

Acteurs et enjeux de l'Internet



par Paul Vidonne

*Docteur d'Etat - Maître de conférences des universités
Expert informatique près la cour d'Appel de Grenoble*

Les quatre premières machines connectées sous l'égide de l'ARPA (Advanced Research Project Agency), il y a trente ans, ne laissaient guère imaginer le devenir d'Internet. Longtemps apanage des militaires, puis des universitaires, Internet s'est étendu aux entreprises et au grand public avec l'apparition, en 1992, des protocoles nécessaires à sa diffusion de masse.

En quittant le domaine universitaire pour celui du grand public, Internet a subi une véritable mutation. La technicité et l'aridité des débuts ont laissé place à la facilité et à la convivialité, ouvrant ainsi la voie à la diffusion universelle des textes, des images et des sons. Mais l'échange scientifique et la gratuité originels doivent désormais cohabiter avec l'échange marchand comme avec les détournements les plus répréhensibles.

La généralité et la rapidité du développement de ce phénomène pose des problèmes nouveaux qui touchent tant l'évolution des structures économiques, l'adaptation du droit que la transformation de la société.

Ces problèmes sont difficiles, car les cerner nécessite une bonne connaissance des acteurs, des structures, du fonctionnement et des techniques de l'Internet.

Par rapport à l'ampleur de ces questions, l'objet de ce papier est bien modeste. Il vise à apporter l'information du technicien sur les rôles et les fonctions des différents acteurs de l'Internet ainsi que sur les techniques mises en œuvre. Il vise aussi à mettre l'accent sur quelques situations sur lesquelles magistrats et experts ont déjà eu, où auront à se pencher, à moins que le projet de loi annoncé apporte toutes les réponses aux questions qui se posent et notamment celle de la responsabilité des fournisseurs d'hébergement qui est aujourd'hui au centre des débats.

Les acteurs et les services de l'Internet dans son ensemble seront présentés dans un premier point. Les second et troisième points seront spécifiquement consacrés au World Wide Web et porteront sur la technique d'une part et le contrôle du contenu des sites d'autre part.

I - LES ACTEURS ET SERVICES DE L'INTERNET

L'Internet relie un ensemble d'utilisateurs et de prestataires qui mettent en œuvre

divers services ou protocoles.

A. Les acteurs

Les acteurs se composent d'utilisateurs et de prestataires. Ces derniers comprennent des prestataires techniques et des éditeurs de contenu.

1. Les utilisateurs

Les utilisateurs sont des personnes physiques qui ont reçu la possibilité d'accéder au réseau et d'utiliser certains de ses services.

Deux types de services doivent être distingués : les services d'accès publics, ouverts à tous les utilisateurs, et les services privés, réservés à un ou plusieurs utilisateurs particuliers.

L'accès aux services privés requiert le plus souvent la disposition de deux éléments composés d'un identifiant et d'un mot de passe comme par exemple l'accès à sa messagerie.

Dès qu'il est connecté au réseau, l'utilisateur peut consulter à l'écran des informations qui comprennent des textes, des images fixes ou animées, des sons ou encore des films. Dans de nombreux cas, l'utilisateur peut demander la copie ou le transfert de ces informations sur son ordinateur et les réutiliser ensuite. De même, l'utilisateur peut demander le transfert sur son ordinateur de fichiers qui lui sont proposés.

Dans l'autre sens, l'utilisateur peut envoyer de l'information sous forme de textes ou de fichiers vers des destinataires qui l'auront implicitement ou explicitement autorisé. L'information expédiée peut être destinée à un utilisateur privé ou devenir immédiatement publique dès sa réception.

2. Les prestataires techniques

Nous nous limiterons cependant ici à la définition des prestataires qui concourent directement au fonctionnement d'Internet : fournisseurs de réseaux de télécommunications, fournisseurs d'accès et fournisseurs d'hébergement.

a) Les fournisseurs de réseaux de télécommunications

Le réseau Internet peut être représenté comme un gigantesque filet qui recouvrirait la terre entière.

Cette architecture, comme celle du découpage de l'information en paquets qui lui est intimement liée, a été mise en place afin de permettre l'acheminement des paquets en tout état de cause et en particulier dans l'hypothèse où un pan entier du réseau aurait été mis hors d'usage à la suite d'un conflit nucléaire.

Les nœuds du réseau sont reliés entre eux par des circuits de communication publics ou privés loués aux opérateurs de télécommunications. Les utilisateurs font appel à ces mêmes opérateurs pour se connecter sur le réseau Internet.

Ces circuits peuvent être partagés ou spécialisés. Les circuits partagés sont occupés pendant un temps donné pour transmettre les informations du réseau Internet et peuvent par ailleurs simultanément ou successivement transmettre de la voix (communications téléphoniques) ou des images (réseaux câblés des villes). Les circuits spécialisés sont exclusivement dédiés au trafic Internet. Il s'agit alors des

liaisons majeures entre les nœuds ou épine dorsale du réseau (Backbone).

Il n'est pas impensable de voir surgir dans le futur d'autres circuits offrant le même service. Tel est le cas des réseaux d'alimentation en électricité en particulier.

b) Les fournisseurs d'accès

Les fournisseurs d'accès sont des prestataires techniques dont les ordinateurs constituent l'architecture informatique du réseau et qui disposent de plages d'adresses par lesquelles ils permettent à leurs clients de rentrer sur Internet. Leurs ordinateurs sont reliés entre eux par l'intermédiaire des réseaux de télécommunications. Leurs clients peuvent être des utilisateurs comme des fournisseurs d'hébergement.

Comme son nom l'indique, Internet, parfois qualifié de "réseau de réseaux", est une architecture qui permet la mise en relation de réseaux informatiques. Sa grande originalité est son caractère totalement décentralisé. Il n'existe pas de point central ou sommet de pyramide qui exercerait une influence hiérarchique sur l'ensemble du réseau.

Tous les ordinateurs du réseau communiquent avec tous les autres et il est toujours possible d'ajouter au réseau Internet un nouvel ordinateur ou un ensemble d'ordinateurs. Cette possibilité repose sur un système original d'identification des ordinateurs qui est l'adressage IP (Internet Protocole).

Les fournisseurs d'accès gèrent la mise en relation de l'ensemble des machines et des réseaux connectés à l'Internet. La quasi totalité des fournisseurs d'accès sont simultanément des fournisseurs d'hébergement, du moins à l'heure actuelle. La réciproque n'est pas vraie.

c) Les fournisseurs d'hébergement

Les fournisseurs d'hébergement sont des prestataires techniques qui proposent à leurs clients les différents services d'Internet. A cette fin, ils disposent d'ordinateurs reliés en permanence au réseau par l'intermédiaire des fournisseurs d'accès.

Quelques grandes entreprises et institutions (universités en particulier) sont leurs propres hébergeurs, mais il s'agit plutôt d'exceptions.

Les contraintes qui pèsent sur les hébergeurs sont des contraintes de nature technique et contractuelles à l'égard de leurs clients.

Une fois les systèmes informatiques en fonctionnement, le fournisseur d'hébergement doit ouvrir des "comptes" pour chacun de ses clients, afin que ces derniers puissent utiliser les services mis en place.

Le métier d'hébergeur est bien un métier spécifique de l'Internet, même si la réalité économique montre des entreprises qui exercent ce seul métier tandis que d'autres fournissent à la fois l'hébergement et l'accès.

3. Les éditeurs de contenu

L'éditeur de contenu combine de manière originale des informations composées de textes, d'images fixes ou animées ou des sons qu'il met à la disposition du public en utilisant les moyens techniques de différents prestataires de l'Internet.

Les éditeurs de contenu peuvent être des sociétés commerciales, des institutions

publiques ou privées ou des individus. La combinaison des informations mises à disposition du public peut être réalisée par l'éditeur lui-même ou, le plus souvent, sous-traitées à une entreprise dont c'est le métier.

Plusieurs services de l'Internet peuvent être utilisés par les éditeurs de contenu pour faire connaître leur "œuvre" au public : le World Wide Web bien connu, les forums de discussions ou encore les listes de diffusion.

a) L'information sur le WEB

La notion d'éditeur de contenu est la plus facile à appréhender concernant le World Wide Web.

Si la notion d'éditeur de contenu est assez évidente pour les informations publiées sur les serveurs du World Wide Web, il reste une particularité qui doit être examinée : celle du "Livre d'Or".

Comme son équivalent papier, le livre d'or résulte de la possibilité donnée au visiteur d'un site de dactylographier une appréciation dans une "fenêtre" qui lui est ouverte. L'éditeur peut alors choisir de rendre immédiatement publiques ces appréciations. Se pose alors la question de savoir qui, de l'éditeur ou du visiteur, est l'éditeur du contenu.

Dans ce cas particulier, nous sommes ramené à la question posée pour les forums de savoir qui doit être considéré comme éditeur de contenu.

b) L'information dans les forums

Les forums sont des lieux de discussion consacrés à des sujets particuliers. Il en existe quelques dizaines de milliers dont les thèmes portent sur des questions extrêmement diverses. Ces forums sont publics, permettant ainsi à chacun d'écrire et de lire tout message.

Le contenu d'un forum peut être mis sous le contrôle d'un "modérateur" dont l'activité consiste à filtrer les messages et à écarter ceux qui ne correspondent pas au sujet ou encore qui comportent des propos illégaux. En fait, assez peu de forums sont "modérés" et beaucoup laissent à la communauté des membres le soin de réagir aux messages inconvenants.

La notion d'éditeur de contenu reste toutefois mal appropriée pour définir une "œuvre" faite de l'échange de très nombreux messages et pose la question de savoir qui, des intervenants du forum ou du modérateur, doit être considéré comme l'éditeur du contenu.

c) Les listes de diffusion

Un cas un peu particulier concerne les "listes de diffusion". Une liste de diffusion est un moyen commode de faire parvenir un message à quelques dizaines ou plusieurs centaines de milliers de destinataires en ne l'adressant qu'à une seule adresse, celle de la liste.

Ces listes sont le plus souvent utilisées comme des moyens de communication permettant des échanges au sein de groupes particuliers "d'abonnés". Dans ce cas, les adhérents ont effectué une démarche volontaire d'inscription et en général tout abonné à la liste a l'autorisation de poster des messages à l'attention de tous les autres. Il en résulte un système d'échanges multipolaires tout à fait intéressant dans

tous les domaines. Il existe des centaines, voire des milliers de listes publiques.

Là encore, le contrôle exercé sur ces listes est très variable. Elles disposent en effet d'un "propriétaire" dont le rôle est d'accepter ou de refuser l'inscription d'un nouveau membre et d'exclure les contrevenants aux règles de fonctionnement. Elles comprennent encore un "modérateur" qui peut recevoir comme charge de donner son accord avant tout envoi de message à la liste.

Mais il existe des listes pour lesquelles ni les inscriptions ni les messages ne sont filtrés et dont la nature privée ou publique est sujette à discussion.

B. Les services

Par convention, nous considérons qu'un service offert par Internet correspond à la mise en œuvre d'un ou plusieurs protocoles IP.

Cette définition permet de faire correspondre l'optique du technicien et celle de l'utilisateur qui connaît au moins deux services : la messagerie et le World Wide Web.

1. Les protocoles

Les spécifications des différents protocoles sont décrits par des notes techniques intitulées RFC ou Request for Comments, qui sont les documents officiels du IETF (Internet Engineering Task Force), consortium chargé d'introduire les procédures pour les technologies Internet. Les principaux protocoles utilisés aujourd'hui sont ceux qui permettent les échanges entre les utilisateurs et les sites World Wide Web, les échanges de messages entre serveurs de messagerie, de relever son courrier électronique, de transférer des fichiers entre machines.

Il existe de nombreux autres protocoles que nous ne citons pas ici.

Le succès des interfaces dites "navigateur" et du World Wide Web tient sans aucun doute au fait que ces logiciels ont su regrouper la plupart de ces protocoles techniques dans une interface unique et conviviale.

2. Informations publiques et privées

Une information est publique quand n'importe quel utilisateur peut en prendre connaissance : l'éditeur de contenu ne peut ni autoriser ni empêcher un utilisateur d'accéder à une telle information.

Sont par essence publiques les informations présentes sur les sites World Wide Web, dans les forums de discussion, ou sur les FTP anonymes.

Une exception concerne les sites World Wide Web, pour lesquels les fournisseurs de contenu peuvent rendre privées certaines parties réservées aux utilisateurs disposant d'un code d'accès.

Sont par essence privés la messagerie ou les FTP privés : seuls les utilisateurs auxquels auront été communiqué un identifiant et un mot de passe pourront avoir accès aux informations.

Après ce tour d'horizon consacré aux différents acteurs et services de l'Internet, il convient de se pencher plus particulièrement sur les sites World Wide Web publics qui attirent actuellement l'attention en raison de plusieurs affaires judiciaires largement médiatisées.

II - LES SITES WORLD WIDE WEB PUBLICS

Le World Wide Web a pris une telle extension qu'il est parfois confondu avec l'ensemble de l'Internet.

Il a été promu tant par les Etats que par les entreprises. Il a été rapidement adopté par le grand public parce qu'il apporte, de manière très conviviale, des services qui répondent à des besoins. Il a su à la fois se substituer à des dispositifs anciens et trouver des champs nouveaux d'application.

C'est également le World Wide Web qui a concentré l'attention des mass-médias tant pour ses réussites que pour ses déviations (pornographie, révisionnisme, pédophilie, atteintes au droit de la propriété).

Pour comprendre cet extraordinaire développement, il faut rappeler quelques-unes de ses modalités de fonctionnement. Nous examinerons tout d'abord l'organisation économique, avant de nous pencher sur la technique.

A. L'hébergement

Les fournisseurs d'hébergement peuvent se ranger en trois grandes catégories selon qu'ils sont "professionnels", "gratuits" et "grand public".

1. Les fournisseurs d'hébergement professionnels

Ces fournisseurs concluent des contrats d'hébergement avec leurs clients auxquels ils facturent, outre les frais de mise en route, les services mis en œuvre pendant une période de temps donné.

Les services ouverts sont en général de bonne qualité et disposent de toutes les fonctionnalités souhaitables.

La clientèle des fournisseurs d'hébergement professionnels est constituée d'entreprises et d'institutions publiques ou privées. Les opérateurs de création de compte, de définition des droits, de mise à disposition de logiciels sont généralement effectuées par le fournisseur d'hébergement.

2. Les fournisseurs d'hébergement gratuits

Ces fournisseurs ouvrent gratuitement des comptes à leurs "clients".

Le plus souvent, cette ouverture de compte est faite sans intervention humaine, par la mise à disposition des outils nécessaires sur le site World Wide Web de l'hébergeur. Le futur "client" est généralement invité à lire une notice attirant son attention sur la législation en vigueur, ses obligations et ses responsabilités.

La vérification de l'existence du "client" est faite par l'envoi d'un mail à l'adresse donnée, avec demande d'un retour qui conditionne la création effective du compte. En fait la seule vérification opérée par cette procédure est celle de la réalité d'une adresse e-mail.

Dès le retour du mail, le compte est créé et le client peut alors mettre immédiatement en œuvre les services qui lui ont été alloués. En général, ces services se composent d'un site Web et de la messagerie, ainsi qu'un moyen de mettre en place des fichiers sur son site.

Par rapport aux hébergements commerciaux, les services sont parfois de moindre qualité et surtout beaucoup plus frustes, les volumes sont limités, les droits réduits à minima.

La "gratuité" de ces sites est très relative, puisque, le plus souvent, l'hébergeur conditionne l'ouverture d'un compte à la possibilité d'insérer un bandeau publicitaire sur une ou plusieurs pages de son client et se réserve le droit de lui envoyer de la publicité à son adresse électronique. Ce mode d'hébergement connaît à l'heure actuelle un développement extrêmement rapide.

Il faut insister sur le risque permanent que prennent ces hébergeurs de se faire prendre par des utilisateurs indécents qui vont aller se faire créer un e-mail gratuit et anonyme sur un premier site, s'en servir pour ouvrir un site Web chez un hébergeur gratuit, et mettre immédiatement à disposition du public des contenus répréhensibles : moins d'une heure suffit pour effectuer ce périple.

3. Les fournisseurs d'hébergement grand public

De nombreux opérateurs, comme par exemple France Télécom Interactive (Wanadoo), AOL ou encore Club-Internet, combinent une offre d'accès et une offre d'hébergement pour un tarif modique à destination des ménages et des très petites entreprises.

L'offre comprend les accès par le réseau téléphonique commuté et par Numéris qui sont de débit limité.

Elle comprend ensuite l'hébergement d'une messagerie pour quelques adresses et l'hébergement d'un site World Wide Web de taille limitée.

Hormis le paiement, l'hébergement fonctionne exactement sur le modèle des hébergements gratuits. On peut même considérer que le paiement ne rémunère que l'accès et que l'hébergement est un service gratuit offert aux seuls abonnés.

Toutefois, ce paiement permet l'identification du client.

B. Les modalités de fonctionnement

Deux éléments doivent être présentés pour comprendre le fonctionnement du World Wide Web : le mode client-serveur et l'hypertexte.

1. Le mode client-serveur

Les informations disponibles sur le Web sont enregistrées dans les répertoires des ordinateurs des fournisseurs d'hébergement. Ces informations sont regroupées en fichiers accessibles à partir de leur adresse.

Le principe de fonctionnement repose sur le principe dit client-serveur.

On dira qu'un ordinateur est "serveur" quand sa fonction est de retourner une sélection d'informations à la requête d'un ordinateur dit client. L'ordinateur serveur doit être capable d'interpréter les requêtes, d'exécuter les instructions et de renvoyer un résultat à l'ordinateur client.

Un ordinateur est dit "client" quand sa fonction est d'envoyer des requêtes à un serveur, d'en recevoir les résultats et de les interpréter. Naturellement, le dialogue nécessite l'usage de langages communs.

La transaction typique du Web est la demande d'une page par l'envoi de son adresse sur le réseau, et la réponse du serveur qui retourne le code de cette page au client demandeur. A l'arrivée, l'ordinateur client doit généralement afficher cette page. A cette fin, le client va interpréter le code reçu, ce qui explique les différences d'affichage d'une même page Web d'un client à l'autre en fonction des matériels et des logiciels utilisés.

Le principe humain fonctionne dans le même sens : un utilisateur doit demander une information pour qu'elle lui soit envoyée.

On a peut-être pas assez mesuré toutes les implications de cette nécessaire initiative de l'utilisateur. Elle porte à conséquence sur la notion de publicité. La publicité est toujours un message non sollicité qui s'insère par force dans un ensemble d'informations visuelles ou auditives recherchées par un individu subissant ainsi les initiatives d'un publicitaire. Sur Internet, l'utilisateur doit effectuer une démarche volontaire pour accéder à une information (bien que la publicité "classique" soit maintenant présente).

2. L'hypertexte

On peut définir l'hypertextualité comme la mise en relation d'un ensemble de textes ou d'informations par des liens permettant à l'utilisateur de naviguer de l'un à l'autre de manière non séquentielle.

L'établissement d'un hypertexte repose sur deux éléments : une ancre d'une part et une cible de l'autre. L'ancre peut être constituée d'un signe, mot ou une phrase d'un texte, ou encore d'une image ou même d'une partie d'image. La cible peut être un texte, un emplacement dans un texte, une image ou encore un son. Il s'agira toujours d'un fichier qui sera interprété selon sa nature à l'arrivée sur l'ordinateur (affiché, joué, stocké, exécuté).

Des millions de fichiers sont liés entre eux pour constituer la fameuse "toile" : le World Wide Web est désormais porteur du projet prométhéen de réunir toute la connaissance du monde en un vaste ensemble d'informations décentralisées et reliées entre elles de manière non hiérarchique.

C. Les moteurs de recherche

L'analyse exhaustive de l'ensemble des sites est naturellement impossible et il ne faudrait céder qu'avec prudence à la tentation taxinomique portant sur un domaine aussi mouvant.

En effet, compte tenu des moyens techniques et des coûts actuellement en vigueur (ces derniers pouvant être nuls) toute entreprise, institution ou individu du monde développé peut disposer de son propre site World Wide Web.

Qualifier l'ensemble de ces sites de monde "virtuel" est une bien malheureuse évolution sémantique de cet adjectif dont l'origine latine est la "virtus". Ce monde n'est ni un monde en puissance au sens des philosophes, ni un monde fictif au sens de la mécanique ou de l'optique.

Le monde du Web n'est que la représentation d'un monde réel, monde dont il fait d'ailleurs partie, comme le peintre des "Ménines" de Vélasquez fait partie du tableau. Chercher à en décrire le contenu est une tâche peu éloignée de celle qui voudrait décrire le monde réel !

C'est pourquoi il faut souligner toute l'importance des "moteurs de recherche".

S'ils n'ont pas fonction de décrire le World Wide Web, les moteurs de recherche permettent de retrouver toute chaîne de caractères (mot, phrase) présente dans l'un quelconque de ses fichiers contenant du texte.

Ils ont été mis en place par quelques grandes entreprises afin de se faire à elles-mêmes de la publicité ou de vendre des espaces publicitaires. La technique consiste à faire visiter régulièrement l'ensemble des sites Web du réseau Internet par des robots qui extraient les textes significatifs et à les indexer dans d'immenses mémoires d'ordinateurs.

Ces sociétés offrent ensuite des sites Web qui permettent aux utilisateurs d'interroger ces mémoires et de connaître les adresses des pages contenant la chaîne de caractère recherchée.

Il en résulte pour notre propos que des mesures allant très loin, comme celles de la fermeture d'un site n'empêcheront pas les utilisateurs de retrouver facilement et rapidement ce site après qu'il ait été transféré vers un nouvel hébergeur situé quelque part dans le monde sous un nouveau nom (affaire Gubbler par exemple).

III - LE CONTRÔLE DES SITES

S'agissant de serveurs en accès public, le contrôle du contenu des sites peut être effectué de manière externe avec des moyens et des limitations assez semblables à ceux du contrôle interne. En effet, grâce aux moteurs de recherche, chacun peut obtenir la liste des pages contenant une chaîne de caractères donnée et les faire afficher.

Les moyens de contrôle interne, à la portée du fournisseur d'hébergement, sont toutefois plus nombreux. Mais leur portée reste limitée compte tenu des volumes d'information en cause et de la volatilité de l'information.

A. Les techniques de contrôle

Même si plusieurs possibilités doivent être mentionnées, il n'existe pratiquement qu'une seule technique de contrôle, celle de la recherche sur chaîne de caractères.

Avant de l'examiner, il faut évoquer la question des sondages pour les distinguer des techniques de contrôle.

1. Sondage et contrôle

La méthode d'investigation par sondage est une technique statistique. Les sondages sont utilisés dans de nombreux domaines, par exemple dans la production pour qualifier des lots par rapport à des spécifications de production. Le prélèvement et l'analyse d'un échantillon permet de savoir si le lot répond ou ne répond pas aux conditions.

Que peut apporter l'analyse par sondage des contenus répréhensibles sur les sites Web ?

La réponse tient tout d'abord à la proportion extrêmement faible de ces sites : quelques dizaines tout au plus sur plusieurs centaines de milliers de sites "gratuits". Du fait de la grande dispersion du phénomène, l'information apportée par le sondage

sera faible à nulle.

Si le sondage indique, avec une forte erreur aléatoire, qu'une certaine proportion des pages présente des contenus répréhensibles, il n'indiquera pas pour autant lesquelles !

Par définition, une enquête par sondage ne peut remplacer une analyse exhaustive.

2. La recherche de chaîne de caractères

Il existe une méthode d'analyse exhaustive qui est celle de la recherche d'une chaîne de caractère (un mot, une phrase) dans un fichier.

Avec des moyens qui restent raisonnables, il est techniquement possible d'établir la liste de tous les fichiers qui contiennent telle ou telle chaîne de caractères.

Cependant, cette méthode retourne souvent beaucoup de bruit et confondra dans un même résultat un site pédophile et un site d'aide à l'enfance malheureuse. Dans tous les cas, il faudra compléter la recherche automatisée par une vérification humaine.

3. Les autres moyens

D'autres méthodes de contrôle peuvent venir compléter la recherche par chaîne de caractères.

Ces autres méthodes reposent sur l'isolement et l'analyse des comportements inhabituels : soudaine augmentation des accès, importance des transferts de fichiers par exemple.

Comme pour la recherche sur chaîne, ces méthodes exigent toujours une vérification humaine : la soudaine augmentation des accès peut tout autant résulter de l'apparition d'un contenu répréhensible attractif que de la citation élogieuse dans une publication à fort tirage !

L'hébergeur peut encore éditer les noms de fichiers et effectuer sur ces derniers des recherches sur chaîne.

B. Efficacité et limites des contrôles

1. Les fichiers images et sons

La première limite des contrôles automatisés concerne les fichiers d'images et de sons pour lesquels il n'existe aucun moyen de recherche automatisé comparable à la recherche de chaîne des fichiers de textes. L'annonce récente de logiciels capables de reconnaître des formes doit être considérée avec circonspection.

Tout au plus, quelques formats de fichier son peuvent inclure un titre ou un nom d'auteur "en clair" sur lesquels une recherche serait possible. Mais cette recherche est assez inutile dans la mesure où, le plus souvent, l'hyperlien qui appelle ces fichiers contient déjà les mêmes indications.

De plus, rien de comparable n'existe concernant les images qui n'ont pas besoin d'hyperlien pour les appeler car elles se représentent elles-mêmes.

2. Les fichiers textes

Pour les fichiers textes, l'auteur qui connaît les techniques de recherche peut assez

facilement déjouer les contrôles en usant de synonymes, de diminutifs ou de périphrases.

Une autre limite doit être mentionnée.

En matière pénale, on peut imaginer qu'il soit possible de dresser une liste de substantifs ou d'adjectifs litigieux. Cette liste ne sera jamais suffisante et entraînera toujours du bruit, mais elle fournira un premier résultat.

En revanche, en matière civile il est rigoureusement impossible de dresser une liste de tous les termes susceptibles de donner lieu à contestation, des noms de marques aux œuvres protégées !

Avec la volatilité déjà examinée, la principale limite aux contrôles provient de volumes d'information extrêmement importants.

C. Les volumes d'information en cause

Il n'est pas possible de faire reposer le contrôle sur une procédure entièrement automatisée : l'intervention humaine s'avère toujours indispensable.

Un calcul des moyens humains nécessaires et donc des coûts de contrôle, peut être esquissé à partir des volumes communiqués par un grand hébergeur français. A partir d'un échantillon, nous avons pu calculer que la vérification sommaire des informations demanderait l'embauche de 60 personnes pour effectuer un contrôle sommaire de tous les fichiers modifiés chaque jour, ce qui, en l'occurrence, reviendrait à tripler l'effectif de cette entreprise !

A cette première vérification, il faudrait ajouter un travail beaucoup plus qualifié, destiné à vérifier les droits concernant les textes et les images, ce qui représente un nombre de postes de travail que nous ne pouvons estimer.

Il faut espérer de la loi à venir que les responsabilités des prestataires techniques soient clairement distinguées de celles des fournisseurs de contenu.