspects complémentaires : la structure des outils de recherche et leurs capacités réelles et potentielles de recherche ; les stratégies recommandées pour l'optimisation de la collecte des données stratégiques ; les actions à valeurs ajoutées pour traiter davantage les résultats à des fins de communication ou de prise de décision.

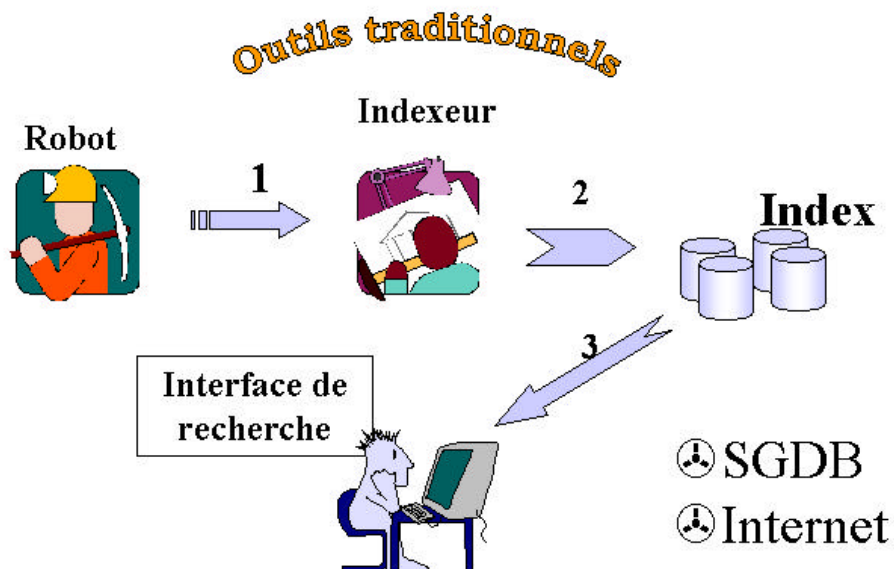
5.1. La structure des outils (répertoires, moteurs, metamoteurs, agents)

Il est peut être important à ce stade d'expliquer sommairement le fonctionnement des outils en amont de la procédure de recherche ; question de distinguer les outils classiques de ceux de la dernière génération de l'intelligence caractérisée.

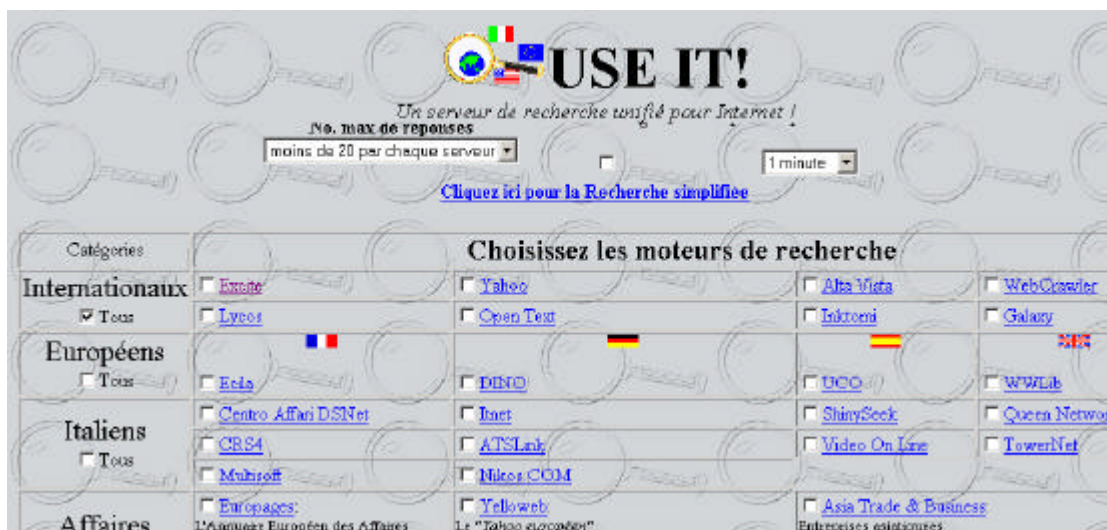
Un outil de recherche (répertoire ou moteur) fonctionne sur le modèle des Systèmes de Gestion des Bases de Données (SGBD). Il se compose de trois éléments essentiels : un robot, un indexeur et une interface de recherche. Le robot parcourt Internet pour ramener à des fréquences régulières et sur une envergure prédéterminée, tous les documents nouveaux et

modifiés qu'il croise dans sa trajectoire. L'indexeur (automatique pour les robots, souvent humain pour les répertoires) se charge d'indexer les documents pour constituer l'Index ; point de départ dans la résolution de toute requête. L'interface de recherche est celle que nous offre chaque outil pour formuler la requête : une arborescence thématique ou une zone formulaire pour saisir nos descripteurs et mots clés.

Système d'Information Classique



Les métamoteurs (i.e. Use It) sont en fait des intermédiaires, des couches interfaces dynamiques qui font le relais entre l'utilisateur et une multitude d'outils. Son rôle est d'économiser le facteur temps de recherche à l'utilisateur.



La catégorie d'outils avancés sont désormais les agents intelligents : outils sophistiqués disposant de beaucoup de ressources d'analyse et de traitement des requêtes. Leur structure de fonctionnement est largement inspirée de l'Intelligence Artificielle et des systèmes experts.

5.1.1. Les répertoires de recherche :

Parmi les astuces et les fonctionnalités avancées des répertoires, il est important de vérifier à tous les niveaux d'approfondissements atteints, la liste des documents proposés qui correspondent à ce niveau de profondeur de la branche suivie. On a tendance à voir des utilisateurs qui confondent d'une part la portée de recherche de la zone formulaire et de l'arborescence du répertoire et d'autre part la distinction entre un répertoire et un moteur. Yahoo est un répertoire et Hotbot est un moteur. Rien ne l'indique à un utilisateur non averti sauf l'analyse du fonctionnement interne de chacun des deux outils. Chacun donne plus d'importance à l'une des deux formes d'indexation : automatique (moteurs) ou humaine (répertoire).

5.1.1. Les moteurs de recherche :

Un moteur de recherche se base essentiellement sur sa partie Index automatique. Ils proposent généralement deux modes de recherche : un mode simple et un mode avancé qui fait appel à des techniques plus pointues de recherche : opérateurs booléens, expressions en langage naturel, envergure de la réponse en nombre de documents restitués et en temps de recherche.

Parmi les précautions souvent négligées dans l'usage des moteurs de recherche, le recours aux documents d'aide à la recherche combinatoire est à prendre en considération. Tous les moteurs n'utilisent pas la même syntaxe. Chaque moteur dispose de sa propre logique d'analyse de la requête multicritère et d'établir un ordre de priorité particulier dans la résolution des équations construites sur des opérateurs booléens.

5.1.1. Les agents intelligents :

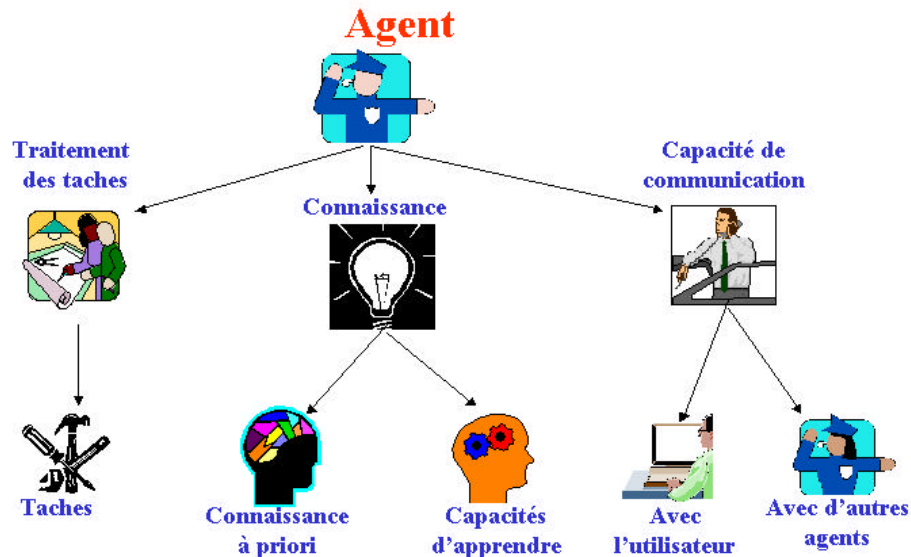
L'une des formes de l'intelligence de certains systèmes est la gestion des profils. Une fois inscrit, l'utilisateur reçoit systématiquement dans sa boîte à lettre ou en se rendant sur le serveur de l'agent une mise à jours des dernières retrouvaille dans le ou les thème(s) proposé(s) lors de son enregistrement.

La structure particulière des agents intelligents en fait des outils capable d'une adaptation et d'une souplesse qui répond de meilleure façon aux attentes des utilisateurs. Parmi leur caractéristiques :

- **L'autonomie** : Ils prennent des initiatives autonomes et exercent un grand contrôle sur ses activités sans intervention humaine.
- **La collaboration** : Ils sont capables de collaborer avec d'autres agents ou avec l'utilisateur pour une meilleure résolution de la requête.
- **La capacité d'apprentissage** : Ils peuvent s'adapter aux besoins de l'utilisateur suite à une analyse des actions passées.
- **Le traitement orienté objet** : Ils retiennent la formulation de l'utilisateur et décident comment procéder à les satisfaire.
- **L'auto-formation** : Ils s'activent automatiquement en réaction à des critères externes : temps, réception de message, appel entrant/sortant etc.

- **La mobilité :** Ils sont capable de se déplacer d'une machine à une autre à travers des plates-formes différentes pour accomplir une action.

LA STRUCTURE DES AGENTS



5.2. La stratégie de recherche

Comme il est déjà avancé, la réussite d'une recherche dépend essentiellement de la stratégie adoptée par l'utilisateur, de son état cognitif et de sa maîtrise de son environnement informationnel.

La recherche d'information sur Internet est souvent appréhendée comme passage obligé par un moteur de recherche vers les ressources sur le Web. Or, le Web ne constitue en réalité qu'une partie infime des ressources réellement disponible sur Internet. A part les catalogues des bibliothèques en ligne, devenus progressivement obsolètes face au document intégral, le commun des mortel tend à oublier ou à ignorer l'existence d'autre source d'IST d'une importance beaucoup plus importante : les News et les archives des News.

Dans les nouvelles formes de communication synchrone et asynchrone, il y a davantage de penchant vers le ludique et le récréatif. Ceci explique la convoitise exponentielle du Chat face aux services asynchrone des listes de diffusion ou des News.

Une bonne stratégie d'information sur Internet suivrait les étapes suivantes :

- Collecte d'opinions sur les outils et les ressources les plus proches à vos besoins ;
- Surveillance régulière des nouveautés: faisant apparaître des points de changement possibles:
 - recherche des rubriques "New" sur les sites de votre choix
 - consultation des sites de nouveautés dans les What's New: exemples: [What's new in the World?](#) ou <http://yahoo.com/new/>, ou encore <http://nctech.fr/NCTech/html/Francais/Nouveautes.html>,
 - inscription à des [listes de diffusion](#),

- revue de presse quotidienne taillée sur vos centres d'intérêt.: [Newspage](#), [CRAYON](#), [Data Times](#), [Excite](#), [Global Interest News Agency](#), [InfoSage](#). Geocities, Netscape Center et bon nombre de prestataires commerciaux offrent des services analogues,
- abonnement à des chaînes transmises en technologies Push comme PointCast, BackWeb et Marimba. Le navigateur Internet Explorer vous propose de recevoir gratuitement un certain nombre de chaînes.(intérêt plutôt grand public)

5.2. Les actions à valeurs ajoutées

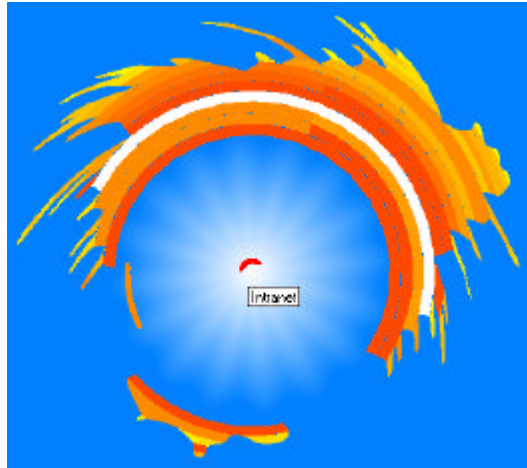
- La traductique : des outils comme www.voilà.fr proposent des services de traduction automatique des documents récupérés.



- Les logiciels d'analyse statistique ou linguistique a posteriori de l'information collectée.: Anchor Page, [Umap](#), [SemioMap](#),

En voici quelques exemples :

- Umap : Le fonctionnement d'[Umap Web](#) sert énormément pour la veille sur Internet. Une fois la requête soumise, Umap récupère le contenu des réponses restituées pour l'analyser en vue d'en faire une cartographie des mots qui permet de mettre en évidence des rapprochements entre certains termes (des « îlots ») sur la carte graphique élaborée par le logiciel.



- [The Brain](http://thebrain.com) (<http://thebrain.com>) est un outil qui permet d'organiser les données selon une approche intuitive basée sur le comportement associatif du cerveau humain.
- ECatch (<http://www.ecatch.com>) est un outil de veille sur Internet. Une fois configurés, les agents intelligents rapatrient automatiquement en tâche de fonds des pages HTML ou des images d'un site ayant subi des modifications.
- Netmind (<http://www.netmind.com>) propose un service gratuit pour être automatiquement averti via la messagerie électronique de toute modification d'une page Web sur l'Internet.

L'effet Internet dans le contexte universitaire : rapports enseignant-enseigné

Qu'en est-il de tout cela dans le contexte universitaire ?

Pourquoi Internet a toutes les chances de réussir comme moyen d'éducation là où tant d'autres inventions ont relativement échoué ?

Comment l'usage de ces outils en particulier et le recours aux services et outils Internet changeraient-ils l'environnement pédagogique et d'enseignement ?

Quelles retombées tous ces outils et ces systèmes auront-ils sur les rapports enseignants-enseignés ?

N'a-t-il pas en fin de compte pour effet d'éloigner de plus en plus étudiants et professeurs des livres ?

N'a-t-il pas aussi l'effet de transformer le travail que représentent l'étude approfondie et la réflexion, et du contact humain direct avec les autres étudiants et professeurs en un monde d'isolement et de déshumanisation des rapports ?

Les quelques exemples suivants argumenteront en faveur des avantages que pourraient engendrer Internet dans le secteur de l'information académique et de l'éducation. Malgré sa

dimension innovatrice, il y a des grandes similitudes avec les systèmes et formes traditionnelles d'enseignement universitaire.

Internet peut donner accès à des sources d'information quasi illimitées, difficilement accessibles par d'autres moyens. L'évolution du réseau fera en sorte qu'il y aurait plus de contrôle de qualité, plus de systèmes d'orientation vers les ressources pertinentes et fiables. L'intelligence artificielle évolue à grand pas pour permettre la gestion des masses de connaissances offertes par les différents médias. Internet sera inéluctablement la bibliothèque du futur, avec beaucoup moins de contraintes de formats de supports, de rapidité d'accès, de contraintes linguistiques, de contraintes d'acquisition etc.

Sur le plan didactique, Internet est tout à donner :

- Ouvrages de références en ligne,
- Edition électronique de cours et de manuels pédagogiques,
- Edition électronique de revues spécialisées et ouvrages de références

L'autre élément d'enrichissement qu'apporte Internet à l'enseignement est la communication elle-même. Le contact et l'échange d'idées et d'opinions entre étudiants et entre professeurs sont parmi les piliers fondamentaux de l'enseignement.

Internet aura tout le mérite de renforcer ce dialogue par son transfert en mode électronique au delà des barrières géographiques et des limites du temps.

Il est certes impossible pour l'instant de dire que l'électronique peut - ou devrait - se substituer au contact humain direct. Il représente néanmoins certaines caractéristiques qui favorisent l'élargissement du champ de l'interaction humaine.

Une autre évolution est par contre en train d'avoir lieu : la mutation du modèle classique dans lequel l'enseignant représente un modèle d'autorité, une figure dominante dans un rapport vertical avec l'étudiant.

Avec l'avènement des NTICs, ce rôle garde une certaine autorité, mais l'enseignant est plutôt enclin à devenir l'accompagnateur, l'aiguilleur vers le savoir et non plus le détenteur absolue de ce savoir.

Les nouvelles technologies éducatives essaient désormais de mettre l'étudiant au centre du dispositif et à lui attribuer davantage de responsabilité avec davantage de moyen d'auto formation dans des structures de groupes d'apprentissage.

.../...